Improved Pyramid Histogram of Oriented Gradients and Statistical Features for Tuberculosis Detection via a Hybrid Classifier Using X-Ray Images

Rathod Dharmesh Ishwerlal
Amity Institute of Information Technology, Amity University, Noida, India | Department of Information Technology, Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS (Deemed to be University), Mumbai, India
dharmesh.phd@gmail.com (corresponding author)

Reshu Agarwal
Amity Institute of Information Technology, Amity University, Noida, India
agarwal.reshu3@gmail.com

K. S. Sujatha
JSS Academy of Technical Education, Noida, India
bsoorya@rediffmail.com

Received: 15 April 2025 | Revised: 5 June 2025, 16 June 2025, and 24 June 2025 | Accepted: 27 June 2025

 DOI: https://doi.org/10.48084/etasr.11513

ABSTRACT

Tuberculosis (TB) is a bacterial infection primarily affecting the lungs, which is diagnosed using methods such as skin tests, Purified Protein Derivative (PPD) test, blood tests, and Chest X-rays (CXRs). Traditionally, diagnostic test results were examined manually by doctors; however, in recent years, the advancements of Machine Learning (ML) have benefited the medical diagnosis process by offering an additional automated classification based on prognostic tests performed. In this study, a novel Hybrid LinkNet-Recurrent Neural Network (LinkNet-RNN) model is proposed for TB detection trained on CXR images. In the models' preprocessing stage, input X-ray images are enhanced using a median filter to reduce noise, while during feature extraction, an Improved Pyramid Histogram of Oriented Gradients (I-PHOG) method is employed to capture gradient information across multiple scales. In the final stage, the LinkNet-RNN model classifies each case as TB or not TB. The model's performance was evaluated through comprehensive experimental analysis, using the TB Chest X-ray Database as the train/test set, and compared against existing prediction models to validate its effectiveness, including standalone LinkNet, standalone RNN, SqueezeNet, LeNet, and Convolutional Neural Networks (CNN).

Keywords-tuberculosis detection; LinkNet-Recurrent Neural Network; X-ray image; Improved Pyramid Histogram of Oriented Gradients (I-PHOG); median filter

I. INTRODUCTION

According to the World Health Organization (WHO) [1], Tuberculosis (TB), caused by *Mycobacterium tuberculosis* [2], was responsible for over 1.5 million deaths worldwide in 2018 [3], and is among the top ten leading causes of death globally [4, 5]. Annually, approximately 10 million people are infected [6], with early symptoms often being mild. Still, without timely detection and treatment, TB can progress, leading to lung tissue necrosis and severe complications such as massive hemoptysis [4]. While antimicrobial drugs can effectively treat TB [2], early detection remains critical.

For early diagnosis of pulmonary TB, Chest X-rays (CXRs) are commonly used as a standard, cost-effective, and widely available diagnostic tool [2, 7]. However, CXRs can be challenging to interpret because TB abnormalities may resemble other lung pathologies [7], which makes the interpretation subjective, depending heavily on the radiologists' expertise [4]. However, advances in Machine Learning (ML) and Deep Learning (DL) have aided TB detection from CXRs by automating image interpretation, reducing inter-reader variability, and enabling diagnostic services in resource-limited settings [6]. Specifically, DL architectures and Convolutional

Neural Networks (CNNs) have shown promising results in differentiating TB from non-TB cases [6, 8].

In this context, this study proposes a DL-based TB detection model using CXR images, with a key contribution being an enhanced Improved Pyramid Histogram of Oriented Gradients (I-PHOG) feature extraction method. The refinement of the model involved incorporating a normalized exponential Laplacian absolute measure to improve gradient magnitude estimation across multiple scales, reducing high-frequency noise influence, and highlighting significant gradient transitions.

A. Literature Review

Several studies have investigated the implementation of ML and DL architectures for TB diagnosis based on CXR images. For instance, authors in [1] introduced a hybrid DL model combining a data-efficient image transformer with ResNet-16. The transformer component leveraged self-attention, while the depth-wise convolution of ResNet-16 enhanced computational efficiency and precision. Using the TBX11K dataset, which includes Healthy, Sick but non-TB, and TB categories, the model effectively reduced False Positives (FP), while resulting in a lightweight (6.9M parameters) and fast (4.79 ms inference time) model. Moreover, authors in [4] proposed an automatic TB detection method employing a CNN with a coordinate attention mechanism. This architecture enhanced cross-channel, directional, and positional information for improved classification accuracy, and utilized transfer learning to accelerate model convergence, resulting in 97.71% Area Under the Curve (AUC), 92.73% accuracy.

Authors in [6] developed a TB detection system using stochastic learning with an Artificial Neural Network (ANN) applied to CXR images. The ANN was optimized by introducing random variations in functions and weights, combined with shuffled training data. Evaluated on the Shenzhen and Montgomery datasets, the system achieved 98.35% accuracy, 98.02% specificity, 96.14% sensitivity, and an F1-score of 95.88%, outperforming conventional models. Additionally, authors in [7] proposed a novel DL framework for classifying TB, non-TB, and healthy patients using 40,000 CXR images, alongside a stacking ML approach for drug-resistant TB diagnosis on 3,037 images. The model incorporated Score-CAM for interpretability and achieved 93.32% accuracy for general classification, 87.48% accuracy for binary, and 79.59% accuracy for multi-class drug-resistant TB detection. Lastly, authors in [8] developed the CBAMWDNet model, which integrates the Convolutional Block Attention Module (CBAM) with Wide Dense Net (WDnet) for TB detection from CXR images. The model achieved 95.7% specificity, 94.28% sensitivity, 98.80% accuracy, 98.50% precision, and a 96.35% F1-score, demonstrating excellent generalization across multiple datasets.

II. DATASET DESCRIPTION

For the purposes of this work, the TB Chest X-ray Database is employed [9]. This dataset was developed collaboratively by the Qatar University (Doha, Qatar), the University of Dhaka (Bangladesh), partner institutions in Malaysia, and the medical experts from Hamad Medical Corporation (Qatar) and Bangladesh. It contains both TB-positive and normal images, including 700 TB images from various publicly available datasets, 2,800 TB images from the NIAID TB Portal Program, and 3,500 normal images from two separate datasets.

To evaluate system performance, experiments were conducted with varying proportions of training data (60/40, 70/30, 80/20, 90/10). For 60% training data, there were 2,520 training images and 1,680 testing images; for 70%, 2,940 training and 1,260 testing images; for 80%, 3,360 training and 840 testing images; and for 90%, 3,780 training and 420 testing images.

III. PROPOSED FRAMEWORK

The novel LinkNet-RNN model proposed for TB detection via CXR images operates in three main stages:

- Preprocessing stage: Apply median filtering to reduce noise in the input X-ray image.
- Feature extraction stage: Retrieve features from the pre-processed image using I-PHOG, hierarchical skeleton features, and statistical features.
- Detection stage: Use the LinkNet-RNN model to detect TB based on the extracted features.

The LinkNet-RNN model architecture and workflow are illustrated in Figure 1.

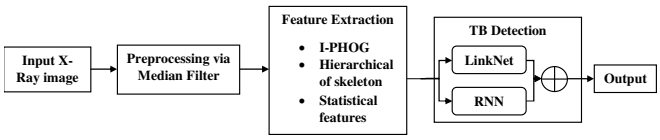


Fig. 1. Architecture of the LinkNet-RNN model.

A. Preprocessing via Median Filter

The median filter is a widely used nonlinear method for noise reduction [10] in input CXR images, denoted as $x_i = x_1, x_2, x_3, \ldots, x_n$. This filter operates by analyzing a local neighborhood around each pixel in the image x_i, typically defined by a kernel. For each pixel, the median value of the pixel intensities within this neighborhood is computed and replaces the original pixel value. In cases with an even number of pixels in the kernel, the median is the average of the two central values. The formula for the median filter is given in (1) [10].

$$\hat{I}(e,f) = Me\{h(r,s)\}, \text{where } (r,s) \in Re_{uv} \quad (1)$$

where Re_{uv} denotes the set of coordinates within the rectangular window centered at (e,f), and $Me\{\cdot\}$ represents the median value in that window. The result is a noise-reduced image, denoted as pr_x (preprocessed image).

B. Feature Extraction

Feature extraction aims to identify discriminative characteristics in pr_x that are useful for TB detection. In this work, we extract I-PHOG, hierarchical skeleton features, and statistical measures.

1) I-PHOG

PHOG enhances image representation by dividing a CXR image into progressively finer grids, capturing spatial information through a pyramid-like structure [11]. This technique uses spatial pyramid matching to analyze gradients and orientations at multiple levels of detail. However, conventional PHOG may not capture global contextual information without incorporating additional features. Therefore, I-PHOG is employed in this research, with Figure 2 depicting the process of the I-PHOG-based feature.

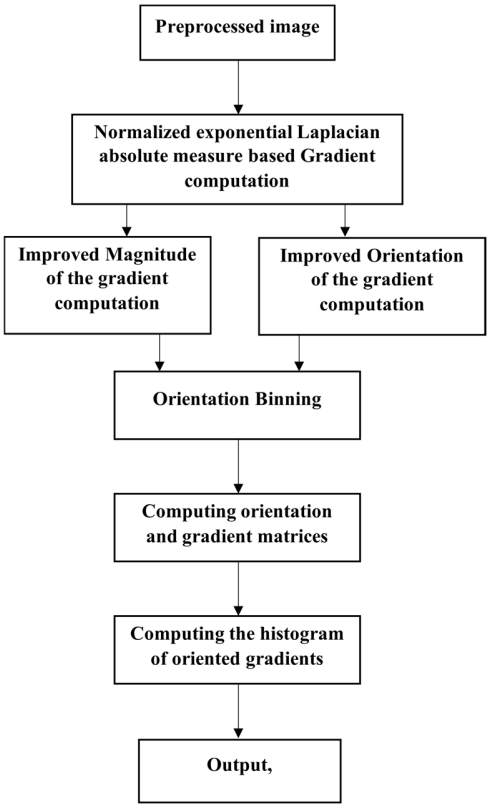


Fig. 2. Process of the I-PHOG-based feature.

To retrieve I-PHOG features from a preprocessed CXR image pr_x the following procedure needs to be followed:

a) Gradient Computation

To determine the gradient magnitude, the I-PHOG technique applies the normalized exponential Laplacian absolute measure. First, exponential and Laplacian kernel masks are applied in both horizontal and vertical directions to the grayscale image [11], as specified in (2):

$$k(p,q) = \left[exp\left(-\frac{\|p-q\|}{2\sigma^2}\right) + exp\left(-\frac{\|p-q\|}{\sigma}\right)\right] \quad (2)$$

The result is then multiplied by the hyperbolic tangent function, as defined in (3):

$$[k(p,q)]' = k(p,q) * tanh(\alpha p^T q + z) \quad (3)$$

Subsequently, the result is normalized using double sigmoid normalization, as described in (4):

$$Sig_k^N = \begin{cases} \frac{1}{1+exp\left(-2\left(k(p,q)\right)^1 - t/r_1\right)} & if[k(p,q)]' < t \\ \frac{1}{1+exp\left(-2\left(k(p,q)\right)^1 - t/r_2\right)} & if[k(p,q)]' \geq t \end{cases} \quad (4)$$

where, r_1, r_2, z, t are parameters and $\alpha = 0.01$.

The horizontal and vertical gradients are then estimated using (5) and (6):

$$R_p = \sqrt{I * D_p + (I * Sig_k^N)} \quad (5)$$

$$R_q = \sqrt{I * D_q + (I * Sig_k^N)} \quad (6)$$

The overall gradient magnitude is computed as in (7) [12]:

$$R_{Gradient} = \sqrt{\left(R_p\right)^2 + \left(R_q\right)^2}, cot\left(R_p + R_q\right) = $$

$$= \left[\frac{cot\, R_p cot\, R_q - 1}{cot\, R_p + cot\, R_q}\right] \quad (7)$$

In conventional image processing, the gradient orientation at a pixel is computed as:

$$\theta = \arctan\left(\frac{R_q}{R_p}\right) \quad (8)$$

where this formulation corresponds to the standard Histogram of Oriented Gradients (HOG) method. In contrast, the I-PHOG approach introduces a correction term based on the cotangent of the combined gradients [12]:

$$R_{orientation} = arctan\left(\frac{R_q}{R_p}\right).cot\left(R_p + R_q\right) \quad (9)$$

Using the cotangent sum identity:

$$\cot(A+B) = \frac{cotA \cdot cotB - 1}{cotA + cotB}$$

and substituting accordingly to (9) results in:

$$R_{orientation} = arctan\left(\frac{R_q}{R_p}\right).\left[\frac{cot\, R_p cot\, R_q - 1}{cot\, R_p + cot\, R_q}\right] \quad (10)$$

which is an equivalent reciprocal form.

b) Creating the Cell Histogram

Each pixel within a defined cell casts a weighted vote for an orientation-based histogram channel, based on values derived from gradient computations of the CXR images. The histogram channels are distributed across an angular range, typically from 0° to 180° or 0° to 360°, depending on whether the gradient is treated as "unsigned" or "signed".

c) Creating Computing Orientation and Gradient Matrices

A matrix is populated with gradient values corresponding to the histogram bins, while another matrix stores the associated gradient magnitudes extracted from the CXR images.

d) Computing HOG

The bin size for HOG remains fixed at K bins (with $K = 8$ in our case). Each bin represents the number of pixels with

orientations within a specific angular range. The CXR image is divided into four regions, and the HOG is computed for each. This process is recursively applied up to a specified depth L, partitioning the image into cells at multiple pyramid levels. Level 0 is represented by a k-dimensional vector corresponding to the K bins of the histogram, while Level 1 corresponds to a $4k$-dimensional vector:

$$k * \sum_{l=0}^{L} 4^l \quad (11)$$

The final I-PHOG feature vector for a CXR image is obtained by concatenating vectors from all pyramid resolutions ($I - PHOG_x$). This approach preserves gradient information across multiple scales, enhances the accuracy of gradient magnitude estimation, reduces high-frequency noise, and emphasizes significant gradient changes.

2) Hierarchical Skeleton Features

The LinkNet-RNN model employs hierarchical skeleton-based features. These features capture the medial axis of shapes within the image, providing a detailed structural representation. The hierarchy consists of broad parent axes outlining the overall shape and finer offshoot axes representing specific components [13]. This comprehensive feature extraction improves the model's ability to identify TB-related anomalies in CXR images.

3) Statistical Features

Statistical features refer to quantitative measures, such as standard deviation, median, and mean, retrieved from the preprocessed image pr_x, to describe its intensity distribution and highlight relevant patterns. Median Med represents the central pixel intensity value, calculated as [14]:

$$Med = median\{pr_x(r_{row}, c_{column}) | (r_{row}, c_{column}) \in W_i\} \quad (12)$$

Standard deviation Std measures the variation in pixel intensities [14]:

$$Std = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(pr_x - mean)} \quad (13)$$

Mean calculates the average pixel intensity [14]:

$$mean = \frac{1}{mn}\sum_{(r_{row}, c_{column}) \in W} pr_x\,(r_{row}, c_{column}) \quad (14)$$

The statistical feature output is denoted as St_x. The complete set of extracted features is therefore defined as:

$$fe_x = [I - PHOG_x, H_x, St_x] \quad (15)$$

C. TB Detection via LinkNet-RNN Model

In the final stage of TB detection, a hybrid DL model combining LinkNet and RNN is employed.

1) LinkNet model

The LinkNet architecture consists of an encoder, a decoder, and a pixel fusion pathway [15]. The process is initiated via input retrieved features fe_x. The encoder, composed of an initial block followed by four encoding blocks, extracts high-level semantic features from CXR images through convolution and max-pooling operations (stride = 2). The decoder, comprising four decoding blocks and a final block, restores the spatial resolution using deconvolution to produce detailed segmentation outputs. The pixel fusion pathway combines shallow encoder features with high-level decoder features through pixel-wise addition, improving TB detection accuracy. The output from this stage is denoted as LN_x. Figure 3 illustrates the full network structure.

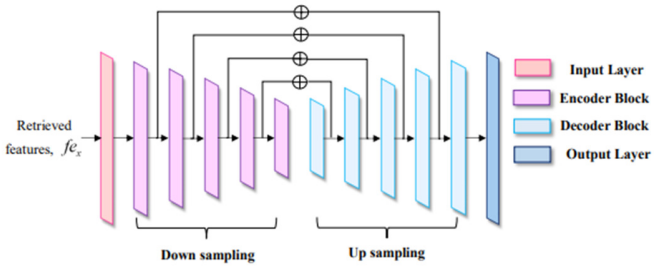


Fig. 3. Architecture of the LinkNet model.

2) Recurrent Neural Network (RNN)

A RNN [16] is an extension of the Feedforward Neural Network (FNN) that incorporates feedback loops, enabling effective processing of sequential data. In this TB detection framework, the RNN processes the retrieved features fe_x to capture temporal dependencies and patterns in sequential or structured feature sets. Its recurrent layers analyze these features over time, enabling the identification of complex patterns associated with TB. Given an input sequence $(a_1, \ldots, a_t)$, the RNN produces an output sequence $(b_1, \ldots, b_t)$, as illustrated in Figure 4, following (16):

$$h_t = sigmoid\left(W^{ha}A_t + W^{hh}h_{t-1}\right) \quad (16)$$

where A_t is the input feature vector at time step t, h_{t-1} is the hidden state from the previous time step, W^{ha} is the weight matrix connecting A_t to the hidden layer, and W^{hh} is the weight matrix connecting h_{t-1} to the current hidden layer. The corresponding output at each time step is computed as:

$$b_t = W^{bh}h_t \quad (17)$$

where W^{bh} is the weight matrix mapping the hidden state to the output space, and b_t represents the model's prediction or confidence score for TB presence within the given sequence. Through this combined architecture, the proposed LinkNet-RNN labels outputs as 0 for normal and 1 for TB, with the final output denoted as $LinkNet - RNN_x$.

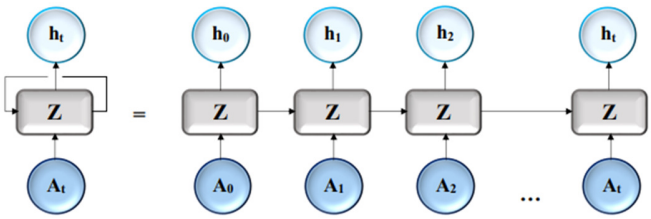


Fig. 4. Architecture of the RNN model.

D. Experimental Setup

The proposed TB detection system was implemented in Python 3.7, while all experiments were conducted on an Intel Core i5-4210U Central Processing Unit (CPU) (1.70 GHz) with 8 GB of Random Access Memory (RAM).

IV. RESULTS AND DISCUSSION

A. Comparative Assessment

The comparative assessment of the proposed model against LinkNet, RNN, SqueezeNet, LeNet, and CNN is based on metrics such as accuracy, precision, sensitivity, specificity, False Negative Rate (FNR), False Positive Rate (FPR), F-measure, Matthews Correlation Coefficient (MCC), and Negative Predictive Value (NPV). MCC is a robust metric for evaluating binary classification performance, incorporating all four components of the confusion matrix: True Positives (TP), True Negatives (TN), FP, and False Negatives (FN). It is computed as:

$$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP+FP) \cdot (TP+FP) \cdot (TN+FP) \cdot (TN+FN)}} \quad (18)$$

The MCC ranges from -1 to +1, where +1 indicates perfect prediction, 0 represents random classification, and -1 indicates complete disagreement between predictions and actual outcomes. Unlike accuracy, MCC provides a balanced measure even for imbalanced datasets. In addition, the NPV quantifies the proportion of TN among all predicted negatives, reflecting the classifier's ability to correctly rule out non-TB cases. It is defined as:

$$NPV = \frac{TN}{TN+FN} \quad (19)$$

The value of NPV lies in the range 0-1 (or 0-100%), where a higher NPV indicates a more reliable model in excluding disease-positive cases. Additional analyses implemented in this study included Receiver Operating Characteristic (ROC) analysis, ablation analysis, and statistical tests. All methods were evaluated on the benchmark TB Chest X-ray Database [9].

Figures 5, 6, and 7 present a detailed comparison of all employed models' performance based on the established metrics and across the four different dataset partition ratios: 60/40, 70/30, 80/20, and 90/10.

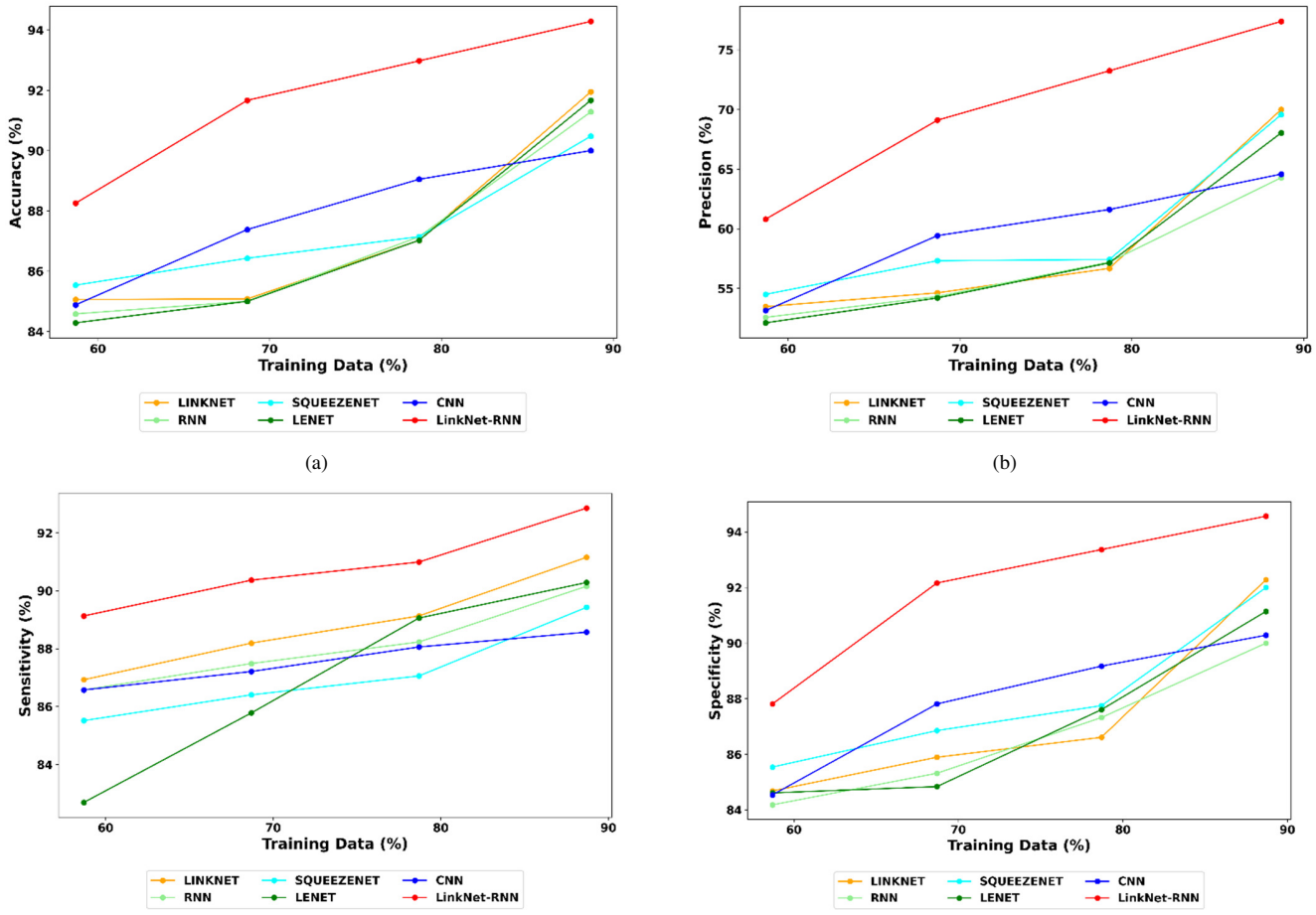
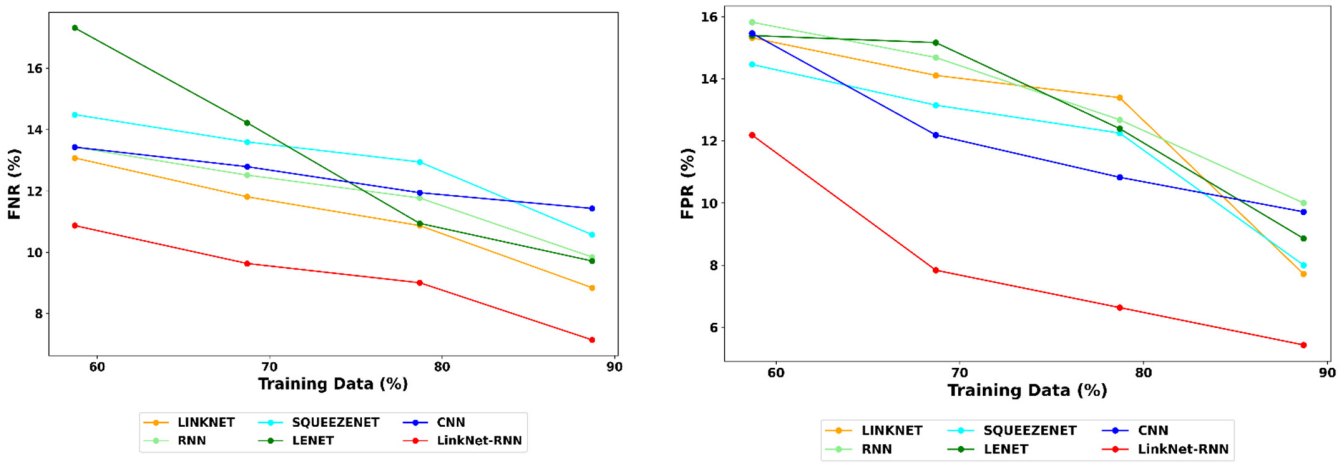
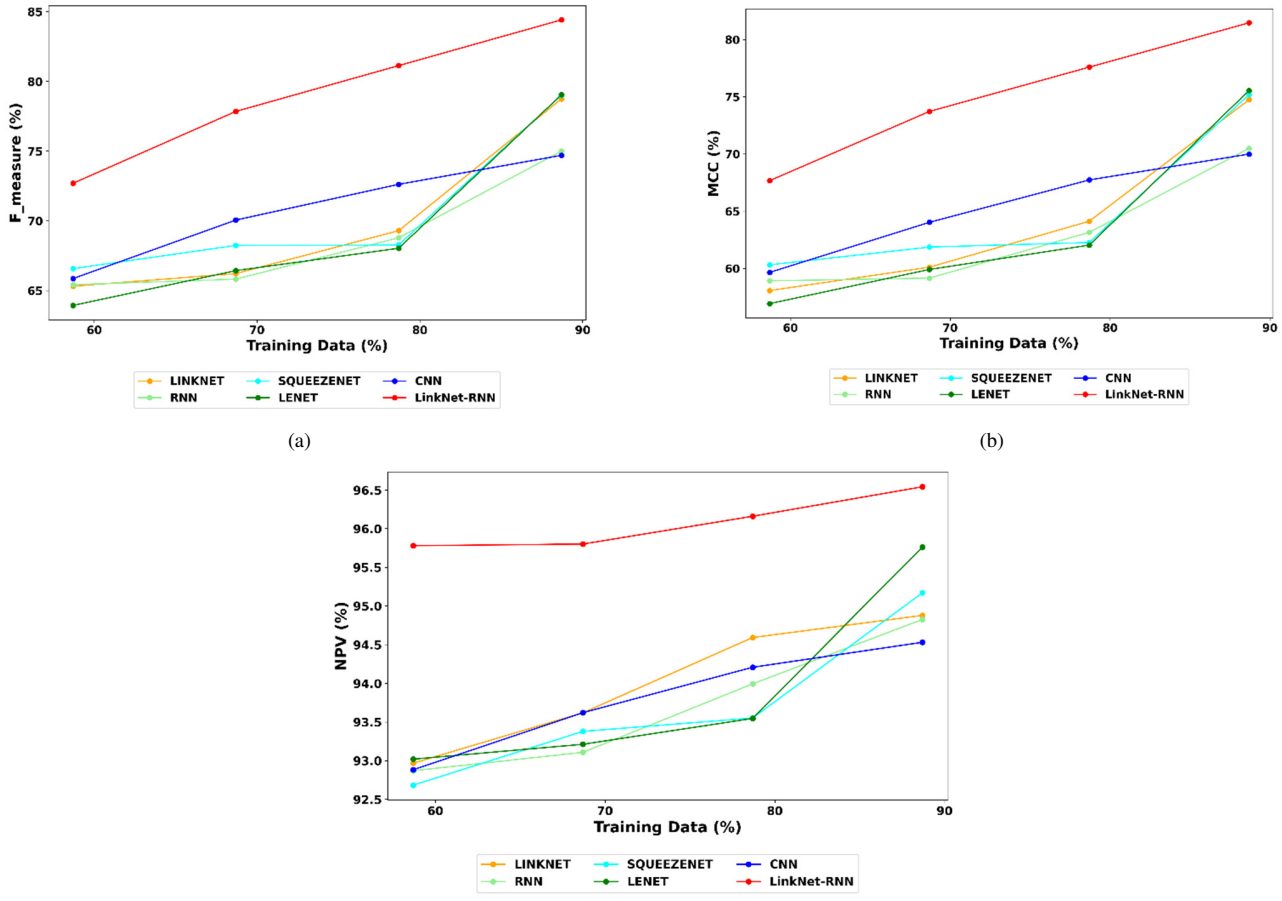


Fig. 5. Evaluation of positive metrics: Contrasting LinkNet-RNN with conventional techniques, (a) accuracy, (b) precision, (c) sensitivity, and (d) specificity.



(a) (b)

Fig. 6. Evaluation of negative metrics: Contrasting LinkNet-RNN with conventional techniques, (a) FNR, and (b) FPR.



(a) (b)

(c)

Fig. 7. Evaluation of other metrics: Contrasting LinkNet-RNN with conventional techniques, (a) F-Measure (F1-score), (b) MCC, and (c) NPV.

With the 60% training data, LinkNet-RNN achieved 88.26% accuracy, outperforming LinkNet (85.36%), RNN (84.98%), SqueezeNet (85.83%), LeNet (84.36%), and CNN (85.18%). At 70% training data, LinkNet-RNN's accuracy increased to 91.98%, while with 80% and 90% training data, LinkNet-RNN achieved accuracies of 93.73% and 94.14%,

respectively, surpassing the highest-performing traditional model, LinkNet, which reached 91.67% at the 90/10 ratio.

FPR comparisons (Figure 6) further highlight the advantages of LinkNet-RNN. At 80% training data, it achieved the lowest FPR (7.467), compared to LinkNet (14.165), RNN (13.854), SqueezeNet (13.365), LeNet (13.473), and CNN (11.859). At 90% training data, LinkNet-RNN reduced its FPR to 5.965, while LinkNet, the next best, achieved an 8.098.

For the F-Measure (or F1-score), the LinkNet-RNN model (Figure 7) again displayed the best performance. At 90% training data, it scored 84.978, outperforming LinkNet (76.532), RNN (73.731), SqueezeNet (76.494), LeNet (76.743), and CNN (73.156).

B. ROC Analysis

The ROC curve provides a graphical representation of each model's diagnostic ability by plotting the True Positive Rate (TPR) (sensitivity) against the FPR at various threshold settings. Figure 8 illustrates the ROC curve analysis, comparing the LinkNet-RNN strategy with the other established models. The LinkNet-RNN achieved the highest TPR of 0.985, while the TPR values for LinkNet, RNN, SqueezeNet, LeNet, and CNN ranged from 0.964 to 0.981.

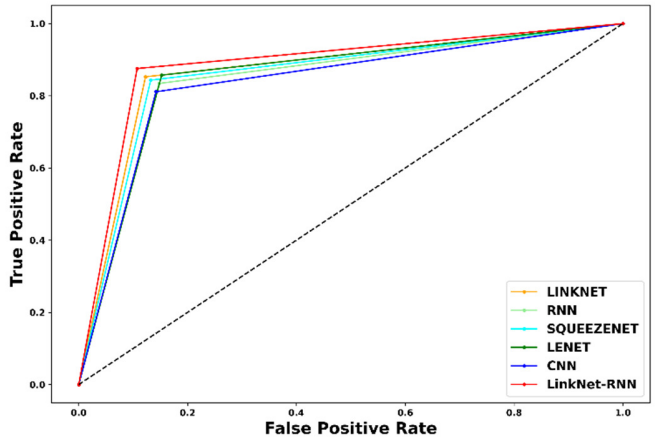


Fig. 8. ROC evaluation on LinkNet-RNN and conventional methods.

C. Statistical Analysis on Accuracy

Table I presents the statistical evaluation of model accuracy calculated across all different training/testing splits. The reported mean, minimum, maximum, standard deviation, and median values summarize the model performance over all iterations.

TABLE I. STATISTICAL EVALUATION OF ACCURACY

Methods	Mean	Minimum	Standard Deviation	Maximum	Median
LinkNet	0.873	0.851	0.028	0.919	0.861
RNN	0.870	0.846	0.027	0.913	0.861
SqueezeNet	0.874	0.855	0.019	0.905	0.868
LeNet	0.870	0.843	0.029	0.917	0.860
CNN	0.878	0.849	0.019	0.900	0.882
LinkNet-RNN	0.918	0.883	0.022	0.943	0.923

The LinkNet-RNN model achieved a mean accuracy of 0.918, which is notably higher than all other models. By comparison, LinkNet and LeNet both achieved 0.873 and 0.870, respectively; RNN also reached 0.870, SqueezeNet 0.874, and CNN 0.878.

D. Ablation Study on LinkNet-RNN

Table II presents an ablation analysis comparing the proposed LinkNet-RNN model with a conventional PHOG-based model. This evaluation isolates the impact of key components within the LinkNet-RNN pipeline, including median filtering-based preprocessing, I-PHOG feature extraction, and the hybrid LinkNet-RNN architecture. As shown, the LinkNet-RNN consistently outperforms the conventional PHOG model across all evaluated metrics. Notably, the LinkNet-RNN achieved higher accuracy (0.917 vs. 0.878), sensitivity (0.904 vs. 0.882), and NPV (0.958 vs. 0.947), along with lower FPR and FNR values, confirming that integrating median filtering-based preprocessing, I-PHOG feature extraction, and the hybrid LinkNet-RNN framework delivered substantial improvements in TB detection performance over conventional approaches.

TABLE II. ABLATION ANALYSIS ON LINKNET-RNN AND MODEL WITH CONVENTIONAL PHOG

Metrics	Model with Conventional PHOG	LinkNet-RNN
Accuracy	0.878	0.917
F-Measure	0.714	0.778
FPR	0.122	0.078
Sensitivity	0.882	0.904
FNR	0.118	0.096
Specificity	0.878	0.922
NPV	0.947	0.958
Precision	0.601	0.691
MCC	0.658	0.737

V. CONCLUSION

This study demonstrates that Tuberculosis (TB) detection from chest X-ray images can be significantly improved through advanced spatial feature extraction and a hybrid deep-learning framework. The proposed LinkNet-Recurrent Neural Network (LinkNet-RNN) model combines spatial and temporal analysis, achieving superior diagnostic accuracy and robustness compared to conventional models, including LinkNet, RNN, SqueezeNet, LeNet, and Convolutional Neural Networks (CNN). By incorporating hierarchical skeleton and statistical descriptors to enhance the Improved Pyramid Histogram of Oriented Gradients (I-PHOG) features, the model captures more faithful representations of lung abnormalities. The integration of an RNN within the spatially enriched LinkNet architecture enables the system to model sequential dependencies and temporal dynamics in image-derived features, capabilities that outperform either architecture individually.

The experimental evaluation, conducted across varying training data proportions (60/40, 70/30, 80/20, 90/10), consistently confirmed the superiority of the LinkNet-RNN approach. At 80% training data, the model achieved the lowest FPR of 7.467, outperforming LinkNet (14.165), RNN (13.854),

SqueezeNet (13.365), LeNet (13.473), and CNN (11.859). This demonstrates that the proposed model not only enhances classification accuracy but also effectively minimizes false diagnoses, a critical factor in clinical applications. Overall, the proposed LinkNet-RNN architecture presents a promising tool for real-world TB screening, offering early, accurate, and reliable diagnosis.

REFERENCES

[1] E. Kotei and R. Thirunavukarasu, "Tuberculosis Detection From Chest X-Ray Image Modalities Based on Transformer and Convolutional Neural Network," *IEEE Access*, vol. 12, pp. 97417–97427, 2024, https://doi.org/10.1109/ACCESS.2024.3428446.

[2] T. Rahman *et al.*, "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020, https://doi.org/10.1109/ACCESS.2020.3031384.

[3] A. F. M. Saif, T. Imtiaz, C. Shahnaz, W.-P. Zhu, and M. O. Ahmad, "Exploiting Cascaded Ensemble of Features for the Detection of Tuberculosis Using Chest Radiographs," *IEEE Access*, vol. 9, pp. 112388–112399, 2021, https://doi.org/10.1109/ACCESS.2021.3102077.

[4] T. Xu and Z. Yuan, "Convolution Neural Network With Coordinate Attention for the Automatic Detection of Pulmonary Tuberculosis Images on Chest X-Rays," *IEEE Access*, vol. 10, pp. 86710–86717, 2022, https://doi.org/10.1109/ACCESS.2022.3199419.

[5] C.-F. Chen *et al.*, "A deep learning-based algorithm for pulmonary tuberculosis detection in chest radiography," *Scientific Reports*, vol. 14, no. 1, Jun. 2024, Art. no. 14917, https://doi.org/10.1038/s41598-024-65703-z.

[6] S. Urooj, S. Suchitra, L. Krishnasamy, N. Sharma, and N. Pathak, "Stochastic Learning-Based Artificial Neural Network Model for an Automatic Tuberculosis Detection System Using Chest X-Ray Images," *IEEE Access*, vol. 10, pp. 103632–103643, 2022, https://doi.org/10.1109/ACCESS.2022.3208882.

[7] T. Rahman *et al.*, "TB-CXRNet: Tuberculosis and Drug-Resistant Tuberculosis Detection Technique Using Chest X-ray Images," *Cognitive Computation*, vol. 16, no. 3, pp. 1393–1412, May 2024, https://doi.org/10.1007/s12559-024-10259-3.

[8] V. T. Q. Huy and C.-M. Lin, "An Improved Densenet Deep Neural Network Model for Tuberculosis Detection Using Chest X-Ray Images," *IEEE Access*, vol. 11, pp. 42839–42849, 2023, https://doi.org/10.1109/ACCESS.2023.3270774.

[9] *Tuberculosis (TB) Chest X-ray Database.* (2020), T. Rahman *et al.*. [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset.

[10] A. Nithya, A. Appathurai, N. Venkatadri, D. R. Ramji, and C. Anna Palagan, "Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images," *Measurement*, vol. 149, Jan. 2020, Art. no. 106952, https://doi.org/10.1016/j.measurement.2019.106952.

[11] A. Saidani and A. K. Echi, "Pyramid histogram of oriented gradient for machine-printed/handwritten and Arabic/Latin word discrimination," in *2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, Tunis, Tunisia, Aug. 2014, pp. 267–272, https://doi.org/10.1109/SOCPAR.2014.7008017.

[12] F. Bimantoro, A. A. Aziz, A. Y. Husodo, A. Musnansyah, A. E. Minarno, and A. Kurniawardhani, "Image Retrieval using Modified Multi Texton and Rotation Invariant Local Binary Pattern," in *2020 International Conference on Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, Lombok, Indonesia, Oct. 2020, pp. 1–5, https://doi.org/10.1109/ICADEIS49811.2020.9277462.

[13] V. Ayzenberg and S. F. Lourenco, "Skeletal descriptions of shape provide unique perceptual information for object recognition," *Scientific Reports*, vol. 9, no. 1, Art. no. 9359, Jun. 2019, https://doi.org/10.1038/s41598-019-45268-y.

[14] B. Krishnakumar and K. Kousalya, "Optimal Trained Deep Learning Model for Breast Cancer Segmentation and Classification," *Information Technology and Control*, vol. 52, no. 4, pp. 915–934, Dec. 2023, https://doi.org/10.5755/j01.itc.52.4.34232.

[15] L. Zhang and X. Liang, "Image Segmentation of Plant Leaves in Natural Environments Based on LinkNet," *Journal of Computing and Electronic Information Management*, vol. 11, no. 3, pp. 67–72, Dec. 2023, https://doi.org/10.54097/jceim.v11i3.15.

[16] M. Islam, Z. Islam, A. Asraf, M. S. Al-Rakhami, W. Ding, and A. H. Sodhro, "Diagnosis of COVID-19 from X-rays using combined CNN-RNN architecture with transfer learning," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 2, no. 4, Oct. 2022, Art. no. 100088, https://doi.org/10.1016/j.tbench.2023.100088.