

Health Technology Innovation

Artificial intelligence for screening drug resistance in tuberculosis

Sucharitha Kannappan Mohanvel¹, Ramalingam Radhakrishnan¹, Prasaanth Balraj³, Niket Singh³, Sachin Danisetty³, Harsh Vats³, Aparna Chaudhury³, Hansraj Choudhary⁴, Priya Rajendran¹, Madhumathi Jayaprakasam⁴, Senthil Devi Vadivel¹, Bershila Preysingh¹, Shakinbanu Thariyasha¹, Sinduja Anjaiyan², Sumalata Chittiboyina⁷, Sowmya Dhawan⁸, Shailesh Lichade¹⁰, Rashmi Kumari¹¹, Rashmi Ratnam¹², Sachin Chandrakar¹³, Mohammed Nasir Khan¹⁴, Ashutosh Kumar Srivastava⁹, Indranath Roy¹⁵, Reena Kingsberry¹⁶, Anand Sridharan⁵, Ranjani Ramachandran⁵, Nishant Kumar⁶, Manjula Singh⁴, Raghuram Rao⁶, Urvashi B. Singh⁶, Chandrasekaran Padmapriyadarsini¹ & Siva Kumar Shanmugam¹

¹Department of Bacteriology, ICMR-National Institute for Research in Tuberculosis, ²Department of Tuberculosis, Intermediate Reference Laboratory, Chennai, Tamil Nadu, ³Department of Machine Learning, Wadhvani Institute for Artificial Intelligence, ⁴Division of Communicable Disease, Indian Council of Medical Research, ⁵Central TB Division, World Health Organization, ⁶Department of Central TB Division, Ministry of Family and Health Welfare, New Delhi, ⁷Department of Tuberculosis, Intermediate Reference Laboratory, Hyderabad, Telangana, ⁸Intermediate Reference Laboratory, State TB Office, Bhopal, ⁹Department of Tuberculosis Culture & DST Lab, Intermediate Reference Laboratory, Indore, Madhya Pradesh, ¹⁰State TB Training & Demonstration Centre & Intermediate Reference Laboratory, Nagpur, Maharashtra, ¹¹Health Department, Intermediate Reference Laboratory, Patna, Bihar, ¹²Department of Microbiology, Intermediate Reference Laboratory, Lucknow, ¹³Intermediate Reference Laboratory, Raipur, Chhattisgarh, ¹⁴Department of TB Culture and DST Lab, Dr. Sampurnanand Medical College, KN Chest Hospital, Jodhpur, Rajasthan, ¹⁵Intermediate Reference Laboratory, Department of Health and Family Welfare, Kolkata, West Bengal, & ¹⁶National Tuberculosis Institute, Bengaluru, Karnataka, India

Received June 15, 2025; Accepted November 16, 2025; Published January 31, 2026

Background & objectives: Central TB division facilitated development of a line probe assay (LPA) artificial intelligence (AI) tool. The tool was developed, trained, and validated for performance by collecting more than 18,000 LPA strips across culture and drug susceptibility Testing (C&DST) laboratories. The Indian Council of Medical Research (ICMR)-National Institute for Research in Tuberculosis (NIRT) evaluated the LPAAI tool independently. The objective was to establish and verify an AI-driven system for automatically interpreting LPA strips, which are employed in tuberculosis drug resistance screening, to improve accuracy, consistency, and scalability across diverse laboratory settings.

Methods: The AI system integrates faster regions convolutional neural network (FR-CNN) for strip detection, detection transformer (DETR) for band localisation, and a hierarchical neural network (HNN) for classification of bands, loci, and drug labels. Independent validation was conducted by ICMR-NIRT

Tuberculosis (TB) caused by *M. tuberculosis* (MTB) is one of the leading causes of death. Globally,

3.2 per cent of individuals with TB were estimated to have multidrug resistant (MDR)/rifampin resistant

using 2810 first-line (FL)-LPA and 241 reflex second-line (SL-LPA) across ten intermediate reference laboratories (IRLs).

Results: AI comparative models demonstrated an accuracy range of 92-100 per cent, with sensitivity between 80-100 per cent and specificity from 86-100 per cent for the *tub*, *rpoB*, *katG*, *InhA*, *gyrA/gyrB*, *rrs*, and *eis* genes. The overall F1 score varies from 0.81 to 1.00, indicating perfect precision and recall.

Interpretation & conclusions: This AI system offers a novel, modular architecture capable of expert-level interpretation of LPA strips. The AI tool performs at par with expert readers and offers a reliable, scalable solution for LPA interpretation. AI tool adoption can reduce interpretation time, enhance result uniformity, and improve treatment delivery across India's TB programme, supporting national goals for TB elimination.

Key words Artificial intelligence - DR - TB - hierarchical neural network - line probe assay

(RR)-TB among new cases and 16 per cent among previously treated ones in 2024¹. Based on the India TB Report 2024, 5.3 per cent MDR-TB were diagnosed in 2023². The End-TB program mandates continued research and development of new tools to fight against the drug resistance of TB at the regional, national, and global levels¹. Currently, under the National Tuberculosis Elimination Program (NTEP), genotypic resistance detection to the first-line (FL) drugs isoniazid (INH) and rifampicin (RIF) and second-line (SL) drugs fluoroquinolone (FQ) and aminoglycosides (AMG) is carried out using line probe assay (LPA) from sputum samples if the patient is smear-positive, and in the case of a smear-negative patient, culture is used^{2,3}. The principle of LPA involves the identification of probes specific to common mutations ('mutation probes') and probes targeting the wild-type sequence ('wild-type probes') on a strip⁴. In FL-LPA, mutations in the *katG* and *inhA* genes associated with INH resistance are detected⁵. In SL-LPA, mutations in the genes *gyrA* and *gyrB*, associated with fluoroquinolone resistance, and the *rrs* gene associated with aminoglycoside resistance are detected^{6,7}. Since machine learning (ML), the vital subdomain of artificial intelligence (AI) has been gaining importance in the last two decades⁸⁻¹⁰, applying it to the interpretation of mutation resistance in LPA could be explored. According to Agrawal *et al*¹¹ (2024), a reconfigurable AI-based system incorporates Faster Region-based Convolutional Neural Network (Faster R-CNN), Detection Transformer (DETR), and a Hybrid Neural Network (HNN), which was advanced to provide end-to-end interpretation in compliance with World Health Organization (WHO) guidelines and was validated through expert consensus. The LPA AI tool, developed by Wadhvani AI¹¹, works on the automatic interpretation of results from the scanned

LPA sheets by an algorithm. An initial validation study to assess the performance of this tool was carried out at 40 laboratories under the supervision of Central TB division. In this study, the overall accuracy of AI interpretation (2641 FL strips and 340 SL strips) was found to be 97 per cent for RIF, 98 per cent for INH and FLQ, and 99 per cent for kanamycin when compared with manual interpretation¹².

Based on these findings, an independent evaluation of this AI tool was conducted by the Indian Council of Medical Research-National Institute for Research in Tuberculosis (ICMR-NIRT). The aim of this evaluation was to assess the feasibility of using the AI solution in interpretation, time reduction, and enhancement of result consistency across states. We hypothesise that the AI system will achieve results similar to those of trained microbiologists concerning band and drug-level interpretation, concurrently enhancing consistency and decreasing turnaround time within high-throughput diagnostic workflows.

Materials & Methods

The study was conducted at ICMR-NIRT during the year 2023-2024 after obtaining the ethical approval for the study from the Institutional Ethics Committee.

AI interpretation mechanism: A classical computer vision technique (Fig. 1) is applied to identify and extract the strips pasted on the LPA sheet. Our AI pipeline, adapted from automatic interpretation of LPA for tuberculosis (2024)¹¹, integrates Faster R-CNN for strip detection, DETR for band localisation, and an HNN for band classification. A transformer-based OCR (TrOCR) module is employed to extract patient or sample identifiers from each sheet. Once strips

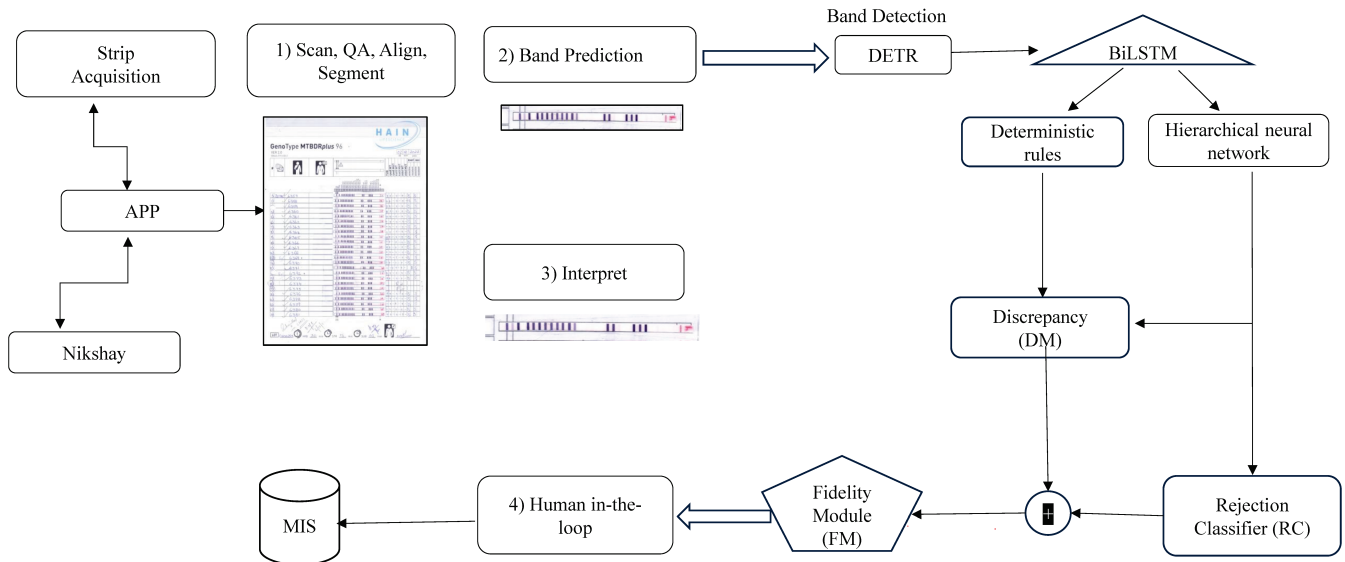


Fig. 1. Tuberculosis (TB) Line Probe Assay (LPA) Artificial Intelligence (AI) solution uses a deep learning-based computer vision model to automate LPA test interpretation. A Convolutional Neural Network (CNN) detects and classifies bands after preprocessing steps, including image enhancement, alignment correction, and segmentation. The model analyses band presence, position, and intensity to map patterns to wild-type and mutant regions, determining drug resistance. Trained on expert-annotated images, it achieves high accuracy, enabling automated, real-time LPA interpretation within diagnostic workflows.

are extracted, as per the deep learning components, a bounding box (BB) is drawn around the band if a band is present in the strip. If two types of objects (foreground and background) are detected, models such as FR-CNN and DETR are applied to customise the architecture. Further, identifying the BB where bands are present, band types such as AC, CC, *rpoB*-Mut1, etc., are associated with each detected band and its corresponding band type. The BBs estimated by the DETR are sorted, and the relative distances of the centroids of the BBs from a reference are computed. Neural networks were trained using the Adam optimiser. Hyperparameters (such as learning rate, sequence length for long short-term memory (LSTM), batch size, and network depth) were optimised using grid search on the validation split. This forms the positional context for the band prediction model, where a sequence-to-sequence translation is carried out to create a bidirectional LSTM layer. Once band prediction is completed, the band probabilities are sent to the HNN and deterministic rules (DET) to produce drug interpretations. Band-level annotations (bounding boxes, presence/absence) were performed using a custom tool, then validated by multiple microbiologists. Final drug calls were derived either by consensus or deterministic rules applied to the band pattern. Some of the strip results are rejected based on HNN by the rejection classifier and DET and HNN by the discrepancy module. The ambiguous bands are

highlighted by the fidelity module (FM) and sent for review by a human in the loop (trained microbiologist) for concordance¹¹.

Study setting: The validation process by ICMR-NIRT was planned in three different phases:

- Phase 1. The LPA sheets at ICMR -NIRT were scanned and interpreted by AI solution.
- Phase 2. The LPA sheets from nine Intermediate reference laboratories (IRLs) (at Kolkata, Nagpur, Patna, Raipur, Telangana, Lucknow, Bengaluru, Chennai, and Jodhpur) were scanned using different types of scanners and sent to ICMR-NIRT for AI interpretation.
- Phase 3. The LPA sheets that were subjected to scanning and AI interpretation at five IRLs (IRL Lucknow, Hyderabad, Bhopal, Chennai, and Nagpur) were cross-verified by ICMR-NIRT.

AI interpretation and validation: The AI validation conducted by Central TB division initially included 20,000 strips for training the microbiologists at 40 laboratories to perform testing and interpretation independently. Based on the consistency of results obtained, a minimum of 10 per cent of the trained set were included for final evaluation (2000 strips). Based on this calculation and accounting for a 20 per cent contamination rate, the minimum sample size required

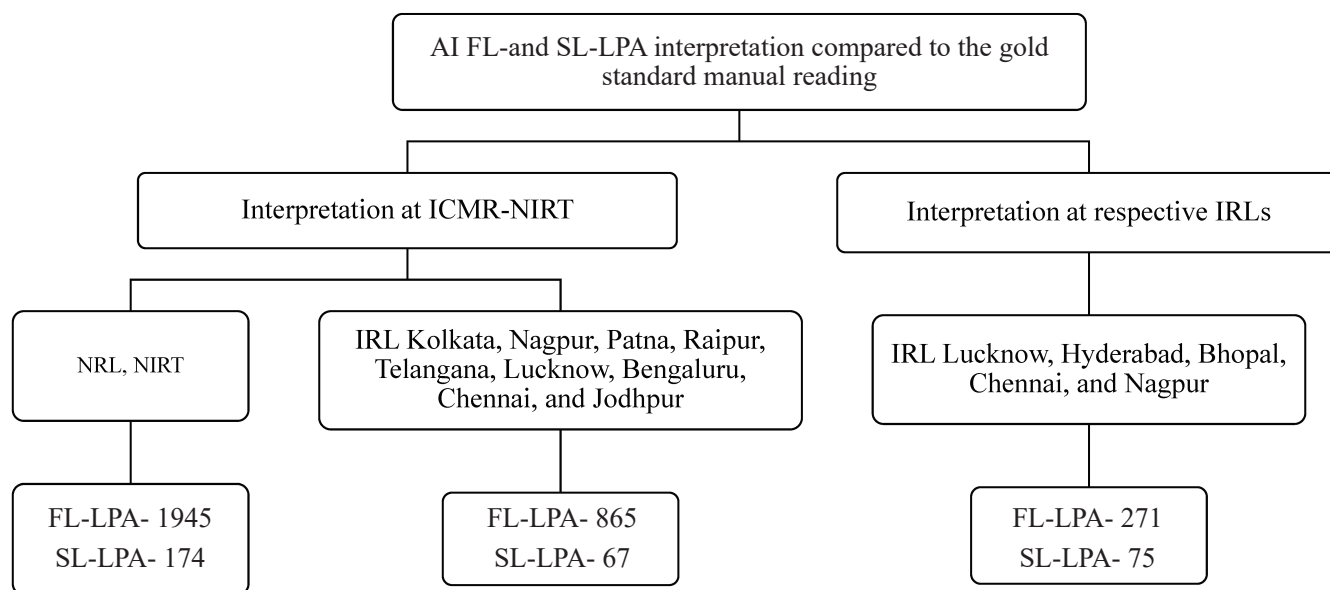


Fig. 2. Flow of three phases of Artificial Intelligence First Line (FL)- and Second Line (SL)-Line Probe Assay (LPA) interpretation compared to the gold standard manual reading. Phase I-LPA strips were interpreted and analysed at the Indian Council of Medical Research-National Institute for Research in Tuberculosis (ICMR-NIRT), Phase II-Intermediate Reference Laboratories (IRLs) LPA strips at ICMR-NIRT and, Phase III- at ICMR-NIRT and the five IRLs.

for this study was 2400 for FL-LPA results and their respective SL-LPA. In phase I, 1945 FL-LPA and 174 SL-LPA results were interpreted and analysed. In Phase II, 865 FL-LPA and 67 SL-LPA, and in Phase III, 271 FL-LPA and 75 SL-LPA, results were included for interpretation and analysis (Fig. 2). The analysis included all the LPA strips with satisfactory internal quality control (IQC) and MTB detection. Sheets with incomplete results or reporting blank/missing interpretations on LPA test results were excluded from the analysis. The minimum scanner specification required for the AI solution is 72 dpi and above; the scanners used in the IRLs across have a range of 260–600 dpi. Three levels of data prediction were used during the analysis; (a) the presence or absence of a band on a given strip; (b) interpretation of gene locus level; (c) the drug level interpretations.

Discordance resolution: Microbiologists finalised the discordance resolution by comparing the LPA sheets with the field (nucleic acid amplification test) NAAT results, which is the usual diagnostic flow in FL-LPA interpretation. Such a manual interface is required for AI interpretation of false positives and negative results. The tubband representing MTB positive is considered regardless of its density or faintness. The LPA sheet, consisting of NAAT results obtained from the field, was verified by microbiologists during AI interpretation in

the case of false-positive results. Lighter wild (WT) bands and mutant (MUT) bands were verified for false negatives and false positives, respectively, during the manual interface.

Statistical analysis: To assess the effectiveness of the AI model, F1 score¹³⁻¹⁵, Cohen's kappa, Fowlkes-Mallows score¹⁶ and Matthew's correlation coefficient (MCC)¹⁷ were calculated and interpreted against the gold standard. By applying Cohen's Kappa, the agreement between the model's predictions and the gold standard was assessed. F1 Score was chosen to balance precision and recall, critical in minimising both false positives and false negatives in clinical diagnostics. MCC provides a robust, balanced evaluation in imbalanced datasets by incorporating all elements of the confusion matrix. Moreover, area under receiver operating curve (AUROC) reflects overall discriminative power, while area under plotted precision recall curve (AUPRC) is more informative when resistant cases (positive class) are rare, making it particularly relevant for gene-level resistance evaluation. To examine the dependability and resilience of the AI algorithm, the data obtained was analysed using the confusion matrix, ROC, and PRC¹⁸. Respective sensitivity (recall), specificity, accuracy, negative predictive value, positive predictive value (precision), and F1 score, along with the respective AUROC and AUPRC were obtained. The

sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) were calculated to evaluate the performance of the AI model¹⁹.

Results & Discussion

ICMR-NIRT evaluation: FL-LPA (n=1945) and SL-LPA (n=174) were extracted from the LPA sheets from the ICMR-NIRT routine diagnostics. Nine IRLs were included in the study, with a sample size of FL-LPA (n=860) and SL-LPA (n=67), respectively. Out of 174 cases from ICMR-NIRT, SL-LPA reported eight invalids and one indeterminate, while in 1945, FL-LPA strips had three reported invalids

Phase I: The NPV score for the *tubb* and the genes *gyrA/gyrB* and *rrs* genes was 1.00; *rpoB*, *KatG*, and *InhA* were 0.99. The PPV scores for the *tubb*, *rpoB*, *KatG*, *InhA*, and *gyrA/gyrB* genes were 0.99-1.00. The accuracy, sensitivity, and specificity were compared for all performances of the *tubb* and the genes *rpoB*, *katG*, *InhA*, *gyrA/gyrB* genes, *rrs*, and *eis*. The accuracy ranges were as follows: overall (0.92-1.00), sensitivity (0.88-1.00), and specificity (0.86-1.00). The overall F1 score range from 0.89 to 1.00, showcasing ideal precision and recall for all the AI comparative model's performance at ICMR-NIRT (Table). A scanner with a resolution of 600 dpi was used at ICMR-NIRT to scan the LPA sheets.

Phase II: For the Nine IRLs, specified alone, the NPV and PPV scores range between 0.99 and 1.00. The accuracy varied between 0.99 and 1.00, the sensitivity between 0.80 and 1.00, and the specificity between 0.98 and 1.00 for the *tubb* and the respective genes. The IRL scanners across had a 260–600 dpi resolution. The F1 score was similar to Phase I, even exhibiting the perfect PR for the AI tool compared with the gold standard.

Phase III: The performance of the AI and manual results at five IRL: NPV ranged from 0.95 to 1.00, and the PPV was 0.78–1.00 for the *tubb* and the *rpoB*, *KatG*, *InhA*, *gyrA*, *gyrB*, and *rrs* genes. The IRLs demonstrated performance levels ranging from 0.88 to 1.00, with sensitivity from 0.81 to 1.00 and specificity of 0.82 and 1.00 for the corresponding *tubb* and genes. The overall F1 score ranged from 0.81 to 1.00, showing ideal precision and recall. According to Kuang *et al*¹⁹ (2022), a high F1 score (close to 1) indicated a balanced trade-off between precision and recall²⁰ (Supplementary Figure).

Considering the minimum scanner specification required for the solution is 72 dpi and above, the scanners used in the five IRLs (Lucknow, Hyderabad, Bhopal, Chennai, and Nagpur) across had a range of 260–600 dpi (Supplementary Table). Based on the above inference, the scanning performed at IRLs and NRLs did not show any difference in the band density of LPA sheets, though there was a difference in the dpi across.

Despite timely outcomes, LPA data interpretation deficiencies persist in many heavily burdened laboratories. Integrating AI-driven systems enhances laboratory workflows for LPA testing, efficiently transmitting data to designated information systems for swift public health interventions. We aim to implement AI-LPA at C&DST laboratories nationwide to achieve the quickest treatment regimen. By utilising an AI solution, this leading study on LPA implementation in India aims to shorten interpretation time and increase the uniformity of the interpretation process across states. The current use of AI and advancements in ubiquitous computing for forecasting respiratory illnesses has opened up a world of potential in the medical industry.

Merged ROC and the PR curve: The AUROC value (0.93–1.00) indicated better model performance for the *tubb* and the considered genes. The AUPRC value (0.99-1.00) for the *tubb* and, *rpoB*, *InhA*, *katG* and, *gyrA/gyrB* genes, indicates perfect precision and recall, while the genes *rrs*, and *eis* values (0.50), which cannot classify these respective genes, model performance at ICMR-NIRT. Moreover, the AUROC value (0.98-1.00) demonstrated the enhanced model performance of the discussed genes. The AUPRC values were 1.00 *gyrA/gyrB* and *eis* genes, which indicated perfect precision and recall, while the *tubb* and the genes *rpoB*, *katG*, *InhA*, *rrs* and *eis* ranged between (0.98-0.50), which indeed indicated better model performance. The value was 1.00 for the *tubb* and, indicating perfect discrimination, while the *rpoB*, *katG*, *InhA*, *gyrA/gyrB*, *rrs*, and *eis* ranges (0.88–0.96) represented a better performance model. The AUPRC value ranged from 0.83-0.95 for the *rpoB*, *katG*, *InhA*, *gyrA/gyrB*, *rrs*, and *eis* except the *tubb* (1.00), which exhibited better model performance for phase III.

Unlike conventional AI systems such as CAD4TB for chest radiographs or CNN-based models for smear microscopy, our system addresses a specific void: the automated analysis of molecular tests using LPA strips. While previous investigations have underscored

Table. Comparative performance of the AI model and manual results

Tub, FL-LPA & SL-LPA	Accuracy			Sensitivity			Specificity		
	A	B	C	A	B	C	A	B	C
Tub detected	0.92	0.99	1.00	1.00	1.00	1.00	0.86	0.98	1.00
<i>rpoB</i>	0.94	0.99	0.98	0.88	0.97	0.94	1.00	1.00	0.99
<i>InhkatG</i>	1.00	0.99	0.97	0.99	0.96	0.90	1.00	1.00	0.98
<i>InhA</i>	0.98	0.99	0.96	0.96	0.80	0.91	1.00	1.00	0.97
<i>gyrA/gyrB</i>	1.00	1.00	0.88	1.00	1.00	0.97	1.00	1.00	0.82
<i>rrs</i>	1.00	1.00	0.92	0.00	1.00	0.81	0.00	1.00	0.95
<i>eis</i>	1.00	1.00	0.96	0.00	0.00	0.91	0.00	0.00	0.97

A, ICMR-NIRT; B, IRLs at ICMR-NIRT; C, IRLs

AI's usefulness in radiology and microscopy, scarce attention has been given to strip-format genetic tests. Thus, our research augments current AI diagnostics by targeting a crucial yet insufficiently automated phase in the drug resistance identification process. Human-in-the-loop review was initiated for 37 per cent of FL-LPA and 32 per cent of SL-LPA strips within the validation dataset; among them, only a small fraction (3–9%) of the total strips called for human correction. Faint bands or visual distortions were the main causes of these overrides, which emphasises the system's durability even as essential quality control is upheld through expert evaluation¹¹.

The model performs reliably with 72 dpi resolution; however, variability in scanners and imaging conditions may affect consistency. Misclassification can occur due to strip misalignment or artifacts. Additionally, training on a specific dataset may limit generalisability, highlighting the need for broader validation. Successful deployment will require compatible scanners (≥ 72 dpi) and reliable computing infrastructure, standardised staff training for AI use and override procedures, and adherence to national regulatory and data governance frameworks. Addressing these factors is essential for safe, equitable, and sustainable integration into TB program workflows. Moreover, the system will ensure privacy, mitigate bias, and maintain accountability through anonymised data, diverse training, monitoring, and human-in-the-loop oversight.

Despite its limitations in terms of manual interface, the background analysis has proven AI's overall performance to be superior when manual intervention is applied. A key characteristic of this model is its efficiency in reducing interpretation time and delivering reliable results across states. Future research is expected to predict algorithm optimisation without a manual interface.

Using AI solutions reduces interpretation time and enhances the consistency of the interpretation process across states. Direct transmission allows public health to access the patient's information and help end TB. In this study, we demonstrated that AI-based systems improve laboratory workflows for patient-wise LPA testing and directly transmit information to designated information systems for public health action. The outcomes of this research yielded an exceptional F1 score, illustrating high levels of precision and recall when comparing the AI tool to the gold standards, suggesting a better fit when used along with the manual interface.

Acknowledgment: Late Ms. Dakshayani, Senior Laboratory technician, for her support in the LPA interpretation process.

Financial support & sponsorship: None.

Conflicts of Interest: None.

Use of Artificial Intelligence (AI)-Assisted Technology for manuscript preparation: The authors confirm that there was no use of AI-assisted technology for assisting in the writing of the manuscript and no images were manipulated using AI.

References

1. World Health Organization. *Global tuberculosis report 2024*. Geneva: World Health Organization; 2024. Available from: <https://iris.who.int/bitstream/handle/10665/379339/9789240101531-eng.pdf?sequence=1>, accessed on August 5, 2025.
2. Central TB Division, Ministry of Health and Family Welfare, Government of India. *India TB report 2024*. New Delhi: Central TB Division; 2024. Available from: https://tbcindia.gov.in/wp-content/uploads/2024/10/TB-Report_for-Web_08_10-2024-1.pdf, accessed on August 5, 2025.
3. World Health Organization. *The use of molecular line probe assays for the detection of resistance to isoniazid and rifampicin*.

- Geneva: World Health Organization; 2016. Available from: <https://www.who.int/tb/publications/molecular-test-resistance/en>, accessed on October 5, 2025.
4. Lin M, Chen YW, Li YR, Long LJ, Qi LY, Cui TT, *et al.* Systematic evaluation of line probe assays for the diagnosis of tuberculosis and drug-resistant tuberculosis. *Clin Chim Acta* 2022; 533 : 183-218.
 5. Mohan S, Pujari V, Meshram P. Drug resistance patterns and treatment outcomes in DR-TB patients at a tertiary care centre in Mumbai. *Indian J Tuberc* 2024; 71 : S10-4.
 6. World Health Organization. *Molecular line probe assays for rapid screening of patients at risk of multidrug-resistant tuberculosis (MDR-TB)*. Geneva: World Health Organization; 2008. Available from: https://www.stoptb.org/sites/default/files/imported/document/BoardDocs/15/2.08-11_Rolling_out_diagnostics_in_the_field/2.08-11.2_Line_Probe_Assays_0.pdf, accessed on September 5, 2025.
 7. World Health Organization. *WHO consolidated guidelines on tuberculosis. Module 3: diagnosis – rapid diagnostics for tuberculosis detection*. Geneva: World Health Organization; 2020. Available from: <https://www.who.int/publications/i/item/9789240000339>, accessed on September 5, 2025.
 8. Hansun S, Argha A, Liaw ST, Celler BG, Marks GB. Machine and deep learning for tuberculosis detection on chest X-rays: Systematic literature review. *J Med Internet Res* 2023; 25 : e43154.
 9. Jamal S, Khubaib M, Gangwar R, Grover S, Grover A, Hasnain SE. Artificial intelligence and Machine learning based prediction of resistant and susceptible mutations in *Mycobacterium tuberculosis*. *Sci Rep* 2020; 10 : 5487.
 10. Kouchaki S, Yang Y, Walker TM, Sarah Walker A, Wilson DJ, Peto TEA, *et al.* Application of machine learning techniques to tuberculosis drug resistance analysis. *Bioinformatics* 2019; 35 : 2276-82.
 11. Agrawal J, Kumar M, Tiwari A, Danisetty S, Dhavala S, Jain N, *et al.* Automatic interpretation of line probe assay test for tuberculosis. *AAAI* 2024; 38 : 21897-904.
 12. World Health Organization. *Line probe assays for detection of drug-resistant tuberculosis: interpretation and reporting manual for laboratory staff and clinicians*. Available from: <https://iris.who.int/bitstream/handle/10665/354240/9789240046665-eng.pdf?sequence=1>, accessed on September 5, 2025.
 13. Park DJ, Park MW, Lee H, Kim YJ, Kim Y, Park YH. Development of machine learning model for diagnostic disease prediction based on laboratory tests. *Sci Rep* 2021; 11 : 7567.
 14. Agarwal A, Sharma P, Alshehri M, Mohamed AA, Alfarraj O. Classification model for accuracy and intrusion detection using machine learning approach. *Peer J Comput Sci* 2021; 7 : e437.
 15. Chicco D, Tötsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min* 2021; 14 : 13.
 16. Dreizin D, Goldmann F, LeBedis C, Boscak A, Dattwyler M, Bodanapally U, *et al.* An automated deep learning method for Tile AO/OTA pelvic fracture severity grading from trauma whole-body CT. *J Digit Imaging* 2021; 34 : 53-65.
 17. Chen PC, Mermel CH, Liu Y. Evaluation of artificial intelligence on a reference standard based on subjective interpretation. *Lancet Digit Health* 2021; 3 : e693-5.
 18. Nijjati M, Ma J, Hu C, Tuersun A, Abulizi A, Kelimu A, *et al.* Artificial intelligence assisting the early detection of active pulmonary tuberculosis from chest X-rays: a population-based study. *Front Mol Biosci* 2022; 9 : 874475.
 19. Yellapu GD, Rudraraju G, Sripada NR, Mamidgi B, Jalukuru C, Fimal P, *et al.* Development and clinical validation of Swaasa AI platform for screening and prioritization of pulmonary TB. *Sci Rep* 2023; 13 : 4740.
 20. Kuang X, Wang F, Hernandez KM, Zhang Z, Grossman RL. Accurate and rapid prediction of tuberculosis drug resistance from genome sequence data using traditional machine learning algorithms and CNN. *Sci Rep* 2022; 12 : 2427.

For correspondence: Dr Siva Kumar Shanmugam, Department of Bacteriology, ICMR-National Institute for Research in Tuberculosis, Chennai 600 031, Tamil Nadu, India
e-mail: shanmugamsiva27@gmail.com