International Journal of Computer Techniques – Volume 12 Issue 3, May - June - 2025

Tuberculosis Detection using Machine Learning

Abhishek Kumar
Dept. of Information Technology
Noida Institute of Engineering and Technology (AKTU)
Uttar Pradesh, India

Roshan Gupta
Dept. of Information Technology
Noida Institute of Engineering and Technology (AKTU)
Uttar Pradesh, India

Soni Kumari
Dept. of Information Technology
Noida Institute of Engineering and Technology (AKTU)
Uttar Pradesh, India

Aman Kumar
Dept. of Information Technology
Noida Institute of Engineering and Technology (AKTU)
Uttar Pradesh, India

ABSTRACT

Tuberculosis (TB) is a highly infectious disease that poses a significant threat to global health, particularly in regions with limited medical resources. Despite advancements in medical science, TB remains a leading cause of mortality worldwide. Early diagnosis and timely treatment are critical in reducing the disease's impact on patients and curbing its spread. Chest X-rays are among the most commonly used diagnostic tools for TB due to their rapid results and affordability. However, interpreting chest X-ray images requires expertise and can be time-consuming, especially given the complexity of identifying subtle abnormalities associated with TB. These challenges highlight the need for innovative solutions to enhance diagnostic accuracy and efficiency.

This study introduces a computer-aided detection system based on the EfficientDet network, enhanced with a fused attention mechanism. The EfficientDet model is known for its strong performance in object detection tasks, and the addition of the fused attention mechanism allows the network to focus on critical features within chest X-ray images. This approach reduces false positives and improves the model's ability to identify TB-related patterns accurately. Extensive testing on the TBX11K dataset demonstrated the system's effectiveness, achieving high accuracy in detecting TB cases. The proposed model's ability to handle large datasets and analyze complex medical images makes it a valuable tool for clinical applications. By integrating such advanced AI solutions, healthcare systems can significantly improve TB diagnosis, reduce the workload of medical professionals, and enable timely treatment, particularly in resource-limited settings. The study highlights the transformative potential of AI in addressing public health challenges and advancing diagnostic technology.

CCS CONCEPTS

• computer vision; • Deep learning; • Image classification;

KEYWORDS

Tuberculosis, deep learning, EfficientDet, attention mechanism

ACM Reference Format:

Xuebin Xu, Guanqi Gong, and Xiangdong Gao. 2023. Lung Tuberculosis Detection Based on Attention EfficientDet. In 2023 6th International Conference on Artificial Intelligence and Pattern Recognition (AIPR) (AIPR 2023), September 22–24, 2023, Xiamen, China. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3641584.3641655

1INTRODUCTION

The lungs play a crucial role in the exchange of gases between the body and the environment. However, their constant exposure to airborne microorganisms and dust particles makes them particularly vulnerable to infections. Tuberculosis (TB) is a chronic, highly infectious lung disease that continues to pose a significant global health challenge. According to the World Health Organization, there were approximately 9.96 million new TB cases and 1.21 million related deaths worldwide in 2019. The disease's vague symptoms and the rise of drug-resistant strains have made diagnosis more difficult and contributed to its increasing prevalence.

Recent advancements in medical diagnostics have explored the use of computer-aided systems to interpret chest X-rays, leveraging neural networks and artificial intelligence to automatically detect diseases. These systems rely on learning key features from data, enabling them to recognize patterns associated with TB. For instance, BV Ginneken and colleagues developed a fully automated system to identify abnormalities in chest X-rays by scoring abnormalities across the entire image, achieving notable results on private datasets. Similarly, Hogeweg and others proposed combining pixel-level texture analysis with additional diagnostic techniques for TB detection. Tan and his team introduced a method that segments lung regions to analyze pixel data and identify abnormalities. While these initial studies significantly advanced computer-aided diagnosis (CAD) systems, early methods often failed to meet practical clinical needs.

One major limitation of current TB detection systems lies in their reliance on datasets that lack comprehensive representation of the diverse stages and forms of TB visible in chest X-rays. This reduces the accuracy of these systems, making them less reliable for clinical use. To address these challenges, this study focuses on convolutional neural networks (CNNs) to improve TB detection. By enhancing the ability of CNNs to accurately identify TB lesions and account for variations in imaging data, the study aims to provide faster and more reliable diagnostic tools. These advancements hold the potential to support healthcare professionals in making timely and precise diagnoses, ultimately improving patient outcomes.

2RELATED WORK

2.1 The Dataset

This study utilizes the TBX11K dataset, a comprehensive tuberculosis chest X-ray dataset made publicly available by the Media Computing Laboratory of Nankai University. Developed by Liu Yun and colleagues, the TBX11K dataset contains a total of 11,200 X-ray images. Among these, 5,000 images depict healthy individuals, 5,000 represent cases with diseases other than tuberculosis, and 1,200 showcase signs of tuberculosis.

For cases with tuberculosis, the dataset provides detailed annotations, including bounding boxes that highlight the affected areas and classifications of the type of tuberculosis present—either active or old tuberculosis. These annotations enhance the dataset's utility for research and model training. Figure 1 illustrates examples of X-ray images for each category included in the TBX11K dataset. This rich and well-annotated dataset serves as a valuable resource for developing and evaluating tuberculosis detection models.

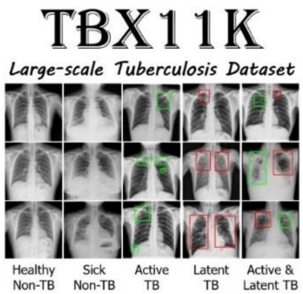

Figure 1: Example of TBX11K dataset

2.2 Data Preprocessing

DICOM is a standardized file format commonly used for medical images, where pixel data is represented in grayscale intensity values. Unlike typical image formats, DICOM files also store sensitive patient information, such as name, gender, and age, which is protected health information (PHI). Viewing these files requires specialized software, and they are not well-suited for direct use in machine learning applications. To make them compatible, DICOM images must be converted to formats like JPG, which are easier to process.

In their original DICOM format, the pixel data is stored as 16-bit integers (int16), while JPG files require 8-bit unsigned integers (uint8). To preserve as much grayscale detail as possible during the conversion, a specific processing method, illustrated in Figure 2, is used.

Additionally, this study employs histogram equalization to enhance the uniformity of the gray-level distribution in the chest X-ray images. This technique redistributes the pixel intensities, enabling the neural network to extract features beyond basic grayscale information. The impact of this process on the pixel intensity distribution is demonstrated in Figure 3, showing the differences in the chest X-ray images before and after applying histogram equalization. This step ensures that the images are optimized for machine learning tasks.

3THE PROPOSED METHOD

3.1 EfficientNet Network

In May 2019, Google introduced EfficientNet, a new deep learning architecture designed to scale convolutional neural networks (CNNs) in a more structured and effective manner. By using a simple yet efficient composite coefficient, EfficientNet achieves significant improvements in both accuracy and efficiency compared to earlier CNN models.

On the ImageNet dataset, the EfficientNet-B7 model delivers exceptional performance, achieving a top-1 accuracy of 84.4% and a top-5 accuracy of 97.1%. Remarkably, it accomplishes this while reducing the number of parameters by 8.4 times and increasing computational speed by 6.1 times compared to previous state-of-the-art convolutional networks.

EfficientNet also demonstrates strong adaptability to other datasets, achieving impressive results with fewer parameters. For example, it achieves 91.7% accuracy on CIFAR-100 and 98.8% accuracy on the Flowers dataset.

The architecture of EfficientNet is organized into nine stages, with each stage designed to perform specific tasks within the network. The first stage begins with... (further details about the first stage would be added based on additional context).

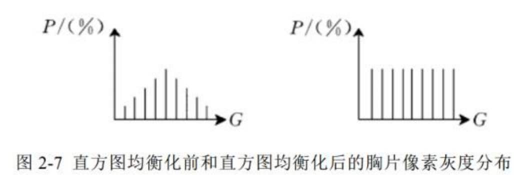

Figure 3: Grayscale distribution of chest film pixels before and after histogram equalization

EfficientNet's architecture is built around the MBConv structure, which is a key component in its design. The MBConv block originates from the Inverted Residual block introduced in the MobileNetV3 network, as described in prior research. EfficientNet enhances this structure by incorporating the Swish activation function and adding a Squeeze-and-Excitation (SE) module to each MBConv block.

The network's final stage includes a 1×1 convolutional layer, an average pooling layer, and a fully connected layer. Within each MBConv block, the first 1×1 convolutional layer is expanded by a factor of nnn, referred to as the multiplicity factor. Additionally, the Depthwise Convolution within the block uses a k3×3k3 \times 3k3×3 or k5×5k5 \times 5k5×5 convolutional kernel, depending on the configuration.

As illustrated in Figure 4, the MBConv structure primarily comprises a 1×1 convolutional layer (with batch normalization and Swish activation), followed by a k×kk \times kk×k Depthwise Convolution (also with batch normalization and Swish activation). The architecture of EfficientNet-B0, the base model, incorporates 3×3 and 5×5 Depthwise Convolutions, SE modules, a standard 1×1 convolutional layer, and a Dropout layer. This combination of elements ensures the model's high efficiency and accuracy while maintaining a relatively lightweight design.

3.2 EfficientDet that integrates attention mechanism

The EfficientDet-D0 algorithm is structured into three main components to enable efficient and accurate detection.

1. **Backbone Feature Extraction Network:** The first component uses EfficientNet-B0 as its backbone for extracting key features from input images. This network processes the input data and outputs features from layers 3 to 7, which serve as the foundation for subsequent analysis.
2. **Bi-Directional Feature Fusion Network:** The second component performs feature fusion by combining features through multiple top-down and bottom-up processes. This fusion integrates information from different layers of the backbone network, enabling the model to capture both high-level and fine-grained details necessary for accurate detection and classification.
3. **Classification Prediction Network:** The final component focuses on target regression and classification, predicting the location and type of objects or lesions in the input images.

This study improves both the backbone feature extraction network and the feature fusion network to enhance the speed and precision of tuberculosis lesion detection and classification. The overall structure of the network is illustrated in Figure 5.

The implementation of the spatial attention module, depicted in Figure 6, begins with processing the input feature map FFF through average pooling and max pooling. These operations generate two feature maps, which are then combined along the channel dimension. A convolution operation reduces the combined map to a single channel while maintaining its spatial dimensions. Finally, a sigmoid function is applied to produce the spatial attention feature MiM_iMi , which is subsequently input into the depthwise convolution operation within the MBConv Block. This mechanism helps the network focus on spatially important regions, improving its detection performance.

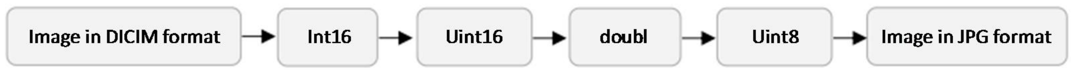

Figure 2: DICOM format image to IPG format image conversion process

Lung Tuberculosis Detection Based on Attention EfficientDet AIPR 2023, September 22–24, 2023, Xiamen, China

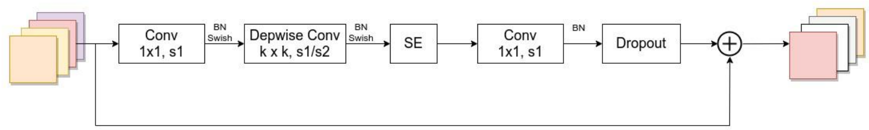

Figure 4: MBConv specific structure

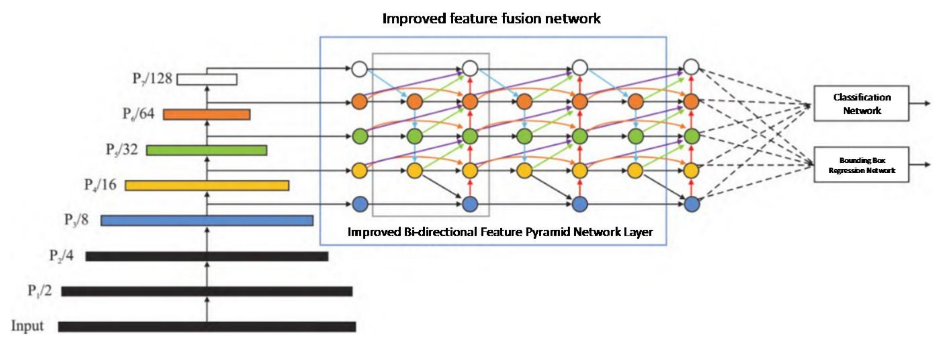

Figure 5: EfficientDet Network Structure

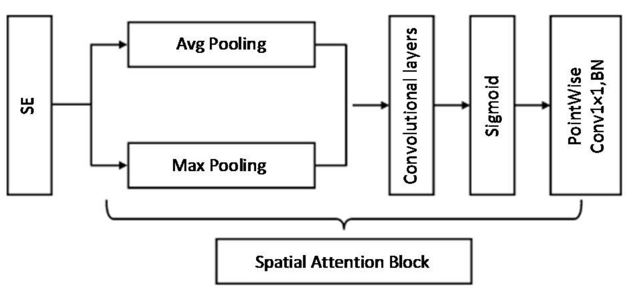

Figure 6: Spatial Attention Module

3.3 Attention EfficientDet-based TB detection network

This chapter presents experiments designed to validate and analyze the proposed attention-based tuberculosis (TB) classification network for chest radiographs. The experimental workflow is illustrated in Figure 7 and follows these steps:

1. **Data Preprocessing:** The chest X-ray dataset is divided into training and testing subsets after undergoing preprocessing.
2. **Data Augmentation:** The training data is enhanced through data augmentation techniques to generate additional training samples.
3. **Model Training:** The augmented training data is used to train the TB classification network for chest X-rays.

4. **Model Testing:** The trained network is then evaluated using the test data to obtain classification results, completing the classification process for chest X-ray images.

To enhance the feature fusion process, this study builds upon pyramid networks (FPNs), which traditionally have a one-way information flow. By contrast, bi-directional feature fusion networks (BiFPNs) enable information to flow both top-down and bottom-up. Using weighted feature fusion, BiFPNs effectively aggregate features across various resolution scales. The output of one BiFPN can be used as the input for the next, allowing for improved feature integration. The paper introduces further improvements to the feature fusion network, as shown in Figure 8.

The top-down feature fusion process in this study involves:

1. **Upsampling and Gaussian Convolution:** The output from the previous layer is upsampled and convolved using a Gaussian filter.
2. **Feature Difference Calculation:** The result of the upsampling operation is subtracted from the current layer's output to generate a detailed feature map. P6det=P6in−Upsample(P7out)⊗g5×5P6_{det} = P6_{in} - \text{Upsample}(P7_{out}) \otimes g_{5 \times 5}P6det =P6in −Upsample(P7out) ⊗g5×5
3. **Detail-Enhanced Features:** The detailed feature map is added back to the current layer's output to enhance feature detail.

Instead of merely upsampling the previous layer and fusing it with the current layer, this method calculates the feature difference using Gaussian convolution and then integrates it into the fusion process. This approach, represented by the blue connecting lines in Figure 8, allows for more precise feature enhancement and improves the overall effectiveness of the TB classification network.

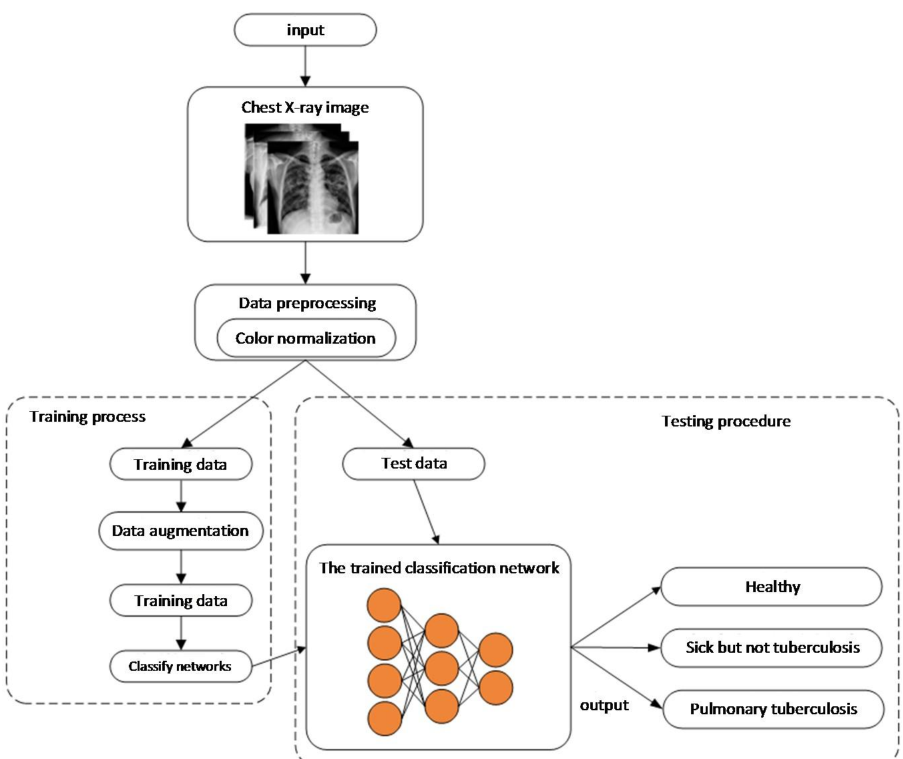

Figure 7: Experimental flowchart of the attention-based mechanism for classification network of pulmonary tuberculosis on chest radiographs

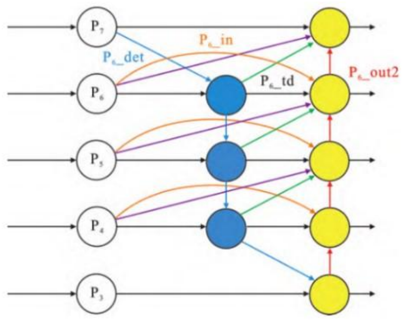

图 4 改进后的 BiFPN 图
Fig. 4 Improved BiFPN diagram

P_i^{out} is the output with added detail features, i∈(3-7).

In this paper, the output PitdP_{itd}Pitd represents the intermediate output after applying a weighted attention mechanism to the feature map PioutP_{iout}Piout . This mechanism assigns different weights to different parts of the feature map, allowing the model to focus more on the important regions. The formula for this mechanism is shown in Equation 3, where the solid blue circle in Figure 8 illustrates this process for the output from the sixth layer, though the process is similar for outputs from other layers.

The weight values ω1\omega_1ω1 and ω2\omega_2ω2 correspond to the importance given to specific features in the map. To avoid division by zero, a small constant ϵ\epsilonϵ is added to the denominator. In this study, the value of ϵ\epsilonϵ is set to 0.001. The "Upsample" operation refers to the process of increasing the resolution of the feature map, allowing finer details to be captured in subsequent layers of the network.

Figure 8: The improved BiFPN diagram

layer, as shown in equation 2).

$$P_7^{out} = Conv\, P_7^{in}$$

$$P_6^{out} = Conv\, P_6^{in} + P_6^{in} - Upsample\, P_7^{out} \otimes g_{5\times5}$$

$$P_3^{out} = Conv\, P_3^{in} + P_3^{in} - Upsample\, P_4^{out} \otimes g_{5\times5} \quad (2)$$

4EXPERIMENTAL ANALYSIS

4.1 Evaluation index

This chapter focuses on the localization and classification of pulmonary tuberculosis lesions on chest X-ray images. Therefore, the evaluation metrics used are consistent with the COCO (Common Objects in Context) dataset. The COCO evaluation metric does not distinguish between AP (Average Precision) and mAP (mean Average Precision), and AP is actually mAP. The main indicators used in this article are AP and AP50. AP is the average value at multiple IoU thresholds, specifically taking values at intervals of

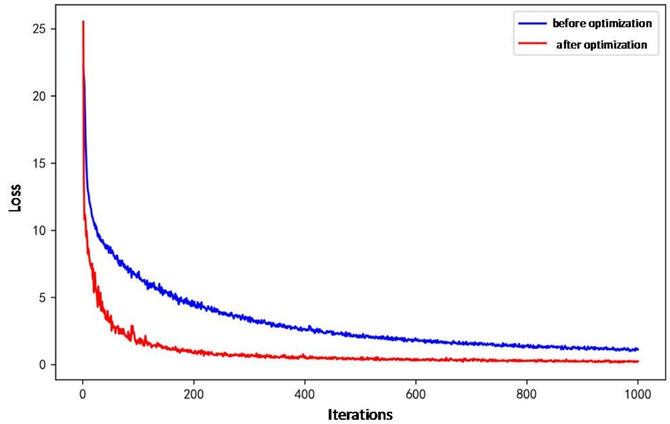

Figure 9: A comparison chart of the changes in the loss values of the network model during the training process before and after the loss function optimization.

0.05 between IoU thresholds of 0.50 to 0.95 and then calculating the average accuracy. AP50 refers to the average accuracy when the IoU threshold is 0.5. AP is essentially the area under the PR curve. The PR curve is a curve drawn with recall rate as the horizontal axis and precision rate as the vertical axis. The calculation methods of recall rate and precision rate are formulas (4) and (5), and AP is calculated by formula (6).

$$Precision = \frac{TP}{+} \qquad (4)$$

$$Recall = \frac{1}{+} \qquad (5)$$

$$AP = \int P(R)dr \qquad (6)$$

$$m \; i=10$$

Where m is the number of data categories, R is the recall rate, and P(R) represents the precision rate when the recall rate is R.

4.2 Analysis of experimental results

The data set used in this chapter's experiment is the chest X-ray images with tuberculosis manifestations in the TBX11K data set. There are a total of 1190 chest X-ray images, including 924 active tuberculosis case images, 212 latent tuberculosis case images, and 54 case images with both active and latent tuberculosis lesions.

Since the original network uses a weighted sum of L1 loss and generalized IoU loss, but when the error value between the predicted value and the true label is small, it may cause gradient disappearance, which may slow down the convergence speed of the network model during training or even stop training, affecting the final performance of the network model. Therefore, this article optimizes it by replacing L1 loss with smooth L1 loss, solving the gradient disappearance problem of L1 loss, and will not cause gradient explosion problems during network model training like L2 loss function. Figure 9 is a comparison chart of the changes in the loss values of the network model during the training process before and after the loss function optimization. Blue is before the loss function optimization and red is after optimization. As can be seen from Figure 9, when using L1 loss function, the network Table 1: The experimental comparison results of the algorithm in this chapter with other mainstream object detection algorithms.

Detection Method	AP(%)	AP50(%)
SSD	28.6	69.2
Faster R-CNN	11.4	37.1
RetinaNet	13.3	40.1
FCOS	13.9	41.7

the algorithm in this article	47.95	73.90

model converges slowly during training and the final loss value obtained is also larger than the loss value after optimization. Therefore, through the optimization of the loss function in this article, it also brings improvement to the final detection performance of the network model. Finally, through experimental simulation and comparison with existing mainstream object detection algorithms, it verifies the superiority of the tuberculosis detection network based on attention EfficientDet proposed in this chapter.The results are shown in Table 1.

5 CONCLUSION

A new method for detecting tuberculosis in chest radiographs based on the attention EfficientDet network is proposed. The method addresses the current problem of low accuracy of TB detection and designs a method to avoid the influence of other lung diseases on the judgment of the detection network by first identifying TB samples and then detecting the lesion regions in TB samples. The method learns the global feature information of the feature map by global average pooling, retains more texture features of the image by local maximum pooling, retains more background information of the image by local average pooling, and enables the model to extract more critical and important information by fusing the attention mechanism, so that the model can make more accurate judgments. The experimental results of the TBX11K dataset show that The AP of the attention-based EfficientDet network method for detecting tuberculosis in chest radiographs reaches 48.35%, and the AP50 reaches 74.20%, which not only meets the requirements of practical applications but also improves the validation accuracy of experimental results.

ACKNOWLEDGMENTS

Thank the teachers for their hard guidance and students for their enthusiastic help.

REFERENCES

[1] Carabalí-Isajar ML et al. (2023) Clinical manifestations and immune response to tuberculosis. World J Microbiol Biotechnol. 39(8)

[2] World Health Organization. Global tuberculosis report 2024. Geneva, Switzerland: World Health Organization, 2024.

[3] Larson, D. B., Harvey, H., Rubin, D. L., et al. (2021). Regulatory frameworks for development and evaluation of artificial intelligence–based diagnostic imaging algorithms: Summary and recommendations. Journal of the American College of Radiology, 18, 413–424.

[4] World Health Organization. Module 2: Screening. WHO operational handbook on tuberculosis: Systematic screening for tuberculosis disease. Geneva: World Health Organization; 2021.

[5] Chang, R. I., Chiu, Y. H., & Lin, J. W. (2023). Two-stage classification of tuberculosis culture diagnosis using convolutional neural network with transfer learning. The Journal of Supercomputing, 76, 8641–8656.

[6] He W, Liu C, Liu D, et al. Prevalence of *Mycobacterium tuberculosis* resistant to bedaquiline and delamanid in China. *J Glob Antimicrob Resist*. 2021;26:290–293.

[7] handra, T. B., et al. (2020). Automatic detection of tuberculosis-related abnormalities in chest X-ray images using hierarchical feature extraction scheme. Expert Systems with Applications, 158, 113514.

[8] Nambiar R, Tornheim JA, Diricks M, et al. Linezolid resistance in *Mycobacterium tuberculosis* isolates at a tertiary care centre in Mumbai, India. *Indian J Med Res*. 2021;154:85–89

[9] Centers for Disease Control and Prevention. (n.d.). *Basic TB facts*.

https://www.cdc.gov/tb/topic/basics/default.htm (accessed December 18, 2023).

[10] Villalobos, N. (n.d.). *Tuberculosis (TB): Symptoms, treatment, diagnosis, and more*. Medical News Today. https://www.medicalnewstoday.com/articles/8856 (accessed December 17, 2023).

[11] Goletti D, Delogu G, Matteelli A, Migliori GB. The role of IGRA in the diagnosis of tuberculosis infection, differentiating from active tuberculosis, and decision making for initiating treatment or preventive therapy of tuberculosis infection. *Expert Rev Anti Infect Ther*. [Year];[Volume(Issue)]:[Pages].

[12] Annotations of lung abnormalities in the Shenzhen chest X-ray dataset for computer-aided screening of pulmonary diseases. *Data*. 2022;7(7):95.

[13] Rahman KK M, Subashini M. A deep neural network-based model for screening autism Spectrum Disorder using the Quantitative Checklist for Autism in Toddlers (QCHAT). *[Journal Name]*. [Year];[Volume(Issue)]:[Pages].

[14] Nyuytiymbiy, K. (2021, April 5). Parameters and hyperparameters in machine learning and deep learning. *Medium*. https://towardsdatascience.com/parameters-andhyperparameters-aa609601a9ac (accessed January 3, 2024).

[15] Junaid MA, Anwar S, Sikander G, Khan MT. Generative adversarial network based chest disease detection and binary mask generation. In: *2023 International Conference on Robotics and Automation in Industry (ICRAI)*; 2023; Peshawar, Pakistan