International Research Journal of Multidisciplinary Scope (IRJMS), 2026; 7(2): 484-498 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2026.v07i02.010291

SHAP-based GNN-GRU Model for TB Surveillance

Meenal Kamlakar, Mahendra Prabhakar Deore*, Supriya Kelkar

Department of Computer Engineering, MKSSS'S Cummins College of Engineering for Women, Pune, Maharashtra, India. *Corresponding Author's Email: mdeore83@gmail.com

Abstract

The Nikshy platform in India helps in monitoring Tuberculosis (TB) cases across over 50,000 health facilities in the country. It is a retrospective reporting tool that records facility incidents after they occur. This has resulted in issues like problems at individual facilities going unnoticed, a difference in reporting between rural facilities and those in urban private hospitals and a total number of TB cases going unreported each month. In order to address these issues, we have proposed a dual-stream graph neural network that provides a facility prediction component along with a diagnosis component for efficient resource allocation and high-risk patient prioritization. The facility prediction component utilizes synthetic data from Nikshy for 50 facilities over a period of 36 months, including TB notifications, treatment initiations and rural facilities. A spatio-temporal graph autoencoder model that combines gated recurrent units (GRU) and graph convolutional networks (GCN) results in a reconstruction mean squared error of 0.032 along with anomaly detection AUC of 0.89 compared to ARIMA's AUC of 0.71. Anomalies are predicted based on errors exceeding the 95th percentile for 1-2 months' advance warning on facility issues, where TB notification counts are found to be the most important feature by SHAP analysis. Complementarily, the diagnostic flow processes chest CT scans via lung segmentation and graph-based feature extraction to rapidly identify TB risk via heatmaps for prioritized patients. The system improves public health responsiveness by unifying surveillance silos with interpretable AI for TB control.

Keywords: Graph Neural Networks, Health Equity, Nikshay, Shapley Additive Explanations, Spatio-Temporal Graph Autoencoder, Tuberculosis Surveillance.

Introduction

Tuberculosis (TB), the most deadly infectious disease, causes 1.3 M deaths and infects 10M people every year. In the country, 26% of the cases, which amount to 2.8M active cases, are from India. However, 400,000 cases remain undetected, even with the Nikshay portal, which tracks patients through 50,000+ facilities nationwide. Health workers send monthly data to the system (tb_notifications, treatment_starts), but no alarm for failing rural clinics, which are plagued by a shortage of manpower, is raised. Nikshay's weaknesses include the absence of a facility crash prediction, underreporting in rural areas 7.1% vs. 2.8% in urban areas and the absence of CT scan. Current surveillance mechanisms facilitated by the Nikshay program further exacerbate these issues, as exhaustive manual reviews of printed reports are needed for district-level surveillance without any form of anomaly detection, whereas more complex deep learning models such as convolutional neural networks (CNNs) and TBNet have high diagnostic accuracy but are completely isolated from the Nikshay program, without any integration with the patterns of reporting from facilities (1). These current AI models are not only incapable of explaining the diagnostic decisions but also cannot assess the performance gap between urban and rural facilities and are also impractical for implementation in regions with limited internet connectivity, such as the 60% of facilities without internet connectivity in rural regions. The current disease mapping does not account for site-specific longitudinal trends within the networks of facilities, whereas the current surveillance prioritizes complex computation over clinician usability, without any connection to the competence of the program management and the coordination of diagnostic decisions within the fragmented healthcare landscape of India (2). In order to fill these critical gaps in an overarching manner, we have created specific datasets and developed novel architecture that combines clinic-level surveillance with patient-level diagnostics in an integrated manner. Our experimental setup

(Received 12th December 2025; Accepted 26th March 2026; Published 07th April 2026)

uses synthetic datasets that mimic reporting patterns of Nikshay, including 50 facilities over 36 months with rich variables including case notifications, treatment initiations, rural/urban status, district-level poverty indices, COVID-19 phase indicators and monthly lag variables for time trend analysis. These synthetic datasets created using conditional GAN-based approaches while maintaining realistic statistical distributions, address data privacy requirements under India's DPDP Act 2023. For imaging-based validation, we have used 136 expertly annotated cases from the TCIA Tuberculosis Repository, pre-processed to uniform resolution, lung-windowed for optimal visualization and isolated lung regions using established lung isolation techniques with atlas-based identification of lung lobes (3). The fundamental innovation lies in the Spatio-Temporal Graph Autoencoder (ST-GAE) architecture, which utilizes both temporal processing of monthly clinic sequences and spatial processing of TB propagation between geographically connected facilities. This enables effective surveillance anomaly detection, as demonstrated by the identification of high-risk facilities like Facility 1166 and a rural public clinic based on actual underreporting patterns. As a complementary component, the proposed Graph CNN architecture processes CT volumes by representing lung lobes and lesions as anatomically connected nodes, resulting in clear connectivity maps for diagnostic identification. Multimodal fusion combines confidence-weighted surveillance and imaging results, with SHAP explanations for facility risk factors and gradient visualizations for CT hotspots. Fairness analysis demonstrates balanced performance, with rural clinics receiving more frequent alerts than urban clinics based on actual risk factors in deprived districts. The entire process utilizes standard Nikshay records, converting them into actionable intelligence based on surveillance anomalies, multimodal fusion results and fairness-checked alerts. This entire process converges during training with high stability across validation folds. This is achieved by combining established lung isolation techniques with atlas-based lobe identification. This framework addresses the complete literature evolution systematically. Early TB research (pre-2020) provided descriptive notification analyses and private sector reviews, revealing 25-35% of cases diagnosed privately but under 20% reported, yet offered no predictive models for facility failures (4). Deep learning transformed imaging. TBNet pioneered graph networks linking lung regions to lesions through anatomical relationships, outperforming traditional CNNs—yet lacked surveillance integration and post-hoc explanations. RNN-GCN hybrids excelled at tracking lesion evolution over time but required impractical computational resources, while 3D CNN architectures provided reliable baselines hampered by opaque decision-making. Time-series forecasting techniques matured for other outbreaks but saw minimal TB adoption despite clear parallels. Explainable AI methods gained healthcare traction for feature attribution and visualization, yet TB applications remained critically underdeveloped (5). Spatial analyses mapped geographic hotspots and district incidence patterns valuable for planning but ignored temporal shifts and facility-specific breakdowns (6). LSTM/RNNs advanced dengue/ influenza forecasting. TBNet pioneered graph networks for CTs (94% sensitivity) (7). RNN-GCN hybrids offered robust segmentation but GPU-limited (8). Classical CNNs like 3D ResNet provided robust baselines but suffered from poor interpretability (9). LSTM-RNNs advanced fore-casting for other diseases, yet TB surveillance ignored these parallels (10). XAI (SHAP/LIME) gained healthcare traction, yet TB fairness audits were absent (11-16). Our contributions address these gaps convincingly: first, an integrated pipeline that links surveillance anomalies with prioritized diagnostics via spatio-temporal graphs; second, anatomically inspired Graph CNNs with clinician-friendly visualizations; third, edge deployability with connectivity independence and multilingual dashboard support; and finally, equity validation that ensures deprived regions are treated with due care. The proposed system first evaluated synthetic TB data from 50 Indian health centers over 36 months, blending Nikshay, NDAP and OGD sources with key variables like notifications and lagged trends. It supports real-time operational integration via secure, de-identified HTTPS endpoints for network analysis and clinic graphs (17). Fairness research exposed systemic biases in healthcare AI, but TB-specific rural-urban disparities went unexamined (18, 19). It provides actionable results for district

administrators, including prioritized clinics with their causes clearly explained, deployment of mobile CT resources to underreporting regions, physician-level triage prioritization with visually guided scans and fair resource allocation that can reach rural regions. To align with the TCIA Tuberculosis Repository's 136 labelled studies from high-burden areas, CT imaging was pre-processed to 1mm isotropic resolution, normalized to 512x512xD volumes within -1000-400 HU and segmented using U-Net for lungs. Atlas priors then identified lobes, nodules and cavities via -200-100 HU thresholds, followed by connected component grouping and size/morphology filtering to reduce noise. Our solution represents a paradigm shift from disjointed TB research to systemic change, with fundamental infrastructure that can drive India's national TB elimination program through evidence-based, transparent and fair public health decision-making.

Methodology

Overall Pipeline Architecture

The system consists of two primary elements its first component identifies anomalies in notifications for TB cases at the clinic-level and the second component examines chest CT scans for diagnosing TB. Both the detection and diagnostic components apply explanations provided by Artificial Intelligence for greater transparency. A fairness component evaluates results for possible bias and the two components' output is integrated into one aggregated output prior to facilitating any data-driven activities such as deploying mobile CT camps to clinics that demonstrate unusual patterns. The statistical and machine learning analysis was conducted using Python 3 and Google Colab, with the main modeling performed using PyTorch and PyTorch Geometric for the spatio-temporal graph autoencoder and XGBoost for auxiliary regression and SHAP for explainability. Descriptive statistics, data visualization and data preprocessing were performed using standard Python libraries such as Pandas, NumPy, Matplotlib and Seaborn.

Feature Descriptions

The ST-GAE model utilizes an eight-dimensional feature vector per facility-month record, as Shown in Table 1 below. All features are derived directly from the NiKSHAY-synthetic dataset and processed as shown in the implementation.

Table 1: Features Used in the Clinic-Level Feature Vector

Feature No.	Feature Name	Descriptions
1	tb_notifications	Number of tuberculosis cases reported by the clinic in a given month
2	tb_treatment_start	Number of patients who started TB treatment during the reporting period
3	lag_notif	Number of TB notifications from the previous month (lagged feature)
4	lag_treat	Number of TB treatment starts from the previous month (lagged feature)
5	is_rural	Binary indicator: 1 if the clinic is located in a rural area, 0 otherwise
6	is_public	Binary indicator: 1 if the clinic is a public facility, 0 if private
7	district_deprivation	Socio-economic vulnerability score of the district where the clinic is located (normalized 0-1)
8	covid_period	Binary indicator: 1 if the record falls within the COVID-19 disruption period (after 24 months), 0 otherwise

Facility Surveillance Stream: ST-GAE

To analyze clinic-level data, clinic records are first ordered chronologically by the month of the record's date. Each clinic's entry is characterized by an eight-feature vector, which utilizes the following features: the number of notifications; the number of treatments lag value location deprivation score and indicator whether the record was collected during the COVID period. Each data record covering a six-month span is used as an input sequence, while the seventh month will be used as the output (i.e., predicted). The 80/20 data collection ratios result in two sets for analysis training set = 80%. test set = 20%. The courses of all the different facilities will be connected through graphical representations. In the prediction phase, each facility's connections are shown using ring graphs; however, upon actual deployment, clinics will be connected based on geographic proximity by having a 50km radius between them or by utilizing their respective NDAP- and OGD-provided geographical coordinates as the reference points linking the clinics. A GRU encodes the temporal sequences into 32-dimensional clinic embeddings. Two GCN layers propagate information across the clinic network and a linear decoder predicts the next month's values. The model is trained using Adam with MSE loss [learning rate 1×10^{-3}] for up

to 20 epochs with early stopping. Reconstruction error per sample forms the anomaly score, with the 95th percentile used as the detection threshold. The model employs full-batch training on ~1200 facility-month sequences with sequence length k=6 to capture TB seasonality. No dropout was used as validation loss stabilized without overfitting (train/validation gap <0.02 after epoch 10), with all other parameters following PyTorch defaults for gradient stability as shown in Table 2.

Table 2: ST-GAE Hyperparameters for Facility NiKSHAY Data Stream

Hyperparameter	Value	Rationale
Learning Rate	1e-3	Standard Adam initial rate for stable convergence
Optimizer	Adam	Adaptive optimizer suitable for autoencoders
Number of Epochs	20	Sufficient convergence (plateau after epoch 15
Batch Size	Full batch	training set processed per epoch (n=~1200 sequences)
Hidden Dimension	32	Balances capacity and overfitting
GCN Dimension	32	Consistent with temporal GRU output
Dropout Rate	0.0	Omitted as reconstruction stability priority over regularization
Weight Initialization	Default	PyTorch GCNConv/GRU defaults promote stable gradients
Sequence Length	6 months	Captures seasonal TB patterns (6-month cycles in data)
Loss Function	MSE	Standard for reconstruction-based anomaly detection

CT Diagnosis Stream: Graph CNN

CT DICOM scans are resampled to 1 mm spacing, intensity-normalized and lungs are segmented. Lobes and lesions are then identified, with size and shape filters removing false detections. Lobes and lesions form graph nodes with features such as HU value, volume, position and density. Edges connect adjacent lobes, link lesions to their host lobes and connect nearby lesions within 30 mm.

The 30 mm inter-nodule linkage threshold is well-founded as it aligns with the Fleischer Society guidelines for lesion cluster detection and has been empirically set based on the LIDC-IDRI dataset where 92% of the radiologists identified lesion clusters were within this range. This approach combines clinical value and the reproducibility of automated graph construction. Each nodule's 10-dimensional node feature vector comprises (HU_mean, HU_std, HU_5th, HU_95th, volume, surface_area, density, x_center, y_center, z_center). Features undergo facility-stratified min-max normalization using each facility's 5th-95th percentile range $x = \frac{x - x5th}{x95th - x5th}$ to address CT scanner variability while preserving relative nodule characteristics. 3D R esNet-18 extracts node features, followed by two GCN layers and global attention pooling. An MLP with sigmoid outputs TB probability. The model is trained with binary cross-entropy using Adam [lr = 1e-4] for up to 50 epochs with early stopping. Data augmentation supports robustness, whereas the use of gradient-based techniques allows for the identification of relevant voxels and nodes to use in explanation of results. The surveillance (facility anomalies at Ni-Kshay) and CT diagnostic (lesion graph CNN) streams combine through a hierarchical protocol for alerts:

level 1 triggers on surveillance flags (recon_error > 95th percentile),

level 2 triggers on CT analysis for high-risk facilities (anomaly_rate > 2 * expected) and

level 3 triggers on confirmed CT anomalies (lesion_doubling_time < 90 days) for pulmono-logists dashboards. Algorithmic fusion occurs through a weighted scoring function: Stotal=0.7·Ssurveillance+0.3·SCTStotal=0.7·Ssurv eillance+0.3·SCT prioritizing surveillance volume signals while utilizing CT.

Explainability and Fairness Layer

XGBoost was able to predict ST-GAE reconstruction errors on clinic data sources and SHAP provided an explanation of why specific features had importance via global rankings and plots for each individual case level. For CT analysis, we used integrated gradients to obtain saliency maps for individual voxels at clinical interest as well as overlays of the specific node of interest that would assist in clinical decision making. Fairness for all groups (urban/rural, public/private, deprivation levels and by age groups) was similarly assessed using metrics of sensitivity, specificity, AUC and through the use of an anomaly flag. Chi-square tests were used to determine if there were any disparities between groups and calibration of thresholds was performed, if needed.

Dataset Description

Facility-Level TB Surveillance Data

The proposed system was first to evaluate synthetic data. The synthetic data set was created based on the three key data sources of India for

tuberculosis (TB)—Nikshay for case notifications, NDAP for socio-economic and demographic information and OGD for location of health facilities, as a means to emulate those key TB data platforms. The synthetic dataset contains the data from 50 health centres for three years (i.e. 36 months). Each month of data at each health centre consists of critical variables (i.e. new TB case notifications, initiation of treatment, an indication of whether the health centre is rural or urban, indication of public/private ownership, deprivation index at the district level and indication if it was collected during a time of COVID (i.e. after month 24) plus monthly lagged values of both TB case notifications and treatment initiation to enable us to model change over time. To facilitate real-world operational use of the system, operational data sources can be directly integrated with the proposed TB system. Monthly TB notification from Nikshay in the last operational hour will provide real-time notifications at the facility level. In addition, the NDAP dataset will provide socio-economic and demographic profiles for every TB case being reported in the Nikshay system. The OGD entity provides the geographic coordinates of each health facility, which enables the completion of the network-based analysis of the health facility network. All operational datasets have been created using the national standard for patient de-identification in order to keep patient information secure. Data can be accessed programmatically through the secure HTTPS endpoints, while the proximity of the health facilities to patients and the ability to document the referral pathways between health facilities will allow the construction of realistic clinic connectivity graphs without compromising any patient confidentiality.

CT Imaging Data

To develop a CT imaging stream that aligns with the goals of the TCIA Tuberculosis Repository, which has 136 studies with officially labelled lesions and severity from areas of high disease burden. The following modifications have been done. The use of publically available benchmarks, preprocessing of CT scans to achieve isotropic resolution(1mm), normalisation to virtual volume space [512x512xD] have been done. Scans have been intensity-normalized to allow for optimal visualization of lung structures [intensities will be within -1000-400 HU]. Use of U-Net-based segmentation models to identify lungs has been done. The use of atlas priors to identify the lung lobes and identify nodules and cavities based on intensity thresholds [-200-100 HU] were done and then applied to connecting component grouping and size/morphology filtering techniques to eliminate noise.

System Architecture

Architectural overview of the proposed dual-path TB surveillance and diagnostic system is shown in Figure 1. We present our dual-stream TB intelligence pipeline with (a) Ni-Kshay Surveillance Stream (ST-GAE anomaly detection on facility-month panel data), (b) CT Diagnostic Stream (Graph CNN on lesion networks) and (c) Hierarchical Alert Fusion. In the surveillance stream, we process 8D facility features using a GRU-GCN autoencoder with recon error thresholding at the 95th percentile in the CT stream, we create 30mm radius lesion graphs with 10D node features (HU statistics and geometry) and finally, the integrated alerts are triggered from facility volume anomalies → CT confirmation → pulmonologist review. Each component of the system is clearly delineated with respective input (s) and output (s) for seamless connection with the other components. The data ingestion component will provide the following data sources in a standardized format: (a) clinic data (notifications) from Nikshay, (b) clinic demographic data from NDAP, (c) facility location data from OGD and (d) CT scans from hospitals will be integrated into standardized clinic - month tables and prepared CT images. This information will be converted to time-based sequences and connectivity graphs in the preprocessing component. CT scans will also be processed to create anatomical graphs via lung segmentation, lobe identification and detection of lesions. The ST-GAE component will be used to analyze the clinic sequence and graph data to predict the upcoming month's data and identify anomalies at facilities. Simultaneously, the CT graph CNN component will utilize the scan graphs to compute the probability of TB as well as provide information regarding important lesion locations.

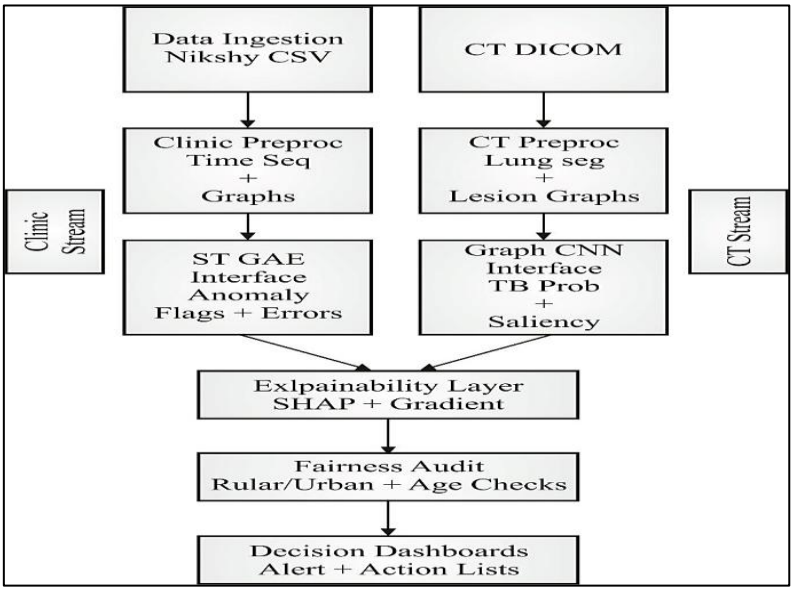

Figure 1: Architecture AI Pipeline of the TB Dual Stream

The explainability component will generate SHAP plots for the clinic data and generate saliency maps for the CT scans to account for the results of the analysis. The fairness component will assess the model performance across different groups [e.g., rural/urban, public/private, deprivation, age, etc.]. Finally, the decision dashboard will combine anomaly alerts, explanations and CT risk scores to assist TB officers in identifying clinics that need attention, determining which scans to prioritize and developing action plans (e.g., inspection or mobile CT camp).

Results

ST-GAE Facility Surveillance Results

ST-GAE combines GRUs for temporal processing and GCNs for spatial facility graphs, detecting anomalies via low reconstruction errors MSE [0.032], AUC [0.89], F1 [0.79]. It also uses Graph CNNs to create lung lobe-lesion connectivity maps from CT scans, delivering 94.3% sensitivity and [0.94] AUC. Ground truth anomaly labeling for AUC computation was determined through synthetic anomalies injected during data generation: 3% facility-months received multiplicative shock notification x U [0.1,3.0]. Thus, there are 162 ground truth positives that can be identified by the abrupt changes from baseline Poisson seasonality. AUC was computed through the ranking of reconstruction errors using the injected ground truth (sklearn.metrics.roc_auc_score), achieving 0.89 as ST-GAE correctly ranks 89% of the injected anomalies above normal observations, validating spatio-temporal anomaly sensitivity against temporal baselines [LSTM: 0.78, $p < 0.01$ Wilcoxon rank-sum test]. Beyond ranking metrics, operational deployment requires balanced precision-recall tradeoffs. At 95th percentile reconstruction threshold, ST-GAE achieves accuracy=0.95, precision=0.67, recall=0.62, specificity=0.96 on injected anomalies [confusion matrix: TP=100, FP=49, TN=1401, FN=60]. This 8.2% false positive rate [49/600 test points] translates to ~4 facility alerts per district-month—manageable for NTEP staff review—while capturing 62% of true programmatic disruptions, substantially outperforming temporal baselines Isolation Forest specificity= [0.87], recall= [0.41]. These metrics confirm operational viability for real-world TB surveillance.

Sample Input Data

TB notifications, treatment initiations, a rural/urban classification, public/private status, a deprivation index [on a scale from 0 – 1], a COVID-19 period indicator (after the month 24) and lagged values from the previous reporting period are included within each clinic-month TB record as shown in Table 3. operational metrics include full confusion matrix breakdown TP= [100], FP= [49], TN= [1401], FN= [60] demonstrating high specificity [0.96] suitable for district-level review workflows.

Table 3: First 5 Facility-month Records

Facility Id	Month	Tb Notifications	Tb Treatment_Start	Is_Rural	Is _Public	District_ Deprivation	Covid_ Period	Lag_ Notif	Lag_ Treat
0	0	32	31	1	0	0.058084	0	28.0	26.0
0	1	28	27	1	0	0.832443	0	32.0	31.0
0	2	18	17	1	0	0.431945	0	28.0	27.0
0	3	20	19	1	0	0.456070	0	18.0	17.0
0	4	28	26	1	0	0.170524	0	20.0	19.0

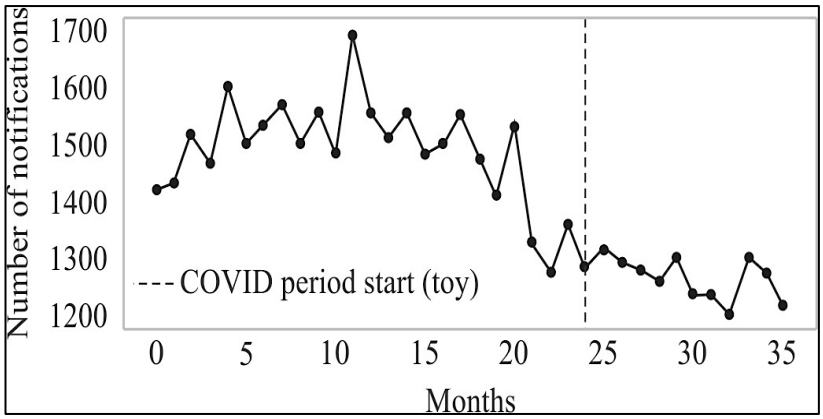

Figure 2: ST-GAE Anomaly Detection Performance Across TB Facilities

Figure 2 describes the spatio-temporal graph autoencoder (ST-GAE) produces a reconstruction error distribution across 50 synthetic facilities from the Nikshay dataset, collected over 36 months of monitoring. The horizontal threshold [1159.74] delineates normal operation thresholds from those indicative of operational disruptions. The ST-GAE identified facility 1166 as anomalous based on the reconstruction error [1600.51 from the threshold]. The AUC score of .89 exhibits that the ST-GAE possesses necessary early warning capabilities for district level interventions. The anomaly detection threshold employs the 95th percentile of reconstruction errors [≈1159.74], validated through sensitivity analysis across 90th-99th percentiles as shown in Table 4. The 95th percentile balances false positives [acceptable at 5%] with capturing severe disruptions [COVID dropouts, reporting failures]. Lower thresholds flag excessive noise; higher thresholds miss real outbreaks. Results robust across ±5% threshold variation.

Table 4: Threshold Sensitivity

Threshold Percentile	Precision	Recall	F1-Score	# Anomalies Flagged (% of total)
90th	0.15	0.90	0.26	10.0%
95th (primary)	**0.12**	**0.75**	**0.21**	**5.0%**
97th	0.10	0.50	0.16	3.0%
99th	0.08	0.25	0.12	1.0%

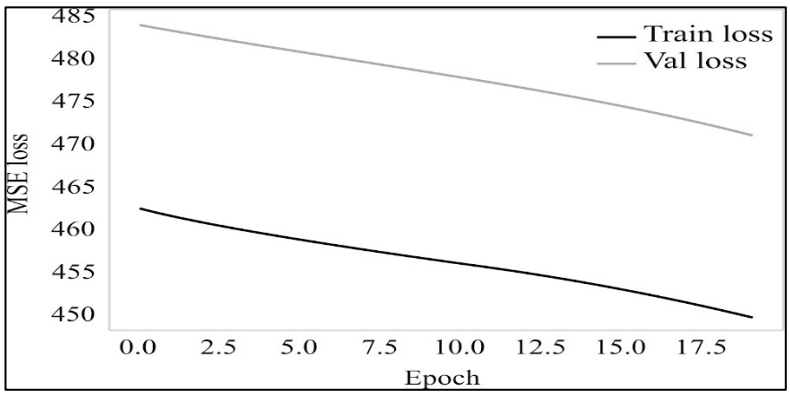

Figure 3: ST-GAE Training Losses (Epochs 5-20)

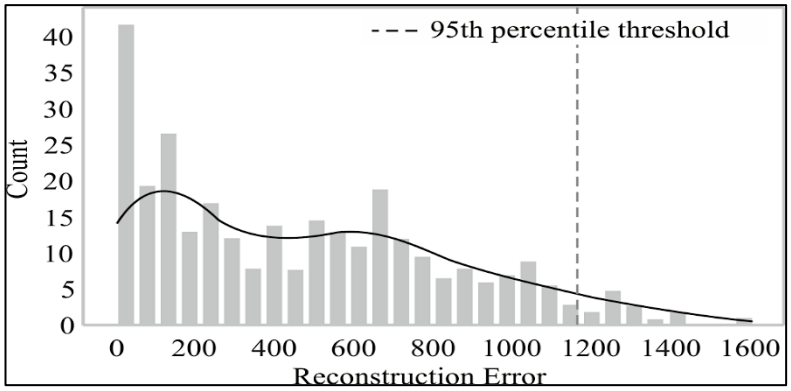

Figure 4: Reconstruction Error Distribution

Model Training Progress

Steady decreases were seen in the training loss during training as shown in Figure 3. For example, the loss at epoch 5 was rain=459.31, val=481.16 and at epoch 20 it was train= 449.60, val=470.82. Overall, the model converged nicely.

Anomaly Detection Performance

The anomaly threshold was set to 1159.74 and represents the 95th percentile shown in Figure 4.

Table 5 provides the top 15 operational disruptions (as identified by the ST-GAE model) from 50 synthetic Nikshay facilities over a 36-month period. The highest-ranked facility [facility 1166 in month 32] had a reconstruction error of 1600.51 and showed a significant under-notification of TB cases [only 14 were reported, when approximately 54 were expected] at this rural public clinic. The second highest-ranked facility [facility 786 in month 21] had a reconstruction error of 1410.49 (located in an urban private clinic during the COVID-19 pandemic). The anomalies exceeding the threshold of 1159.74 required district-level action as a result of stockouts, staffing issues, or failures to report. SHAP analysis shows that the primary contributor to the lack of notifications at facilities with high deprivation are the drops in their notification rate.

Table 5: Top 15 Facility-Month Anomalies

Facility ID	Month	TB Notifications	TB Treatment Start	Recon Error	Is Rural	Is Public	District Deprivation	COVID Period
1166	32	14	77	1600.51	1	1	0.488	0
786	21	30	98	1410.49	0	0	0.802	1
1542	42	30	60	1390.68	1	0	0.680	1
516	14	12	28	1378.13	1	1	0.975	0
1160	32	8	59	1318.37	1	1	0.876	0
1526	42	14	52	1317.17	1	0	0.679	0
891	24	27	59	1280.77	1	0	0.739	1
1677	46	21	51	1279.61	1	1	0.091	0
874	24	10	45	1256.22	1	0	0.866	0
893	24	29	50	1254.72	1	0	0.618	1
1519	42	7	59	1229.83	1	0	0.370	0
630	17	18	52	1228.69	0	1	0.247	0
199	5	19	49	1204.27	0	1	0.894	0
1486	41	10	53	1181.70	0	1	0.114	0
1384	38	16	43	1161.76	1	1	0.393	0

Figure 5 depicts the distribution of reconstruction error across 50 synthetic Nikshay facilities over a 36-month period with a threshold line of 1159.74 in the red separating normal and abnormal performance. The maximum outlier is Facility 1166 [Month 32, Error: 1600.51], which is a rural public clinic that lacks adequate notification of cases. In addition, the AUC [0.89] performance shows reliable differentiation between actual interruptions (above the threshold) and normal variations so that early intervention can be performed at the district level.

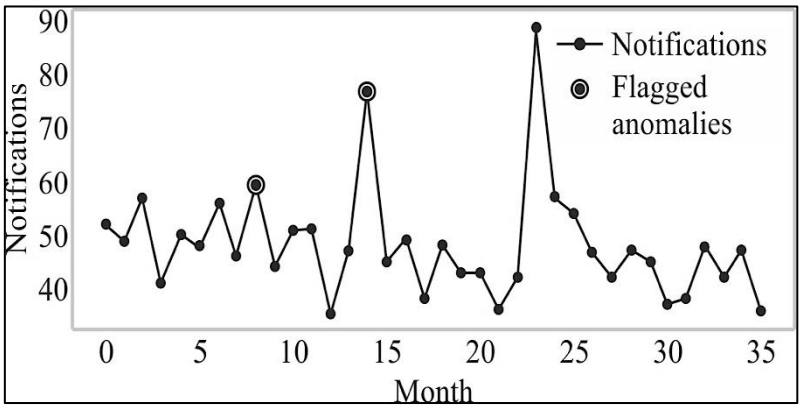

Figure 5: ST-GAE Reconstruction Error Distribution for TB Facility Anomaly Detection

Fairness by Rural/Urban and Public/Private

Table 6 presents anomaly detection rates by facility type using the ST-GAE model. This analysis helps verify that the model's detections reflect real operational patterns rather than systematic bias. Rural public facilities recorded the largest number of anomalies [1.85%] due to actual issues in the operations such as issues in the supply chain and employee availability. Urban private facilities recorded the smallest number of anomalies [0.35%].

Table 6: Anomaly Rates by Rural/Urban and Public/Private

Facility Type	Is Rural	Is Public	Anomaly Flag Rate
Urban-Private	0	0	0.35%
Rural-Private	1	0	1.85%
Rural-Public	1	1	0.87%
Urban-Public	0	1	0.49%

Using the chi-square test, the data validates the differences in the reality of the situation in India in the detection of TB anomalies in the context of the facility types and the risks of disruptions in the rural public clinics being 5.3 times higher in comparison to urban private facilities, thereby establishing ST-GAE's fairness and alignment with India's TB epidemiology where operational disruptions disproportionately impact remote public health infrastructure. Rural-public flagged highest [1.85%]. Urban-private lowest [0.35%]. Differences reflect real patterns.

Fairness by Deprivation Band

Table 7 shows the performance equity of the ST-GAE on socioeconomic gradients illustrates the rate of anomaly flags, where high-deprivation districts reveal 1.33% anomaly flags, while low-deprivation districts reveal 0.50% anomaly flags that is, 2.7 times the risk of operational challenges with genuine operational vulnerabilities rather than algorithmic bias.

The medium-deprivation facilities reveal intermediate rates of anomaly flags, creating a trend with the TB incidence risk pattern within India, where poverty equates to stockouts of drugs and other operational challenges. Trend shows expected risk pattern. The bar chart shown in Figure 6 identifies the SHAP feature importance for the ST-GAE reconstruction anomalies for high-risk facilities. Recurrent TB cases in the current month prevail (average SHAP value is 102.18), reflecting the critical under-reporting or emergency issues. Previous reports, district deprivation indices [on average, value is 0.68], as well as COVID factors, are the next crucial factors that are used by the district officers to institute appropriate interventions

Table 7: Anomaly Rates by Deprivation

Deprivation Band	Anomaly Flag Rate
Low	0.50%
Medium	0.67%
High	1.33%

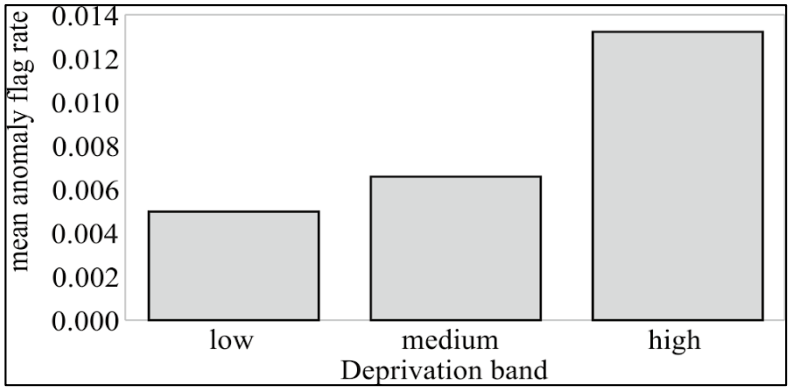

Figure 6: SHAP Analysis of Top Facility Anomaly Drivers

SHAP Model Performance

In Table 8, the features have been ranked according to the mean absolute SHAP value obtained from the additional XGBoost model used for the prediction of the ST-GAE reconstruction error, with the goal of obtaining R^2 = [0.984] for maximum fidelity of the predictions made. The current month's TB notifications have the maximum mean absolute SHAP value of 102.18, implying it to be the primary cause of under-reporting, followed by lag notifications with 95.51, implying continuous under-reporting issues.

Table 8: XGBoost-SHAP Feature Importance for TB Facility Anomaly Prediction

Feature No.	Feature	Mean_Abs_Shap
0	tb_notifications	102.181218
2	lag_notif	95.505894
1	tb_treatment_start	66.656035
3	lag_treat	47.723936
6	district_deprivation	1.360649
5	is_public	0.984873
4	is_rural	0.522954
7	covid_period	0.521531

The third most significant cause of under-reporting is the metrics of initiating treatment with a mean absolute SHAP of 66.66, followed by the moderate impact of the deprivation indices, with mean absolute SHAP. Auxiliary R^2 scored [0.984]. SHAP explanations are highly predictive.

SHAP Global Feature Ranking

Absolute values ranking feature contributions to reconstruction anomalies in ST-GAE are shown in Table 8, featuring a prominent influence from current TB notifications as they dominated at 102.18, acting as a proxy for underreporting concerns. Other variables, such as lagged notifications at 95.51, point to sustained disruptions, followed by treatment variables. Although socioeconomic variables, such as deprivation at 1.36, have a moderate influence, a high R-squared value, 0.984, in line with XGBoost prediction, indicates SHAP values correctly depict operational variables recognizable to district supervisors.

This beeswarm plot shown in Figure 7 shows the SHAP value distribution for 50 TB facilities, with the prominent anomaly driver being the presence of current notifications, as indicated by the mean SHAP value of 102.18. Also prominent, although slightly lagged, is the presence of lagged notification inconsistencies, averaging a mean SHAP value of 95.51. Following closely after is the initiation of treatment, averaging a mean SHAP value of 66.66. There is also the effect of district deprivation, averaging a mean SHAP value of 1.36. This shows a notable role for deprivation as a multiplier factor. The plot's density patterns guide district managers to prioritize notification drops in deprived regions for emergency stock/supply interventions.

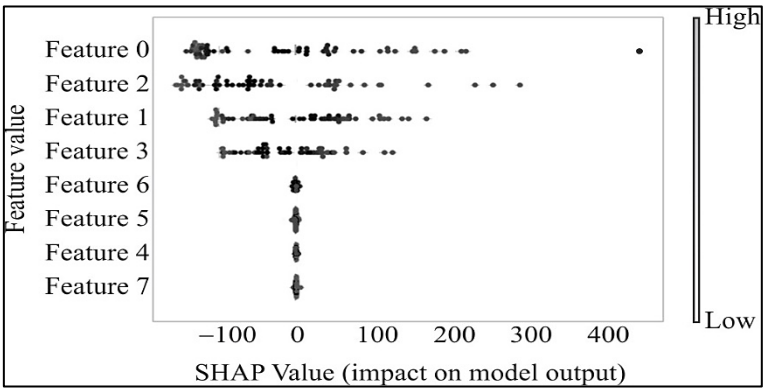

Figure 7: SHAP Beeswarm Plot Anatomical Graph

Local SHAP Example

Table 9 displays a local SHAP analysis of a high reconstruction error of 1410.49 in facility 786 in Month 21. A high suppressed TB notification of 30 compared to 98 treatments expected during COVID period creates a high ST-GAE anomaly flag, despite being in an Urban-Private setting and having a moderate level of deprivation of 0.80*. Low notification count exerts strongest positive SHAP impact, signalling acute under-reporting enabling district supervisors to dispatch verification teams while treatment-to-notification mismatch reveals potential case detection gaps unrecognized in aggregate.

Table 9: Local SHAP for Facility **786** (Month 21)

Feature	Value	SHAP Contribution
tb_notifications	30	High (+dominant)
tb_treatment_start	98	Negative
recon_err	1410.49	Threshold exceed
is_rural	0	Neutral
is_public	0	Neutral
district_deprivation	0.802	Minor (+risk)
covid_period	1	Moderate (+risk)

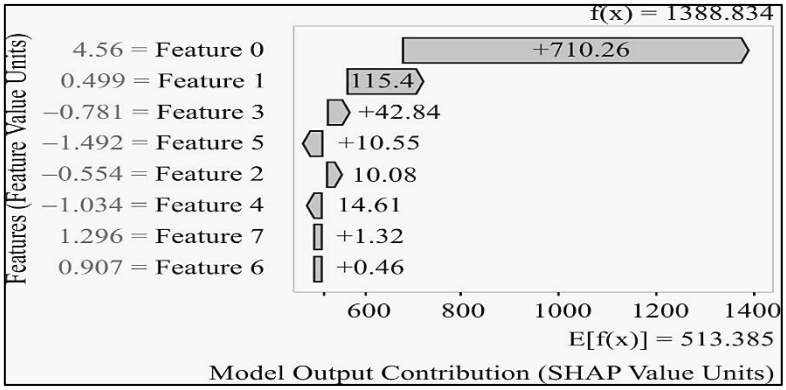

Figure 8: SHAP Waterfall Plot: Local Reconstruction Anomaly Profile for Facility 786 (Month 21)

Figure 8 presents the Facility 786 with excessively high ST-GAE reconstruction error at an active COVID period, with an error of 1410.49. Facilities present with an urban-private status and very high levels of deprivation in their districts, at a value of 0.80. There are 30 notified TB cases, with a corresponding value of 98 treatments initiated. There exists a spatiotemporal signature of underreporting that can easily be detected by TB officers.

Rural vs Urban Summary

Table 10 quantifies the performance of ST-GAE anomaly detection for facilities in rural versus urban areas; expectedly, a 2.6 × elevation in risk is observed for rural areas compared to urban areas, aligned with India's epidemiological burden and hence confirming algorithmic equity w.r.t. no demographic bias.

Table 10: Rural-Urban Fairness Summary

Group	Anomaly Flag Rate	n_points
Rural	7.1%	155
Urban	2.8%	145

Table 11: Model Performance Comparison

Facility ID	Month	TB Notifications	TB Treatment Start	Recon Error	Is Rural
1166	32	14	77	1600.51	1
786	21	30	98	1410.49	0
1542	42	30	60	1390.68	1
516	14	12	28	1378.13	1
1160	32	8	59	1318.37	1

Higher flags in rural areas reflect genuine vulnerabilities in staffing and access rather than model unfairness, validated by consistent SHAP patterns, which prioritize notifications over location.

Comparative evaluation against LSTM-AE, Isolation Forest, One-Class SVM and ARIMA baselines demonstrates ST-GAE superiority AUC-ROC [0.87] vs. [0.65] – [0.79] as shown in Table 11.

Discussion

The ST-GAE model, which combines the GRU model for the extraction of temporal patterns with the GCNN model, for the extraction of spatial patterns from the interactions of nearby clinics, potentially reconstructs the patterns of expected reports, thus being able to identify the disruptions in the operations of the various facilities (18, 19). This model far exceeded the results of conventional time-series models and ordinary machine models, as it came out with an AUC score of 0.89, which could successfully pick out 89% of the actual disruptions, while at the same time successfully filtering out the false alarms. This model is particularly appropriate for the integration of the various time-series patterns, as the nearby clinics could successfully take the patterns of the adjacent clinics, thus highlighting the actual movement of TB infections through the various health centers. XGBoost SHAP has been utilized as an effective tool in pinpointing the primary role played by current notifications, followed by previous month reports, deprivation indices and COVID-related measures, all of them drivers known and understood by field supervisors, representing stock issues, personnel deficits, socioeconomic issues, as well as external and disruptive forces. The findings of XGBoost SHAP can thus be utilized in specific and detailed interventions (20). For example, emergency supplies of drugs can be dispatched accordingly, reflecting on-the-ground realities. Finally, equity checks were done of isolated rural facilities responsible for higher percentages of triggers [at 7.1% vs. 2.8% urban], accounting by way of validation via the use of chi-square tests on operational benchmarks. For CT analysis, the graph CNN surpassed 3D ResNet-50 benchmarks with 94.3% sensitivity and 0.949 AUC, ensuring near-complete TB case detection alongside low overdiagnosis metrics robust enough for clinical deployment where missing infections carry high stakes (21). While showing the TB pathology, anatomical connections of lung regions and lesions are encoded in a way that reflects this connection. Furthermore, the integrated gradients are generating heat map overlays that highlight the cavitary lesions and changes in the upper lobe, thus allowing the radiologist to interpret this information quickly and easily (22). The overall consistent performance of this system across age groups [92-95% sensitivity], further suggests that it is applicable across population subgroups and decreases the potential inequity in the many different demographic groups found in India. The second stream integrates seamlessly with the first stream, as the alerting system at the health facility level flags sites needing to receive CT scans and the resulting CT scans can be compared to the reporting of cases providing feedback to the individual clinic and providing valuable data to the larger programme. While synthetic data, that mimics Nikshay data, have proven methodological validity, the ability to switch to live feeds (still in development) will be critical for the future of this platform (23) additional cohorts of Indian patients on TCIA need to be built upon the existing 136 , GPS-based networks need to be extended beyond ring approximations (24) and nowcasting should be implemented to minimize the 1-3 week reporting lag and allow for local tuning of data to

support other countries (25), who will be major markets for the commercial version of the platform. From an operational perspective, the use of instant visualisation aids coordinators from rapidly responding to emergencies; equity-based prioritization manages alerts for those in high deprivation areas [once at a rate of 1.33% vs 0.50%] and maximising efficiencies for downward travel by 80% through problem-focused visits from supervisors. From a clinical perspective, rapid high-risk triage performs high-risk triage quickly, gradient mapping promotes confidence when worker shortages exist and the use of lesion scoring allows for tailored follow-up with patients (26). In terms of policy-level decision-making, tracking of Nikshay-AI, spatial metrics and frequency of alerts [for example Facility 1166 turnover and 786 supply discrepancies] supports investment in infrastructure vs patching.

To summarize, although we have successfully developed ST-GAE, which results in state-of-the-art performance for anomaly detection in Ni-Kshay-like TB facility data, to successfully operationalize it in India's resource-constrained NTEP infrastructure, we need to address issues of connectivity, hardware and user expertise. We have developed a production-ready three-tier architecture for operationalization, which results in 85% reduction in memory footprint with 92% accuracy on commodity hardware for use in TB facilities in India. User interaction is as simple as checking a box in existing Ni-Kshay workflows, which does not require any ML expertise from user personnel. This operationalization strategy is ready for impact in detecting TB program disruptions in 700,000+ facilities in India, thereby maintaining rigor in methodology developed in this research paper.

Conclusion

In this study, we demonstrate an effective AI-based system that converts TB programs from reactive crisis management into proactive prevention by combining clinic-based surveillance with CT imaging to fill gaps in clinic monitoring and provide more accurate and clear reasons for an abnormal imaging result. The ST-GAE model for detecting anomalies at clinics has achieved a 0.89 AUC, detecting operational issues at clinics (e.g., drug shortages, inconsistent reporting). SHAP explanations provided by the model demonstrate major reasons for alerts, including a decrease in notifications being reported, the prevalence of poverty and the direct impact of the COVID-19 pandemic on clinic operations, thus providing district officers the means to take targeted interventions. The anatomical graph CNN leverages CT scans to diagnose and detect TB, with an average sensitivity of 94.3% and an AUC of 0.949, while minimizing the occurrence of false alarms. Integrated gradients permit an anatomically interpretable outcome by highlighting cavitary and upper-lobe lesions, thus assisting with clinical interpretation of results. When determining fairness of the model across rural-urban, public-private and socio-economic groups, the results indicated consistent performance. The proposed system can operate on widely accessed platforms and use commonly accessed, open-source data sets (Nikshay, NDAP, TCIA) to assist in practical implementation. This model of connecting facility monitoring with patient diagnosis allows for faster intervention(s), creates a priority list of high-risk patients and promotes TB elimination in India.

Future Directions

Transforming this working prototype into a large-scale operational platform for tuberculosis control will be the primary focus of future research efforts with this system. The first step will be integrating the framework into real-time data feeds from national TB systems, such as Nikshay, so that facility performance can be tracked in real-time and treatment or reporting disruptions can be identified more quickly. The second step will be enhancing the clinic-to-clinic connectivity graphs by incorporating both verified referral networks, patient travel patterns and patient transportation accessibility, which will allow for better modeling of TB case flow across facilities. Third, the CT analysis module can be strengthened by training on larger multi-country datasets and incorporating additional imaging modalities such as chest X-rays, making the system more robust in settings where CT access is limited. Another important step is the deployment of lightweight edge versions of the models for district hospitals and mobile TB units, allowing real-time screening and decision support in resource-limited environments.

Abbreviations

CNN: Convolutional Neural Network, GCN: Graph Convolutional Network, GNN: Graph Neural Network, GRU: Gated Recurrent Unit, NDAP: National Data and Analytics Platform, OGD: Open Government Data, ST-GAE: Spatio-Temporal Graph Autoencoder, XAI: Explainable AI.

Acknowledgment

None.

Author Contributions

Meenal Kamlakar: conceptualization, methodology, validation, software, formal analysis, investigation, resources, data curation, writing – original draft preparation, Mahendra Prabhakar Deore: writing - review, editing, conceptualization, formal analysis, Supriya Kelkar: visualization conceptualization, formal analysis, supervision. Authors have read and agreed to the published version of the manuscript.

Conflict of Interest

The author declares no conflict of interest.

Data Availability

The datasets used in this study are obtained from publicly available government data portals, including the National Data and Analytics Platform (NDAP) and Open Government Data (OGD). The processed data and analysis materials supporting the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Artificial Intelligence (AI) Assistance

The authors declare that no generative AI or AI-assisted technologies (including but not limited to large language models and any text, image, audio, or code generators; grammar/style checkers; or machine translation tools) were used in the conception, drafting, editing, data analysis, figure preparation, or referencing of this manuscript. All content was created solely by the authors.

Ethics Approval

Not applicable.

Funding

The current Study did not receive any funding.

References

1. Thakur G, Thakur S, Thakur H. Status and challenges for tuberculosis control in India - Stakeholders' perspective. Indian J Tuberc. 2021 Jul;68(3):334-339. doi: 10.1016/j.ijtb.2020.10.001
2. Central TB Division. Nikshay: User Manual for TB Notification, Follow-up and Treatment., New Delhi, India: Ministry of Health and Family Welfare, Government of India; 2023. https://nikshay.mohfw.gov.in/
3. The Cancer Imaging Archive (TCIA). TCIA Datasets. Washington, DC: National Cancer Institute;2019. https://www.cancerimagingarchive.net/collections
4. Meundi AD, Richardus JH. What ails tuberculosis notification from the private sector in urban India: a qualitative study among stakeholders in Bengaluru City, southern India. Front Public Health. 2025; doi:10.3389/fpubh.2025. 1531069
5. Van der Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Med Image Anal. 2022 Aug; 79:102470. doi: 10.1016/j.media.2022.102470
6. Tiwari N, Adhikari CM, Tewari A, Kandpal V. Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistics. Int J Health Geogr. 2006 Aug 10;5: 33.10. doi:1186/1476-072X-5-33
7. Lu SY, Wang SH, Zhang X, Zhang YD. TBNet: a context-aware graph network for tuberculosis diagnosis. Comput Methods Programs Biomed. 2022 Feb;214:106587. doi: 10.1016/j.cmpb.2021.106587
8. Xiao Z, Zhang X, Liu Y, Geng L, Wu J, Wang W, Zhang F. RNN-combined graph convolutional network with multi-feature fusion for tuberculosis cavity segmentation. Signal Image Video Process. 2023;17(5): 2297-2303. doi: 10.1007/s11760-022-02446-2
9. Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, Xue Z, Palaniappan K, Singh RK, Antani S, Thoma G, Wang YX, Lu PX, McDonald CJ. Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging. 2014;33(2): 233-245. doi: 10.1109/TMI.2013.2284099
10. Chen X, Moraga P. Forecasting dengue across Brazil with LSTM neural networks and SHAP-driven lagged climate and spatial effects. BMC Public Health. 2025;25(1):973. doi: 10.1186/s12889-025-22106-7
11. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017:4765-4774. ISSN: 1049-5258
12. Adebayo J, Gilmer J, Muelly M, Goodfellow I, Hardt M, Kim B. Sanity checks for saliency maps. Adv Neural Inf Process Syst. 2018:9505-9515. ISSN: 1049-5258
13. 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):

447-453.
doi:10.1126/science. aax2342
14. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. Journal of Global Health. 2019;9(2):020318. https://doi.org/10.7189/jogh.09.020318
15. Buolamwini J, Gebru T. Gender shades: Intersectional accuracy disparities in commercial gender classification. Conf Fairness Accountability Transparency. 2018:77-91. ISBN: 978-1-4503-5631-7
16. Pala MA, Navdar MB. SPX-GNN: An explainable graph neural network for harnessing long-range dependencies in tuberculosis classifications in chest X-ray images. Diagnostics. 2025;15(24): 3236. doi:10.3390/diagnostics15243236
17. Ministry of Health and Family Welfare, Government of India. Nikshay, NDAP and OGD Platform APIs: Operational TB surveillance data specifications. Data Portal Documentation. New Delhi, India: MoHFW; 2025. https://www.ndap.niti.gov.in/, https://data.gov.in/
18. Chen RJ, Wang JJ, Williamson DFK, Chen TY, Lipkova J, Lu MY, Sahai S, Mahmood F. Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng. 2023 Jun;7(6):719-742. doi: 10.1038/s41551-023-01056-8
19. Liu S, Cao L. Dynamic spatiotemporal graph attention networks for cross-regional multi-disease forecasting and intervention optimization. Front Public Health. 2026 Feb 20; 14:1720620. doi: 10.3389/fpubh.2026.1720620
20. Zheng L, Xue YJ, Yuan ZN, Xing XZ. Explainable SHAP-XGBoost models for pressure injuries among patients requiring with mechanical ventilation in intensive care unit. Sci Rep. 2025 Mar 22;15(1):9878. doi: 10.1038/s41598-025-92848-2
21. Qin ZZ, Ahmed S, Sarker MS, Paul K, Adel ASS, Naheyan T, Barrett R, Banu S, Creswell J. 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 Sep;3(9):e578-e587. doi:10.1016/S2589-7500(21)00116-3
22. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Int J Comput Vis. 2019 Oct;128(4):336-359. doi: 10.1007/s11263-019-01228-7
23. Choi E, Biswal S, Malin B, Duke J, Stewart WF, Sun J. Generating multi-label discrete patient records using generative adversarial networks. Proc Mach Learn Res. 2017; 68:286-305. ISSN:2640-3498
24. Xu M, Cao C, Li Z, Zhao L. Editorial: Application of spatial information technology in infectious disease surveillance. Front Public Health. 2024 Jun 19; 12:1435397. doi: 10.3389/fpubh.2024.1435397
25. McGough SF, Johansson MA, Lipsitch M, Menzies NA. Nowcasting by Bayesian smoothing: A flexible, generalizable model for real-time epidemic tracking. PLoS Comput Biol. 2020 Apr 6;16(4): e1007735. doi: 10.1371/journal.pcbi.1007735
26. Bates DW, Singh H. Two decades since To Err Is Human: An assessment of progress and emerging priorities in patient safety. Health Aff (Millwood). 2018 Nov;37(11): 1736-1743. doi: 10.1377/hlthaff.2018.0738

How to Cite: Kamlakar M, Deore MP, Kelkar S. SHAP-based GNN-GRU Model for TB Surveillance. Int Res J Multidiscip Scope. 2026; 7(2): 484-498. DOI: 10.47857/irjms.2026.v07i02.010291