PENINSULA PUBLISHING LLC

EDRAAK
Vol. **(2026)**, 2026, **pp**. 75–84
ISSN: 3078-8412

Research Article

A Machine Learning-Based Diagnostic Framework for Accurate Tuberculosis Detection

Hina Afzal [1], Ghulam Mustafa[1,*], Asif Nawaz[1], Muhammad Rizwan Rashid Rana[2], Tariq Ali[1]

[1] *University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi, Pakistan*

[2] *Department of Robotics & Artificial Intelligence, Shaheed Zulfiqar Ali Bhutto Institute of Science & Technology, Islamabad, Pakistan*

ARTICLEINFO

Article History
Received 1 Mar 2026
Revised: 15 Apr 2026
Accepted 14 May 2026
Published 03 Jun 2026

Keywords
Tuberculosis,
Diagnosing,
Machine Learning,
Deep Learning,
Patients.

ABSTRACT

Tuberculosis (TB) is a highly infectious disease with a profound impact on global health, making early and accurate diagnosis crucial for preventing its spread. In recent years, machine learning has emerged as a powerful diagnostic tool capable of enhancing disease detection and optimizing healthcare processes. This study presents a machine learning-based diagnostic model designed to improve the precision and efficiency of TB diagnosis. The proposed framework utilizes a comprehensive dataset comprising clinical and laboratory information, including patient demographics, symptoms, and diagnostic test results. The model operates through several key stages: data collection, preprocessing, feature extraction, feature fusion union, classification, and distribution. Experimental evaluation across multiple datasets demonstrates the model's robustness and superior performance. On the TB-DS-I dataset, it achieved precision, recall, accuracy, and F-measure scores of 94.84%, 94.56%, 95.85%, and 94.32%, respectively. For the TB-DS-II dataset, it attained precision of 96.74%, recall of 96.5%, F-measure of 96.15%, and accuracy of 97.02%. Similarly, the TB-DS-III dataset yielded outstanding results with 97.83% precision, 97.54% recall, 97.2% F1 score, and 98.03% accuracy, confirming the model's effectiveness and reliability in accurately classifying individuals as TB-positive or TB-negative.

1. INTRODUCTION

Tuberculosis (TB), caused by Mycobacterium tuberculosis, ranks among the most severe illnesses on earth. M. Tuberculosis (M. TB) is a highly contagious illness with the potential to be fatal, affecting approximately 25% of the population [1]. TB is an airborne disease that typically affects the human lungs. According to a report from the World Health Organization (WHO), in 2020, 9.9 million individuals were afflicted by this condition. According to the WHO, the Southeast Asia region constitutes a significant burden of TB cases, accounting for 44% of all cases in Africa (24%), followed by the Western Pacific (18%), the Mediterranean (8%), and Europe and America (3%) [2].

Relying solely on symptom-based screening is insufficiently sensitive in the initial task of mass tuberculosis screening. Consequently, the majority of patients are medically identified when defects on chest X-rays become visible [3]. However, in the present situation, a successful and precise diagnostic approach is crucial for the timely and effective treatment of tuberculosis. This can potentially save lives and help control the spread of the disease. Some of the TB symptoms are shown in Figure 1.

The fact that TB is treatable with medications. A high death frequency indicates that TB instances are often eliminated unnoticed or remain unknown until they have advanced [4]. Chest X-ray (CXR) and sputum smear microscopy are the two most common techniques for TB diagnosis. CXR demonstrate greater sensitivity compared to verbal screening in detecting cases of TB. CXR is a practical approach for TB detection; however, it is not without drawbacks [5]. Expert personnel are required to analyse CXR pictures to diagnose tuberculosis. Lungs are affected by tuberculosis in a variety of ways. Symptoms of tuberculosis include penetration, synthesis, and depression.

*Corresponding author email: gmustafa@uaar.edu.pk
DOI: https://doi.org/10.70470/EDRAAK/2026/006

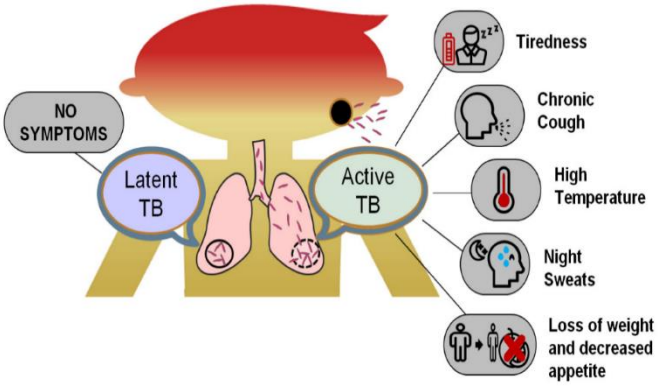

Fig. 1. TB symptoms

Additionally, TB can be categorized as latent or active. Therefore, the best prevention strategy remains early detection and diagnosis [6]. In recent years, approaches for automatic TB diagnosis and classification have been applied to enhance the accuracy of disease detection, assisting healthcare professionals in making better decisions. Early systems primarily relied on machine learning techniques for automatic categorization, but they required multiple intermediate processes such as feature selection, feature extraction, and classification [7–10].

Deep learning (DL) has further advanced the field, enabling models to automatically learn hierarchical features from raw medical data without extensive manual intervention. In TB research, convolutional neural networks (CNNs) have demonstrated strong performance in analyzing chest X-rays, CT scans, and other imaging modalities to detect subtle patterns often missed by human experts. Similarly, recurrent neural networks (RNNs) and transformer-based architectures have enabled the integration of sequential patient data, such as medical histories and laboratory results, for more comprehensive assessment. Despite these promising developments, most existing studies focus narrowly on imaging data, leaving out other critical clinical and demographic variables. Moreover, there remains a need for more robust and generalizable frameworks that can combine heterogeneous data sources while ensuring reliability across diverse datasets.

Motivated by these challenges, this study proposes a comprehensive machine learning framework that unifies the entire diagnostic pipeline, including data collection, preprocessing, feature extraction, feature fusion, classification, and distribution. Unlike prior work that emphasizes imaging alone, the proposed approach integrates multimodal data such as demographics, symptoms, and laboratory test outcomes to provide a more holistic diagnosis. Through its robust feature fusion strategy, the model demonstrates exceptional diagnostic performance across three benchmark datasets, consistently achieving high accuracy, precision, recall, and F-measure. By offering an automated and scalable solution, the proposed framework not only advances the state of TB detection research but also holds practical value for deployment in real-world, resource-limited healthcare settings where timely and reliable diagnosis is critical.

2. LITERATURE REVIEW

The literature has assembled a number of conventional machine learning and deep neural networks to distinguish tuberculous CXR pictures from non-tuberculous CXR pictures successfully. Here we have a number of ideas for possible research on TB or other lung diseases very recently reports.

Govindarajan et al. conducted a TB evaluation with chest X-rays for digital diagnosis [11]. The authors focused on the development of this method for tuberculosis diagnosis. Han et al. (2019) reported a new approach for automated recognition of cavity imaging signs through lung CT images (12). The study focused on creating a classification technique that involved the use of hybrid resampling and multi-feature fusion. The authors were able to fuse the features from Convolutional Neural Network CNN, HOG, and LBP. The authors found superior results with a fusion method compared to using a single feature detector method. The study had implications towards lung cavity sign detection using CT scan images.

Cao et al. proposed a model with a staged CNN to detect lung nodules [13]. In the first stage, they used a U-Net as a separation network. Detect lung nodules. In the next stage, they integrated a two-stage TSCNN (Three-Tree Structure CNN) using a dual pooling structure for false-positive reduction in the classification phase. Momeny et al. investigated the MTB categorization images with a deep-extended CNN [14]. Their strategy is based on the second least square coarse chaos and used hybrid sharing with a Greedy Auto augment method. Lu et al. presented TBNet, a graph-based neural network that processes contextual information to mediating tuberculosis diagnosis [15]. This effort contributes to tuberculosis diagnosis and presents a distinctive approach to increase accuracy.

Dey et al. put forward a heterogeneous convolutional neural network topology to classify TB in chest X-ray images [16]. The study's objective was to increase the accuracy of detecting tuberculosis. The authors proposed an optimization method using the meta-heuristics to optimize ambiguous measurements, thus improving the training procedure of the network. The

authors have also used pre-trained algorithms like DenseNet-121, VGG-19, and ResNet-50 to obtain features providing initial prediction estimations from the Network/module being fully connected. As a collective approach by numerous authors, this provides value to the ever-increasing body of knowledge on early detection of tuberculosis through the use of Chest X-ray images.

Rahman et al. used deep pre-trained networks (VGG19, ResNet101, and DenseNet-201) as feature extractors to detect TB chest X-ray images [17]. The study procedure mainly included detecting features from X-ray images and the authors used the XGBoost network for classification purposes, in which the X-ray images were classified as either being TB-positive or healthy. This collaborative work by multiple authors proved that using a deep network that has been pre-trained along with the XGBoost algorithm was a valuable tool for detecting TB from chest X-ray images. On the other hand, Iqbal et al. developed the MDA-Net (a network with a multiscale dual attention network) for breast lesion segmentation in ultrasound images [18]. This model has a dual attention mechanism where it combines lesion-based attention and channel-based attention closer to how actual images are being segmented. Furthermore, they developed a multiscale lamination block in order to recover more structures and properties from ultrasound images. This collaborative project illustrated the success of the MDA-Net for breast lesion segmentation in ultrasound images.

Kukker et al. applied a reinforcement learning paradigm to classify TB, a type of respiratory infection, as well as healthy CXR images [19]. They researched altering uncertainty, with query learning approaches, and the use of preliminary processing with wavelets. The authors created a classifier TB and the seriousness of the pulmonary stage of TB. Their research is a primary approach for tuberculosis and pneumonia classification that uses an adapted fuzzy Q-learning approach. Khatibi et al. also introduced a new multi-instance learning algorithm for correcting TB via Chest X-ray images [20]. This study introduced CNNs, complex networks, and also for layer batch networks, accurate recognition of tuberculous and healthy X-ray images. The method followed involved. The authors obtained multiple different patches from the images, cauterizing and then formulating to a feature vector. Then they performed clustering for all patches globally, which meant assigning cleanliness ratings for the corresponding clusters and combining each merge, for each image.

Abdar et al. proposed UncertaintyFuseNet, a system that successfully utilizes a combination of features, to classify COVID-19 CXR images accurately [21]. The study proposed the UncertaintyFuseNet, which includes a hierarchical combination of features that was maintained in uncertain conditions. The presented approach showed an ensemble Monte Carlo dropout to minimize uncertainty from a feature fusion perspective, as it eliminates noise problems. The UncertaintyFuseNet proved effective for accurately classifying COVID-19 CXR images while also outperforming unseen datasets. Wang et al. presented FGCNet, a classification approach for COVID-19. The paper is published in Information Fusion, employing graph convolutional networks (GCN) and deep feature fusion (DFF) methodologies [22]. The authors separately fused image-level features as well as features from the respectively connection-aware CNN and GCN networks. The FGCNet methodology effectively fused underlying deep features for accurate COVID-19 classification.

Bamrah et al. studied tuberculosis in homeless populations in the United States during a period from 1994 to 2010 [23]. The study assessed the use of posterior-anterior chest radiographs (CXR) as an economical and lower-radiation screening method for tuberculosis. The researchers pointed out that CXR results are available immediately which can be a vital feature for screening tuberculosis from the homeless population. Bobak et al. conducted a study furthering AI/ML for Tuberculosis diagnosis with transcriptomic indicators with merged datasets [24]. Bobak proposed a networked framework including several openly accessible conveying array datasets that were used to validate a gene pattern for tuberculosis identification. The study tested three machine learning classifiers, with the Random Forest classifier providing the most efficient run with an accuracy of 86.46%.

The literature shows that machine learning and deep learning methods have been shown to markedly improve the detection of tuberculosis and related lung diseases, highlighting methods such as hybrid feature fusion, reinforcement learning, and neural architecture search are showing the best results. Each of these approaches are improving the accuracy and effectiveness of diagnostics, but a significant number of these works are based on a single data source, or have very little robustness across datasets. This constitutes an urgent need for a multimodal and effective diagnostic framework, which considers the following elements: patient's clinical, demographic or laboratory data and image-based features. Lastly, the gaps in existing literature have prompted this study to develop a more widespread machine learning modelling method to improve diagnosis reliability and hold real-life applicability.

3. PROPOSED METHODOLGY

This section describes the conceptual framework for diagnosing tuberculosis with a machine-learning model. The framework is based on combining, or fusing, image data and text data, utilizing advanced feature extraction techniques followed by feature fusion and classification. The proposed model is shown in figure 2. For image data (X-ray images), features are extracted using the VGG network, and for text-based features in the form of textual test reports, Glove embeddings are used to extract key features from the report text to build the feature vectors. The features from image data and text-based data are

fused to create a single vector, that is used in a transfer learning model to classify the tuberculosis presence, and finally, federated learning is accounted for to geometric the underlying model training even in a decentralized manner from separate datasets to improve privacy and ensure robustness.

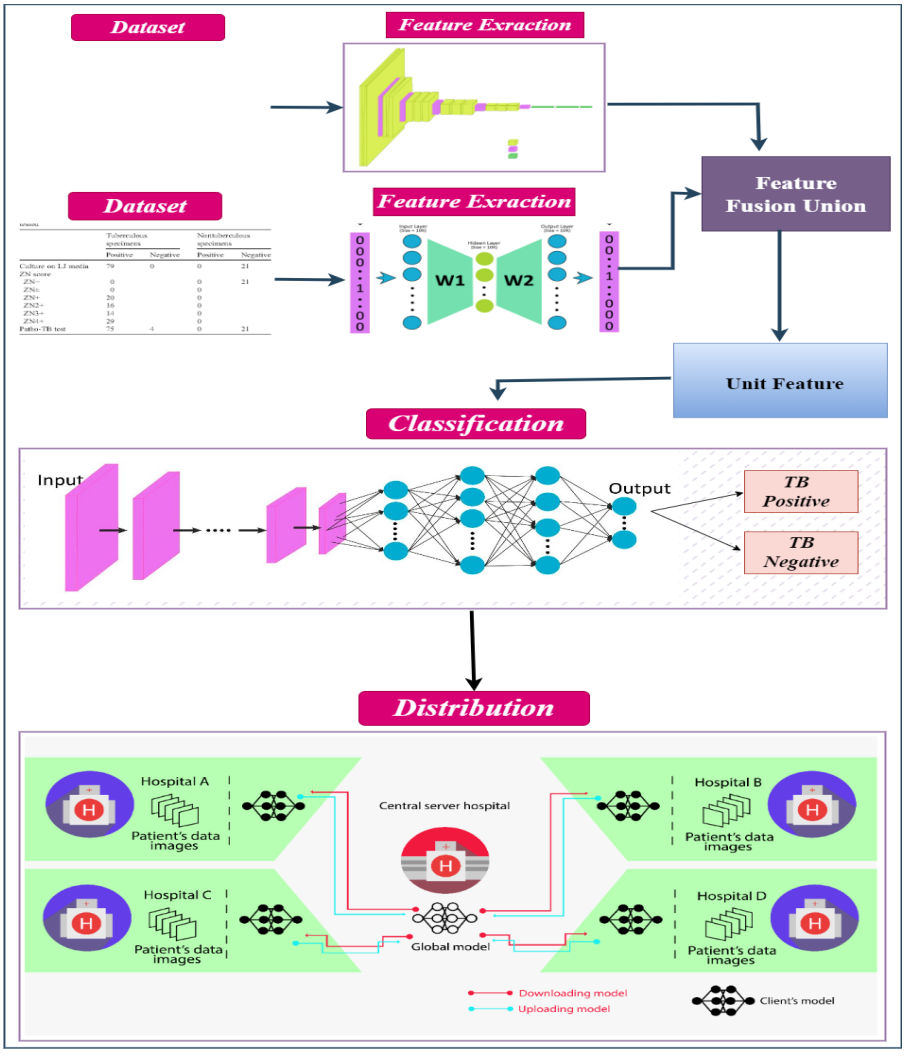

Fig. 2. Proposed Methodology

3.1 Preprocessing

This section outlines the preprocessing framework for tuberculosis (TB) diagnosis using a multimodal machine learning approach that integrates chest X-ray images and textual clinical reports. The objective is to transform raw data into a standardized and noise-reduced format suitable for effective feature extraction, fusion, and classification.

3.1.1Preprocessing Steps for Image Dataset of Tuberculosis

Preprocessing of chest X-ray images is essential to ensure consistency and enhance diagnostically relevant features. All images are resized to a uniform resolution (e.g., 224 × 224 pixels) to match the input requirements of convolutional neural networks such as VGG-16. Pixel values are normalized to the range [0, 1] to stabilize training and improve convergence. To improve model generalization, data augmentation techniques—including rotation, flipping, and scaling are applied to generate diverse training samples. Noise reduction methods, such as Gaussian filtering, are used to remove artifacts, while contrast enhancement techniques like histogram equalization or CLAHE improve the visibility of critical structures.

Edge sharpening is applied to highlight fine anatomical details. Lung segmentation is performed to isolate the region of interest, reducing background interference and enabling the model to focus on relevant features. Additionally, non-diagnostic elements (e.g., labels or borders) are removed through cropping. To ensure dataset quality, all images undergo annotation verification through cross-checking procedures. Finally, the dataset is divided into training, validation, and test

sets with balanced class distributions, followed by a quality assurance step to confirm the correctness of preprocessing operations.

3.1.2Preprocessing Steps for Test Report Dataset

Textual clinical reports are preprocessed to convert unstructured data into a structured representation suitable for embedding. The process begins with tokenization, where text is divided into individual words or tokens. Stop words (e.g., "and", "the", "is") are removed to reduce noise and retain meaningful content. Text normalization is applied to standardize terminology, including handling abbreviations and synonyms commonly found in clinical notes.
This step ensures consistency across the dataset. Additional noise removal techniques are employed to eliminate irrelevant characters and inconsistencies. The cleaned and standardized text is then prepared for feature extraction using word embedding techniques such as GloVe.

3.2 Feature Extraction

Feature extraction transforms unstructured data into structured representations suitable for machine learning models. In tuberculosis (TB) diagnosis from chest X-ray images, this step focuses on identifying meaningful patterns and structures that distinguish TB-positive from TB-negative cases. Effective feature extraction enhances both model accuracy and computational efficiency by capturing relevant visual and semantic information.

3.2.1 Feature Extraction of Image Dataset Using VGG Net.

In this study, the VGG network—specifically the VGG-16 architecture is used as the primary feature extractor for chest X-ray images. Pre-trained on the ImageNet dataset, VGG-16 provides robust initialization by learning fundamental visual patterns such as edges, textures, and contours, which are transferable to medical imaging tasks via transfer learning. Before feature extraction, all images are resized to 224×224 pixels to match the model's input requirements. Pixel normalization is then applied to scale intensity values to the range [0,1], improving training stability and reducing illumination variations.
The convolutional layers of the pre-trained VGG-16 are utilized as hierarchical feature detectors. Each convolutional layer applies a set of learnable kernels to the input feature maps to capture spatial dependencies and textural characteristics. The convolution operation can be mathematically expressed as:

$$F_{i,j}^{(k)} = \sum_m \sum_n X_{i+m,j+n} \cdot K_{m,n}^{(k)} + b^{(k)} \quad (1)$$

Where $F_{i,j}^{(k)}$ is the activation position (i,j) in the k^{th} feature map, X represents the input feature map, $K^{(k)}$ is the convolution kernel and $b^{(k)}$ denotes the bias. Non – linear activation functions such as ReLU are applied each convolution to introduce non-linearity and enhance representational capacity given by

$$f(x) = max(0, x) \quad (2)$$

This operation allows the model to learn complex, non-linear patterns essential for identifying subtle abnormalities in medical images. For the purpose of feature extraction, the fully connected layers of the VGG network are removed, and only the convolutional base is retained. The output of the final convolutional block, denoted as F_{conv}, constitutes a high-dimensional tensor containing spatially enriched feature maps. To prepare these for downstream classification, the tensor is flattened into a one-dimensional vector:

$$f = Flatten\ (F_{conv}) \in \mathbb{R}^d \quad (3)$$

where d denotes the dimensionality of the extracted features. Each image is thus represented as a high-dimensional vector capturing both low-level and high-level visual information, which is then used for classification. The VGG-16 architecture offers advantages such as a structured deep design and small 3×3 filters, enabling effective capture of both local and global image features. Its hierarchical representations significantly improve diagnostic performance in TB detection.

3.2.2 Feature Extraction of Test Report Dataset Using Glove Network.

Feature extraction from clinical text reports is essential for incorporating semantic information into the diagnostic model. The GloVe (Global Vectors for Word Representation) model converts textual data into continuous vector representations that preserve semantic relationships between words. The training objective of GloVe aims to minimize the difference between the dot product of two-word vectors and the logarithm of their co-occurrence probability, expressed as:

$$J = \sum_{i=1}^{V} \sum_{j=1}^{V} f(X_{ij})\ (w_i^T\ w_j + b_i + b_j - log X_{ij})^2 \quad (4)$$

where V is the vocabulary size, X_{ij} is the co-occurrence frequency between words i and j, w_i and w_j are the word and context vectors, b_i and b_j are bias terms, and $f(X_{ij})$ is a weighting function to balance the influence of rare and frequent words. After training on a large corpus such as Wikipedia or Common Crawl, GloVe provides dense word vectors $w_i \in \mathbb{R}^d$, where d represents the embedding dimension (commonly 50, 100, 200, or 300). For this study, a 100-dimensional GloVe model is used to provide an optimal balance between semantic richness and computational efficiency.

Each clinical report$R$, consisting of N words $\{w_1, w_2, \dots, w_n\}$ is converted into a sequence of word vectors $\{v_1, v_2, \dots, v_n\}$ where each $v_i \in \mathbb{R}^{100}$. To generate a single fixed-size feature vector representing the report, a simple yet effective averaging method is applied:

$$V_R = \frac{1}{N}\sum_{i=1}^{N} v_i \quad (5)$$

This averaged embedding acts as a semantic centroid, capturing the overall contextual meaning of the report. Alternatively, TF-IDF weighted averaging can be employed to emphasize more informative terms, formulated as:

$$V_R = \frac{\sum_{i=1}^{N} TF\text{-}IDF\ (w_i).v_i}{\sum_{i=1}^{N} TF\text{-}IDF(w_i)} \quad (6)$$

The resulting report-level vector V_R is then used as the input to a classification model, which distinguishes between TB-positive and TB-negative cases based on the extracted textual features. The classification function can be represented as:

$$y = \sigma\ (W v_R + b) \quad (7)$$

where W and b denote the weight matrix and bias of the classifier, respectively, and $\sigma(\cdot)$ is the sigmoid activation function that outputs a probability y∈[0,1], indicating the likelihood of TB-positivity.

3.3 Feature Fusion

The integration of features from both the test report dataset and the image dataset constitutes an important point in the establishment of TB diagnosis model completeness and performance. By merging the detailed information from the textual descriptions and the image patterns, we intend to produce a single feature to portray the multi-dimensional characteristics of TB diagnosis. The initial process is to fuse both the feature vectors from the test report dataset from the GloVe network feature vectors and the image dataset from the VGG network feature vectors. Feature vectors are the characteristics derived from the textual and visual modalities, which contain complementary information regarding TB-related anomalies. It is essential to point out that prior to us being able to fuse the different modalities, the feature vectors must be transferred into a common feature space so that they can be fused and compared in an interpretable manner.

A common approach is concatenation, which involves simply stacking the feature vectors along the feature dimension, yielding a single long fused vector. Feature vectors can also be combined by using simple arithmetic operations such as addition or multiplication. The goal of these fusion strategies is to preserve the unique characteristics of each modality as well as its additive properties. Reducing the dimension results in increased computational efficiency and a reduction of overfitting. After fusion, the fused feature vectors may be normalized, so that each feature contributes equally in later analysis. For example, Z-score normalization or min-max scaling may be used in order to rescale the features to similar dimensions, allowing for fair comparison of features and avoiding metrics dominating the fused metric. Methods such as mutual information and recursive feature elimination may also assist in determining and retaining only the relevant and different features in a subset. Reducing the dimension to a subset improves model interpretability and generalizability, and only a subset of features allows for further dimension reduction.

This process ensures that the representation obtained contains the needed information from both modalities, ensuring the two modalities are compatible for further analysis. The mixed features obtained from the combination process of the test report and images are accumulated. The purpose of the accumulation is to combine the useful information that each of these modalities brings to the table separately into a single representation. There are many ways to go about accumulating or combining features, such as an average, a weighted sum, or continuous feature concatenation. This unit representation feature stage is the critical interface between the fused test report and image data sets and the second classification stage. Just as it is needed in managing the fusion of the test report images appropriately, it is equally important in the aspect of accumulating or unit representation to bring the fused features together. By having a unit representation, the model/may now assume to provide models a common unified format to work from that also taps into the information from test report images, thereby improving TB diagnostic accuracy

3.4 Classification

Transfer learning is a widely adopted deep learning strategy that allows the reuse of knowledge learned from large-scale datasets for new, domain-specific tasks [30]. In tuberculosis (TB) diagnosis from chest X-ray images, transfer learning enables the adaptation of pre-trained convolutional neural networks (CNNs) to effectively recognize disease patterns, despite the limited availability of labeled medical data. In this study, the ResNet50 architecture is utilized as the feature extractor and classification backbone. ResNet, or Residual Network, introduces skip connections that overcome the vanishing gradient problem, allowing efficient training of very deep networks. The core operation of a residual block can be mathematically represented as:

$$y_l = \overline{J^{-}}(x_l, W_1) + x_l \quad (8)$$

Where x_l and y_l denote the input and output of the l^{th} layer, W_l represents the learnable parameters and $\overline{J^{-}}(x_l, W_1)$ is the residual mapping learned through convolution, batch normalization and ReLU activation. This formulation enables gradient flow access layers more stable and accurate learning.

The pre-trained ResNet50 model, originally trained on the ImageNet dataset, is adapted for TB classification by retaining its convolutional base for hierarchical feature extraction and replacing the final fully connected layer with a Global Average Pooling (GAP) layer followed by a dense classification layer. The classification layer outputs the probability of TB presence using a sigmoid activation function, expressed as:

$$y = \sigma\left(W.GAP\left(F_{conv}\right) + b\right) \quad (9)$$

Where F_{conv} represents the final convolutional feature map, W and b are the weights and bias of the classification layer. Model optimization is performed through backpropagation by minimizing the binary cross-entropy loss, which measures the difference between the predicted and actual class labels:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log\left(\hat{y}_i\right) + (1 - y_i)\log\left(1 - \hat{y}_i\right)] \quad (10)$$

Where y_i and $\hat{y}_i$ denote the true and predicated labels of the i^{th} X-ray image, respectively and N is the total number of samples. To enhance model generalization, data augmentation techniques such as rotation, flipping, and contrast adjustment are applied, while dropout and weight decay regularization help prevent overfitting. By leveraging the residual learning capability of ResNet and the efficiency of transfer learning, the proposed approach achieves high diagnostic accuracy, enabling reliable differentiation between TB-positive and TB-negative cases from chest X-ray images.

4. RESULTS

In this section, the experimental results are presented and the performance of the proposed machine learning model for tuberculosis (TB) diagnosis is evaluated. The model integrates chest X-ray images and clinical test reports to enable a comprehensive and accurate diagnostic assessment.

4.1 Dataset Description

Three different types of datasets have been used for the analysis of the proposed model for tuberculosis (TB) diagnosis. The TB-DS datasets contain three separate collections to help develop models for diagnosing TB. TB-DS-I consists of 58,402 fully annotated chest X-ray images obtained from the National Library of Medicine (NLM), Kaggle, and GitHub. The chest X-ray images were taken from many different medical devices; therefore, they are represented as JPG (224x224 pixels) cases from both TB-positive and TB-negative studies, depicting a wide range of ages and demographics. TB-DS-II consists of over 35,000 de-identified chest X-ray images that include radiology reports (de-identified) from open-sourced sites. All the radiology reports were verified by medical professionals in the respective field, are de-identified, and were identified in their own study cases and vary in length and present possible clinical outcomes. TB-DS-III combines information from both images and associated reports from the Mimic-CXR radiology report content and presents a multimodal dataset with over 35,000 cases that contain a combination of high-quality chest X-ray images and textual reports. TB-DS, as a dataset, allows clinicians to use these two modalities to fully comprehend the interaction between the images of the recognized TB.

4.2 Baseline Approaches

To assess the performance and reliability of the proposed machine learning model designed for tuberculosis (TB) diagnosis, the following baseline methods were employed:

- Baseline 1 [31]: Vats et al. introduced a cascaded model incorporating incremental learning techniques to identify and pinpoint cases of tuberculosis using chest X-ray imagery. Their approach combines incremental learning with cascaded models to enhance the performance and adaptability of TB detection systems.
- Baseline 2 [21]: Abdar et al. proposed Uncertainty FuseNet, an advanced uncertainty-aware model designed for hierarchical feature fusion, which incorporates ensemble Monte Carlo dropout techniques to enhance the accuracy of COVID-19 detection.
- Baseline 3 [14]: Momeny et al. studied MTB categorization using a deep extended CNN. Their approach involved the mean square error and hybrid sharing, incorporating the Greedy Auto augment technique.

4.3 Results

Experimental validation of our proposed machine learning model for tuberculosis (TB) diagnosis that uses both X-ray images and reports generates significant improvement over baseline methods. In this subsection, we provide detailed results assessing the performance of our proposed model on three datasets (TB-DS-I, TB-DS-II, and TB-DS-III) and precision, recall, F1-score, accuracy, specificity, and sensitivity metrics. The results as shown in figure 3 demonstrate the capability of our model in accurately classifying TB-positive and TB-negative cases in varying clinical contexts. The model's performance on the TB-DS-I dataset, evaluated with 94.84% precision, 94.56% recall, 95.85% accuracy, and 94.32% F1-score, was a tremendous performance, presenting a large and diverse dataset of chest X-ray images. For the second TB-DS-II dataset, containing the reports in addition to the images, resulted in a similar impressive performance with 96.74% precision, 96.5% recall, 96.15% F1-score, and 97.02% accuracy. TB-DS-III multimodal datasets, which used X-ray images and reports,

provided even higher performance metrics than did TB-DS-II, with a precision of 97.83%, a recall of 97.54%, an F1-score of 97.2%, and an overall accuracy of 98.03%.

These matrices show that the model consistently performed well at correctly classifying TB-positive and TB-negative subjects, and the number of misclassifications was small lending further evidence to the claims of the diagnostic accuracy of the proposed diagnostic tool. The results were compared to a simple baseline using various conventional methods, namely convolutional neural networks (CNN), BiLSTM, and a combination of CNN, and LSTMs. The results showed that our model consistently outperformed conventional methods on all performance metrics, especially in the regard to integrating multimodal data, specifically in the case of X-ray images and radiology reports, and based on advanced feature extraction. In addition, we also studied and presented whether federated learning improved performance and privacy. Our results indicated that federated learning improved the model's ability for accuracy while being able to solve the privacy problem, in terms of a model that can learn across independent healthcare institutions without exposing sensitive patient data. This made our model suitable for deployment in real-world medical settings that err on the side of privacy.

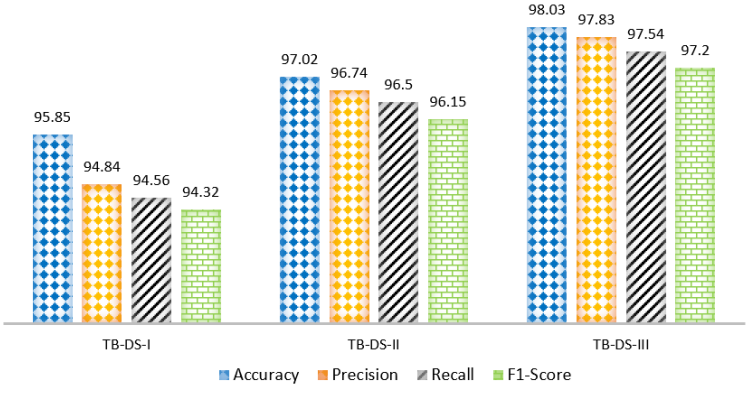

Fig. 3. Results on Different Datasets

These matrices show that the model consistently performed well at correctly classifying TB-positive and TB-negative subjects, and the number of misclassifications was small lending further evidence to the claims of the diagnostic accuracy of the proposed diagnostic tool. The results were compared to a simple baseline using various conventional methods, namely convolutional neural networks (CNN), BiLSTM, and a combination of CNN, and LSTMs. The results showed that our model consistently outperformed conventional methods on all performance metrics, especially in the regard to integrating multimodal data, specifically in the case of X-ray images and radiology reports, and based on advanced feature extraction. In addition, we also studied and presented whether federated learning improved performance and privacy. Our results indicated that federated learning improved the model's ability for accuracy while being able to solve the privacy problem, in terms of a model that can learn across independent healthcare institutions without exposing sensitive patient data. This made our model suitable for deployment in real-world medical settings that err on the side of privacy. The graphical representation of the data can be seen in Figure 4.

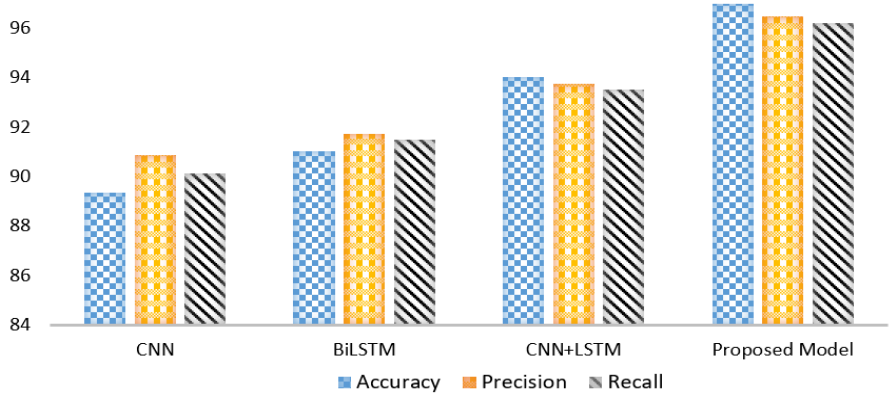

Fig. 4. Comparison of Proposed Model with CNN, BiLSTM and CNM+LSTM

The performance of the proposed machine learning model and compared to three baseline models in accuracy, precision, and recall, as illustrated in Figure 5. Baseline 1 achieved moderate results with an accuracy of 91.36%, a precision of 90.87%, and a recall of 90.32%. Although Baseline 1 performed reasonably, it was outperformed by the more complex methods. Baseline 2 improved significantly with an accuracy of 95.02%, a precision of 95.74%, and a recall of 95.5%. However, it still could not outperform the proposed model. Baseline 3 performed slightly better than baseline one with an accuracy of

94.12%, a precision of 93.84%, and a recall of 93.6%; however, it could not surpass the proposed model. In general, the proposed model performed better than all three baselines with an accuracy of 96.96%, a precision of 96.47%, and a recall of 96.2%. These results illustrate the superior performance of the proposed model was primarily the result of extracting better features and using multimodal data that dramatically decreased false positives and false negatives. The comparison analysis showed that the proposed model's strong accuracy, precision, and recall came from its ability to successfully integrate clinical and laboratory data and offer increased dependability and efficiency for TB diagnosis.

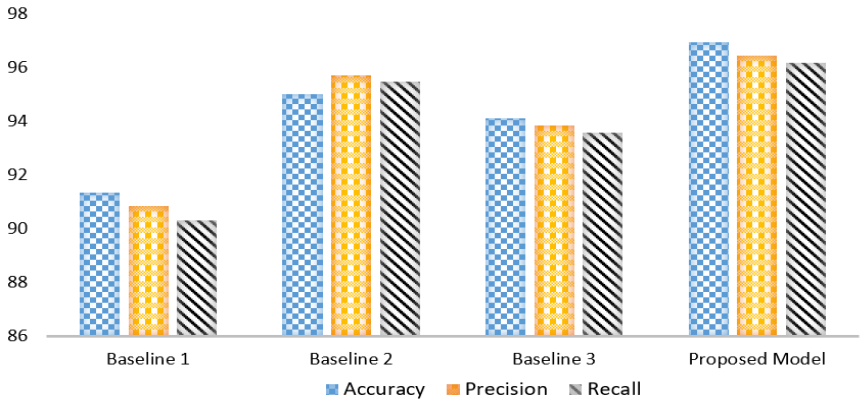

Fig. 5. Comparison of Proposed Model with baseline approaches

5. CONCLUSION

The research provided a noval machine-learning model for diagnosing tuberculosis (TB), with a combination of clinical and laboratory data, such as demographics, symptoms, and diagnostic test data. The model, which utilized advanced techniques regarding feature extraction, fusion, and classification, provided superior accuracy, compared to existing methods. Based on the results, the new model demonstrated consistency in identifying both TB-positive and TB-negative cases without significant misclassifications. Additionally, the comparative analysis that included baseline models demonstrated the efficacy of our new model, particularly its capabilities to perform TB diagnosis leveraging multimodal data. The medical utility of our model was substantiated not only by strong performance metrics but through fewer false positives and true positives. Furthermore, using federated learning techniques ensured that privacy concerns were addressed and successfully applied machine-learning techniques while minimizing privacy, and nevertheless maintained a high degree of diagnostic performance. Thus, presenting a high-performance model existent and suitable for high-focus communities of practitioners that can be developed in a distributed manner, without the restrictions of centralizing data gathering. Ultimately, the results from this study, convey the strong potential of machine learning models to help dynamically shape the future of TB diagnosis that are reliable, efficient, scalable, and clinically applicable tools that can augment existing healthcare pathways. Future work will further enhance the proposed model by incorporating additional data modalities, such as genomic and treatment history data, to improve diagnostic precision. Additionally, efforts will be made to optimize the model for real-time clinical use, ensuring faster and more efficient decision-making.

Funding:

The authors declare that no financial aid or sponsorship was received from any external agencies or institutions for this study. All research activities were independently carried out.

Conflicts of Interest:

The authors declare no conflicts of interest.

Acknowledgment:

The authors are sincerely grateful to their institutions for their invaluable guidance and technical support.

References

[1] H. Qureshi, Z. Shah, M. A. Z. Raja, M. Y. Alshahrani, W. A. Khan, and M. Shoaib, "Machine Learning Investigation of Tuberculosis with Medicine Immunity Impact," Diagn. Microbiol. Infect. Dis., vol. 110, no. 3, Art. no. 116471, 2024.

[2] A. Dixit, L. Freschi, R. Vargas, M. I. Gröschel, M. Nakhoul, S. Tahseen, et al., "Estimation of Country-Specific Tuberculosis Resistance Antibiograms Using Pathogen Genomics and Machine Learning," BMJ Glob. Health, vol. 9, no. 3, Art. no. e01353, 2024.

[3] G. Tamura, G. Llano, A. Aristizábal, J. Valencia, L. Sua, and L. Fernandez, "Machine-Learning Methods for Detecting Tuberculosis in Ziehl-Neelsen Stained Slides: A Systematic Literature Review," Intell. Syst. Appl., Art. no. 200365, 2024.
[4] S. Feng, S. Wang, C. Liu, S. Wu, B. Zhang, C. Lu, et al., "Prediction Model for Spinal Cord Injury in Spinal Tuberculosis Patients Using Multiple Machine Learning Algorithms: A Multicentric Study," Sci. Rep., vol. 14, no. 1, Art. no. 7691, 2024.
[5] Y. Xiao, Y. Chen, R. Huang, F. Jiang, J. Zhou, and T. Yang, "Interpretable Machine Learning in Predicting Drug-Induced Liver Injury Among Tuberculosis Patients: Model Development and Validation Study," BMC Med. Res. Methodol., vol. 24, no. 1, Art. no. 92, 2024.
[6] L. Li, H. Huang, and X. Jin, "AE-CNN Classification of Pulmonary Tuberculosis Based on CT Images," in Proc. 9th Int. Conf. Inf. Technol. Med. Educ. (ITME), 2018, pp. 39–42.
[7] M. Singh, G. V. Pujar, S. A. Kumar, M. Bhagyalalitha, H. S. Akshatha, B. Abuhaija, et al., "Evolution of Machine Learning in Tuberculosis Diagnosis: A Review of Deep Learning-Based Medical Applications," Electronics, vol. 11, no. 17, Art. no. 2634, 2022.
[8] M. Oloko-Oba and S. Viriri, "A Systematic Review of Deep Learning Techniques for Tuberculosis Detection from Chest Radiograph," Front. Med., vol. 9, Art. no. 830515, 2022.
[9] E. Showkatian, M. Salehi, H. Ghaffari, R. Reiazi, and N. Sadighi, "Deep Learning-Based Automatic Detection of Tuberculosis Disease in Chest X-Ray Images," Pol. J. Radiol., vol. 87, Art. no. e118, 2022.
[10] V. Acharya, G. Dhiman, K. Prakasha, P. Bahadur, A. Choraria, S. Prabhu, et al., "AI-Assisted Tuberculosis Detection and Classification from Chest X-Rays Using a Deep Learning Normalization-Free Network Model," Comput. Intell. Neurosci., vol. 2022, no. 1, Art. no. 2399428, 2022.
[11] S. Govindarajan and R. Swaminathan, "Analysis of Tuberculosis in Chest Radiographs for Computerized Diagnosis Using Bag of Keypoint Features," J. Med. Syst., vol. 43, pp. 1–9, 2019.
[12] G. Han, X. Liu, H. Zhang, G. Zheng, N. Q. Soomro, M. Wang, and W. Liu, "Hybrid Resampling and Multi-Feature Fusion for Automatic Recognition of Cavity Imaging Sign in Lung CT," Future Gener. Comput. Syst., vol. 99, pp. 558–570, 2019.
[13] H. Cao, H. Liu, E. Song, G. Ma, X. Xu, R. Jin, and C. C. Hung, "A Two-Stage Convolutional Neural Networks for Lung Nodule Detection," IEEE J. Biomed. Health Inform., vol. 24, no. 7, pp. 2006–2015, 2020.
[14] M. Momeny, A. A. Neshat, A. Gholizadeh, A. Jafarnezhad, E. Rahmanzadeh, M. Marhamati, and Y. D. Zhang, "Greedy Autoaugment for Classification of Mycobacterium Tuberculosis Image via Generalized Deep CNN Using Mixed Pooling Based on Minimum Square Rough Entropy," Comput. Biol. Med., vol. 141, Art. no. 105175, 2022.
[15] S. Y. Lu, S. H. Wang, X. Zhang, and Y. D. Zhang, "TBNet: A Context-Aware Graph Network for Tuberculosis Diagnosis," Comput. Methods Programs Biomed., vol. 214, Art. no. 106587, 2022.
[16] S. Dey, R. Roychoudhury, S. Malakar, and R. Sarkar, "An Optimized Fuzzy Ensemble of Convolutional Neural Networks for Detecting Tuberculosis from Chest X-Ray Images," Appl. Soft Comput., vol. 114, Art. no. 108094, 2022.
[17] M. Rahman, Y. Cao, X. Sun, B. Li, and Y. Hao, "Deep Pre-Trained Networks as a Feature Extractor with XGBoost to Detect Tuberculosis from Chest X-Ray," Comput. Electr. Eng., vol. 93, Art. no. 107252, 2021.
[18] A. Iqbal and M. Sharif, "MDA-Net: Multiscale Dual Attention-Based Network for Breast Lesion Segmentation Using Ultrasound Images," J. King Saud Univ. Comput. Inf. Sci., vol. 34, no. 9, pp. 7283–7299, 2022.
[19] A. Kukker and R. Sharma, "Modified Fuzzy Q Learning-Based Classifier for Pneumonia and Tuberculosis," IRBM, vol. 42, no. 5, pp. 369–377, 2021.
[20] T. Khatibi, A. Shahsavari, and A. Farahani, "Proposing a Novel Multi-Instance Learning Model for Tuberculosis Recognition from Chest X-Ray Images Based on CNNs, Complex Networks, and Stacked Ensemble," Phys. Eng. Sci. Med., vol. 44, pp. 291–311, 2021.
[21] M. Abdar, S. Salari, S. Qahremani, H. K. Lam, F. Karray, S. Hussain, and S. Nahavandi, "UncertaintyFuseNet: Robust Uncertainty-Aware Hierarchical Feature Fusion Model with Ensemble Monte Carlo Dropout for COVID-19 Detection," Inf. Fusion, vol. 90, pp. 364–381, 2023.
[22] S. H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, and Y. D. Zhang, "COVID-19 Classification by FGCNet with Deep Feature Fusion from Graph Convolutional Network and Convolutional Neural Network," Inf. Fusion, vol. 67, pp. 208–229, 2021.
[23] S. Bamrah, R. S. Yelk Woodruff, K. Powell, S. Ghosh, J. S. Kammerer, and M. B. Haddad, "Tuberculosis Among the Homeless, United States, 1994–2010," Int. J. Tuberc. Lung Dis., vol. 17, no. 11, pp. 1414–1419, 2013.
[24] C. A. Bobak, A. J. Titus, and J. E. Hill, "Comparison of Common Machine Learning Models for Classification of Tuberculosis Using Transcriptional Biomarkers from Integrated Datasets," Appl. Soft Comput., vol. 74, pp. 264–273, 2019.
[25] M. R. R. Rana, A. Nawaz, T. Ali, A. S. Alattas, and D. S. AbdElminaam, "Sentiment Analysis of Product Reviews Using Transformer Enhanced 1D-CNN and BiLSTM," Cybern. Inf. Technol., vol. 24, no. 3, pp. 112–131, 2024.
[26] J. Lin and Q. Chen, "An Intelligent Sensor Data Preprocessing Method for OCT Fundus Image Watermarking Using an RCNN," CMES-Comput. Model. Eng. Sci., vol. 138, no. 2, 2024.
[27] H. J. D. Koresh, "Impact of the Preprocessing Steps in Deep Learning-Based Image Classifications," Natl. Acad. Sci. Lett., pp. 1–3, 2024.
[28] N. J. Dabanloo, S. M. J. Hosseini, and K. Maghooli, "Detection of COVID-19 Using VGG-16 Neural Network and Classification of Chest X-Ray Images," J. Mar. Med., vol. 6, no. 1, pp. 64–71, 2024.
[29] L. Abualigah, Y. Y. Al-Ajlouni, M. S. Daoud, M. Altalhi, and H. Migdady, "Fake News Detection Using Recurrent Neural Network Based on Bidirectional LSTM and GloVe," Soc. Netw. Anal. Min., vol. 14, no. 1, Art. no. 40, 2024.
[30] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A Comparison Review of Transfer Learning and Self-Supervised Learning: Definitions, Applications, Advantages, and Limitations," Expert Syst. Appl., vol. 242, Art. no. 122807, 2024.
[31] S. Vats, V. Sharma, K. Singh, A. Katti, M. M. Ariffin, M. N. Ahmad, et al., "Incremental Learning-Based Cascaded Model for Detection and Localization of Tuberculosis from Chest X-Ray Images," Expert Syst. Appl., vol. 238, Art. no. 122129, 2024.