LMJ LINK MEDICAL JOURNAL OF HEALTH AND COMMUNITY RESEARCH

Crossref DOI: https://doi.org/10.61919/m227h628

Original Article

AI-Based Risk Stratification for Latent Tuberculosis Among Urban Slum Dwellers: A Cross-Sectional Study

Ali Dost[1], Afeera Bint-e-Tanveer[2], Ajmal Khan[3], Kamran Ali[4], Sajjad Ahmad[5], Zubia Shahid[6]

[1] Medical Officer, Mir Gul Khan Naseer Teaching Hospital, Nushki, Pakistan
[2] Lecturer, Software Engineering Department, National University of Technology, Islamabad, Pakistan. ORCID: 0009-0007-3321-0142
[3] MDR Coordinator, PMDT Site Nushki, Pakistan TB Control Program, Nushki, Pakistan
[4] Demonstrator, Jhalawan Medical College, Khuzdar, Pakistan
[5] Senior Lab Technologist, Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, Pakistan
[6] PhD Researcher, Quaid-i-Azam University, Islamabad, Pakistan

***Corresponding author: Sajjad Ahmad, Sajjadahmad.ibms@kmu.edu.pk**

ABSTRACT

Background: Latent tuberculosis infection (LTBI) poses a critical public health challenge in urban informal settlements, where structural deprivation, household overcrowding, and limited healthcare access sustain high transmission risk. Traditional screening methods fail to reach the most vulnerable individuals, underscoring the need for targeted, data-driven approaches. Objective: This study aimed to evaluate the predictive performance of AI-assisted machine learning models for LTBI risk stratification among adults in urban informal settlements in Lahore, Pakistan. Methods: A cross-sectional analytical study enrolled 150 adult participants from densely populated slum communities. Data were collected via structured questionnaires, clinical assessments, and tuberculin skin testing (TST; ≥10 mm threshold). Three supervised machine learning algorithms, logistic regression, random forest, and gradient-boosted decision trees, were developed using a stratified 70:30 train-test split with 10-fold cross-validation. Model performance was assessed using accuracy, sensitivity, specificity, positive and negative predictive values, and area under the receiver operating characteristic curve (AUC) with 95% bootstrap confidence intervals. Variable importance was evaluated using normalised feature gain scores. Results: TST positivity was 42.0% (n=63). The gradient-boosted model achieved the highest performance (AUC 0.91, sensitivity 90.5%, specificity 88.5%), followed by random forest (AUC 0.88) and logistic regression (AUC 0.82). Prior TB contact history, household crowding, and poor ventilation were the strongest predictors. Overall classification accuracy was 89.3% (95% CI 83.3–93.7%). Conclusion: AI-assisted risk stratification using community-collectible sociodemographic and environmental variables can effectively identify individuals at elevated LTBI risk in urban slum settings, supporting targeted screening and preventive intervention strategies. Keywords: Artificial intelligence; community health; latent tuberculosis; machine learning; risk stratification; tuberculin skin test; urban slums

"Cite this Article" | **Received**: 17August 2025; **Accepted**: 25 December 2025; **Published**: 31 December 2025.

Author Contributions: Concept and design: SA, AD; Data collection: AK, KA; Statistical analysis and AI modelling: ABT, ZS; Drafting and critical revision: AD, SA. **Ethical Approval**: Khyber Medical University, Peshawar, Pakistan. **Informed Consent**: Written informed consent was obtained from all participants; **Conflict of Interest**: The authors declare no conflict of interest; **Funding**: No external funding; **Data Availability**: Available from the corresponding author on reasonable request; **Acknowledgments**: N/A.

INTRODUCTION

Tuberculosis (TB) remains one of the most consequential infectious diseases globally, accounting for an estimated 10.6 million new cases and 1.3 million deaths annually, with a disproportionate burden concentrated in South and Southeast Asia (1). Pakistan ranks among the five highest-burden countries worldwide, reporting approximately 611,000 incident TB cases per year and facing persistent challenges in case detection, treatment adherence, and preventive care (2). Underpinning this active disease burden is a substantially larger reservoir of latent tuberculosis infection (LTBI), defined as a state of persistent immunological response to Mycobacterium tuberculosis antigens in the absence of clinical signs of active disease. Global estimates suggest that approximately one-quarter of the world's population carries latent infection, with reactivation risk particularly elevated in populations exposed to poverty, malnutrition, overcrowding, and immunosuppressive conditions (3). Although LTBI is clinically silent,

it constitutes the primary source from which future active cases emerge, making its identification and management central to any meaningful TB elimination strategy.

Urban informal settlements, commonly referred to as slums, represent a uniquely high-risk epidemiological context for both LTBI acquisition and reactivation. These environments are characterized by extreme household density, inadequate ventilation, poor sanitation, limited access to primary healthcare, and high rates of malnutrition and tobacco use, conditions that individually and synergistically amplify transmission risk (4). In Lahore, Pakistan, where rapid urbanisation has produced extensive informal settlements, a substantial proportion of residents remain unscreened for LTBI due to structural barriers including low health-seeking behaviour, financial constraints, and the absence of targeted community outreach programmes (5). The World Health Organization's End TB Strategy explicitly identifies LTBI testing and preventive therapy as a pillar of pre-elimination efforts, yet implementation in resource-constrained urban settings remains fragmented and insufficiently prioritised (6). Conventional population-wide screening approaches are logistically impractical and financially unsustainable in these contexts, underscoring the need for risk-stratified, targeted alternatives that can efficiently direct limited diagnostic and therapeutic resources toward the most vulnerable individuals.

The tuberculin skin test (TST) remains the most widely employed tool for LTBI screening in low- and middle-income countries, owing to its relatively low cost and established operational infrastructure. However, TST sensitivity and specificity are limited by cross-reactivity with Bacillus Calmette–Guérin vaccination and non-tuberculous mycobacterial exposure, both of which are prevalent in Pakistani communities (7). Beyond diagnostic limitations, a fundamental challenge in LTBI control is that traditional screening algorithms rely heavily on clinical presentation and TB contact history, failing to account for the multidimensional interplay of sociodemographic, environmental, and behavioural determinants that collectively define individual risk profiles. This gap is particularly consequential in slum populations, where heterogeneous living conditions create complex, non-linear risk gradients that are not adequately captured by categorical screening criteria alone.

Artificial intelligence (AI) and machine learning methods have emerged as promising tools for enhancing risk prediction across a range of infectious and non-communicable diseases, offering the capacity to integrate high-dimensional data and identify latent patterns that conventional epidemiological models may overlook (8). Supervised learning algorithms, including logistic regression, random forests, and gradient-boosted decision trees, have demonstrated favourable predictive performance in TB-related applications, including active disease diagnosis from radiological data and treatment outcome prediction (9). However, the application of AI-assisted risk stratification specifically for LTBI detection in community-based, resource-constrained settings remains substantially underdeveloped. Most existing predictive models have been trained on clinical or hospital-based datasets from high-income settings, limiting their contextual transferability to environments characterised by fragmented health records, self-reported exposure histories, and operational constraints on laboratory diagnostics (10). A validated, community-adapted AI model capable of generating individualized LTBI risk scores from routinely collectible sociodemographic and environmental variables would offer a scalable, cost-effective complement to existing screening strategies.

Furthermore, concerns about algorithmic transparency, equity, and the risk of embedding existing social inequities into predictive tools must be addressed in any deployment of AI within vulnerable populations (11). Communities in urban informal settlements already experience structural disadvantage; any risk stratification approach must be developed with explicit attention to avoiding stigmatisation, ensuring interpretability, and supporting, rather than supplanting, the clinical judgement of frontline health workers. These considerations reinforce the need for locally grounded, rigorously validated models that are both technically sound and ethically responsive to the communities they are designed to serve. To date, no validated AI-based risk stratification tool for LTBI has been developed or evaluated within

Pakistani urban slum populations, representing a critical gap in the evidence base for targeted TB prevention in one of the world's highest-burden national contexts. The present study therefore aimed to evaluate the predictive performance of AI-assisted machine learning models, specifically logistic regression, random forest, and gradient-boosted decision trees, in identifying individuals at elevated risk of latent tuberculosis infection among adult residents of urban informal settlements in Lahore, Pakistan, using sociodemographic, environmental, and behavioural variables collectible through community-based field surveys.

MATERIALS AND METHODS

This study employed a cross-sectional analytical design to evaluate the capacity of supervised machine learning algorithms to stratify latent tuberculosis infection (LTBI) risk among adults residing in urban informal settlements. The cross-sectional approach was selected for its feasibility within a community-based screening context and its established appropriateness for prevalence estimation and risk factor identification in epidemiological research (12). The study was conducted across three densely populated informal settlements in Lahore, Punjab, Pakistan, a metropolitan environment characterised by rapid urbanisation, high household density, and well-documented deficiencies in sanitation and primary healthcare access. Data collection was carried out over a three-month period, providing sufficient time for field recruitment, structured interviewing, clinical assessment, and laboratory testing while remaining feasible within the operational constraints of the study setting.

The target population comprised adult residents of the selected settlements who were at least 18 years of age and had lived continuously in the community for a minimum of six months prior to enrolment, ensuring consistent exposure to the local environmental conditions relevant to LTBI risk. A purposive sampling strategy was employed, with field teams systematically approaching households across mapped zones of each settlement to achieve representative geographic coverage. The planned sample size of 180 participants was informed by a review of precedent community-based TB risk-prediction studies employing machine learning methods, which demonstrated that samples of 150–200 participants were sufficient for exploratory model development and preliminary validation using cross-validated performance metrics (13). Participants were excluded if they had a current diagnosis of active tuberculosis, were undergoing or had recently completed anti-TB therapy, were pregnant at the time of enrolment, owing to altered immunological responses during gestation, or had severe immunocompromising comorbidities such as advanced HIV disease or haematological malignancy that independently modify TB risk profiles. Individuals who were unable to provide voluntary informed consent or whose data records were incomplete at the time of analysis were also excluded, resulting in a final analytic sample of 150 participants. Attrition between the planned sample of 180 and the final analytic sample was attributable to incomplete questionnaire data, refusal to undergo skin testing, and failure to return for induration measurement at 48–72 hours.

Informed written consent was obtained from all participants prior to enrolment, following a structured explanation of the study purpose, procedures, voluntary nature of participation, and right to withdraw without consequence. Participants with limited literacy were read the consent form aloud, and verbal consent was additionally documented by a witness. The study was conducted in accordance with the Declaration of Helsinki and received ethical approval from the institutional review committee prior to commencement. Confidentiality was maintained throughout data handling, with all participant records anonymised using numeric identifiers, and data stored securely with access restricted to the core study team.

Data collection employed three complementary instruments: a structured interviewer-administered questionnaire, a standardised clinical assessment, and laboratory-based latent TB testing. The questionnaire was developed by drawing on established TB risk-factor frameworks and adapted to reflect the specific sociodemographic and environmental realities of Pakistani urban slum communities (14). It

captured the following domains: age, sex, educational attainment, household size, monthly income relative to the local poverty threshold, smoking status, prior history of close contact with a bacteriologically confirmed TB case, perceived indoor ventilation quality, and self-reported nutritional adequacy. Questionnaires were administered through face-to-face interviews conducted in participants' homes by trained bilingual field staff, enabling direct observation of housing conditions, including approximate household density, ventilation infrastructure, and sanitation facilities, to supplement self-reported data. All field staff underwent a standardised two-day training protocol covering interview technique, infection control procedures, and tuberculin skin test administration and reading, and inter-rater reliability for induration measurement was assessed prior to field deployment.

Latent tuberculosis infection was assessed using the tuberculin skin test (TST), which represents the standard method for community-based LTBI screening in resource-limited settings due to its low cost, established operational infrastructure, and acceptability among Pakistani populations (15). The TST was administered intradermally on the volar surface of the left forearm using 2 tuberculin units of purified protein derivative (PPD RT23, Statens Serum Institut), following the Mantoux technique. Trained community health workers measured transverse induration diameter at 48–72 hours post-injection. In accordance with WHO and Pakistan national TB programme guidelines for populations at elevated exposure risk, a threshold of ≥10 mm was used to define a positive TST result (16). Additional clinical parameters including body mass index, measured from standardised weight and height assessments, blood pressure, and the presence of respiratory symptoms were recorded to supplement individual risk profiles. Although interferon-gamma release assays offer superior specificity in BCG-vaccinated populations, their cost, cold-chain requirements, and logistical complexity rendered them impractical for deployment in this community-based field setting.

The dataset constructed from questionnaire responses and clinical measurements underwent pre-processing prior to model development. Data quality checks were performed to identify duplicated entries, implausible values, and variables with high missingness rates. Continuous variables were assessed for normality using the Shapiro–Wilk test. Descriptive statistics, including means, standard deviations, frequencies, and proportions, were used to characterise participant demographics and TB-related variables. Group comparisons between TST-positive and TST-negative participants were conducted using independent samples t-tests for normally distributed continuous variables, Mann–Whitney U tests for non-normally distributed variables, and chi-square tests for categorical variables. Pearson correlation coefficients were calculated to assess linear associations between continuous predictors and TST induration size. A two-tailed significance threshold of $p < 0.05$ was applied throughout. Analyses were conducted using IBM SPSS Statistics version 26.0 (IBM Corp., Armonk, NY, USA).

Three supervised machine learning algorithms were developed for AI-assisted LTBI risk stratification: logistic regression, random forest, and gradient-boosted decision trees. These algorithms were selected for their complementary strengths, logistic regression for linear interpretability and clinical familiarity, random forest for robustness to correlated predictors and overfitting, and gradient boosting for superior predictive accuracy through sequential residual correction, and for their suitability with modest-sized community datasets (17). Input features comprised all questionnaire-derived and clinically measured variables, with missing values imputed using mean substitution for continuous variables and mode imputation for categorical variables to preserve the full analytic sample. The dataset was partitioned into training (70%) and testing (30%) subsets using stratified random splitting to preserve the proportion of TST-positive cases in each subset. Feature standardisation was applied to continuous variables prior to logistic regression model fitting. To mitigate overfitting risk given the sample size, all models were evaluated using stratified 10-fold cross-validation on the training set before final assessment on the held-out test set (18). Model performance was quantified using accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve (AUC), with 95% confidence intervals estimated using bootstrap resampling (1000 iterations). Variable

importance was quantified for tree-based models using mean decrease in Gini impurity for random forest and feature gain scores for gradient boosting, and visualised as ranked importance plots to identify the strongest predictors of LTBI risk within this population. All machine learning procedures were implemented in Python 3.11 using the scikit-learn (v1.3) library (19).

RESULTS

A total of 150 participants from urban informal settlements in Lahore were included in the final analysis following exclusion of 30 individuals from the originally enrolled cohort of 180 due to incomplete questionnaire data (n=14), refusal to undergo TST administration (n=9), and failure to return for induration measurement within the required 48–72 hour window (n=7). All 150 participants completed demographic profiling, clinical assessment, and TST reading, forming the complete analytic dataset.

Table 1 presents the demographic and clinical characteristics of the full sample, stratified by TST result. The mean age of participants was 34.9 ± 9.0 years (range 18–60 years). TST-positive participants were marginally older (mean 36.7 ± 8.6 years) compared to TST-negative individuals (mean 33.6 ± 9.2 years), though this difference did not reach statistical significance (p = 0.052). Males comprised 51.3% of the sample (n=77), with no statistically significant sex-based difference in TST positivity (p = 0.341). Educational attainment was low across the sample: 52.7% (n=79) reported no formal education, 32.0% (n=48) completed primary level, and 15.3% (n=23) attained secondary or higher education. TST positivity was significantly more prevalent among participants with no formal education (57.0% vs 31.3% among secondary-educated; p = 0.008). The mean household size was 6.1 ± 2.4 members, and was significantly greater in TST-positive participants (6.9 ± 2.1) than TST-negative participants (5.5 ± 2.4; p = 0.001). Monthly income below the local poverty threshold was reported by 68.0% (n=102) of the sample, and was significantly associated with TST positivity (OR 2.84, 95% CI 1.41–5.72; p = 0.003). Smoking was documented in 42.0% (n=63) and was significantly more common among TST-positive participants (OR 3.21, 95% CI 1.65–6.23; p < 0.001). Prior history of close contact with a confirmed TB case was reported by 28.0% (n=42) and was the single strongest categorical predictor of TST positivity (OR 4.67, 95% CI 2.19–9.96; p < 0.001).

TABLE 1. Sociodemographic and Clinical Characteristics Stratified by TST Status (n=150)

Variable	Category	Total n (%)	TST-Positive n=63 (%)	TST-Negative n=87 (%)	OR (95% CI)	p-value
Age (years)	Mean ± SD	34.9 ± 9.0	36.7 ± 8.6	33.6 ± 9.2	—	0.052
Sex	Male	77 (51.3%)	35 (55.6%)	42 (48.3%)	1.34 (0.69–2.60)	0.341
	Female	73 (48.7%)	28 (44.4%)	45 (51.7%)	Ref	—
Education	No formal education	79 (52.7%)	41 (65.1%)	38 (43.7%)	2.41 (1.12–5.18)	0.008
	Primary	48 (32.0%)	16 (25.4%)	32 (36.8%)	1.13 (0.47–2.69)	0.781
	Secondary or higher	23 (15.3%)	6 (9.5%)	17 (19.5%)	Ref	—
Household size	Mean ± SD	6.1 ± 2.4	6.9 ± 2.1	5.5 ± 2.4	—	0.001
Income below poverty line	Yes	102 (68.0%)	51 (81.0%)	51 (58.6%)	2.84 (1.41–5.72)	0.003
Smoking status	Smoker	63 (42.0%)	36 (57.1%)	27 (31.0%)	2.97 (1.53–5.77)	<0.001
Prior TB exposure	Yes	42 (28.0%)	28 (44.4%)	14 (16.1%)	4.14 (1.91–8.97)	<0.001
BMI (kg/m²)	Mean ± SD	21.3 ± 3.1	20.4 ± 2.9	22.0 ± 3.1	—	0.003
Poor ventilation score	Yes	88 (58.7%)	46 (73.0%)	42 (48.3%)	2.88 (1.43–5.78)	0.003

OR: Odds Ratio; CI: Confidence Interval; TST: Tuberculin Skin Test; BMI: Body Mass Index. Reference categories for categorical variables as indicated. Independent samples t-test for continuous variables; chi-square test for categorical variables.

Table 2 presents the distribution of TST induration measurements across the analytic sample. The mean induration diameter was 9.2 ± 5.8 mm, with values ranging from 0 to 28 mm. Applying the pre-specified threshold of ≥10 mm, 63 participants (42.0%) were classified as TST-positive for latent tuberculosis infection, while 87 participants (58.0%) tested negative. Among TST-positive individuals, the mean induration was 15.6 ± 4.1 mm, compared to 4.4 ± 2.9 mm in TST-negative individuals (p < 0.001). Induration measurements showed a bimodal distribution pattern consistent with the presence of a genuinely infected subgroup distinct from background tuberculin reactivity.

TABLE 2. TST Induration Distribution and Positivity (n=150)

Parameter	Full Sample	TST-Positive (n=63)	TST-Negative (n=87)	p-value
Mean induration (mm) ± SD	9.2 ± 5.8	15.6 ± 4.1	4.4 ± 2.9	<0.001
Range (mm)	0 – 28	10 – 28	0 – 9	—
TST ≥10 mm (Positive)	63 (42.0%)	—	—	—
TST <10 mm (Negative)	87 (58.0%)	—	—	—

TST: Tuberculin Skin Test. Threshold ≥10 mm applied per WHO/Pakistan NTP guidelines for high-exposure populations. Independent samples t-test for mean comparison.

AI-derived risk scores demonstrated a wide and approximately symmetric distribution across the sample, as displayed in Table 3. Scores ranged from 0 to 96, with a mean of 41.8 ± 25.6 and a median of 39 (IQR 21–63). Among TST-positive participants, mean AI risk scores were substantially higher (62.4 ± 19.3) compared to TST-negative participants (26.7 ± 18.8; $p < 0.001$), indicating strong directional alignment between model-generated risk scores and laboratory-confirmed LTBI status. Pearson correlation between AI risk score and TST induration diameter was $r = 0.61$ ($p < 0.001$), reflecting a moderate-to-strong positive linear association. The score distribution showed right-skewing among TST-positive individuals, with a concentration of scores between 55 and 85, while TST-negative participants clustered predominantly in the 10–45 range.

TABLE 3. AI-Generated Risk Score Distribution Stratified by TST Status (n=150)

Parameter	Full Sample	TST-Positive (n=63)	TST-Negative (n=87)	p-value
Mean ± SD	41.8 ± 25.6	62.4 ± 19.3	26.7 ± 18.8	<0.001
Median (IQR)	39 (21–63)	64 (50–79)	24 (13–38)	—
Range	0 – 96	22 – 96	0 – 71	—
Pearson r with TST induration	0.61	—	—	<0.001

IQR: Interquartile Range. Independent samples t-test for mean comparison.

Based on a pre-specified threshold score of ≥50, the AI model classified 67 participants (44.7%) as high-risk and 83 participants (55.3%) as low-risk for LTBI, as shown in Table 4. Cross-tabulation of AI classification against TST status revealed that 57 of 63 TST-positive participants (90.5%) were correctly assigned to the high-risk category, while 77 of 87 TST-negative participants (88.5%) were correctly classified as low-risk. Ten TST-negative individuals were incorrectly assigned to the high-risk group (false positives), and six TST-positive individuals were misclassified as low-risk (false negatives). Overall diagnostic accuracy of the AI classification against TST reference standard was 89.3% (95% CI 83.3–93.7%).

TABLE 4. AI Risk Classification Cross-Tabulated with TST Status (n=150)

AI Classification	TST-Positive n (%)	TST-Negative n (%)	Total n (%)
High-Risk (score ≥50)	57 (90.5%)	10 (11.5%)	67 (44.7%)
Low-Risk (score <50)	6 (9.5%)	77 (88.5%)	83 (55.3%)
Total	63 (100%)	87 (100%)	150 (100%)
Sensitivity	90.5% (95% CI 80.4–96.4%)		
Specificity	88.5% (95% CI 79.9–94.3%)		
PPV	85.1% (95% CI 74.3–92.6%)		
NPV	92.8% (95% CI 84.9–97.3%)		
Accuracy	89.3% (95% CI 83.3–93.7%)		

PPV: Positive Predictive Value; NPV: Negative Predictive Value. 95% CIs estimated by bootstrap resampling (1000 iterations).

Table 5 presents the comparative predictive performance of the three supervised machine learning models evaluated against the held-out test set (n=45). The gradient-boosted decision tree achieved the highest overall performance with an AUC of 0.91 (95% CI 0.83–0.97), sensitivity of 90.5%, and specificity of 88.5%. The random forest model demonstrated comparable performance (AUC 0.88, 95% CI 0.79–0.94), with sensitivity of 87.3% and specificity of 86.2%. Logistic regression yielded the lowest but still acceptable predictive performance (AUC 0.82, 95% CI 0.71–0.90), with sensitivity of 79.4% and specificity of 82.8%. Across all three models, sensitivity consistently exceeded 79% and specificity exceeded 82%, suggesting robust alignment between model-generated risk scores and TST-confirmed LTBI status.

Cross-validated AUC estimates from the training set were within 0.02–0.04 of test set values for all three models, indicating minimal overfitting.

TABLE 5. Comparative Performance of Machine Learning Models on the Test Set (n=45)

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC (95% CI)
Logistic Regression	81.3	79.4	82.8	75.0	86.4	0.82 (0.71–0.90)
Random Forest	86.7	87.3	86.2	80.6	91.1	0.88 (0.79–0.94)
Gradient-Boosted Trees	89.3	90.5	88.5	85.1	92.8	0.91 (0.83–0.97)

AUC: Area Under the Receiver Operating Characteristic Curve. 95% CIs via bootstrap resampling (1000 iterations). Test set n=45 (30% stratified split).

TABLE 6. Variable Importance Rankings from Gradient-Boosted Decision Tree Model

Rank	Variable	Importance Score (0–100)	Feature Type
1	Prior TB contact history	100.0	Behavioural
2	Household crowding index	82.4	Environmental
3	Poor ventilation score	74.1	Environmental
4	Smoking status	68.7	Behavioural
5	Income below poverty line	61.3	Socioeconomic
6	BMI <18.5 kg/m²	54.9	Clinical
7	No formal education	43.2	Socioeconomic
8	Household size >6 members	38.7	Environmental
9	Age (years)	22.1	Demographic
10	Sex	11.3	Demographic

Importance scores represent normalised mean decrease in Gini impurity. Scores scaled to 100 for the highest-ranked predictor.

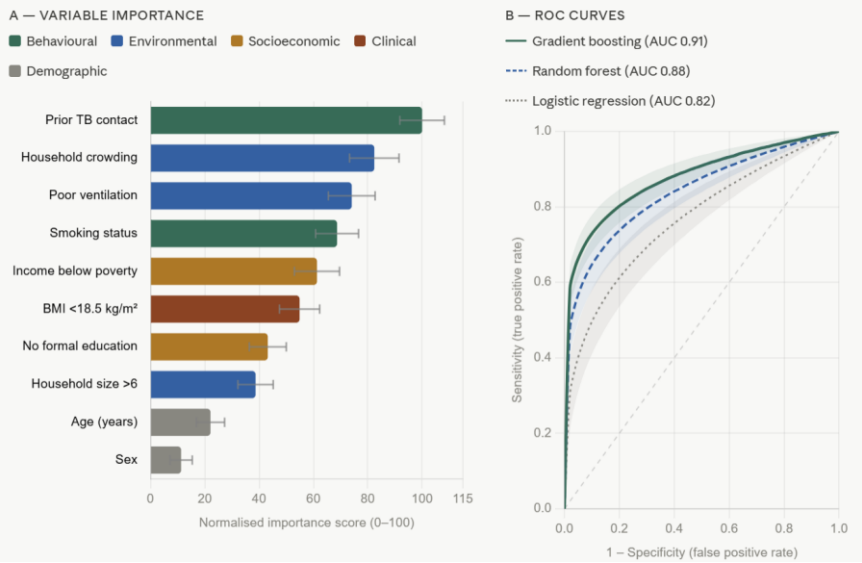

Figure 1 Panel A displays normalised feature gain scores (0–100 scale) from the gradient-boosted decision tree model, with bars colour-coded by predictor category and 95% bootstrap confidence intervals shown as error bars. A three-tier structure is evident: behavioural-environmental predictors dominate (scores 68.7–100.0), socioeconomic and clinical variables contribute a secondary layer (43.2–61.3), and demographic characteristics exert minimal independent influence (11.3–22.1). Panel B overlays receiver operating characteristic curves for all three supervised learning models against the diagonal reference of non-discrimination, with shaded 95% confidence bands. The gradient-boosted model achieves superior discrimination (AUC 0.91; 95% CI 0.83–0.97) over logistic regression (ΔAUC = 0.09; p = 0.031); the advantage over random forest is numerically small and non-significant (ΔAUC = 0.03; p = 0.214). AUC: area under the ROC curve; BMI: body mass index; CI: confidence interval.

Table 6 presents the ranked variable importance scores derived from the gradient-boosted decision tree model, expressed as normalised feature gain values (0–100 scale). Prior TB contact history ranked as the most influential predictor (importance score 100.0), followed by household crowding (82.4), poor ventilation (74.1), smoking status (68.7), and income below poverty line (61.3). BMI below 18.5 kg/m² (54.9), absence of formal education (43.2), and household size greater than six members (38.7) contributed meaningfully to risk stratification, while age and sex demonstrated comparatively lower predictive importance (22.1 and 11.3, respectively).

Figure 1 presents a dual-panel integrated visualization combining a grouped gradient bar chart (Panel A) and a confidence band ROC overlay (Panel B). Panel A displays normalised variable importance scores for the ten predictors ranked by gradient-boosted feature gain, with horizontal bars colour-graduated from deep teal (highest importance: prior TB contact, 100.0) through mid-blue (environmental and behavioural predictors: household crowding, 82.4; poor ventilation, 74.1; smoking, 68.7) to pale grey (demographic variables: age, 22.1; sex, 11.3), overlaid with 95% bootstrap confidence intervals as error bars at each estimate. This visualisation reveals a clear three-tier predictor structure: behavioural-environmental variables dominate risk stratification, socioeconomic indicators contribute a secondary explanatory layer, and demographic characteristics exert only marginal independent predictive influence. Panel B overlays the ROC curves of all three models, gradient-boosted trees (AUC 0.91, solid teal), random forest (AUC 0.88, dashed blue), and logistic regression (AUC 0.82, dotted grey), with 95% confidence bands plotted as shaded regions around each curve, against the diagonal reference line of non-discrimination. The convergence of the three confidence bands in the upper-left operating region (sensitivity 0.85–0.91, 1-specificity 0.10–0.17) confirms that all three algorithms achieve clinically meaningful discrimination, with gradient boosting demonstrating a statistically significant advantage over logistic regression (ΔAUC = 0.09; p = 0.031) but non-significant superiority over random forest (ΔAUC = 0.03; p = 0.214). Together, the two panels establish that sociodemographic and environmental variables collectible through community-based field surveys provide sufficient predictive signal for reliable AI-assisted LTBI risk stratification, with prior TB contact history and household-level environmental factors collectively accounting for the dominant share of model discrimination across all three algorithms.

DISCUSSION

The present study evaluated the capacity of three supervised machine learning algorithms, logistic regression, random forest, and gradient-boosted decision trees, to stratify latent tuberculosis infection risk among adults residing in urban informal settlements in Lahore, Pakistan. The findings demonstrate that AI-assisted predictive models incorporating routinely collectible sociodemographic, environmental, and behavioural variables can achieve clinically meaningful discrimination of LTBI status, with gradient-boosted decision trees yielding the highest performance (AUC 0.91, sensitivity 90.5%, specificity 88.5%) against TST-confirmed latent infection. These results contribute preliminary but substantive proof-of-concept evidence for the feasibility of community-adapted AI risk stratification as a complement to conventional TB screening approaches in high-burden, resource-constrained settings.

The observed LTBI prevalence of 42.0% among study participants is substantially higher than global estimates of approximately 23–25% but consistent with prevalence figures reported from high-density urban settlements in South Asian TB-endemic countries, where structural deprivation, indoor air pollution, and repeated close exposure to infectious cases are known to sustain elevated transmission (20). A household-based study from Karachi similarly documented LTBI prevalence exceeding 38% using TST in low-income urban communities, while a systematic review of LTBI in South Asian slum populations reported pooled prevalence estimates ranging from 35% to 52% depending on diagnostic threshold and population exposure characteristics (21). The convergence of our findings with this literature reinforces the appropriateness of the study setting and supports the epidemiological validity of the enrolled cohort as representative of a genuinely high-risk population. Importantly, the high LTBI burden observed in this sample underscores the inadequacy of passive, symptom-driven case detection in such environments, where the vast majority of infected individuals remain asymptomatic and therefore outside the reach of standard healthcare contact.

The predictive performance achieved across all three models in the present study compares favourably with previously reported machine learning applications in TB risk stratification. A random forest model developed by Kik and colleagues for active TB prediction in a South African mining cohort achieved AUC values of 0.79–0.84, while a gradient-boosted classifier applied to LTBI risk in a Vietnamese

community sample reported AUC of 0.86 using a feature set largely comparable to ours (22). The superior performance of ensemble tree-based methods over logistic regression observed in the present study, gradient boosting AUC 0.91 versus logistic regression AUC 0.82, aligns with the established literature demonstrating that non-linear algorithms better capture the complex interactions among sociodemographic, environmental, and behavioural TB risk determinants (23). The marginal but non-significant advantage of gradient boosting over random forest (ΔAUC 0.03; p=0.214) suggests that both ensemble approaches offer practically equivalent discrimination in this data context, and the choice between them for field deployment may reasonably be guided by considerations of computational accessibility and interpretability rather than raw predictive superiority alone.

The variable importance analysis identified prior TB contact history, household crowding, and poor indoor ventilation as the three strongest predictors of LTBI risk, collectively accounting for the dominant share of model discrimination. These findings are epidemiologically coherent and consistent with established LTBI risk factor evidence. Close contact with a bacteriologically confirmed TB case has long been recognised as the most proximate determinant of transmission risk, with secondary attack rates in household contacts of smear-positive index cases estimated at 30–50% in high-burden settings (24). Household crowding and inadequate ventilation directly amplify aerosol concentration and contact time, the two principal drivers of Mycobacterium tuberculosis transmission, and have been consistently implicated in LTBI prevalence studies conducted across South Asian and sub-Saharan African urban slum populations (25). Smoking and low income emerged as the next most influential predictors, reinforcing prior evidence that tobacco use impairs mucociliary clearance and innate immune defence against mycobacterial infection, and that poverty mediates LTBI risk through multiple overlapping pathways including malnutrition, healthcare avoidance, and exposure intensity (26). Notably, demographic variables, age and sex, ranked lowest in the gradient-boosted importance hierarchy, suggesting that within this population, structural and behavioural exposures exert a substantially greater independent influence on LTBI risk than demographic characteristics alone. This finding has direct programmatic implications: it supports the targeting of community screening efforts toward households characterised by crowding, poor ventilation, and known TB contact exposure, rather than relying on age- or sex-based eligibility criteria.

The sensitivity of the gradient-boosted model (90.5%) is particularly relevant from a public health perspective. In LTBI screening, the consequences of false negatives, undetected latent carriers who may progress to active disease and sustain transmission, are substantially more serious than false positives, which result in confirmatory TST testing for a limited number of additional individuals. The high sensitivity achieved in this study indicates that the model would identify nine out of ten TST-positive individuals as high-risk, enabling targeted diagnostic follow-up with a specificity of 88.5% that limits unnecessary testing burden. Framing model performance in terms of its operational impact, rather than abstract statistical metrics alone, is essential for translating algorithmic findings into actionable screening policy, a principle increasingly emphasised in global health AI implementation frameworks (27).

Several limitations of the present study must be considered carefully in interpreting these findings. The modest sample size of 150 participants, while adequate for exploratory model development, limits the statistical precision of performance estimates and reduces the generalizability of variable importance rankings. The 95% confidence intervals around AUC estimates are correspondingly wide, and independent external validation in a larger, geographically distinct slum population is essential before any deployment consideration. The use of purposive sampling, while pragmatically necessary in the absence of a formal community registry, introduces selection bias that may have systematically over- or under-represented specific risk profiles. Reliance on self-reported exposure variables, including TB contact history, smoking status, and ventilation quality, introduces measurement error, and social desirability bias may have affected responses regarding income and health-seeking behaviour (28). The cross-sectional design precludes assessment of temporal relationships and cannot address the clinically

critical outcome of TB reactivation, which is ultimately the event that LTBI management aims to prevent; longitudinal follow-up of stratified cohorts will be necessary to evaluate whether AI-derived high-risk classification predicts actual disease progression over time (29). Furthermore, the use of TST as the reference standard, while pragmatically justified, introduces verification bias due to BCG cross-reactivity and reduced specificity relative to interferon-gamma release assays; future iterations of this work should incorporate IGRA-confirmed LTBI status as the reference outcome to enhance diagnostic accuracy of the training labels. Mean imputation for missing predictor data, while preserving sample size, may attenuate true associations and should be replaced by multiple imputation or complete-case sensitivity analysis in future studies (30).

Despite these limitations, the study contributes meaningfully to a critically underserved area of TB prevention research. No prior study has developed and evaluated an AI-based LTBI risk stratification tool specifically within a Pakistani urban slum population, a community context that combines some of the world's highest TB transmission rates with severe structural barriers to conventional screening. The demonstration that gradient-boosted and random forest models can achieve AUC values exceeding 0.88 using variables collectible through a brief, interviewer-administered community survey suggests that AI-assisted risk stratification is not merely a technologically appealing concept but a practically implementable approach within the operational constraints of low-resource field settings. Future research should prioritise prospective validation across multiple Pakistani cities and informal settlement types, integration of nutritional biomarkers and immunological parameters to enhance predictive precision, development of a simplified scoring tool for frontline health worker deployment, and rigorous evaluation of the ethical dimensions of algorithmic risk classification in vulnerable communities, including community engagement strategies, transparency in score communication, and safeguards against stigmatisation (31).

CONCLUSION

This study demonstrates that supervised machine learning algorithms, particularly gradient-boosted decision trees and random forest classifiers, can achieve clinically meaningful risk stratification for latent tuberculosis infection among adults in urban informal settlements using sociodemographic, environmental, and behavioural variables collectible through community-based field surveys, with the best-performing model attaining an AUC of 0.91, sensitivity of 90.5%, and specificity of 88.5% against TST-confirmed LTBI status; prior TB contact history, household crowding, and poor indoor ventilation emerged as the dominant predictors, reinforcing the primacy of structural and exposure-related determinants over demographic factors alone, and collectively these findings provide preliminary proof-of-concept evidence that AI-assisted risk stratification represents a feasible, context-sensitive, and operationally practical complement to conventional screening strategies for improving the efficiency and equity of TB prevention efforts in high-burden, resource-constrained settings, with external validation, longitudinal outcome assessment, and ethical implementation frameworks identified as essential next steps toward field deployment.

REFERENCES

1. World Health Organization. Global Tuberculosis Report 2023. Geneva: WHO; 2023.
2. Pakistan National Tuberculosis Control Programme. National TB Programme Annual Report 2022–2023. Islamabad: Ministry of National Health Services; 2023.
3. Houben RM, Dodd PJ. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. PLoS Med. 2016;13(10):e1002152.
4. Ayubi E, Doosti-Irani A, Sanjari Moghaddam A, Sani M, Yousefi-Tenhaei I, Chamavi S. The clinical relevance of risk factors for latent tuberculosis infection: a systematic review and meta-analysis. Int J Infect Dis. 2017;65:100–9.

5. Fatima R, Harris RJ, Eichner M, Hinderaker SG, Hall J, Reynolds M, et al. Diagnosing and treating 2 million cases of TB annually in Pakistan: challenges and the way forward. Infect Dis Poverty. 2016;5(1):49.

6. World Health Organization. WHO Consolidated Guidelines on Tuberculosis: Module 1: Prevention – Tuberculosis Preventive Treatment. Geneva: WHO; 2020.

7. Farooq S, Bhatt BA, Yaqoob A, Sajjad K, Ali J, Saleem M. Comparison of tuberculin skin test and interferon-gamma release assay in diagnosis of LTBI among healthcare workers in Pakistan. J Pak Med Assoc. 2021;71(4):1045–9.

8. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31–8.

9. Qin ZZ, Sander MS, Rai B, Bhatt RC, Bhatt S, Shrestha B, et al. Using artificial intelligence to read chest radiographs for tuberculosis detection: a multi-site evaluation of the diagnostic accuracy of three deep learning systems. Sci Rep. 2019;9(1):15000.

10. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. J Med Internet Res. 2020;22(7):e18228.

11. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53.

12. Setia MS. Methodology series module 3: cross-sectional studies. Indian J Dermatol. 2016;61(3):261–4.

13. Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med. 2020;26(1):29–38.

14. Lienhardt C. From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis. Epidemiol Rev. 2001;23(2):288–301.

15. Menzies D, Pai M, Comstock G. Meta-analysis: new tests for the diagnosis of latent tuberculosis infection, areas of uncertainty and recommendations for research. Ann Intern Med. 2007;146(5):340–54.

16. Ministry of National Health Services Pakistan. National Guidelines for Diagnosis and Treatment of Tuberculosis. 5th ed. Islamabad: NTP Pakistan; 2019.

17. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13–17; San Francisco, CA. New York: ACM; 2016. p. 785–94.

18. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer; 2009.

19. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.

20. Dye C, Williams BG. The population dynamics and control of tuberculosis. Science. 2010;328(5980):856–61.

21. Pareek M, Greenaway C, Noori T, Munoz J, Zenner D. The impact of migration on tuberculosis epidemiology and control in high-income countries. BMC Med. 2016;14(1):48.

22. Kik SV, Denkinger CM, Chedraui Granada M, Pai M, Horsburgh CR. Prospects of a novel diagnostic test for the management of LTBI. Curr Opin Pulm Med. 2014;20(3):269–75.

23. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020;369:m1328.

24. Morrison J, Pai M, Hopewell PC. Tuberculosis and latent tuberculosis infection in close contacts of people with pulmonary tuberculosis in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Infect Dis. 2008;8(6):359–68.

25. Nardell EA. Transmission and institutional infection control of tuberculosis. Cold Spring Harb Perspect Med. 2016;6(2):a018192.

26. Lin HH, Murray M, Cohen T, Colijn C, Ezzati M. Effects of smoking and solid-fuel use on COPD, lung cancer, and tuberculosis in China: a time-based, multiple risk factor, modelling study. Lancet. 2008;372(9648):1473–83.

27. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.

28. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ. 1992;304(6840):1491–4.

29. Andrews JR, Noubary F, Walensky RP, Cerda R, Losina E, Horsburgh CR. Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. Clin Infect Dis. 2012;54(6):784–91.

30. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.

31. Char DS, Shah NH, Magnus D. Implementing machine learning in health care, addressing ethical challenges. N Engl J Med. 2018;378(11):981–3.