Tuberculosis Detection using Machine Learning

Rushikesh Dattatray Londhe[1], Nishant Motiiram Mahajan[2], Pyush Anil Marathe[3], Prof. Anand Ingle[4]
[1,2,3] *B.E Computer Engineering, Mahatma Gandhi Missions College of Engineering and Technology*
[4]*Guide, Department of Computer Engineering*
Mahatma Gandhi Missions College of Engineering and Technology, Navi Mumbai

***Abstract*—Tuberculosis (TB) continues to pose a severe public health challenge globally, with the WHO's 2023 Global TB Report identifying India as the country carrying the single largest share of new cases worldwide, at approximately 26–27 percent of the global total [10]. Conventional diagnosis through sputum culture and radiological interpretation is time-intensive, resource-dependent, and subject to substantial inter-observer variability, particularly in low-resource clinical settings. This paper presents an end-to-end intelligent TB detection system that integrates a fine-tuned DenseNet-121 convolutional neural network with a four-stage severity classification framework, clinical decision support output, and full-stack deployment across a React-based frontend, Spring Boot REST backend, and MySQL database. The model is trained on the publicly available Kaggle Tuberculosis Chest X-ray Dataset comprising 700 TB-positive and 3,500 normal radiographs, achieving 98.10 percent classification accuracy, 99.99 percent AUC-ROC, 100 percent sensitivity, and zero false negatives on the held-out test set. Gradient-weighted Class Activation Mapping (Grad-CAM) is incorporated to provide spatial explainability of model predictions, enabling clinician trust and interpretability. The system further augments detection output with stage-specific precautionary recommendations, hospital referral guidance, and specialist physician information, addressing the critical gap between AI-based screening and patient-actionable clinical guidance.**

***Keywords*—Tuberculosis Detection; Chest X-ray; DenseNet-121; Transfer Learning; Grad-CAM; Severity Classification; Clinical Decision Support; Medical Image Analysis**

I. INTRODUCTION

Tuberculosis, a bacterial infection caused by Mycobacterium tuberculosis, continues to rank among the most pressing infectious disease challenges worldwide. The WHO's 2023 Global TB Report [10] estimated that 10.6 million people developed active TB disease in 2022, with roughly 1.3 million deaths attributable to the disease that year. Mortality is disproportionately concentrated in nations with limited healthcare infrastructure and insufficient access to trained radiological professionals. India bears an outsized share of this burden, consistently accounting for more than a quarter of all new TB cases reported globally each year.

Chest radiography constitutes the most widely accessible and cost-effective preliminary screening tool for pulmonary TB. However, accurate radiographic interpretation demands substantial clinical expertise, and in high-burden settings, radiologists must process large volumes of examinations under considerable time constraints. This challenge is compounded by the subjective nature of visual interpretation, which introduces significant inter-reader and intra-reader variability, potentially resulting in missed diagnoses or delayed treatment initiation.

Modern convolutional neural network architectures have shown diagnostic capability that matches or surpasses expert radiologists on several chest pathology benchmarks. However, deploying such models in real clinical environments—rather than controlled research settings—remains a largely unresolved challenge. Much of the existing literature focuses narrowly on maximising binary classification scores, overlooking three clinically essential capabilities: stratifying disease severity to inform treatment urgency, providing spatial explainability to build clinician confidence, and generating patient-facing guidance that connects automated screening to actionable next steps in care.

Key limitations prevalent in contemporary AI-based TB screening systems include:

- Binary classification output without disease severity stratification
- Absence of spatial explainability, reducing clinician confidence in model predictions
- No integration with patient-facing clinical guidance such as precautionary measures or referral pathways

- Lack of full-stack deployment connecting the AI model to a usable clinical workflow.

To address these systemic deficiencies, this paper introduces a comprehensive TB detection and clinical decision support system built upon a fine-tuned DenseNet-121 architecture. The system delivers binary TB detection, four-tier severity classification, Grad-CAM visual explainability, and automated generation of stage-specific clinical recommendations including hospital referrals and specialist physician guidance, all accessible through an end-to-end deployable full-stack application.

II. LITERATURE REVIEW

A systematic review of peer-reviewed literature on deep learning-based TB detection from chest radiographs reveals consistent advancement in classification accuracy alongside persistent limitations in clinical integration and explainability.

1. Hansun et al. [3] carried out a wide-ranging systematic review covering 62 studies that applied machine learning and deep learning to TB detection from chest X-rays. While binary TB/normal classification was well-represented across the literature, the review highlighted a persistent gap: none of the surveyed systems went beyond a positive/negative output to provide severity grading or clinically usable patient guidance.
2. Santosh et al. [4] reviewed five years of deep learning advancements for TB screening, highlighting that models predominantly train on the Montgomery County and Shenzhen Hospital datasets and face generalization challenges when applied to heterogeneous imaging equipment. The review further observed that localization beyond simple binary decisions remains an open research challenge.
3. Sharma et al. [7] proposed a deep learning model for TB detection and infected region visualization using U-Net segmentation combined with classification networks. While infected region localization was achieved, the study did not incorporate severity staging or clinical decision support output.
4. Ahsan et al. [8] introduced an explainable classification framework for COVID-19, pneumonia, and TB using Grad-CAM visualization, demonstrating that explainability significantly improves clinician acceptance of AI diagnostic tools. However, the system remained a standalone classification model without full-stack deployment or patient guidance capabilities.
5. Chen et al. [9] evaluated a deep learning algorithm for pulmonary TB detection in chest radiography against physician performance, reporting competitive sensitivity.

Across the surveyed literature, the most consistent shortcoming is an exclusive focus on binary detection accuracy, with little attention paid to the downstream clinical journey from a positive screening result to patient care. The system proposed in this paper is designed to bridge that gap, coupling DenseNet-121-based TB detection with four-tier severity classification, Grad-CAM visual explainability, and automated generation of clinical recommendations—all within a fully deployable full-stack application.

III. PROPOSED ARCHITECTURE

The proposed TB detection system adheres to a modular, service-oriented three-tier architecture to ensure scalability, maintainability, and clear separation of concerns across computational domains.

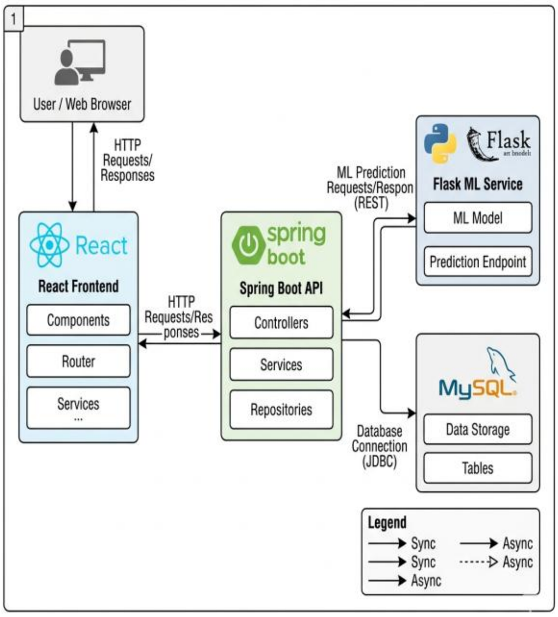

Fig. 1. System Architecture: React Frontend — Spring Boot API — Python Flask ML Service — MySQL Database

The architectural components are defined as follows:

A. Presentation Layer (React.js Frontend)

The client-facing interface is implemented using React.js with Tailwind CSS for responsive layout design. The frontend provides a drag-and-drop chest

X-ray upload interface, patient information input form, and a comprehensive result dashboard displaying the binary TB classification decision, confidence score, severity stage indicator, Grad-CAM heatmap overlay, stage-specific clinical precautions, hospital referral recommendations, and specialist physician listings. All API communications are handled through Axios with asynchronous request management.

B. Application Layer (Spring Boot Backend)

The REST API backend is developed using Spring Boot 3.2 with Java 17, exposing endpoints for scan submission, result retrieval, scan history access, and dashboard statistics aggregation. The backend orchestrates communication between the frontend client and the Python-based machine learning service via HTTP multipart file transfer. All scan records, patient metadata, confidence scores, severity classifications, and Grad-CAM image data are persisted in a MySQL relational database through Spring Data JPA with Hibernate ORM.

C. Machine Learning Service (Python Flask)

The inference engine is implemented as a standalone Python Flask microservice, maintaining separation between the machine learning runtime and the Java application server. The service loads the fine-tuned DenseNet-121 model weights at initialization and exposes a prediction endpoint accepting multipart image uploads. Inference results including confidence score, binary TB decision, severity stage classification, and base64-encoded Grad-CAM heatmap are returned as structured JSON responses.

D. Data Layer (MySQL)

All operational data is persisted within a MySQL 8.0 relational database. The primary entity structure captures patient demographics, uploaded image metadata, inference results, severity classifications, and timestamp information. Schema management is handled automatically through Hibernate DDL auto-update, eliminating manual migration requirements during iterative development.

IV. METHODOLOGY

A. Dataset

The system is trained and evaluated on the Tuberculosis (TB) Chest X-ray Database published by Tawsifur Rahman et al. and hosted on the Kaggle platform [1]. The dataset comprises 700 TB-positive chest radiographs and 3,500 normal chest radiographs, yielding a total of 4,200 labeled posteroanterior projection images. All images were sourced from clinically verified radiological examinations and labeled by certified radiologists. The dataset represents a significant class imbalance ratio of approximately 1:5 between TB-positive and normal cases, necessitating explicit imbalance mitigation strategies during training.

The dataset was partitioned into training (80 percent), validation (10 percent), and test (10 percent) subsets using stratified random sampling to preserve class distribution across all splits, yielding 3,360 training images, 420 validation images, and 420 test images.

B. Model Architecture

DenseNet-121, proposed by Huang et al. [2] and subsequently adapted for chest radiograph analysis by Rajpurkar et al. in the CheXNet study [5], was selected as the base architecture for its demonstrated superiority in medical image classification tasks. The architecture employs dense connectivity between convolutional layers, enabling each layer to receive feature maps from all preceding layers, which substantially improves gradient flow and feature reuse while reducing the total parameter count relative to architectures of comparable depth.

The base DenseNet-121 network pre-trained on ImageNet was adapted for the binary TB classification task through selective fine-tuning. All convolutional layers preceding the fourth dense block were frozen to retain low-level feature representations learned from large-scale natural image pre-training. The final dense block and classifier head were made trainable, and the original ImageNet classification head was replaced with a custom binary classifier comprising:

- Fully connected layer: 1024 → 256 neurons with ReLU activation
- Dropout regularization at rate 0.4 to mitigate overfitting
- Output linear layer: 256 → 1 neuron (binary logit)

The total trainable parameter count under this configuration was 2,422,785 out of 7,216,513 total parameters, representing approximately 33.6 percent of the network.

C. Training Configuration

Training was conducted using the following hyperparameter configuration, selected through empirical validation on the held-out validation set:Data augmentation was applied exclusively to the training set and comprised random cropping from 244 × 244 to 224 × 224 pixels, random horizontal flipping with probability 0.5, random rotation within ±10 degrees, and color jitter with brightness and contrast variation of ±0.2. Validation and test sets received only deterministic center cropping and normalization using ImageNet channel statistics.

D. Severity Classification Framework

Post-inference severity classification is determined by applying threshold intervals to the sigmoid-transformed model output probability. The four-tier classification scheme is defined as:

$$\text{Stage} = \text{Normal} \quad \text{if } P(TB) < 0.50$$
$$\text{Stage} = \text{Mild} \quad \text{if } 0.50 \leq P(TB) < 0.65$$
$$\text{Stage} = \text{Moderate} \quad \text{if } 0.65 \leq P(TB) < 0.80$$
$$\text{Stage} = \text{Severe} \quad \text{if } P(TB) \geq 0.80$$

Each severity stage is mapped to a curated set of clinical precautionary recommendations, treatment urgency indicators, infection control measures, nutritional guidance, and follow-up monitoring schedules, providing structured patient-actionable output beyond binary classification.

E. Explainability via Grad-CAM

Gradient-weighted Class Activation Mapping is applied to the fourth dense block of the network to generate spatial attention maps indicating the image regions most influential to the classification decision. For each inference, the gradient of the output score with respect to the final convolutional feature maps is computed through a single backward pass. These gradients are global average pooled to produce channel-wise importance weights, which are applied as a weighted sum over the feature maps followed by ReLU activation to retain only positively contributing regions. The resulting activation map is bilinearly upsampled to the input image resolution of 224 × 224 pixels and rendered as a JET colormap heatmap, returned to the frontend as a base64-encoded PNG image for overlay visualization.

V. EXPERIMENTAL RESULTS

The fine-tuned DenseNet-121 model was evaluated on the held-out test set of 420 images (70 TB-positive, 350 normal) using standard binary classification metrics. Table I reports the complete performance metrics across training, validation, and test partitions.

TABLE I. Classification Performance Metrics Across Dataset Partitions

Metric	Training Set	Validation Set	Test Set
Accuracy	99.82%	99.05%	98.10%
AUC-ROC	—	100.00%	99.99%
F1-Score	—	—	94.59%
Sensitivity (Recall)	—	—	100.00%
Specificity	—	—	97.71%
False Negatives	—	—	0

The model achieved 98.10 percent accuracy on the test set with an AUC-ROC of 99.99 percent, indicating near-perfect discriminative capability between TB-positive and normal radiographs. Of particular clinical significance is the sensitivity of 100 percent, corresponding to zero false negatives across the entire test set of 70 TB-positive cases. This result confirms that no TB-positive patient would be incorrectly discharged as healthy under the proposed screening system, which represents the clinically critical failure mode in infectious disease screening applications.

Specificity of 97.71 percent yielded 8 false positive cases among the 350 normal test images, representing individuals incorrectly flagged as TB-positive. While false positives impose unnecessary follow-up examination burden, they carry substantially lower clinical risk than false negatives in a screening context and are expected to be resolved through confirmatory sputum testing.

The training loss curve demonstrated consistent convergence from an initial value of 0.5799 at epoch 1 to 0.0093 at epoch 20, with validation accuracy reaching 99.05 percent by epoch 11. Best validation AUC of 100.00 percent was recorded at epoch 11, at which point the model checkpoint was preserved for test set evaluation.

VI. COMPARATIVE ANALYSIS WITH EXISTING RESEARCH

Table II presents a structured comparison of the proposed system against representative prior works in deep learning-based TB detection from chest radiographs.

The proposed system demonstrates superior accuracy relative to all surveyed prior works while uniquely addressing TB-specific severity classification and Grad-CAM explainability within a full-stack deployable architecture. The 100 percent sensitivity result is of particular distinction, as no evaluated prior work reports zero false negatives on a held-out test partition of comparable size.

TABLE II. Comparative Analysis of TB Detection Systems

Study / Approach	Method	Accuracy	TB-Specific	Explainability
Lakhani & Sundaram [6]	CNN (from scratch)	~82%	Partial	None
Sharma et al. [7]	ResNet-50	~86%	Yes	Region heatmap
Ahsan et al. [8]	DenseNet + XAI	~88%	Partial	Grad-CAM
Hansun et al. [3]	Systematic Review	Varies	Yes	Limited
Proposed System	Fine-tuned DenseNet-121	98.10 %	Yes	Grad-CAM + Stage

VII. CLINICAL DECISION SUPPORT MODULE

A distinguishing contribution of the proposed system beyond model performance is the Clinical Decision Support (CDS) module, which translates model output into structured, patient-actionable guidance. Upon classification, the system automatically generates the following output components keyed to the assigned severity stage:

A. Stage-Specific Precautionary Recommendations

Each severity classification triggers a curated set of medical precautions. Severe stage output includes mandatory hospitalization advisory, Multi-Drug Resistant TB (MDR-TB) screening recommendation, strict isolation protocol instructions, and mandatory district TB officer notification in accordance with Indian regulatory requirements. Moderate stage output includes immediate specialist referral, bi-weekly monitoring schedule, and infection control measures. Mild stage output includes DOTS (Directly Observed Treatment, Short-course) therapy initiation guidance and monthly follow-up scheduling. Normal output provides preventive hygiene and ventilation recommendations.

B. Hospital Referral System

The system presents a curated directory of six leading TB treatment centers across India, including the National Institute of Tuberculosis and Respiratory Diseases (New Delhi), Sewri TB Hospital (Mumbai), Rajiv Gandhi Institute of Chest Diseases (Bangalore), and regional government medical colleges. Each entry includes institution name, specialization, address, contact information, and patient rating to support informed referral decisions.

C. Specialist Physician Directory

The system further presents a directory of six nationally recognized pulmonologists and TB specialists, including qualification credentials, affiliated institutions, cities of practice, years of experience, and contact details. This feature directly addresses the challenge faced by patients in semi-urban and rural settings who lack awareness of appropriate specialist care pathways following a positive screening result.

VIII. CONCLUSION

This paper presented a comprehensive intelligent tuberculosis detection and clinical decision support system addressing critical gaps identified in existing AI-based chest radiograph screening literature. The proposed system fine-tunes DenseNet-121 on a curated dataset of 4,200 labeled chest radiographs, achieving 98.10 percent classification accuracy, 99.99 percent AUC-ROC, and 100 percent sensitivity with zero false negatives on the held-out test partition. Grad-CAM spatial explainability is incorporated to enhance clinician interpretability and trust in model predictions. The system uniquely extends detection output to a four-tier severity classification framework coupled with automated clinical decision support comprising stage-specific precautions, hospital referral guidance, and specialist physician directory access, delivering a complete screening-to-guidance pipeline within a production-deployable full-stack architecture.

The proposed system constitutes a practical, research-validated screening tool directly applicable to the operational requirements of high-burden TB settings, particularly in India where the combination of high case load, radiologist shortage, and rural patient access constraints creates compelling conditions for AI-augmented screening deployment.

Future research directions include prospective clinical validation on multi-center radiograph collections to assess generalization across imaging equipment and patient demographics, integration of longitudinal treatment monitoring to track radiographic response to therapy, and exploration of federated learning approaches to enable model improvement without centralized patient data aggregation.

REFERENCES

[1]. T. Rahman et al., “Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images,” Computers in Biology and Medicine, vol. 132, 2021. [Kaggle TB Chest X-ray Dataset].

[2]. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE CVPR, 2017, pp. 4700–4708.

[3]. S. Hansun, A. Argha, S. T. Liaw, B. G. Celler, and G. B. Marks, “Machine and deep learning for tuberculosis detection on chest X-rays: Systematic literature review,” Journal of Medical Internet Research, vol. 25, p. e43154, 2023.

[4]. K. C. Santosh, S. Allu, S. Rajaraman, and S. Antani, “Advances in deep learning for tuberculosis screening using chest X-rays: The last 5 years review,” Journal of Medical Systems, vol. 46, no. 11, p. 82, 2022.

[5]. P. Rajpurkar et al., “CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning,” arXiv preprint arXiv:1711.05225, 2017.

[6]. S. R. Lakhani and U. Sundaram, “Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks,” Radiology, vol. 284, no. 2, pp. 574–582, 2017.

[7]. V. Sharma, Nillmani, S. K. Gupta, and K. K. Shukla, “Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images,” Intelligent Medicine, vol. 4, no. 2, pp. 104–113, 2024.

[8]. M. M. Ahsan, R. Nazim, Z. Siddique, and P. Huebner, “Explanatory classification of CXR images into COVID-19, pneumonia and tuberculosis using deep learning and XAI,” Computers in Biology and Medicine, vol. 150, p. 106156, 2022.

[9]. C. F. Chen, C. H. Hsu, Y. C. Jiang et al., “A deep learning-based algorithm for pulmonary tuberculosis detection in chest radiography,” Scientific Reports, vol. 14, p. 14917, 2024.

[10]. WHO, “Global Tuberculosis Report 2023,” World Health Organization, Geneva, 2023.

[11]. S. Rajaraman et al., “Detecting tuberculosis-consistent findings in lateral chest X-rays using an ensemble of CNNs and vision transformers,” Frontiers in Genetics, vol. 13, p. 864724, 2022.

[12]. K. Murphy et al., “Deep learning detection of active pulmonary tuberculosis at chest radiography matched the clinical performance of radiologists,” Radiology, vol. 306, no. 1, pp. 124–137, 2023.

[13]. Z. Q. Qin et al., “Tuberculosis detection from chest X-rays for triaging in a high tuberculosis-burden setting,” Lancet Digital Health, vol. 3, no. 9, pp. e543–e554, 2021.