

Original Research Article

A retrospective multi-source clinical validation of Lenek intelligent radiology assistant: an artificial intelligence-based chest radiograph screening and triage system for high-burden pulmonary and cardiac conditions in India

Vinit Singh¹, Ananya Jhamb^{1*}, Shiladitya Sil², Sidesh Kumar³, Chirag Agrawal³, Anant Pareek³, Aditya Gautam⁴, Gurunath Parale⁵, Shuchi Singh⁶, Deepak Padmanabhan⁷

¹Jawaharlal Nehru Medical College, KLE Academy of Higher Education and Research, Belagavi, Karnataka, India

²Department of Oral Medicine and Radiology, North Bengal Dental College and Hospital, Thiknikata, West Bengal, India

³Lenek Technologies Private Limited, Pune, Maharashtra, India

⁴Department of Respiratory Medicine, Uttar Pradesh University of Medical Sciences, Saifai, Uttar Pradesh, India

⁵Department of Cardiology, Ashwini Hospital, Solapur, Maharashtra, India

⁶Department of Radiodiagnosis, Apollomedics Super Speciality Hospital, Lucknow, Uttar Pradesh, India

⁷Department of Electrophysiology, Narayana Institute of Cardiac Sciences, Bengaluru, Karnataka, India

Received: 14 March 2026

Revised: 20 April 2026

Accepted: 24 April 2026

*Correspondence:

Ananya Jhamb,

E-mail: ananyajhamb@lenektech.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: A critical radiologist shortage exists in India, leading to delayed chest radiograph (CXR) interpretation. This leads to disease progression, higher morbidity, and mortality. Artificial intelligence-based CXR interpretation by Lenek Intelligent Radiology Assistant (LIRA) is a promising solution. This study aims to establish the screening and triaging capabilities of LIRA by assessing its accuracy in detecting abnormalities and pathologies in CXRs from geographically diverse institutions.

Methods: We conducted a retrospective multi-source validation of the diagnostic accuracy of LIRA for the detection of general abnormalities, tuberculosis, consolidation, pleural effusion, pneumothorax, and cardiomegaly. De-identified chest radiographs were input into LIRA models. The obtained interpretations were compared to the established ground truth reporting for the calculation of sensitivity, specificity, and AUROC with 95% CI for individual pathologies across varying probability thresholds.

Results: LIRA demonstrated high sensitivity for general abnormality detection (AUROC 0.93-0.986, 84.4-97.1% sensitivity, 88.9-92.4% specificity) and tuberculosis triaging (Shenzhen and Montgomery: 88.5-89.7% sensitivity, 89.9-90.5% specificity; Jaypee: 98.7% sensitivity, 63.6% specificity). For consolidation (AUROC 0.884-0.895, 96.4-96.9% sensitivity, 70.8-77.1% specificity), pleural effusion (AUROC 0.942-0.967, 79.7-99.1% sensitivity, 81.2-87.7% specificity), pneumothorax (AUROC 0.87, 90.6-94.8% sensitivity, 79.5-82.7% specificity) and cardiomegaly (AUROC 0.883, 95.1% sensitivity, 81.6% specificity), the model exhibited commendable accuracy as well.

Conclusions: The diagnostic performance of LIRA was consistent across various pathologies and chest radiographs from diverse geographic locations, with particular strengths in abnormality detection and tuberculosis screening. The risk-stratified triaging and high sensitivity of LIRA make it a reliable adjunct solution to address radiologist shortages, reduce turnaround times, and support India's tuberculosis elimination goals.

Keywords: Artificial intelligence in radiology, Chest-radiography, Tuberculosis screening, AI-driven triage, Access to healthcare, Screening tool

INTRODUCTION

Chest radiography is the most widely performed imaging investigation for cardiopulmonary diseases worldwide. Chest X-Rays (CXR) serve as the first-line diagnostic modality for tuberculosis, pneumonia, pleural diseases, pneumothorax, and cardiomegaly. Driven by the country's substantial burden of tuberculosis, chronic respiratory diseases, and cardiovascular conditions, India performs millions of chest radiographs annually. However, due to radiologist shortages, this burden of CXRs does not meet the timely and quality interpretation.¹ The Indian healthcare infrastructure today caters to a population exceeding 1.4 billion. India experiences a severe diagnostic bottleneck with an estimated radiologist-to-population ratio of 1:100,000, which adversely affects patient outcomes.² This problem further exaggerates in tier-2, tier-3, and rural healthcare settings, where there is a larger shortage of radiologists, and patients often travel long distances to reach a tertiary care center, resulting in delayed diagnostics.³

India contributes nearly one-quarter of the global tuberculosis burden with approximately 2.6 million cases annually and has set an ambitious target to eliminate tuberculosis by 2030, five years ahead of the global Sustainable Development Goal timeline.^{4,5} Delayed radiological interpretation of conditions like tuberculosis, consolidation, and pneumothorax leads to disease progression, increased morbidity, higher mortality rates, and continued disease transmission in communities.^{6,7} Deep learning-based image analysis by Artificial Intelligence (AI) is a promising solution to address these challenges.

These AI algorithms can rapidly analyze chest radiographs and flag abnormal studies, therefore assisting clinicians in prioritizing cases that require urgent attention.^{8,9} AI tools have demonstrated high diagnostic accuracy; however, many remain limited to single-disease detection or lack validation across heterogeneous datasets that are representative of real-world clinical practice.^{10,11} LIRA (Lenek Intelligent Radiology Assistant) (Lenek Technologies Private Limited, Pune, India) is an AI-enabled radiological computer-assisted triage, interpretation, and reporting software application.

It is designed to assist qualified healthcare professionals in the review and analysis of adult CXR scans (AP/PA views) for the presence of abnormalities, multiple high-impact thoracic and cardiac pathologies, including tuberculosis, pulmonary consolidation, pleural effusion, pneumothorax, and cardiomegaly. LIRA also has a risk-stratified output for tuberculosis suspicion, where the predictions are scored by a quantitative "Tuberculosis Score" (0-100%) and categorized by a color-coded scoring system into red (high likelihood), yellow (indeterminate), and green (low likelihood) zones, thus enabling rational triage and referral decisions in resource-limited settings. This is a multi-source retrospective study aimed at the clinical validation

of the screening and triaging capabilities of LIRA across diverse publicly available CXR datasets from international institutions (National Institutes of Health (NIH); Department of Health and Human Services, Montgomery County; Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen; CheXpert, Stanford Hospital; Society for Imaging Informatics in Medicine, American College of Radiology (SIIM-ACR)) and Indian institutions (Jaypee University of Information Technology) to assess the diagnostic accuracy, generalizability, and clinical feasibility of LIRA-CXR interpretation and reporting as a solution to bridge the radiology reporting gap and support tuberculosis elimination efforts in India and other high-burden, low-resource environments.

METHODS

Study design and ethical considerations

This was a multi-source retrospective clinical validation study aimed to evaluate the diagnostic accuracy of LIRA in screening and triaging of CXRs for general abnormalities, tuberculosis, consolidation, pleural effusion, pneumothorax, and cardiomegaly. The study protocol was designed in adherence to the Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015 guidelines. Ethical approval was not required for this study as it involved the retrospective analysis of de-identified, publicly available CXR datasets (NIH, CheXpert, Montgomery, Shenzhen, SIIM-ACR, and Jaypee University of Information Technology).

No human subjects were directly recruited, and no identifiable private information was accessed. The study was conducted in compliance with the Health Insurance Portability and Accountability Act (HIPAA) and the Indian Council of Medical Research (ICMR) National Ethical Guidelines for Biomedical and Health Research.

Study period and study site

This study was conducted between February 2026 and March 2026. It utilized publicly available, open-access datasets and did not involve data collection from any specific hospital or clinical setting. Therefore, no single geographic location is applicable.

Study protocol

The study used de-identified CXRs obtained from geographically diverse publicly available CXRs from international institutions, which were input into the LIRA Models for interpretation.

The obtained LIRA interpretation of the CXRs was then compared to the established ground truth. The ground truth (reference standard) was established by consensus labelling of CXRs by board-certified radiologists. Based on this comparison, the performance metrics were calculated to establish the diagnostic accuracy of LIRA.

Eligibility criteria

This study included frontal CXR's (Posteroanterior (PA) or Anteroposterior (AP) views) from adult patients (>18 years). Images of the pediatric population, non-CXR modalities (CT, MRI), and images with significant artifacts or metallic implants obscuring the lung fields were excluded from the study.

Data acquisition

The clinical validation was performed on five distinct CXR datasets from global benchmark institutes to establish the diagnostic accuracy of the LIRA across geographically diverse patient populations and imaging equipment.

National institutes of health (NIH, USA): We used a subset of the NIH ChestX-ray8 dataset for general abnormality detection, comprising 1,344 chest radiographs (653 abnormal, 691 normal) with ground truth established by consensus of three American Board of Radiology-certified radiologists.¹² Another subset of 810 chest radiographs labelled by five board-certified radiologists was used for specific pathologies (83 consolidation, 226 pleural effusion, 136 pneumothorax, 81 cardiomegaly).

CheXpert (USA): It is a large public dataset consisting of 224,316 chest radiographs of 65,240 patients collected from Stanford Hospital between 2002 and 2017.¹³ A subset consisting of 234 studies was labelled by radiologists, among which 202 of these images are of AP/PA view, which were used for testing here (32 consolidation, 64 pleural effusion).

Montgomery county (USA) and Shenzhen (China): These datasets were created by the National Library of Medicine in collaboration with the Department of Health and Human Services, Montgomery County, Maryland, USA and Shenzhen No. 3 People's Hospital, Guangdong Medical College, Shenzhen, China.¹⁴ These served as the primary benchmarks for Tuberculosis (TB) detection in our study. The Shenzhen dataset consisted of 662 images: 269 of confirmed TB positive cases, and 326 of confirmed non-TB. The Montgomery dataset consisted of 137 images: 58 of confirmed TB positive, and 79 of confirmed TB negative.

Society for imaging informatics in medicine - American college of radiology (SIIM-ACR): A large-scale open-source dataset comprising 3,286 pneumothorax-positive and 8,296 pneumothorax-negative images was utilized for pneumothorax validation.¹⁵

Jaypee university of information technology (India): A publicly available dataset from Jaypee University of Information Technology with 154 images (77 confirmed TB positive, 77 confirmed TB negative) was used to assess the TB model's sensitivity in the Indian demographic.¹⁶

Lenek intelligent radiology assistant (LIRA) overview

LIRA (Lenek Intelligent Radiology Assistant) v1.0 is an AI-enabled radiological computer-assisted triage, interpretation, and reporting software as a Medical Device (SaMD). It is classified as a Class B Medical Device under India's Medical Device Rules (MDR) 2017. It is designed to assist qualified healthcare professionals in the review and analysis of adult chest X-ray (CXR) scans (AP/PA views). LIRA is intended for use by trained Radiologists, Referring Physicians, and Radiographers as an assistive tool. It aids in the triage and prioritization of CXR studies by identifying and highlighting potentially critical findings, thereby reducing reporting turnaround times in teleradiology and hospital Picture Archiving and Communication System (PACS) environments, thereby streamlining radiological workflows. The software analyzes CXR studies to identify the presence of general abnormalities, calculate a Tuberculosis (TB) probability score (0-100%) for screening purposes, and detect specific critical findings, including Consolidation, Pneumothorax, Pleural Effusion, and Cardiomegaly. It also produces color heatmap overlays that highlight image regions potentially relevant to the identified findings.

Abnormality detection model

The Abnormality detection model reports under a binary classification of "Abnormal/Yes" or "Normal/No". "Abnormal/Yes" indicates that the model has detected features that deviate from what it considers normal. This suggests that the study may warrant further review. "Normal/No" means that the AI model did not detect features it considers abnormal.

Tuberculosis (TB) triage model

For every radiograph that LIRA screens, the model assesses specific 'Tuberculosis tags' (radiological indicators of active or prior infection) such as cavitation, infiltration, nodules, and pleural effusion to derive a quantitative TB probability score between 0-100% to assess the probability of a patient harbouring TB (Figure 1 A). Based on the quantitative score, LIRA also gives a risk-stratified output for TB suspicion using a color-coded zone system, providing an immediate visual assessment of TB suspicion (Figure 1B).

Interpretation of color coding by LIRA

The Green Zone (Score 0-X%) indicates a low level of suspicion for TB. This means that, according to the AI analysis, cases in this category are unlikely to be indicative of Tuberculosis. The Red Zone (Score Y-100%) signifies a high level of suspicion for TB. This means that the AI model has identified features consistent with the disease. The Yellow Zone (Score X+1 to Y-1%) represents an intermediate level of suspicion. Interpretation requires further consideration in conjunction with the output from the "Abnormality Detection Model", which is a crucial

determinant in refining the suspicion level for Tuberculosis when the TB model yields a Yellow Zone (Figure 2 A).

Pathology detection model for other pathologies

LIRA also includes models for detecting specific pathologies such as Consolidation, Pleural Effusion, Pneumothorax, and Cardiomegaly. The system uses a binary classification model ("Yes/Present" or "No/Not Present") accompanied by heatmap overlays to help radiologists in localizing the pathology. A heatmap is an overlay on the medical image that visually highlights the specific regions the AI model identified as most indicative of the detected pathology (Figure 2 B). These models are designed as assistive tools to provide supplementary information to radiologists.

Reference standard and ground truth establishment

The reference standard was strictly defined for each dataset to ensure reliability. For the NIH dataset, labels were determined by the consensus of three U.S. board-certified radiologists. For the CheXpert validation set, labels were provided by board-certified radiologists using the Fleischner Society definitions. For the Shenzhen, Montgomery, and Jaypee datasets, labels were based on microbiological confirmation and expert clinical diagnosis as provided in the source dataset documentation.

Statistical analysis and performance metrics

Diagnostic accuracy of LIRA was assessed by calculating Sensitivity, Specificity, and the Area Under the Receiver Operating Characteristic (AUROC), with 95% confidence intervals (CI) calculated using De-Long's method for AUROC and Wilson score method for sensitivity and specificity. For the TB triage model, sensitivity was defined as the proportion of true positive cases classified correctly into "Red" or "Yellow" categories. Specificity was defined as the proportion of true negatives classified correctly into the "Green" category. These metrics were calculated across multiple specified probability thresholds (e.g., 0.4, 0.5, 0.6) to identify the optimal operating point for a triage setting with high sensitivity to minimize missed diagnoses.

RESULTS

Results of abnormality detection model

The abnormality detection model achieved excellent diagnostic performance across the NIH and the Montgomery datasets. On the NIH dataset (1,344 images), AUROC was 0.986 (95% CI: 0.98-0.992). At the threshold of 0.5, the model exhibited 97.1% sensitivity (CI: 96.2-98.0%) and 88.9% specificity (CI: 87.2-90.5%). Higher thresholds (0.6) improved specificity to 93.6% while maintaining sensitivity above 95%. In the Montgomery dataset (137 images), the calculated AUROC was 0.93 (CI:

0.88-0.971). At threshold 0.5, sensitivity was 84.4% (CI: 78.4-90.5%) with a specificity of 92.4% (CI: 86.5-96.0%) (Figure 3).

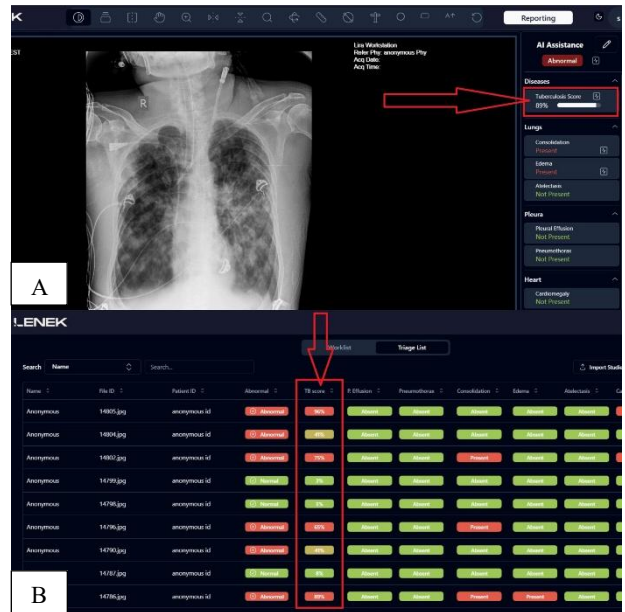


Figure 1: (A) the TB score generated by LIRA for abnormal CXRs with probable TB and (B) LIRA interface and the unique three-zone triage framework for TB screening.

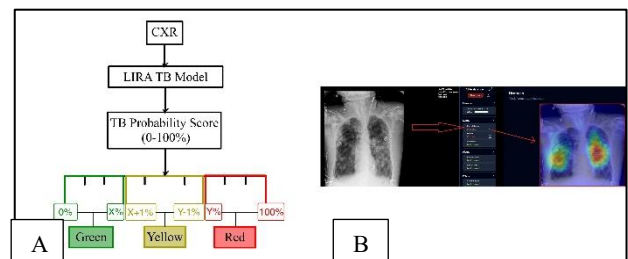


Figure 2: (A) color coding by LIRA for tuberculosis triaging and (B) a positive case of lung consolidation in a chest X-ray (PA view).

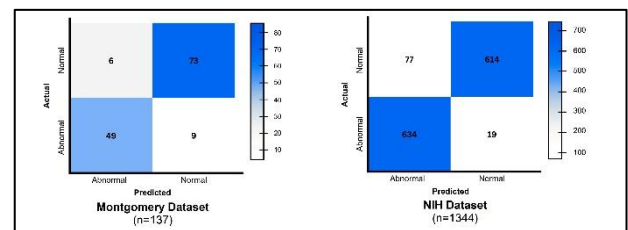


Figure 3: Confusion matrix for abnormality detection model.

Results of tuberculosis triage model

The TB model demonstrated consistently high sensitivity across three geographically diverse datasets using the three-zone triage classification system. In the Shenzhen

Dataset (595 images), LIRA correctly identified 238 of 269 TB-positive cases (red: 216, yellow: 22) and achieved a sensitivity of 88.5%. Specificity reached 90.5% with 295 of 326 TB-negative cases appropriately classified into the “Green” category. In the Montgomery Dataset (137 images), the model correctly classified 52 of 58 TB-positive cases (red: 50, yellow: 2) with a sensitivity of 89.65%. Calculated specificity was 89.9% (correctly identified 71 of 79 TB-negative cases as green), demonstrating consistent performance despite a smaller sample size. The model demonstrated the highest sensitivity (98.7%, correctly identified 76 of 77 TB-positive cases) in the Jaypee Dataset (154 images), though specificity was 63.6% (49 of 77 TB-negative cases classified “green”), possibly reflecting different disease severity or imaging characteristics in this Indian population (Figure 4).

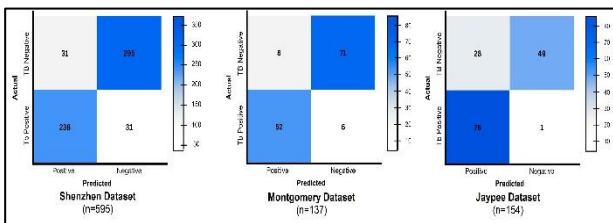


Figure 4: Confusion matrices for tuberculosis triage model.

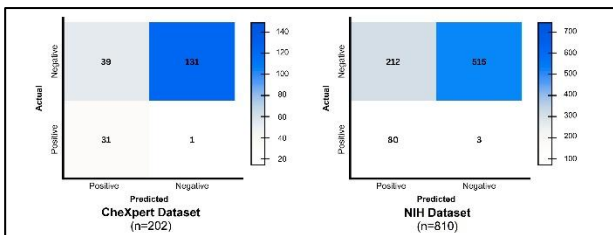


Figure 5: Confusion matrices for consolidation detection model.

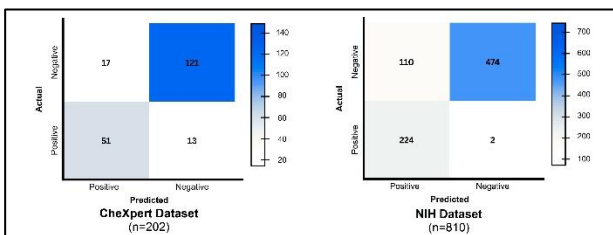


Figure 6: Confusion matrices for pleural effusion detection model.

Results of consolidation detection model

In the CheXpert Dataset (202 images), the calculated AUROC was 0.894 (CI:0.85-0.936). Sensitivity reached 96.9% (CI:94.5-99.3%) with a specificity of 77.1% (CI:71.2-82.8%) at the threshold of 0.55. In the NIH Dataset (810 images), the model achieved an AUROC of

0.885 (CI:0.863-0.907), with a sensitivity of 96.4% (CI:95.1-97.7%) and a specificity of 70.8% (CI:67.7-74.0%) at a threshold of 0.55 (Figure 5).

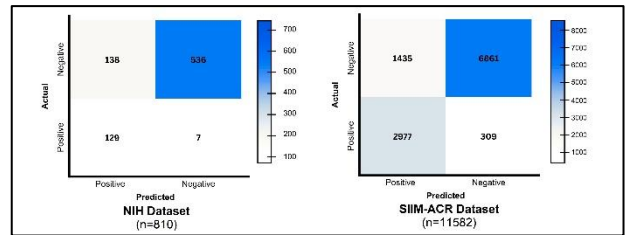


Figure 7: Confusion matrices for pneumothorax detection model.

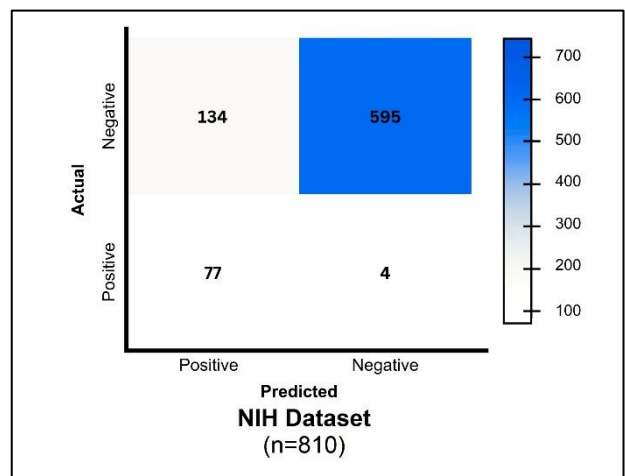


Figure 8: Confusion matrix for cardiomegaly detection model.

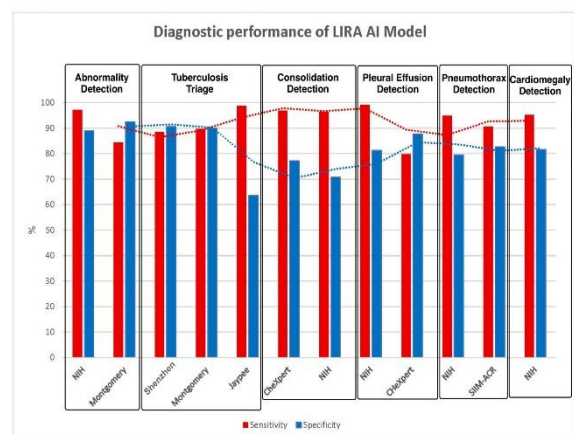


Figure 9: Sensitivity and specificity of LIRA AI model across chest X-ray datasets.

Results of pleural effusion detection model

In the NIH Dataset (810 images), LIRA achieved a high AUROC of 0.967 (CI:0.95-0.98), with a sensitivity of 99.1% (CI:98.47-99.76%) and a specificity of 81.2% (CI:78.5-83.8%) at a threshold of 0.45. In the CheXpert

Dataset (202 images), AUROC was 0.942 (CI:0.91-0.974), with a sensitivity of 79.7% (CI:74.1-85.2%) and a specificity of 87.7% (CI:83.1-92.2%) at a threshold of 0.45 (Figure 6).

Results of pneumothorax detection model

In the NIH Dataset (810 images), an AUROC of 0.872 (CI:0.849-0.895), sensitivity of 94.8% (CI:93.3-96.4%), and specificity of 79.5% (CI:76.7-82.3%) were achieved at a threshold of 0.38.

In the SIIM-ACR Dataset (11,582 images), an AUROC of 0.867 (CI: 0.86-0.873), sensitivity of 90.6% (CI: 90.13-91.2%), and specificity of 82.7% (CI: 82.0-83.38%) were observed at a 0.38 threshold. The large sample size provided narrow confidence intervals, demonstrating stable performance of the model (Figure 7).

Results of cardiomegaly detection model

In the NIH Dataset (810 images), AUROC=0.883 (CI:0.861-0.905), sensitivity of 95.1% (CI:93.5-96.55%), and specificity of 81.6% (CI:78.95-84.28%) were achieved at a threshold of 0.39 (Figure 8).

Summary of findings

A comprehensive summary of the diagnostic performance of LIRA across all studied pathologies and datasets is presented below. These metrics highlight the model's consistent sensitivity and specificity across diverse geographic populations and imaging equipment, supporting its utility in high-burden settings (Figure 9, Table 1).

Table 1: Diagnostic performance of LIRA AI model across chest X-ray datasets.

Sr no	Parameter	Dataset	Sample size (positive/negative)	AUROC (95% CI)	Optimum threshold	Sensitivity (%)	Specificity (%)	Intended clinical role
1	Abnormality detection	NIH (USA)	1344	0.986 (0.98-0.992)	0.5	97.1	88.9	Rapid screening and workflow reduction
2	Abnormality detection	Montgomery (USA)	137 (58/ 79)	0.93 (0.88-0.971)	0.5	84.4	92.4	Primary triage
3	Tuberculosis	Shenzhen (China)	595 (269/ 326)	NA	0.2,0.6	88.5	96.9	TB screening
4	Tuberculosis	Montgomery (USA)	137 (58/ 79)	NA	0.2,0.6	89.65	98.7	TB screening
5	Tuberculosis	Jaypee (India)	154 (77/ 77)	NA	0.2,0.6	98.7	85.7	TB screening
6	Consolidation	CheXpert (USA)	202 (32/ 170)	0.894 (0.85-0.936)	0.55	96.9	77.1	Pneumonia triage
7	Consolidation	NIH (USA)	810 (83/ 727)	0.885 (0.86- 0.91)	0.55	96.4	70.8	ED prioritization
8	Pleural effusion	NIH (USA)	810 (226/ 584)	0.967 (0.95- 0.98)	0.45	99.1	81.2	Urgent case flagging
9	Pleural effusion	CheXpert (USA)	202 (64/ 138)	0.942 (0.91-0.974)	0.45	79.7	87.7	In-patient triage
10	Pneumothorax	NIH (USA)	810 (136/674)	0.872 (0.8490.89)	0.38	94.8	79.5	Emergency detection
11	Pneumothorax	SIIM-ACR	11582 (3286/8296)	0.867 (0.86-0.873)	0.38	90.6	82.7	Emergency triage
12	Cardiomegaly	NIH (USA)	810 (81/729)	0.883 (0.8610.90)	0.39	95.1	81.6	Chronic disease flag

DISCUSSION

This multi-source clinical validation study shows that LIRA achieves comparable diagnostic performance across heterogeneous datasets and imaging environments.

This supports its intended role as a clinical screening and triage tool in resource-limited healthcare settings. These findings have important implications for catering to systemic gaps in India's healthcare delivery systems and for supporting national tuberculosis (TB) elimination efforts.

Addressing healthcare workforce shortages and diagnostic delays

India faces a critical shortage of radiologists, resulting in substantial diagnostic delays. Such delays adversely affect patient outcomes. With approximately one radiologist per 100,000 population, compared to 10 to 15 per 100,000 in high-income countries, the imbalance between imaging acquisition and interpretation capacity continues to widen.^{1,2} This deficit translates into prolonged turnaround times in terms of reporting, often ranging from several

days in district hospitals to weeks in primary health centers, during which treatable conditions may worsen, and infectious diseases may continue to spread.³ By cutting reporting delays from days to minutes, LIRA's capacity to provide near-real-time preliminary interpretations radically alters this paradigm. When deployed as a first-line screening tool, LIRA can prioritize studies with critical findings for immediate radiologist review while also clearing likely to be normal examinations.

This focused allocation of scarce specialized resources allows radiologists to completely focus on complex cases, accelerate triage decisions in emergencies, and minimize delays in outpatient follow-up.^{8,17} The economic impact of such efficiency gains is substantial. Diagnostic delays lead to higher healthcare costs, prolonged hospital stays, advanced disease management, loss of productivity and livelihood, and increased transmission of communicable diseases. By facilitating earlier diagnosis and timely intervention, AI-assisted screening redirects healthcare expenditure toward prevention and early treatment. Such approaches are more cost-effective than managing advanced disease.^{18,19}

Tuberculosis screening, triage, and elimination strategy

Tuberculosis continues to remain one of India's most challenging public health diseases. Despite ongoing efforts under the National Tuberculosis Elimination Program (NTEP), India continues to contribute approximately 27-28% of global TB cases, with an estimated incidence of 2.6 million and a mortality of nearly 450,000 deaths annually.^{4,5} Delayed diagnosis, under-reporting, and failure to follow-up remain the primary factors driving high incidence and mortality rates. Amidst all this, LIRA's TB detection model continues to demonstrate consistently high sensitivity across diverse and inclusive datasets, making it well-suited for deployment in community screening initiatives and primary health centers where specialist expertise is limited.

The three-zone classification framework provides a pragmatic approach wherein: redzone cases indicate rapid referral for confirmatory molecular testing (CBNAAT/TrueNat), yellow-zone cases undergo further clinical assessment or repeat imaging, and green-zone cases are reassured. This risk-based triage strategy is designed to overcome the limitations of current TB screening methods. It was noted that while symptom-based screening lacks sensitivity for early or subclinical illness, universal chest radiography without triage remains resource-intensive.

LIRA addresses this gap by identifying and prioritizing radiographs most likely to show active disease, enabling targeted molecular testing. In doing so, it reduces workload and turnaround time for radiologists and clinicians, especially in emergency and resource-limited settings. This significantly enhances the efficiency and cost-effectiveness of existing screening workflows.^{20,21}

Scalable deployment of LIRA may prove beneficial for achieving the stage of TB elimination across India's diverse and heterogeneous healthcare system.

Its minimal infrastructure requirements, being basic computing hardware and good internet connectivity, further align with India's expanding mobile health and telemedicine initiatives. This ensures smooth deployment of the model in remote areas and supports easy scalability across the country. Moreover, integration with national TB information systems and databases could facilitate real-time surveillance, identification of high-risk groups, and continuous monitoring of screening outcomes.²² Additionally, LIRA supports contact tracing and screening in crowded settings such as households, schools, prisons, army camps, and medical facilities. Rapid AI-assisted screening of these high-risk groups can help detect secondary cases earlier and interrupt transmission, while the three-zone design guarantees the efficient use of confirmatory testing resources.²³

Risk stratification and rapid referral pathways

LIRA demonstrates high sensitivity for acute and potentially life-threatening conditions, which enables effective risk stratification and rapid referral. Its strong performance in detecting pleural effusion (99.1% sensitivity on NIH data) and pneumothorax (94.8% sensitivity) allows it to serve as an early warning system, prompting alerts for immediate therapeutic intervention. In core medical emergencies such as massive pleural effusion or tension pneumothorax, rapid identification is critical, and AI-assisted triage makes sure that critical findings like these are not delayed by routine reporting protocols.⁸ With respect to cardiomegaly, early detection leads to timely referral to cardiology and initiation of heart failure management protocols. With cardiovascular disease prevalence on the rise in India due to aging, hypertension, diabetes, and rheumatic heart disease, early intervention can significantly lower hospitalizations and mortality rates.²⁴ LIRA's capacity to detect cardiomegaly supports screening in high-risk groups and facilitates longitudinal monitoring. The horizontal deployment of a single AI system capable of detecting multiple pathologies offers clear operational advantages over disease-specific tools. This optimizes the radiology worklist and reduces turnaround times for several clinical scenarios at once, rather than focusing on a single disease. Therefore, implementing one integrated solution simplifies training, workflow integration, and quality assurance while providing broad screening coverage across a multiplicity of clinical scenarios.⁸

Generalizability and real-world considerations

As evidenced by its consistent performance across geographically diverse clinical datasets (NIH, Montgomery, CheXpert, Shenzhen, SIIM-ACR, Jaypee), LIRA demonstrates generalizability across diverse populations, imaging equipment, and treatment settings.

This breadth of validation addresses critical ethical concerns regarding dataset diversity and algorithmic bias. These discrepancies likely represent intrinsic differences between carefully curated research datasets and real-world clinical practices. Research datasets usually comprise high-quality images obtained under standardized protocols and interpreted by expert radiologists, whereas real-world data includes portable radiographs, variable positioning, motion artifacts, and reports generated under time constraints.²⁵ The comparatively lower specificity of AI reporting may also be shaped by multiple contextual factors, such as higher disease prevalence, local reporting thresholds, widespread use of portable imaging, and population-specific characteristics such as prior tuberculosis, chronic lung disease, and malnutrition, resulting in radiographic variations. Despite this, LIRA maintained high sensitivity on Indian datasets, which aligns with its intended primary use as a screening and triage tool. In public health screening, prioritizing sensitivity is crucial. False negatives carry greater consequences than false positives, as missed cases may go unchecked and contribute to disease progression and transmission. Whereas false positives can be resolved through subsequent confirmatory testing.¹⁷ Further optimization of real-world performance may be achieved through ongoing refinement using more diverse Indian datasets, especially from district hospitals and primary health centers.

Integration into clinical workflow and human-AI cooperation

Clinical impact cannot be guaranteed by technical validation alone. Successful implementation and adoption require seamless workflow integration, clinician acceptance, and clearly defined human-AI collaboration models.²⁶ LIRA's rapid and robust processing enables deployment at multiple points in the care delivery system, including primary health center screening, emergency department triage, outpatient pre-screening, and radiologist work list prioritization. Each application necessitates tailored workflows, training, and quality assurance protocols.¹⁷ Positioning AI as clinical decision support rather than autonomous diagnosis is crucial for ethical application and professional acceptance. Diagnostic autonomy and responsibility still remain with clinicians, while AI only contributes to speed, consistency, and prioritization.²⁷ To further enhance clinician trust and usability, transparency must be maintained. This can be achieved through building more explainable AI models through objective methods like visual explanations, confidence scores, and feedback mechanisms.²⁸

Screening of abnormalities and triaging of pathologies

In resource-limited settings like India, high-burden pulmonary and cardiac conditions such as tuberculosis, pneumonia, and interstitial lung diseases impose significant financial and logistical challenges on patients. The deployment of AI-based chest radiograph (CXR)

triage systems at the primary care level offers a transformative approach to early screening and triaging protocols. This can facilitate rapid identification and prioritization of abnormalities in the general population during community outreach, thereby enabling efficient triaging of pathologies in emergency and hospital scenarios.

These systems have the potential to guide the patients towards the most appropriate diagnostic and treatment pathways at the outset.^{19,29} Such early interventions not only minimize unnecessary expenditures on advanced imaging or referrals but also reduce the burden of travel and alleviate the loss of income due to redundant hospital visits. This particularly helps the rural or low-income populations who may otherwise endure prolonged waits and visits to tertiary centers. LIRA's high sensitivity and specificity could support scalable implementation to alleviate overall healthcare system pressures and improve equitable access to timely care.

Limitations and future directions

The applicability of this study is limited by its retrospective design; prospective trials are necessary to evaluate real-world impact on turnaround times, referral patterns, treatment initiation, and patient outcomes. Strong proof of therapeutic benefit and cost-effectiveness would come from randomized studies comparing AI-augmented and traditional workflows.^{25,30} Future research should also focus more on implementation challenges, economic analyses, and clinician acceptance. Continuous model training and refinement using diverse Indian data and research into optimal human-AI collaboration will be critical for safe, effective, and scalable deployment.

CONCLUSION

LIRA exhibits multi-label performance across various datasets with consistently high sensitivities. This data strongly supports our model's intended role as an AI-based chest radiograph screening and triage tool. This clinical validation study highlights the potential of deep-learning-based AI models to address critical diagnostic gaps in India. LIRA can potentially reduce diagnostic delays, enable early detection of preventable diseases by facilitating rapid, reliable screening, risk stratification, and prioritization of high-risk cases, and encourage timely referrals to appropriate levels of care within the healthcare delivery system. For tuberculosis screening, its high sensitivity and three-zone risk-based framework make it suitable for community outreach and primary healthcare centers. This aspect directly aligns with the national TB elimination priorities and strategies while simultaneously optimizing resource utilization.

However, retrospective validation alone is not sufficient for adequate adoption. Future prospective clinical studies are required to establish the real-world impact of LIRA on patient outcomes, workflow integration, and cost-

effectiveness of this intervention. Successful implementation will depend on appropriate clinician training, workflow integration tailored to specific needs, steady monitoring of the performance delivered, and regulatory oversight. Moreover, to address the observed specificity variations in this study, continuous refinement of the model using diverse Indian datasets will be required. Lastly, with suitable governance and human-AI collaboration, systems such as LIRA have the potential to expand screening and diagnostic capacity, address the burden of healthcare disparities, and meaningfully contribute to India's public health mission.

ACKNOWLEDGEMENTS

The authors express sincere gratitude to the developers of the public and open-access datasets used in this study, also acknowledging the work of the National Institutes of Health (NIH) for the Chest X-ray dataset, the Stanford CheXpert team, the Society for Imaging Informatics in Medicine - American College of Radiology (SIIM-ACR) Pneumothorax dataset collaborators, Jaypee University of Information Technology, and the teams behind the Montgomery County and Shenzhen datasets from the U.S. National Library of Medicine. These openly available resources were instrumental in validating the AI model. The authors also acknowledge the contributions of the Lenek Technologies team, including AI developers, engineers, and data scientists who supported model development, deployment, and technical aspects of the validation process. The datasets used in this study are publicly available from the repositories cited.

Funding: No funding sources

Conflict of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: Financial Relationships: Sidesh Kumar; Chirag Agrawal declare(s) employment and stock/stock options from Lenek Technologies Private Limited, Co-founders and shareholders of Lenek Technologies Private Limited. Shiladitya Sil declare(s) personal fees from Lenek Technologies Private Limited, Clinical Implementation Consultant at Lenek Technologies Private Limited. Vinit Singh; Ananya Jhamb declare(s) employment from Lenek Technologies Private Limited, Medical Research Associates at Lenek Technologies Private Limited. Gurunath Parale; Deepak Padmanabhan declare(s) non-financial support from Lenek Technologies Private Limited, Official Clinical Advisor at Lenek Technologies Private Limited. Aditya Gautam; Shuchi Singh declare(s) nonfinancial support from Lenek Technologies Private Limited, Clinical Collaborators with Lenek Technologies Private Limited. Anant Pareek declare(s) non-financial support from Lenek Technologies Private Limited, Research Intern at Lenek Technologies Private Limited. Intellectual Property Information: Lenek Technologies Private Limited has filed for copyright registration for the proprietary software and algorithms associated with the Lenek Intelligent Radiology Assistant (LIRA) referenced in this work. Authors Sidesh Kumar and Chirag Agrawal are

co-founders and shareholders of Lenek Technologies Private Limited, the entity that owns the relevant intellectual property. Other relationships: All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required

REFERENCES

1. Pande T, Pai M, Khan FA, Denkinger CM. Use of chest radiography in the 22 highest tuberculosis burden countries. *Eur Respir J.* 2015;46(6):1816-9.
2. Arora R. The training and practice of radiology in India: current trends. *Quant Imaging Med Surg.* 2014;4(6):449-50.
3. Sreeramareddy CT, Qin ZZ, Satyanarayana S, Subbaraman R, Pai M. Delays in diagnosis and treatment of pulmonary tuberculosis in India: a systematic review. *Int J Tuberc Lung Dis.* 2014;18(3):255-66.
4. WHO. Global Tuberculosis Report 2025. Geneva: World Health Organization. 2025. Available at: <https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/globaltuberculosis-report-2025>. Accessed on 14 February 2026.
5. Central TB Division. Ministry of Health and Family Welfare, Government of India; India TB Report 2025. 2025. Available at: https://tbcindia.mohfw.gov.in/wpcontent/uploads/2024/10/TBReport_for-Web_08_10-2024-1.pdf. Accessed on 14 February 2026.
6. Balasubramnian A, Francis PT, Leelamoni K, Rakesh PS, Lalu JS. Diagnostic and treatment delay among new pulmonary tuberculosis patients in Southern India: A cross-sectional study. *Indian J Public Health.* 2022;66(1):S60-5.
7. McIntosh N, Becher JC, Cunningham S, Stenson B, Laing IA, Lyon AJ, et al. Clinical diagnosis of pneumothorax is late: use of trend data and decision support might allow preclinical detection. *Pediatr Res.* 2000;48(3):408-15.
8. Annarumma M, Withey SJ, Bakewell RJ, Pesce E, Goh V, Montana G. Automated Triaging of Adult Chest Radiographs with Deep Artificial Neural Networks. *Radiology.* 2019;291(1):196-202.
9. Sheikh A, Anderson M, Albala S, Casadei B, Franklin BD, Richards M, et al. Health information technology and digital innovation for national learning health and care systems. *Lancet Digit Health.* 2021;3(6):383-96.
10. Çalli E, Sogancioglu E, Van Ginneken B, Van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: A survey. *Med Image Anal.* 2021;72:102125.
11. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest

- radiographs: A cross-sectional study. *PLoS Med.* 2018;15(11):1002683.
12. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-Ray8; hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2017;3462-71.
 13. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. *Proc AAAI Conf Artif Intell.* 2019;33:590-7.
 14. Jaeger S, Candemir S, Antani S, Wang YX, Lu PX, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med Surg.* 2014;4:475-7.
 15. Zawacki A, Wu C, Shih G, Elliott J, Fomitchev M, Hussain M, et al. SIIM-ACR Pneumothorax Segmentation. Kaggle. 2019.
 16. Chauhan A, Chauhan D, Rout C. Role of Gist and PHOG features in computer-aided diagnosis of tuberculosis without segmentation. *PLoS One.* 2014;9(11):112980.
 17. Qin ZZ, Ahmed S, Sarker MS, Paul K, Adel ASS, Naheyan T, et al. Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. *Lancet Digit Health.* 2021;3:543-54.
 18. Chatterjee S, Das P, Stallworthy G, Bhambure G, Munje R, Vassall A. Catastrophic costs for tuberculosis patients in India: Impact of methodological choices. *PLOS Glob Public Health.* 2024;4:0003078.
 19. Raval D, Parmar D, Saha S, Sarkar R, Wadhwa M, Pandya A, et al. Cost-effectiveness analysis of AI-assisted chest X-ray interpretation tools for TB screening: a rapid HTA. *Front Digit Health.* 2025;7:1629127.
 20. WHO. WHO Consolidated Guidelines on Tuberculosis. Module 2: Screening - Systematic Screening for Tuberculosis Disease. 2021. Available at: <https://www.who.int/publications/i/item/9789240022676>. Accessed on 14 February 2026.
 21. Nash M, Kadavigere R, Andrade J, Sukumar CA, Chawla K, Shenoy VP, et al. Deep learning, computer-aided radiography reading for tuberculosis: a diagnostic accuracy study from a tertiary hospital in India. *Sci Rep.* 2020;10:210.
 22. Shah H, Patel J, Yasobant S, Saxena D, Saha S, Sinha A, et al. Capacity Building, Knowledge Enhancement, and Consultative Processes for Development of a Digital Tool (Ni-kshay SETU) to Support the Management of Patients with Tuberculosis: Exploratory Qualitative Study. *J Med Internet Res.* 2023;25:e45400.
 23. Menon S, Koura KG. Artificial intelligence for tuberculosis control: a scoping review of applications in public health. *J Glob Health.* 2025;15:04192.
 24. Jan B, Dar MI, Choudhary B, Basist P, Khan R, Alhalmi A. Cardiovascular Diseases Among Indian Older Adults: A Comprehensive Review. *Cardiovasc Ther.* 2024;2024:6894693.
 25. Suleman MU, Mursaleen M, Khalil U, Saboor A, Bilal M, Khan SA, et al. Assessing the generalizability of artificial intelligence in radiology: a systematic review of performance across different clinical settings. *Ann Med Surg (Lond).* 2025;87:8803-11.
 26. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. *Eur Radiol.* 2020;30:552532.
 27. Blezek DJ, Olson-Williams L, Missert A, Korfiatis P. AI Integration in the Clinical Workflow. *J Digit Imaging.* 2021;34(6):1435-46.
 28. Abbas Q, Jeong W, Lee SW. Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges. *Healthcare.* 2025;13:2154.
 29. Sridharan S, Seah XHA, Venkataraman N, Tirukonda PS, Jeyaratnam RP, John S, et al. Real-world evaluation of an AI triaging system for chest Xrays: a prospective clinical study. *Eur J Radiol.* 2024;181:111783.
 30. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ.* 2020;368:m689.

Cite this article as: Singh V, Jhamb A, Sil S, Kumar S, Agrawal C, Pareek A, et al. A retrospective multi-source clinical validation of Lenek intelligent radiology assistant: an artificial intelligence-based chest radiograph screening and triage system for high-burden pulmonary and cardiac conditions in India. *Int J Res Med Sci* 2026;14:2449-58.