



Automated Multimodal Edge Intelligence for Tuberculosis Screening in Low-Resource Settings: A Review and Deployment-Readiness Taxonomy

Avani G. Shahane^{1,2}, Snehal Bhosale^{3*}, Shabana Urooj⁴

¹ Department of Computer Science and Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India

² Computer Sc. & Engineering (AI and ML), D.K.T.E. Society's Textile & Engineering Institute, Ichalkaranji 416115, India

³ Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India

⁴ Department of Electrical Engineering, College of Engineering, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

Corresponding Author Email: snehal.bhosale@sitpune.edu.in

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/jesa.jesa.590805>

ABSTRACT

Received: 17 June 2026

Revised: 9 August 2026

Accepted: 20 August 2026

Available online: 31 August 2026

Keywords:

computer-aided detection, edge AI, low-resource settings, multimodal screening, tuberculosis, uncertainty calibration

Screening for tuberculosis (TB) remains difficult in resource-limited areas, where access to radiology, laboratory services, stable electricity and connectivity is constrained; automated point-of-care systems are therefore encouraged. This review covers automated systems that use chest radiography, cough audio, symptoms and clinical metadata for screening, highlighting edge inference, workflow automation, and offline deployment. The search was conducted across five databases with a structured search string, explicit inclusion and exclusion criteria, structured data extraction, and deployment-oriented quality assessment. The synthesis suggests that radiographic models have the strongest evidence base, cough and metadata models can extend access and triage when radiography is delayed, and intermediate or late fusion designs are reported to tolerate missing or low-quality inputs better than early fusion, although systematic comparative evidence for this remains limited. Compact convolutional networks, quantized transformers, and hardware-assisted inference reduce latency and guarantee privacy in edge-compatible architectures. There are still ongoing gaps in paired multimodal data, limited prospective validation, calibration drift, and poor cost-effectiveness data. The contribution is a structured classification, consolidated from established categories, and a deployment-readiness framework that connects the selection of modalities, fusion strategy, edge deployment, handling of uncertainty, missing-modality resilience, explainability, and public-health workflow integration into a system-level design reference.

1. INTRODUCTION

The global burden of tuberculosis (TB) continues to demand screening approaches that require less infrastructure than facility-based pathways, so that remote and marginalised populations can be reached. Active case-finding with ultraportable radiography and artificial intelligence (AI) shows that screening can be carried into remote settings, yet it also exposes the operational difficulty of establishing such services in the field [1].

Finding missed cases is central to TB control, because individuals outside established diagnostic pathways require faster, more affordable, and more accessible routes to presumptive diagnosis [2]. The World Health Organization (WHO) recognises symptom screening, chest radiography, computer-aided detection (CAD) software, WHO-approved rapid molecular tests, and C-reactive protein as complementary screening and triage tools [3]. The 2025 WHO policy statement on CAD reinforces this framework: six CAD products were found to meet WHO performance standards for

TB screening in people aged 15 years and older, although CAD is not yet recommended for children and adolescents younger than 15 [4]. The Stop TB Partnership provides complementary operational guidance on implementing CAD with ultra-portable X-ray systems in screening and triage programmes [5]. An automated screening system should therefore connect screening, triage, and confirmatory testing rather than terminate at a single model prediction.

This is typically done in conjunction with a symptom history, chest X-ray, and confirmatory microbiological investigations. While symptom screening is low cost, it can yield a low detection rate of the disease and a high rate of false positives; AI-assisted chest radiography can therefore improve triage in community screening [6]. Analyses of digital chest X-ray images in hospitalised groups also suggest that AI can aid in triage, but that the benefits rely on integration with confirmation testing [7]. It is crucial to validate externally, choose a high threshold for sensitivity, and consider the population context-in the multi-site study, a TB-detecting radiographic AI achieved a sensitivity of 87% and specificity

of 70% at a high threshold; neither the AI nor the radiologists met the WHO threshold for sensitivity in that study population [8].

Apart from radiology, cough-based analysis is an important low-cost channel because cough can be recorded on mobile phones before radiology is available [9]. The most rigorous multi-country evidence to date comes from the CODA TB DREAM Challenge, in which algorithms developed on 2,143 adults from seven high-burden countries showed that cough-only models were modest (area under the receiver operating characteristic curve (AUROC) of 0.69-0.74), whereas cough-plus-clinical models performed markedly better (AUROC 0.78–0.83; best 80% sensitivity and 73.8% specificity), approaching but not yet meeting the WHO target accuracy for a TB screening test [10]. Against this benchmark, a single high-burden-setting study using speech foundation models reported an AUROC increase from 85.2% to 92.1%, with sensitivity of 90.3% and specificity of 73.1% at a representative threshold, although the authors stated that the model was still awaiting broader validation [9].

Multimodal fusion in radiographic settings is supported by early evidence: a single-centre retrospective study reported that combining clinical data with a deep-learning-based chest X-ray detection algorithm yielded an AUROC of 0.924, with specificity of 81.4% at a sensitivity of 90%; given the single-centre design and limited sample, this result should be read as preliminary evidence of feasibility rather than established proof of a generalisable multimodal benefit [11]. Vision-language models (VLMs) can be built with this as well, connecting imaging features and clinical text, but with a computational cost that should be considered in deployment [12].

The issues addressed in the present review, therefore, do not only involve diagnostic accuracy but also deployment. Studies on user acceptance of CAD indicate that user acceptance is

influenced by workflow integration, interpretability and trust when using the CAD system [13]. Implementation studies in India also show that AI tools are not just a product of model performance but are also influenced by the connectivity, training, equipment, and programme design of the tools [14].

Technologically, multimodal machine learning forms the basis for integrating images, audio, symptoms, and metadata into more powerful screening [15]. Despite all this, known pitfalls of TB computer-aided diagnosis-dataset shift, weak external validation, and limited domain adaptation-remain [16]. Infrastructure, hardware and quality control still need to be in place to ensure success; when they are, population-scale deployments illustrate the high throughput that automated radiographic screening can achieve [17].

Offline and edge processing offer a route to move screening intelligence from the cloud to the clinic, the mobile van, and community devices. Edge deployment on low-cost hardware indicates that on-device inference is feasible for radiographic TB screening [18]. Compact convolutional models likewise show that radiographic screening can be accurate while being engineered for clinical deployment under resource limitations [19].

This review examines automated multimodal TB screening systems operating under offline and edge constraints and organises them within a deployment-oriented classification. The contribution is fourfold: a classification of modalities, fusion strategies, deployment modes, automation functions, and human-machine interaction, consolidated from categories already established in the literature; a comparative analysis of representative systems; a deployment-readiness framework that operationalises these categories for low-resource settings; and a mapping of research gaps to a future-research roadmap. The overall system that frames this review is shown in Figure 1.

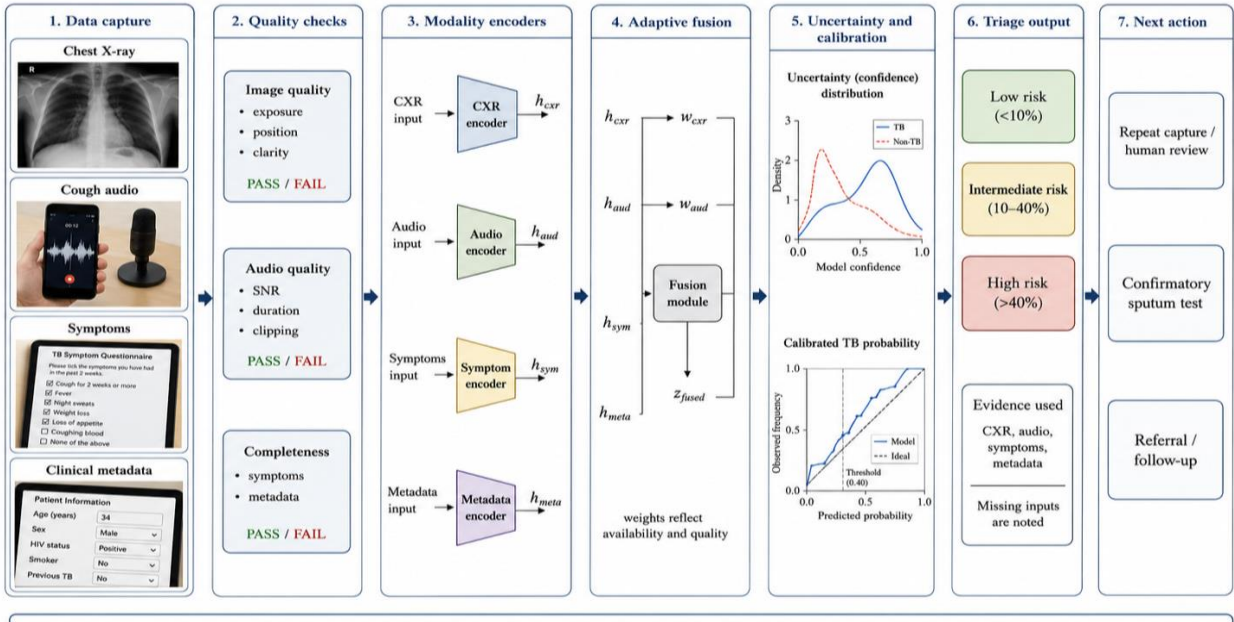


Figure 1. Architecture of an automated multimodal tuberculosis (TB) screening system

2. REVIEW METHODOLOGY

This review followed a transparent, reproducible protocol so that the synthesis can be replicated and extended. The

search covered five databases - PubMed, Scopus, Web of Science, IEEE Xplore, and arXiv-for studies published between January 2020 and May 2026, reflecting the dispersion of relevant work across medical, public-health, and computer-

science literature.

Four concept groups were combined with Boolean operators. The first group captured the disease ("tuberculosis" OR "TB"); the second captured methods ("AI" OR "machine learning" OR "deep learning" OR "CAD"); the third captured modality and fusion ("multimodal" OR "image" OR "chest X-ray" OR "cough" OR "audio" OR "symptom" OR "metadata"); and the fourth captured deployment ("edge computing" OR "offline inference" OR "point-of-care" OR "mobile screening" OR "low-resource"). Within each group, terms were combined with OR, and the four groups were combined with AND.

Inclusion and exclusion criteria are summarised in Table 1. Record selection followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 approach. The searches returned 1,412 records (PubMed 312, Scopus 428, Web of Science 296, IEEE Xplore 214, arXiv

162). After removal of 386 duplicates, 1,026 records were screened by title and abstract, and 858 were excluded. The remaining 168 full texts were assessed against the criteria in Table 1, and 123 were excluded: 58 lacked deployment relevance, 37 were not specific to AI, and 28 provided insufficient technical detail. The 45 remaining studies, together with methodological and guideline sources, form the basis of the synthesis. Studies were retained when they addressed TB or related respiratory screening, AI interpretation of imaging or cough audio, clinical-metadata modelling, multimodal fusion, or low-resource and point-of-care deployment. Studies were excluded when they addressed only laboratory microbiology without screening relevance, were not specific to AI, or lacked sufficient technical detail to judge deployment potential. The selection process, with record counts at each stage, is shown in Figure 2.

Table 1. Inclusion and exclusion criteria

Dimension	Inclusion	Exclusion
Topic	TB or related respiratory screening	Laboratory microbiology without screening relevance
Method	AI, ML, deep learning, or CAD	Non-AI statistical or descriptive studies
Modality	Imaging, cough audio, symptoms, metadata, or fusion	Studies with no analysable modality
Deployment	Reports edge, offline, or point-of-care relevance	No deployment or feasibility detail
Period	January 2020 to May 2026	Published before January 2020
Language	English full text available	Full text unavailable

Note: TB: tuberculosis; AI: artificial intelligence; ML: machine learning; CAD: computer-aided detection.

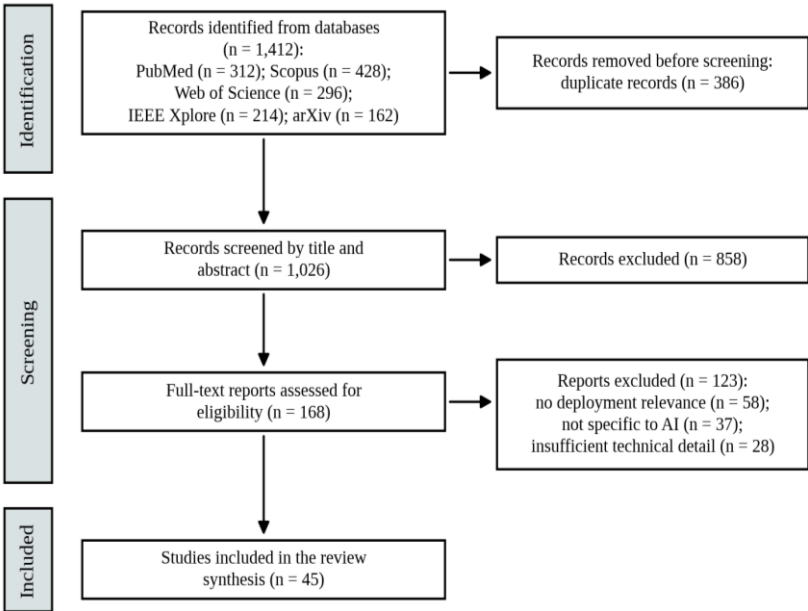


Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of record identification, screening and inclusion

A structured data extraction form was recorded for each included study, including information on the modality or modalities, dataset/population, model family, validation design, reported metrics, deployment target, and feasibility in the edge. Appraisal of quality was done for deployment-relevant construct dimensions: external or prospective validation, sample size, risk of bias, availability of code or data, and deployment evidence reported. A narrative and technical synthesis was used instead of a meta-analysis since the studies are quite varied in their modalities, datasets, validation environments, metrics, and target hardware.

The synthesis categories were derived from the recently published bibliometric study of multimodal data fusion in the

field of smart healthcare [20] and the limitations of edge-AI, including inference latency, energy, and communication cost [21]. A medical multimodal AI study for treating cross-cutting issues of modality missingness, variable input quality and workflow integration was guided by scoping reviews [22] and a multimodal deep-learning review gave consistent definitions of early, intermediate and late fusion [23].

3. CLASSIFICATION OF MULTIMODAL TUBERCULOSIS SCREENING SYSTEMS

TB screening is a triage procedure to identify people

needing confirmatory testing; diagnosis is confirmed by microbiological identification of *Mycobacterium tuberculosis*. Screening is most useful in health care settings where confirmatory testing, radiology services, and specialist radiologists may be expensive, not local, or unavailable. Presumptive diagnosis can then be helpful in automated screening if the rate of false negatives and false positives is low.

The different types of systems can be distinguished by five axes: (1) modality, (2) the way the modalities are combined, (3) the deployment architecture of the system, (4) the automation function performed by the system, and (5) the way that humans interact with the system. Modality combinations specify the type of signals that the model is tuned to; the fusion strategy will describe the process used to combine signals; the deployment architecture will describe where inference is achieved; the automation function will indicate whether the system does triage, prioritisation, or referral support; the interaction mode will describe how the outputs are communicated to clinicians or community health workers. These five axes consolidate categories that are already well established in the multimodal-learning and edge-computing literature [21, 23]; they are used here as an organising classification for TB screening systems rather than as a claim of a new taxonomy.

For pulmonary TB, chest radiography is the most widely used modality, and lesions can be localised and evaluated automatically, making it suitable for a wide range of input modalities with only a small number of annotations, which can cause the loss of accuracy with image variability and lead to the development of hybrid transformer-convolutional neural network (CNN) architectures and customised loss functions [24]. In healthcare, multimodal learning is often beneficial as complementary data streams reinforce each other [25]. Cough sounds, symptoms, and clinical metadata provide accessible context, while AI imaging provides efficiency, not in lieu of clinical judgment [26]. Explainable models are required for practitioners to identify when they are confident, uncertain or when the quality of the data is poor [27] and confirmatory molecular and urine testing demonstrate that triage is enhanced when AI output is supported by practical confirmatory pathways, especially for people living with HIV [28]. Previous experience with broader respiratory AI indicates that validation, usability and reliability are key factors for the adoption of AI, rather than novelty [29]. The main inputs to the screening process, and the roles they typically automate, are summarised in Table 2; the confirmatory molecular pathway is listed only as the end-point to which screening must link, since it is a diagnostic rather than a screening modality.

Table 2. Overview of tuberculosis (TB) screening modalities and automation roles

Modality	Evidence Maturity	Edge Feasibility	Main Failure Mode	Reviewer Expectation
Chest X-ray	High	Medium	Device availability, image quality, threshold shift	External validation and threshold calibration [8].
Cough audio	Emerging	High	Noise, microphone variation, respiratory confounding	Country-level validation and cough-plus-clinical fusion [10].
Symptoms	Medium	Very high	Low specificity	Treat as triage support, not standalone diagnosis [6].
Clinical metadata	Medium	Very high	Missing or inconsistent records	Peer-reviewed validation, fairness and missing-data handling [15].
Molecular or sputum confirmation (diagnostic, not screening)	Confirmatory	Low to medium	Cost, consumables, laboratory access	Link AI screening to confirmatory pathway [28].

Note: 1. The final row is a confirmatory diagnostic pathway rather than a screening modality; it is shown only as the end-point to which automated screening must link. 2. TB: tuberculosis; CXR: chest X-ray; AI: artificial intelligence; HIV: human immunodeficiency virus.

Table 3. Classification of multimodal fusion and deployment strategies

Category	Core Idea	Best-Fit Setting	Advantage	Limitation
Early fusion	Inputs merged before modelling	Controlled data capture	Simple single model	Weak with missing modalities
Intermediate fusion	Latent features merged	Hospital or mobile van	Rich cross-modal interaction	Higher training complexity
Late fusion	Independent scores combined	Rural clinics	Works when one input is absent	Needs score calibration
Cloud-assisted	Central server inference	Connected hospitals	Supports large models	Connectivity and privacy burden
Hybrid edge-cloud	Local triage plus periodic sync	Programmatic screening	Balances autonomy and audit	Requires maintenance plan
Edge-offline	On-device inference	Remote clinics and vans	Low latency and privacy	Limited compute and updates

The most basic multimodal pairing is radiography and symptoms, as radiography gives visual cues and symptoms provide low-cost clinical context. Age, sex, HIV status, smoking history and previous TB treatment are known, and age and sex alone do not yield enough information for risk stratification, so radiography is used with the other information. Cough-and-symptom systems are useful when

radiography is not available or is limited; more sophisticated systems include radiography, cough audio and notes for contextual screening [6].

The second design decision is the fusion strategy. Early fusion merges raw parts prior to representation learning, which might be unstable in case of modalities that are of different scale or quality. Intermediate fusion uses separate encoders

and combines latent features, thus yielding richer cross-modal learning. Late fusion is a combination of per-modality scores and is better suited when the inputs are missing or low quality; adaptive fusion weights modalities based on their availability and quality. The fusion and deployment strategies are organised in Table 3.

The types of deployment architecture include cloud-assisted, hybrid edge-cloud, and offline edge. Cloud-assisted

systems can handle large models and centralized auditing, but are susceptible to outages. Hybrid systems perform local screening and periodic training/auditing in the cloud. Offline systems maintain local inferences and work well in environments that have electricity, internet, and expert access issues [18]. A decision matrix comparing these three architectures is presented in Figure 3.

	Cloud-assisted	Hybrid edge-cloud	Offline edge
Latency	Low	Medium	High
Connectivity dependence	High	Medium	Low
Privacy	Medium	High	High
Local autonomy	Low	Medium	High
Compute capacity	High	Medium	Low–Medium
Maintenance complexity	High	Medium	Low
Operating cost	High	Medium	Low
Best-fit setting	Connected hospitals	Programmatic screening	Remote clinics
Recommended when	high compute and central oversight	balanced autonomy and audit	low-latency screening with unreliable network

Figure 3. Edge deployment decision matrix comparing architectures

4. COMPARATIVE ANALYSIS OF EXISTING SYSTEMS

The systems with the best evidence base for automated TB screening are radiography-dominated systems. Chest images can be used to detect TB patterns using visualisation studies [30] that generate interpretable cues and compact convolutional networks that can be used for fast TB detection and make imaging models more deployable on mobile devices [31].

Some recent small multimodal radiology models suggest that, when evaluated using clinically relevant metrics, radiology interpretation can be made more accessible [32]. In addition, the use of multimodal models for chest X-ray interpretation (M4CXR-style) demonstrates that a single model can be used for several interpretation tasks and not just binary classification [33].

The differences between the use of AI in high-resource and African contexts highlight the need for decisions on the deployment of AI to be made contextually, rather than based on model performance alone [34]. Pilot testing for respiratory imaging in low-resource settings also indicates variability across different settings, populations, and disease burdens from the training distribution [35].

Comparative evaluation of automated radiographic algorithms is important because algorithms can perform differently between countries and populations: in an independent head-to-head evaluation of five CAD products on 3,927 participants from seven countries, the area under the curve ranged from 0.774 to 0.819, specificity at 90% sensitivity ranged from 64.8% to 73.8%, and sensitivity was lower among females and people living with HIV while specificity was lower among males and people previously treated for TB [36]. Supportive evidence for this is provided

by prospective multi-site validation, looking at AI detection and multiple external datasets [8], and screening studies in South Africa and Lesotho, which highlight the importance of understanding performance in relation to local burden, threshold selection, and screening goals [37].

The findings of the AI software reviews for the diagnosis of radiographic TB are generally positive regarding the use of AI as a screening tool and show variation between the study designs and metrics [38]. Wider evaluations of the diagnostic performance also show the value of radiographic AI, but with an emphasis on linkage to confirmatory strategies and local implementation [39].

Cough- and symptom-based systems have importance because many communities do not have access to radiography at first triage. Speech foundation models were used to build a cough-sound system which demonstrated useful triage in high-burden regions [9] and VLMs for chronic TB represent a complementary approach where radiographic reading is augmented by findings from the language model [40].

There is growing evidence for cost-effectiveness. The value of using AI to support the interpretation of radiographs was identified to be dependent on disease burden, device cost, referral pathway, and cohort size in a health-technology assessment (HTA) of the technology [41]. A preprint has additionally described a neural-network model that identifies people with TB from easily obtainable demographic and clinical characteristics [42]; pending peer review and external validation, such models may come to support low-cost triage where imaging or cough recording is not possible, but they should not yet be treated as validated screening tools.

So, it's not just a battle between convolutional networks and transformers. Biomedical VLMs indicate that larger models can be beneficial for reasoning about the semantic content of images, yet do not necessarily exhibit desirable aspects of

deployment [43], and vision-language pretraining on chest X-ray images over time demonstrates the benefits of leveraging visual data along with textual data [44]. Uncertainty-aware vision-language diagnostics are also applicable to field decisions, where a confidence estimate is given along with the prediction [45]. Ensemble convolutional networks and field-programmable gate array (FPGA)-based accelerators can reduce the inference delay of imaging tasks, which are relevant to rapid point-of-care screening [46]; deep learning for the detection of bacilli using microscopy can assist confirmatory workflows, but is not a replacement for community screening [47]. The potential for combining medical reasoning with

image interpretation has been demonstrated by large-language-model methods in the context of respiratory disease [48]; robustness analysis of the comparison between vision transformers and the latest convolutional networks has helped to confirm the significance of out-of-distribution testing [49]. The representative systems compared are listed in Table 4. Table 4 also records whether code or data were released: among these systems, only the CODA TB DREAM Challenge provides data access to registered researchers, and none of the others reported public code or data, which restricts independent benchmarking and is itself a deployment-readiness weakness.

Table 4. Comparative analysis of representative tuberculosis (TB) screening AI systems

Study	Model or System	Inputs	Deployment Target	Code or Data Availability	Key Interpretation
Li [18]	MobileNetV3 with MindSpore	CXR	Ascend 310 edge chip	Not reported	Prototype demonstration of low-cost on-device inference; not yet clinically validated.
Ganapathy et al. [12]	Vision-language model	CXR and clinical notes	High-end edge or server	Not reported	High precision but heavier compute burden.
Ma et al. [9]	Speech foundation model	Cough and metadata	Smartphone or portable capture	Not reported	Cough-plus-clinical fusion improves triage where X-ray access is limited.
Jaganath et al. [10]	Challenge-benchmarked cough algorithms	Cough audio and clinical data	Smartphone or portable capture	Challenge data available to registered researchers	Cough-plus-clinical fusion outperforms cough-only models across seven countries.
Worodria et al. [36]	Multiple CAD algorithms	CXR	Portable X-ray programmes	Not reported	Specificity at 90% sensitivity ranged from 64.8% to 73.8%, with heterogeneity across countries and subgroups.
Parihar et al. [42]	Neural-network triage	Demographic and clinical data	Low-cost digital workflow	Not reported	Preprint-stage evidence for low-cost triage; peer-reviewed external validation still needed.
Nawaz et al. [46]	Ensemble CNN accelerator	CXR	FPGA-supported inference	Not reported	Relevant for fast point-of-care screening.

Note: TB: tuberculosis; AI: artificial intelligence; CXR: chest X-ray; CAD: computer-aided detection; CNN: convolutional neural network; FPGA: field-programmable gate array.

Table 5. Deployment-readiness criteria for low-resource settings

Criterion	What to Report	Why It Matters
Sensitivity and specificity	Values at clinically meaningful thresholds	TB screening prioritises avoiding missed cases
Calibration	Brier score, calibration curve, threshold transfer	Prevents misleading risk scores across sites
Missing-modality robustness	Performance when CXR, audio, or metadata is absent	Low-resource systems face incomplete data
Latency	Inference time on phone, Raspberry Pi, Jetson, or edge chip	Triage decisions are needed during the visit
Memory and model size	Size in MB, parameters, quantization level	Determines feasibility on offline devices
Energy and power	Battery use and thermal stability	Essential for mobile vans and rural clinics
Privacy	Local storage, encryption, and sync policy	Protects sensitive medical data
Human-machine interface	Risk label, explanation, and next-step referral	Converts AI output into usable workflow
Regulatory status	CE marking, FDA clearance, WHO policy or prequalification recognition	Determines procurement eligibility and lawful clinical use

Note: CXR: chest X-ray; AI: artificial intelligence; CE: European conformity marking; FDA: United States Food and Drug Administration; WHO: World Health Organization.

5. DEPLOYMENT-READINESS FRAMEWORK FOR LOW-RESOURCE SETTINGS

Offline and edge deployment must be considered as a fundamental design criterion and not an added afterthought. Edge deployment of computer vision and medical diagnostic applications reveals limitations due to memory, energy, inference latency, model size, and hardware availability [50].

The clinical effectiveness and operational feasibility are thus important practical considerations for an edge model. The proposed deployment-readiness criteria are summarized in Table 5.

The first is latency. For mobile vans and screening sessions, a quick inference is needed in order to provide a triage evaluation during the patient visit. The second point is computational cost and footprint: A model that requires

powerful graphics processing unit (GPU) servers may be attractive in benchmark terms but impractical in clinics that lack such resources and often face an unstable power supply. The third is privacy: on-device inference reduces the need to transfer chest X-rays, cough recordings and clinical data to remote servers.

These properties can be achieved by model compression. First, quantization lowers numerical precision so that the model uses less memory and less compute; second, pruning eliminates unnecessary parameters or channels; third, knowledge distillation moves capability from a large network to a smaller deployable one. Efficient convolutional networks can then be combined with device runtimes or hardware accelerators to realise fast inference.

Deployment readiness also depends on the specific hardware platform, because the same compressed model behaves differently across devices. Central processing unit (CPU)-only single-board computers such as the Raspberry Pi offer the lowest cost but the least computational headroom, so aggressive quantization is usually unavoidable; Jetson-class boards add an embedded GPU that supports larger models at a higher energy cost; and dedicated accelerators, such as the Ascend 310 chip used for radiographic TB screening [18] or FPGA-based pipelines [46], offer low latency within tight power budgets but tie the system to a specific toolchain. Quantization interacts with accuracy across all of these targets: post-training 8-bit quantization typically costs little accuracy when it is calibrated on representative data, but the degradation can grow under the distribution shift and class imbalance that are common in field screening. Deployment-ready reporting should therefore state accuracy, latency, memory and energy per hardware platform after compression, not only for the full-precision development model [21, 50].

The associated risks are also of significance. Domain shift is a decrease in performance after training on a different population or imaging device. Calibration risk exists because the predicted probability does not necessarily align with the true disease risk, which is why reporting of the Brier score, calibration curves and threshold transfer across sites is important. The missing-modality risk is when a cough is not present in the audio, metadata or imaging modality. The risks that arise in such areas as rural environments and mobile

programmes are prevalent and are generally equipped with fallback strategies, per-modality confidence as well as quality indicators in deployment-ready systems.

Regulatory clearance is a further readiness dimension that technical criteria alone do not capture. Commercially deployed radiographic CAD products illustrate the available pathways: CAD4TB and qXR are CE-marked medical devices under the European regulatory framework, qXR additionally holds United States Food and Drug Administration (FDA) 510(k) clearances for a set of chest X-ray findings, and the 2025 WHO policy statement identified six CAD products that meet WHO performance standards for TB screening in people aged 15 years and older [4]. National programmes increasingly require such recognition for procurement, and the Stop TB Partnership implementation guidance treats regulatory status as part of product selection [5]. By contrast, most research prototypes reviewed here, including the edge systems, report no clearance of any kind, which is precisely what separates a feasibility demonstration from a deployable medical device. Regulatory status is therefore included as an explicit criterion in Table 5.

A deployment-ready system also contains a monitoring and update loop, as shown in Figure 4. In field use, drift detection can be made concrete by monitoring input and output distributions on the device: population-stability-index or Kolmogorov–Smirnov statistics computed on image-quality indices, cough signal-to-noise ratios and predicted-risk scores flag drift, while the abnormality-call rate and the proportion of AI-positive cases that are bacteriologically confirmed serve as label-free performance proxies. Checks of this kind can run automatically over a rolling window of recent cases, with alert limits fixed at commissioning; an exceedance triggers a threshold review in which the operating point is re-derived from a recent, locally verified sample. Any model update is then validated locally against a fixed site-specific reference set – several hundred consecutive cases with known confirmatory outcomes is a practical minimum – before release, and the previous version is restored if sensitivity, specificity or calibration degrades. The main contribution of this review to system engineering is the transformation of this static model into an automated screening system that can be maintained safely over time.

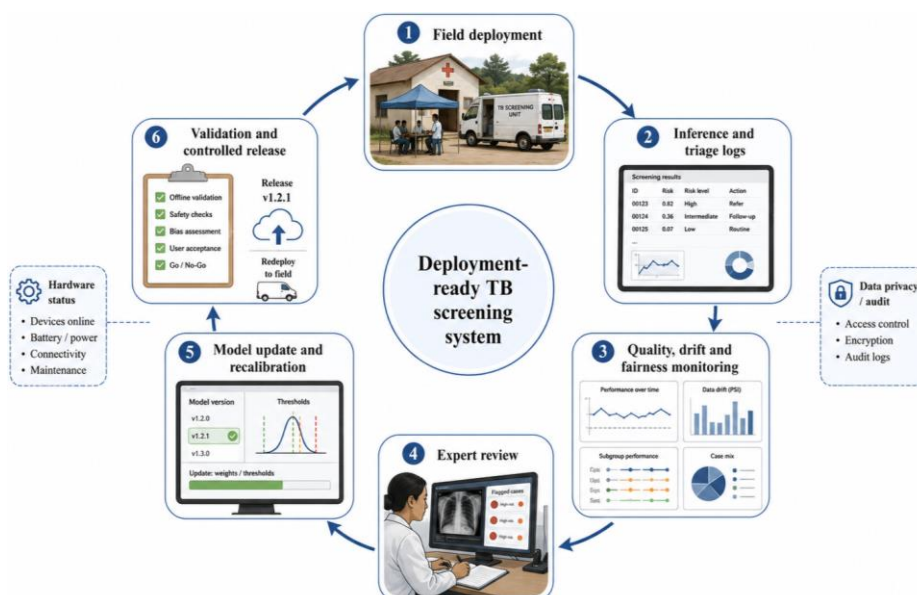


Figure 4. Field monitoring and model update loop

6. RESEARCH GAPS AND FUTURE DIRECTIONS

The first is the lack of paired multimodal TB datasets. Publicly available TB data resources are heavily skewed towards single modalities and almost never pair chest radiography, cough sounds, symptoms, clinical data, and microbiological results for the same patients, which makes it hard to compare fusion strategies or to establish the most useful set of signals in the field.

The second gap is for robustness to missing modalities. Common pitfall of multimodal models is that all modalities are assumed to be present but data on screening are often incomplete. Adaptive fusion, uncertainty-aware routing and modality dropout in training would be good to investigate further to ensure screening can continue without loss of a signal.

The third gap is human-centred explanation. Evaluating multimodal AI emphasizes the need for output interpretability, validation, and relevance to the decision [51]. The explanation in TB screening needs to be understood not just by the machine-makers, but also by radiographers, nurses, community health workers and programme managers.

The fourth gap is the need to validate multimodal large language and VLMs in domain-specific settings. The growth rates in these applications are reported in surveys, but they need to consider reliability and hallucination as well as clinical safety for healthcare applications [52].

The fifth gap is equity and population-level generalisability. The potential of AI for latent TB risk stratification has been highlighted in machine-learning studies of latent tuberculosis infection (LTBI), which demonstrate the importance of diverse immunological, demographic and clinical representation in training and testing [53]. The performance of screening models should be assessed by age, sex, HIV status, geographic region, radiographic platform, and by symptom presentation, especially for subclinical TB and HIV comorbidity sub-groups.

Lastly, the field is embarking on a new phase of medical AI-one that is generative and powerful in its ability to support medical decisions, but is also set to demand greater validation, governance and accountability [54]. Future studies should focus on prospective field trials, federated or privacy-preserving learning, open calibration benchmarks for edge TB screening, cost-effectiveness analysis and implementation science. This is the system-learning and human-machine theme of this journal [55] that is aligned with establishing such automated and monitored screening systems.

7. CONCLUSIONS

The automated multimodal AI for TB screening can integrate complementary information from chest radiography, cough audio, symptoms, metadata, and confirmatory tests. This review demonstrates that the best evidence is currently available with radiographic systems, and with a lack of radiography and radiologists, access and triage consistency become better with audio, symptom and metadata models. Across the included studies, intermediate and late fusion are consistently reported to tolerate missing or degraded inputs better than early fusion; because this review found no systematic head-to-head comparison of fusion strategies under controlled missing-data conditions, that pattern should be read as a design expectation to be confirmed by future

benchmarking rather than as an established finding.

In practical terms, accuracy is not the only property that matters; deployability matters as well. To be efficient, edge-compatible, calibrated, privacy-preserving, human-centered, and workflow-integrated, a low-resource system must meet these criteria. Edge and offline computing are not engineering conveniences; they are clinical enablers for rural clinics, mobile screening vans and community health-worker programmes.

One of the biggest gaps in the current body of evidence is the lack of paired multimodal data sets, prospective field validation and reporting on calibration. This prohibits direct comparison among systems. It is proposed that the future TB screening systems evolve towards adaptive fusion, open multimodal benchmarks, local validation and cost-aware deployment, while being monitored and updated in the proposed deployment-readiness loop. This vision would see demonstration models mature into robust, deployment-resilient automated screening systems in the settings where the burden of TB is greatest.

REFERENCES

- [1] John, S., Abdulkarim, S., Usman, S., Rahman, M.T., Creswell, J. (2023). Results from TB screening using ultraportable X-ray and artificial intelligence in remote populations in northeast Nigeria. Research Square. <https://doi.org/10.21203/rs.3.rs-2714909/v1>
- [2] Byrne, R.L., Wingfield, T., Adams, E.R., et al. (2024). Finding the missed millions: Innovations to bring tuberculosis diagnosis closer to key populations. BMC Global and Public Health, 2(1): 33. <https://doi.org/10.1186/s44263-024-00063-4>
- [3] World Health Organization. (2021). WHO consolidated guidelines on tuberculosis. Module 2: Screening-systematic screening for tuberculosis disease. World Health Organization, Geneva. <https://www.who.int/publications/i/item/9789240022676>.
- [4] World Health Organization. (2025). Use of computer-aided detection software for tuberculosis screening: WHO policy statement. World Health Organization, Geneva. <https://www.who.int/publications/i/item/9789240110373>.
- [5] Stop TB Partnership. (2021). Screening and Triage for TB Using Computer-Aided Detection (CAD) Technology and Ultra-Portable X-Ray Systems: A Practical Guide. Stop TB Partnership.
- [6] John, S., Abdulkarim, S., Usman, S., Rahman, M.T., Creswell, J. (2023). Comparing tuberculosis symptom screening to chest X-ray with artificial intelligence in an active case finding campaign in northeast Nigeria. BMC Global and Public Health, 1(1): 17. <https://doi.org/10.1186/s44263-023-00017-2>
- [7] Biewer, A.M., Tzelios, C., Tintaya, K., et al. (2024). Accuracy of digital chest X-ray analysis with artificial intelligence software as a triage and screening tool in hospitalized patients being evaluated for tuberculosis in Lima, Peru. PLOS Global Public Health, 4(2). <https://doi.org/10.1371/journal.pgph.0002031>
- [8] Kazemzadeh, S., Kiraly, A.P., Nabulsi, Z., et al. (2024). Prospective multi-site validation of AI to detect

- tuberculosis and chest X-ray abnormalities. *NEJM AI*, 1(10). <https://doi.org/10.1056/aioa2400018>
- [9] Ma, N., Mirheidari, B., Brown, G.J., et al. (2025). AI-enabled tuberculosis screening in a high-burden setting using cough sound analysis and speech foundation models. *arXiv preprint arXiv:2509.09746*. <https://doi.org/10.48550/arxiv.2509.09746>
 - [10] Jaganath, D., Sieberts, S.K., Raberahona, M., et al. (2025). Accelerating cough-based algorithms for pulmonary tuberculosis screening: Results from the CODA TB DREAM challenge. *Open Forum Infectious Diseases*, 12(10): ofaf572. <https://doi.org/10.1093/ofid/ofaf572>
 - [11] Choi, S.Y., Choi, A., Baek, S., Ahn, J.Y., Roh, Y.H., Kim, J.H. (2023). Effect of multimodal diagnostic approach using deep learning-based automated detection algorithm for active pulmonary tuberculosis. *Scientific Reports*, 13(1): 19794. <https://doi.org/10.1038/s41598-023-47146-0>
 - [12] Ganapathy, A., Shastry, P., Kumarasami, N., et al. (2025). Vision-language models for acute tuberculosis diagnosis: A multimodal approach combining imaging and clinical data. *arXiv preprint arXiv:2503.14538*. <https://doi.org/10.48550/arxiv.2503.14538>
 - [13] Creswell, J., Vo, L.N.Q., Qin, Z.Z., et al. (2023). Early user perspectives on using computer-aided detection software for interpreting chest X-ray images to enhance access and quality of care for persons with tuberculosis. *BMC Global and Public Health*, 1(1): 30. <https://doi.org/10.1186/s44263-023-00033-2>
 - [14] Vijayan, S., Jondhale, V., Pande, T., et al. (2023). Implementing a chest X-ray artificial intelligence tool to enhance tuberculosis screening in India: Lessons learned. *PLOS Digital Health*, 2(12). <https://doi.org/10.1371/journal.pdig.0000404>
 - [15] Warner, E., Lee, J., Hsu, W., et al. (2024). Multimodal machine learning in image-based and clinical biomedicine: Survey and prospects. *International Journal of Computer Vision*, 132(9): 3753-3769. <https://doi.org/10.1007/s11263-024-02032-8>
 - [16] Liu, Y., Wu, Y.H., Zhang, S.C., Liu, L., Wu, M., Cheng, M.M. (2023). Revisiting computer-aided tuberculosis diagnosis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4): 2316-2332. <https://doi.org/10.1109/tpami.2023.3330825>
 - [17] Munjal, P., Mahrooqi, A.A., Rajan, R., et al. (2025). Population-scale cross-sectional observational study for AI-powered TB screening on one million CXRs. *npj Digital Medicine*, 8(1): 418. <https://doi.org/10.1038/s41746-025-01832-7>
 - [18] Li, H.Y. (2025). Pulmonary tuberculosis edge diagnosis system based on MindSpore framework: Low-cost and high-precision implementation with Ascend 310 chip. *arXiv preprint arXiv:2502.14885*. <https://doi.org/10.48550/arxiv.2502.14885>
 - [19] Rahman, K.K.M., Zulaikha, S., Dhafer, B., Ahmed, R. (2025). Advancing tuberculosis screening: A tailored CNN approach for accurate chest X-ray analysis and practical clinical integration. *Intelligence-Based Medicine*, 11: 100196. <https://doi.org/10.1016/j.ibmed.2024.100196>
 - [20] Chen, X.L., Xie, H.R., Tao, X.H., Wang, F.L., Leng, M.M., Lei, B.Y. (2024). Artificial intelligence and multimodal data fusion for smart healthcare: Topic modeling and bibliometrics. *Artificial Intelligence Review*, 57(4): 91. <https://doi.org/10.1007/s10462-024-10712-7>
 - [21] Himeur, Y., Sayed, A.N., Alsalemi, A., Bensaali, F., Amira, A. (2023). Edge AI for Internet of energy: Challenges and perspectives. *Internet of Things*, 25: 101035. <https://doi.org/10.1016/j.iot.2023.101035>
 - [22] Schouten, D., Nicoletti, G., Dille, B., et al. (2025). Navigating the landscape of multimodal AI in medicine: A scoping review on technical challenges and clinical applications. *Medical Image Analysis*, 105: 103621. <https://doi.org/10.1016/j.media.2025.103621>
 - [23] Shafizadegan, F., Naghsh-Nilchi, A.R., Shabaninia, E. (2024). Multimodal vision-based human action recognition using deep learning: A review. *Artificial Intelligence Review*, 57(7): 178. <https://doi.org/10.1007/s10462-024-10730-5>
 - [24] Bougourzi, F., Dornaika, F., Nakib, A., Taleb-Ahmed, A. (2024). Emb-TrAttUnet: A novel edge loss function and transformer-CNN architecture for multi-class pneumonia infection segmentation in low annotation regimes. *Artificial Intelligence Review*, 57(4): 90. <https://doi.org/10.1007/s10462-024-10717-2>
 - [25] Krones, F.H., Marikkar, U., Parsons, G., Szmul, A., Mahdi, A. (2024). Review of multimodal machine learning approaches in healthcare. *Information Fusion*, 114: 102690. <https://doi.org/10.1016/j.inffus.2024.102690>
 - [26] Khalifa, M., Albadawy, M. (2024). AI in diagnostic imaging: Revolutionising accuracy and efficiency. *Computer Methods and Programs in Biomedicine Update*, 5: 100146. <https://doi.org/10.1016/j.cmpbup.2024.100146>
 - [27] Nasarian, E., Alizadehsani, R., Acharya, U.R., Tsui, K.L. (2024). Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. *Information Fusion*, 108: 102412. <https://doi.org/10.1016/j.inffus.2024.102412>
 - [28] Bjerrum, S., Yang, B., Åhsberg, J., et al. (2024). Parallel use of low-complexity automated nucleic acid amplification tests and lateral flow urine lipoarabinomannan assays to detect tuberculosis disease in adults and adolescents living with HIV. *Cochrane Database of Systematic Reviews*, 2024(5). <https://doi.org/10.1002/14651858.cd016070>
 - [29] Annan, R., Qingge, L. (2025). Artificial intelligence in COVID-19 research: A comprehensive survey of innovations, challenges, and future directions. *Computer Science Review*, 57: 100751. <https://doi.org/10.1016/j.cosrev.2025.100751>
 - [30] Natarajan, S., Sampath, P., Arunachalam, R., et al. (2023). Early diagnosis and meta-agnostic model visualization of tuberculosis based on radiography images. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-49195-x>
 - [31] Capellan-Martin, D., Gomez-Valverde, J.J., Bermejo-Pelaez, D., Ledesma-Carbayo, M.J. (2023). A lightweight, rapid and efficient deep convolutional network for chest X-ray tuberculosis detection. In *Proceedings of the IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, Cartagena, Colombia, pp. 1-5. <https://doi.org/10.1109/isbi53787.2023.10230500>

- [32] Zambrano Chaves, J.M., Huang, S.C., Xu, Y., et al. (2025). A clinically accessible small multimodal radiology model and evaluation metric for chest X-ray findings. *Nature Communications*, 16(1): 3108. <https://doi.org/10.1038/s41467-025-58344-x>
- [33] Park, J., Kim, S., Yoon, B., Hyun, J., Choi, K. (2025). M4CXR: Exploring multitask potentials of multimodal large language models for chest X-ray interpretation. *IEEE Transactions on Neural Networks and Learning Systems*, 36(10): 17841-17855. <https://doi.org/10.1109/tnnls.2025.3587687>
- [34] Babarinde, O., Ayo-Farai, O., Maduka, C.P., Okongwu, C.C., Ogundairo, O., Sodamade, O. (2023). Review of AI applications in healthcare: Comparative insights from the USA and Africa. *International Medical Science Research Journal*, 3(3): 92. <https://doi.org/10.51594/imsrj.v3i3.641>
- [35] Togunwa, T.O., Babatunde, A.O., Fatade, O.E., Olatunji, R., Ogbole, G., Falade, A.G. (2025). Detection of pneumonia in children through chest radiographs using artificial intelligence in a low-resource setting: A pilot study. *PLOS Digital Health*, 4(9). <https://doi.org/10.1371/journal.pdig.0000713>
- [36] Worodria, W., Castro, R., Kik, S.V., et al. (2024). An independent, multi-country head-to-head accuracy comparison of automated chest X-ray algorithms for the triage of pulmonary tuberculosis. *medRxiv*. <https://doi.org/10.1101/2024.06.19.24309061>
- [37] Nzimande, N., Murphy, K., Reither, K., et al. (2025). Performance of CAD4TB artificial intelligence technology in TB screening programmes among the adult population in South Africa and Lesotho. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, 40: 100540. <https://doi.org/10.1016/j.jctube.2025.100540>
- [38] Han, Z.L., Zhang, Y.Y., Li, J., et al. (2025). A systematic review and meta-analysis of artificial intelligence software for tuberculosis diagnosis using chest X-ray imaging. *Journal of Thoracic Disease*, 17(5): 3223. <https://doi.org/10.21037/jtd-2025-604>
- [39] Hansun, S., Argha, A., Bakhshayeshi, I., et al. (2025). Diagnostic performance of artificial intelligence-based methods for tuberculosis detection: Systematic review. *Journal of Medical Internet Research*, 27: e69068. <https://doi.org/10.2196/69068>
- [40] Shastry, P., Muthulur, S.C., Kumarasami, N., et al. (2025). Advancing chronic tuberculosis diagnostics using vision-language models: A multimodal framework for precision analysis. *arXiv preprint arXiv:2503.14536*. <https://doi.org/10.48550/arxiv.2503.14536>
- [41] Raval, D., Parmar, D., Saha, S., et al. (2025). Cost-effectiveness analysis of AI-assisted chest X-ray interpretation tools for TB screening: A rapid HTA. *Frontiers in Digital Health*, 7: 1629127. <https://doi.org/10.3389/fdgth.2025.1629127>
- [42] Parihar, D.S., van Vüren, J.J., Niesler, T., et al. (2025). Neural network-based identification of easily-obtainable demographic and clinical characteristics to identify people with tuberculosis. *medRxiv*, 2025-10. <https://doi.org/10.1101/2025.10.23.25338536>
- [43] Tong, R., Liu, J.Q., Wang, T., et al. (2025). Does bigger mean better? Comparative analysis of CNNs and biomedical vision-language models in medical diagnosis. *arXiv* [arXiv:2510.00411](https://arxiv.org/abs/2510.00411)
- <https://doi.org/10.48550/arxiv.2510.00411>
- [44] Cho, Y., Kim, T., Shin, H., Cho, S., Shin, D.M. (2024). Pretraining vision-language model for difference visual question answering in longitudinal chest X-rays. *arXiv preprint arXiv:2402.08966*. <https://doi.org/10.48550/arxiv.2402.08966>
- [45] Catak, F.O., Kuzlu, M., Patrick, T. (2024). Improving medical diagnostics with vision-language models: Convex hull-based uncertainty analysis. *arXiv preprint arXiv:2412.00056*. <https://doi.org/10.48550/arxiv.2412.00056>
- [46] Nawaz, M., Shehab, M., Rana, M.R.R., Qureshi, B., Khan, Z., Babar, M.A. (2025). IMPACT-TB: Integrated medical imaging and AI for precise tuberculosis detection. *Engineering Reports*, 7(8). <https://doi.org/10.1002/eng2.70356>
- [47] Samuel, R.D.J., Kanna, B.R. (2025). Enhancing tuberculosis diagnosis: A deep learning-based framework for accurate detection and quantification of TB bacilli in microscopic images. *Tuberkuloz ve Toraks*, 73(3): 165-177. <https://doi.org/10.5578/tt.2025031113>
- [48] Song, M.Y., Wang, J.R., Yu, Z.H., et al. (2024). PneumoLLM: Harnessing the power of large language model for pneumoconiosis diagnosis. *Medical Image Analysis*, 97: 103248. <https://doi.org/10.1016/j.media.2024.103248>
- [49] Springenberg, M., Frommholz, A., Wenzel, M., Weicken, E., Ma, J., Strodthoff, N. (2023). From modern CNNs to vision transformers: Assessing the performance, robustness, and classification strategies of deep learning models in histopathology. *Medical Image Analysis*, 87: 102809. <https://doi.org/10.1016/j.media.2023.102809>
- [50] Xu, Y.W., Khan, T., Song, Y., Meijering, E. (2025). Edge deep learning in computer vision and medical diagnostics: A comprehensive survey. *Artificial Intelligence Review*, 58(3): 93. <https://doi.org/10.1007/s10462-024-11033-5>
- [51] Kaczmarczyk, R., Wilhelm, T.I., Martin, R., Roos, J. (2024). Evaluating multimodal AI in medical diagnostics. *npj Digital Medicine*, 7(1). <https://doi.org/10.1038/s41746-024-01208-3>
- [52] Li, S.R., Wong, K.W., Wang, G.J., Duong, T.T. (2025). A systematic review of multimodal large language models on domain-specific applications. *Artificial Intelligence Review*, 58(12): 383. <https://doi.org/10.1007/s10462-025-11398-1>
- [53] Li, L.S., Yang, L., Li, Z., Ye, Z.Y., Zhao, W.G., Gong, W.P. (2023). From immunology to artificial intelligence: Revolutionizing latent tuberculosis infection diagnosis with machine learning. *Military Medical Research*, 10(1): 58. <https://doi.org/10.1186/s40779-023-00490-8>
- [54] Fahrner, L.J., Chen, E., Topol, E.J., Rajpurkar, P. (2025). The generative era of medical AI. *Cell*, 188(14): 3648. [https://www.cell.com/cell/abstract/S0092-8674\(25\)00568-9](https://www.cell.com/cell/abstract/S0092-8674(25)00568-9)
- [55] Euldji, R., Batel, N., Rebhi, R., et al. (2022). Optimal design and performance comparison of a combined ANFIS-PID with back stepping technique, using various meta-heuristic algorithms to solve wheeled mobile robot trajectory tracking problem. *Journal Européen des Systemes Automatisés*, 55(3): 281-298. <https://doi.org/10.18280/jesa.550301>