University of Ibadan
Journal of Science and Logics in ICT Research

Efficient Tuberculosis Detection Using Chest X-ray Images with Deep Learning Algorithms

[1]Obakunle Olusesan A., [2]Abiola Oladimeji A. and [3]Akinola Solomon O.
University of Ibadan, Ibadan, Nigeria
[1]oobakunle2383@stu.ui.edu.ng [2]oladimejiarowolo@yahoo.co.uk [3]so.akinola@ui.edu.ng

Abstract

Tuberculosis is a threat to the existence of the human race due to its substantial mortality rate and it has become a significant public health concern, which if detected at an early stage could reduce the death rate globally. Harnessing the potential of machine learning to combat the low detection rate of tuberculosis detection by traditional methods and promote a faster and more accurate diagnosis of the disease. An online Dataset comprising 11,200 Chest X-ray (CXR) images of different categories of patients that are healthy, Sick but not infected, and those infected with Tuberculosis with their corresponding bounding box annotations were used for this research, and feature engineering was carried out on the dataset for effective data cleaning, Image resizing, normalization, and augmentation to increase the quality of data during data segmentation. The dataset was divided into training, validation, and testing sets using the RestNet50 model which demonstrated a good performance in classification, achieving impressive precision, and recall of 98%, and 96% respectively, and YOLOv8 was also used with 68% precision, 65% recall, and 68% mean average precision respectively which showed that the model needs improvement to further accurately detect regions infected with tuberculosis.

Keywords: Tuberculosis Detection, Chest X-ray, YOLOv8 and RestNet50

1. Introduction

Tuberculosis (TB) is one of the leading 10 causes of mortality globally, and the main cause from one infectious agent, placing it above HIV/AIDS. The World Health Organization (W.H.O) announced an expected 10.6 million new infections and 1.4 million fatalities, owing to TB in 2021. TB is a communicable illness triggered by Mycobacterium tuberculosis, that is transmitted via the air by coughing and sneezing from an infected individual, the germs are ejected from the body, when breathed in by an uninfected individual, even in minor amounts, these bacteria may be transferred and lead to TB [1].

Symptoms of TB include weariness, prolonged coughing, high fever, and loss of weight are designated active TB (ATB) sufferers. Conversely, TB patients without symptoms are termed latent TB (LTB) individuals who cannot transfer the illness to other individuals but have a significant chance of contracting TB if they fail to live a healthy way of life. The illness normally infects the lungs, also known as pulmonary TB, but when it impacts other sections of the body including lymph nodes, the brain, and bones, it causes extrapulmonary tuberculosis (EPT), [2].

W.H.O.'s 2022 reported that most persons who had TB were found in undeveloped areas, Southeast Asia accounted for 45% of the cases, Africa, 23%, and the western Pacific region accounted for 18%. Europe, the Americas, and the Eastern Mediterranean accounted for 2.2%, 2.9%, and 8.1% respectively. Several methods of detecting TB infections have been explored over time and the World Health Organization endorses the practice of both sputum smear microscopy and swift molecular examinations like the Xpert MTB/RIF assay, which can spot Tuberculosis and resistance to medication within two hours [3].

Obakunle Olusesan A., Abiola Oladimeji A. and Akinola Solomon O. (2024). Efficient Tuberculosis Detection Using Chest X-ray Images with Deep Learning Algorithms, *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR)*, Vol. 12 No. 2, pp. 40 – 48.

2. Artificial Intelligence in TB Diagnosis Detection

Numerous studies have utilized AI for Tuberculosis diagnostics, confronting the difficulty of detecting and assessing disease. Hwang, *et. al.*, [4] reported that accurate detection of active pulmonary tuberculosis from chest radiographs (CRs) plays a vital role in the timely diagnosis and screening of this infectious disease. An automated system capable of analyzing CRs can potentially streamline the tuberculosis screening workflow and enhance diagnostic accuracy. By leveraging advanced techniques like deep learning and neural networks, such a system could assist radiologists and healthcare professionals in identifying the characteristic patterns associated with active tuberculosis lesions on chest X-rays. Artificial neural networks (ANN) began to alter TB diagnosis in the early 1990s owing to algorithms for identifying patterns. The model demonstrated excellent performance and outcomes, showing that ANN has significant promise in computer-aided diagnosis of lung illnesses [5].

The primary problem encountered by all these models was diagnostic characteristics to conform with precise prognosis. In addition, more effective research was carried out to support AI growth, resulting in deep learning (DL). In early 2010, deep learning proved to be a beneficial idea when combining neural networks with algorithms (genetic algorithm and fuzzy logic) and sole hidden layer feed-forward neural networks became available for TB diagnosis. The power of certain deep learning algorithms to forecast and assess complicated and varied data has brought a fresh ray of optimism about overcoming Tuberculosis-related issues, like resistance to medication.

There are various machine learning techniques with a varied diversity of methodologies employed in TB diagnosis. Many machine learning approaches utilized for the diagnosis of Tuberculosis include supervised learning, unsupervised learning, semi-supervised learning, and transfer learning. The selection of the ML approach relies on the type of demand, and each learning technique has its strengths and limits. Supervised learning is a procedure where the process of acquiring knowledge incorporates a collection of input data (X) and output data (Y), and the learning phase seeks to look for a trend in the provided data that would be associated with the wanted output data [6]. Computer-Aided Design system suitable for the diagnosis of Tuberculosis integrates complex characteristics with hand-crafted features. Ayaz, *et. al.*, [7] employed supervised learning and pre-trained CNN architectures to identify TB in chest X-ray pictures which decreased the time required for preprocessing and diagnostic efficiency, resulting in early detection of the disease.

The machine learning model that had been trained exhibited high precision and could discover key characteristics for the diagnosis of Tuberculosis that radiotherapists misrepresented. This approach might also considerably lower the demand for vast, ordered data even when the database labels are lacking or unreliable [8]. A specific kind of semi-supervised learning is active learning (AL), in which the model may fight the issue of insufficient labeled training data in different ways [9].

Hijazi *et al.*, [10] performed Tuberculosis screening on Chest X-rays (CXRs) and canny edge detected images using two convolutional neural networks (VGG 16 and InceptionV3) using an object-level annotated dataset of CXRs which allowed them to determine the area of pulmonary problems suggestive of TB. From their work, the detection rate was low and it consumed a lot of processing time to get results. Due to these limitations, the evaluation outcomes of the model like accuracy, sensitivity, and specificity needed improvement.

The focus of this paper is to optimize the TB diagnostics process through the utilization of deep learning techniques to the problem of automated screening of CXRs for TB. The goal is to create a highly efficient tool that will assist doctors and nurses in getting a diagnosis and commencing therapy, if required, quickly.

*2.1 **YOLO Algorithm***

The real-time object detection capabilities of YOLO (You Only Look Once) have proven to be highly valuable in autonomous vehicle systems. These capabilities allow for the rapid identification and tracking of a wide range of objects, including automobiles and people Tuberculosis model detection is in Figure 1, [11, 12]. Action recognition is one of the various domains where these capabilities have been

applied in video sequences for surveillance [13, 14], sports analysis [15] and human-computer interaction [16].

Within the realm of medicine, YOLO has been applied in the diagnosis of cancer [17], skin segmentation [18] and pill recognition [19], resulting in increased precision in diagnosis and more effective therapies. Remote sensing can be used for item recognition and categorization in telescopic and aerial photos, assisting in land use mapping, urban development, and surveillance of the environment.

Alsaffar *et. al.* [20] established a system for the automated categorization and identification of TB in medical X-ray imaging. Three alternative classification approaches were employed to test the technique: Logit regression, proximity-based neighbors, and SVM. Then two categorization scenarios were employed: the construction of datasets for training and test, then Cross-validation. To conduct the categorization, characteristics were retrieved utilizing deep learning and the RESNET50 neural network in Figure 2. The scenario with the most favorable outcomes was achieved when the training and testing datasets were produced with a precision better than 0.85 whereas the classification approach that displays the highest performance in the two scenarios used in this study is SVM.

3. Research Methodology

Several conventional methods of detecting TB infections have been explored over time, including the use of diagnostic tools such as sputum smear microscopy, rapid molecular tests, and the Interferon-Gamma Release Assay. Recent advancements in machine learning, artificial intelligence, and deep learning have led to the development of algorithms that can help in disease detection. To support medical practitioners in the early diagnosis of Tuberculosis, we explored the option of developing an efficient Tuberculosis detection model using YOLOv8 from Figures 1 and 2.

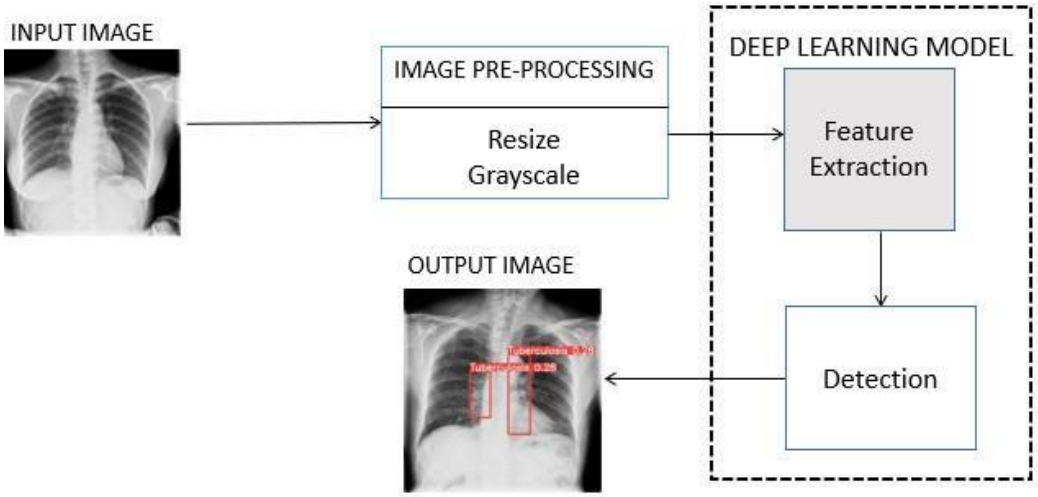

Figure 1: YOLOv8 Tuberculosis Detection Model

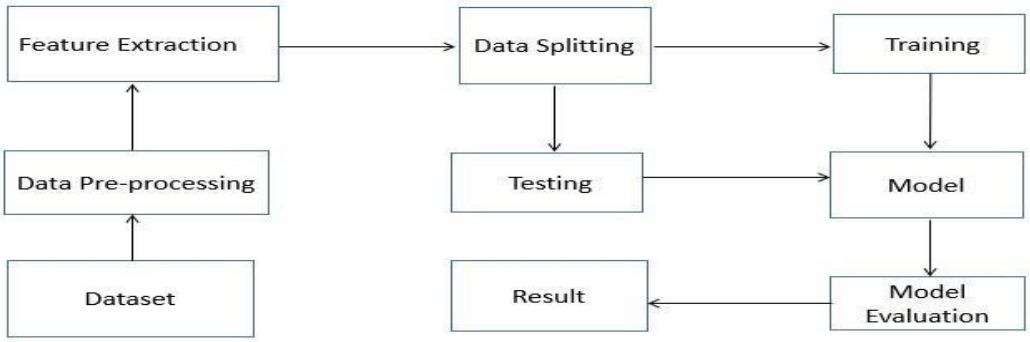

Figure 2: ResNet50 Classification Model

3.1 RestNet50 Classification Model Process

*The **ResNet50** model is a powerful convolutional neural network architecture that has been widely used for image classification tasks and the model is based on the following steps:*

i. Data collection: A dataset of X-ray images was collected. The images was obtained from Kaggle, an online repository of datasets used in data science and artificial intelligence.

ii. Pre-processing: The images were pre-processed to meet the standard required by the pre-trained model such as resizing, augmentation, etc.

iii. **Data Splitting**: The pictures were separated into training, validation, and testing consisting of 7182, 798, and 420 X-ray images respectively representing 85% for training, 10% for validation, and 5% for testing.

iv. The architecture of the model was set to perform the classification task before training begins.

v. Validation and test datasets were used to evaluate the trained model and assess its performance.

vi. The Performance of the model was calculated using standard metrics.

3.2 YOLOV8 Model Process

You Only Look Once (YOLO) version 8 is an advanced object detection model that has gained popularity due to its real-time performance and accuracy which has the following steps in its analysis:

i. **Input Data**: Active TB X-ray images with corresponding bounding box annotations for tuberculosis parts were used as input images.

ii. **Pre-processing**: The images and their bounding box annotations were pre-processed to meet the standard required by the pre-trained model such as normalization, resizing, augmentation, etc.

iii. **Data Splitting**: The 800 images were split into training, and validation, consisting of 690, and 110 X-ray images respectively representing about 86% for training, and 14% for validation.

iv. The model's configuration was attuned to perform the detection task before training starts.

v. The validation dataset was used to assess the trained model and assess its performance.

vi. The Performance of the model was calculated using standard metrics like precision and recall.

3.3 Dataset

TBX11K dataset comprises 11200 X-ray pictures with accompanying bounding box annotations for tuberculosis (TB) regions. All photos have a size of 512x512. There are five distinct groups in this dataset, i.e., Healthy, Sick but Non-TB, Active TB, Latent TB, and Uncertain TB. Both the Healthy and Sick categories had 3,800 X-ray images each while the Active TB category had 800 X-ray images, bringing the total X-ray images in these categories to 8,400 X-ray images which were subjected to pre-processing and feature analysis techniques such as Image Resizing, Image Normalization, and Image Augmentation were carried out on the dataset. The dataset was split into training and testing sets with 7,182 images used for training, 798 for validation, and 420 images for testing respectively.

Figure 3 shows an overview of different categories of X-ray images found in the dataset. The dataset contained the bounding box annotations for TB X-rays in XML format as well as MSCOCO style in JSON format.

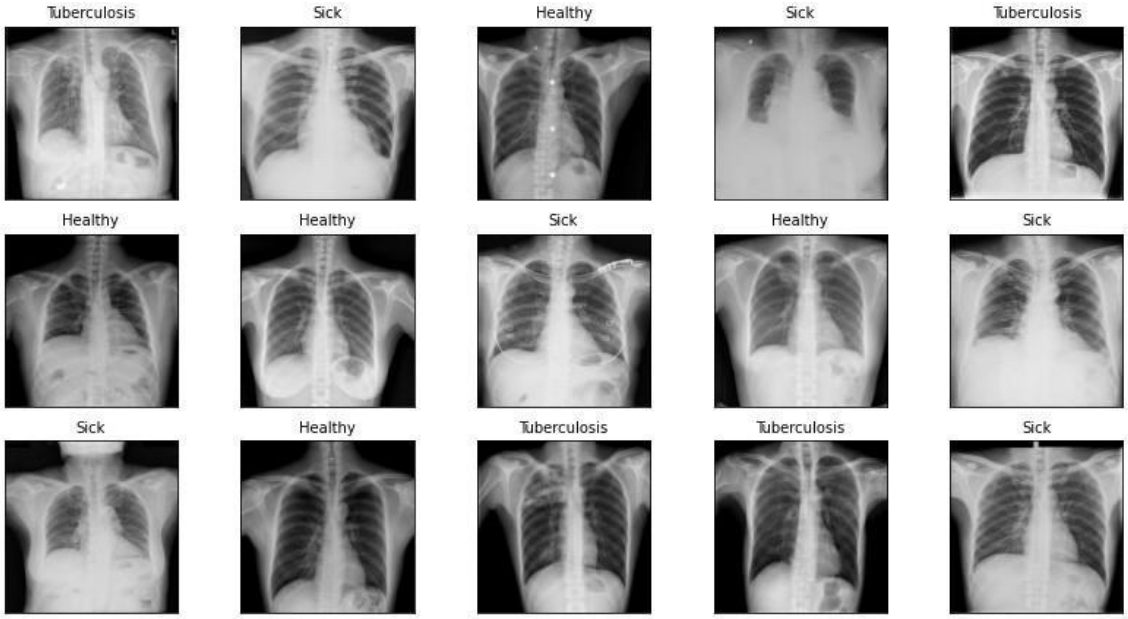

Figure 3: TBX11K Dataset obtained from Kaggle (Liu *et al*., 2020).

3.4 Tuberculosis Detection Model Methodology

The YOLOv8n model version was adopted in this study, which takes an input X-ray image of 640x640 pixels. The training and test datasets were mounted from the drive, which was used for training and testing the model. The 800 Tuberculosis X-ray images dataset was augmented, 690 was used for training and 110 used as validation for model learning and testing. The optimization algorithm, Adam, was used to modify the weights for the network depending on the training data, and the model was fine-tuned by adjusting the hyper-parameters such as the Epoch, learning rate, and batch size.

Precision, Recall and mean Average Precision are the fundamental measures used in evaluating Tuberculosis detection models. Precision measures the accuracy of positive predictions; it can be seen as a measure of quality with more relevant results returned than irrelevant ones while Recall measures the completeness of positive predictions, it can be seen as a measure of quantity with most of the relevant results returned which may include irrelevant ones. F1 score is a widely employed performance indicator for assessing the effectiveness of classification models. Mean Average Precision (mAP), also known as Average Precision is the generally used parameter for measuring the effectiveness of object identification models. It assesses the average precision throughout every category, giving a single number to compare various models.

4. Results

The classification results as shown in Figure 4, have a precision value of 98% and, a recall value of 96%, which indicates that 98% of the X-ray images predicted as 'Healthy' were indeed belonging to the Healthy category. This means the model has a high precision, which implies that it correctly identified the majority of X-ray images while minimizing the false positive rate.

Figure 5 shows the precision of the YOLOv8 model across different epochs, it had a precision value of 68%. This means the model performed averagely, which implies that to a decent extent, it was able to correctly detect regions on the X-ray images where Tuberculosis had infected while minimizing the false positive rates.

Figure 6 shows the recall of the YOLOv8 model across different epochs, it had a recall value of 65%. This means the model performed averagely, which implies that to a decent extent, it was able to correctly detect regions on the X-ray images where Tuberculosis had infected while minimizing the false negative rate.

	precision	recall	f1-score	support
0.0	0.9810	1.0000	0.9904	206
1.0	0.9909	0.9864	0.9886	220
2.0	0.9744	0.9048	0.9383	42
accuracy			0.9850	468
macro avg	0.9821	0.9637	0.9724	468
weighted avg	0.9850	0.9850	0.9849	468

Figure 4: Precision, Recall, and F1-score of RestNet50 Model

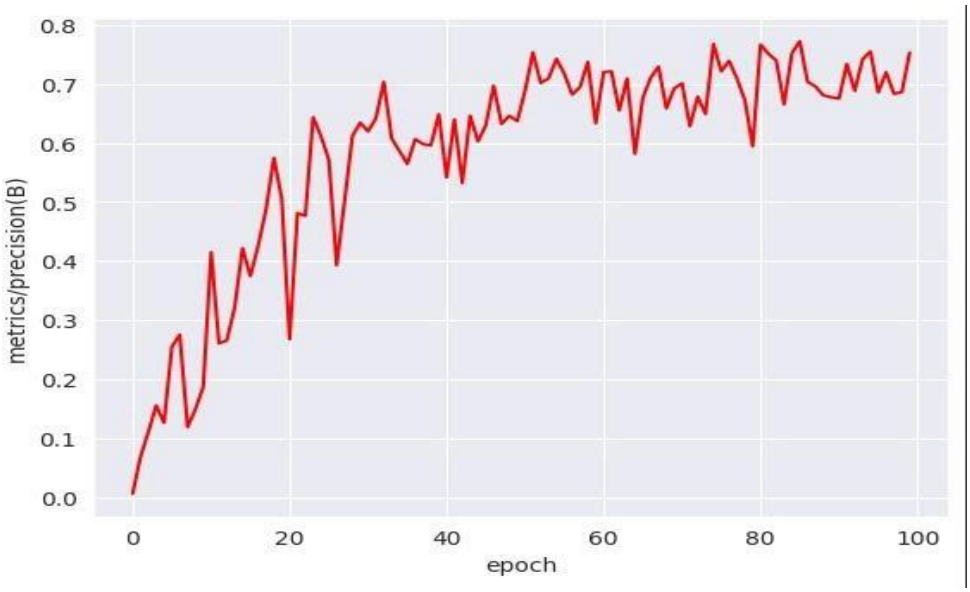

Figure 5: YOLOv8 Precision graph

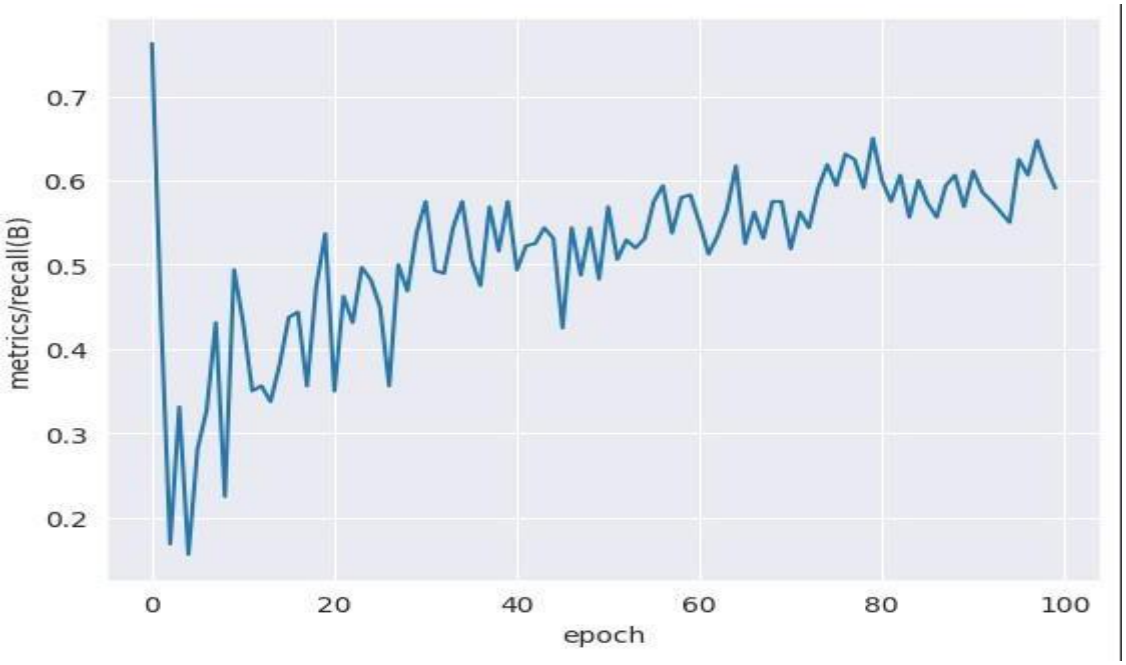

Figure 6: YOLOv8 Recall graph

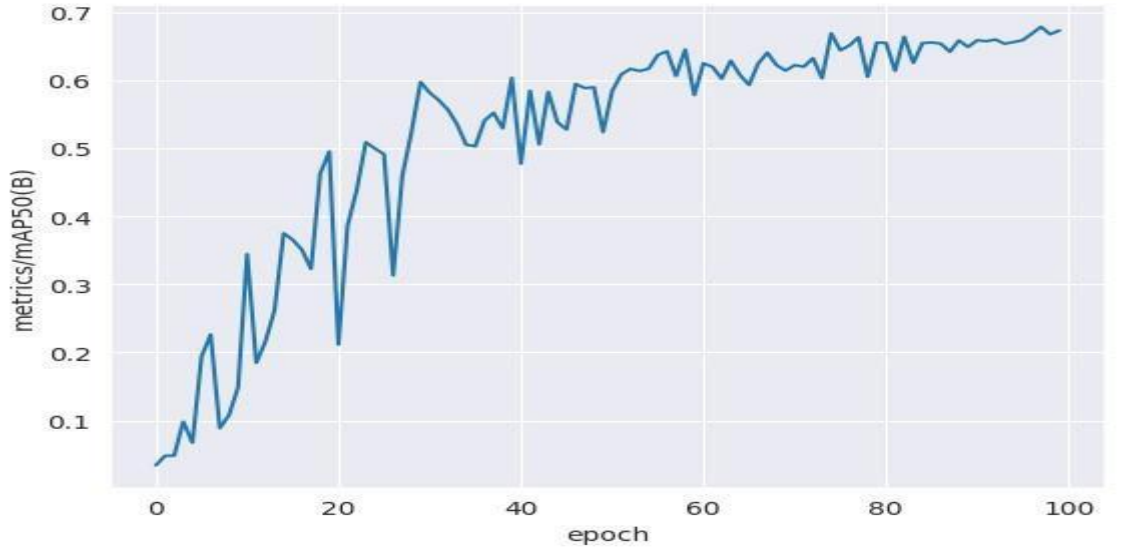

Figure 7: YOLOv8 Mean Average Precision graph

Figure 7 shows the mean Average Precision of the YOLOv8 model across different epochs, it had a mAP value of 68%. This means the model performed averagely, which implies that to a decent extent, it was able to correctly detect regions on the X-ray images where Tuberculosis had infected.

Figure 8 portrays the parameters set and the programming command for training the YOLOv8 model.

Figure 9 reveals an X-ray image with bounding boxes drawn on it to show the region infected by Tuberculosis and detected by the YOLOv8 model and their corresponding confidence score.

4.1 Discussion of Results

The metrics used to benchmark the performance of the models are Precision, Recall, and Accuracy. They are used to evaluate the relevance of tuberculosis classification and detection.

The ResNet50 algorithm demonstrated a good performance in classification, achieving an impressive accuracy of 98%. This high accuracy indicates that the model was proficient at correctly classifying 98% of the images in the dataset, showcasing its ability to make precise and reliable predictions.

```
!yolo task=detect mode=train model=yolov8s.pt data=/content/datasets/Tuberculosis-2/data.yaml epochs=25 imgsz=800 plots=True

Image sizes 800 train, 800 val
Using 2 dataloader workers
Logging results to runs/detect/train2
Starting training for 25 epochs...

      Epoch    GPU_mem   box_loss   cls_loss   dfl_loss  Instances       Size
       1/25      5.92G      2.603       3.93      2.389         21        800: 100% 94/94 [01:17<00:00,  1.22it/s]
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100% 5/5 [00:04<00:00,  1.21it/s]
                   all        160        359      0.188      0.255      0.113     0.0319

      Epoch    GPU_mem   box_loss   cls_loss   dfl_loss  Instances       Size
       2/25      7.17G      2.167      2.468      1.981         19        800: 100% 94/94 [01:13<00:00,  1.27it/s]
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100% 5/5 [00:02<00:00,  1.75it/s]
                   all        160        359     0.0472     0.0919     0.0239    0.00673
```

Figure 8: Training the YOLOv8 model

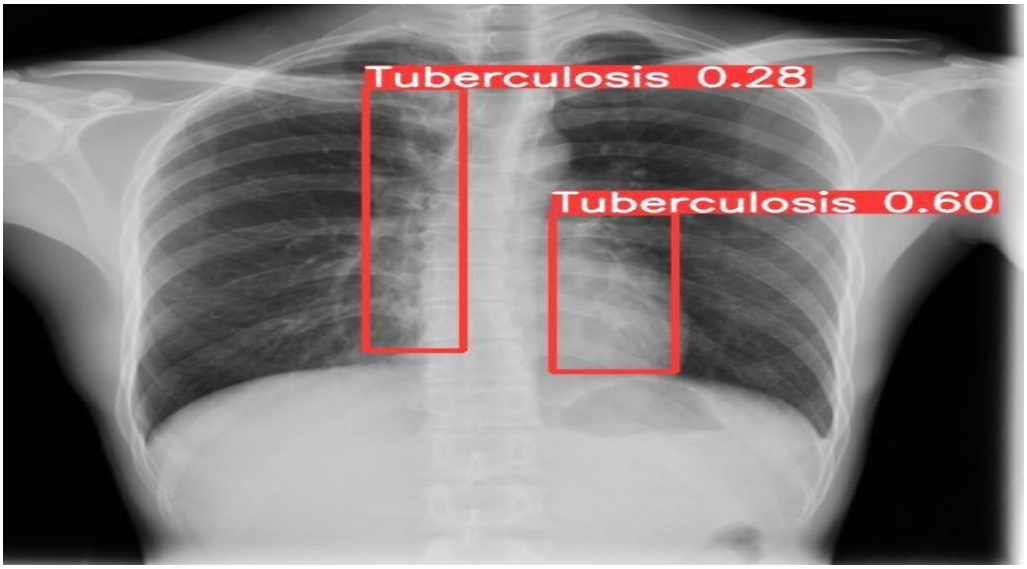

Figure 9: YOLO Result

The precision of the ResNet algorithm was 98%, indicating that 98% of the images predicted as true were indeed true while minimizing false positives. Its ability to effectively handle imbalanced datasets and identify positive instances with minimal false positives and false negatives makes it a suitable choice for classification problems.

The YOLOv8 algorithm was utilized to make detections on the tuberculosis X-ray dataset. The model demonstrated average performance in detecting the exact location of tuberculosis on X-ray images, achieving a precision of 68%, recall of 65%, and 68% mean average precision. The YOLOv8 algorithm tends to do better with larger datasets based on the nature of the model, as it works better with larger datasets.

5. CONCLUSION

In this study, powerful machine-learning models for tuberculosis detection during medical diagnosis were proposed. By providing accurate and timely diagnosis, we aim to empower medical practitioners with the tools needed to make informed decisions during diagnosis. While our models exhibit promising results, we acknowledge the ever-evolving nature of image analysis and the need for continuous improvements.

Depending on online sources for relevant X-ray images as the dataset appeared to be insufficient, largely due to the unavailability of a large amount of annotated Tuberculosis X-ray images on the web, the results of this research were limited which made it a major drawback. In the future, both online and physical datasets can be combined to provide enough data to work with. Also, investigating the severity of Tuberculosis after detection can provide additional details to medical practitioners and give them a sense of urgency to treat infected patients.

References

[1] Howard, E. L. (2023).Tuberculosis Harvard Health https://www.health.harvard.edu/a_to_z/tuberculosis-a-to-z. Accessed, May 2023.

[2] Singh, M., Pujar, G. V., Kumar, S. A., Bhagyalalitha, M., Akshatha, H. S., Abuhaija, B., Alsoud, A. R., Abualigah, L., Beeraka, N. M., and Gandomi, A. H. (2022). Evolution of Machine Learning in Tuberculosis Diagnosis: A Review of Deep Learning-Based Medical Applications. Electronics, 11(17), 2634. https://doi.org/10.3390/electronics11172634

[3] World Health Organization. (2019). WHO policy on the use of new rapid molecular tests for the diagnosis of tuberculosis.

[4] Hwang, S., Hijazi, M. H. A., Hwa, S. K. T., Bade, A., Yaakob, R., and Jeffree, M. S. (2019). Ensemble deep learning for tuberculosis detection using chest X-ray and canny edge detected images. *IAES International Journal of Artificial Intelligence, (IJ-AI).* Vol. 8, No. 4, December 2019, pp. 429~435, ISSN:

2252-8938, DOI: 10.11591/ijai.v8.i4.pp429-435

[5] Asada, N., Doi, K., MacMahon, H., Montner, S. M., Giger, M. L., Abé, C., and Wu, Y. (1990). The potential usefulness of an artificial Neural network for differential diagnosis of interstitial lung diseases: Pilot study. Radiology 1990, 177, 857–860

[6] Raymond, J. L., and Medina, J. F. (2018). Computational Principles of Supervised Learning in the Cerebellum. Annu. Rev. Neurosci. 2018, 41, 233–253.

[7] Ayaz, M., Shaukat, F., and Raja, G. (2021). Ensemble learning-based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors. Phys. Eng. Sci. Med. 2021, 44, 183–194.

[8] Kim, T. K., Yi, P. H., Hager, G. D., and Lin, C. T. (2020). Refining dataset curation methods for deep learning-based automated tuberculosis screening. J. Thorac. Dis. 2020, 12, 5078–5085.

[9] Reker, D., Schneider, P., Schneider, G., and Brown, J. B. (2017). Active learning for computational chemogenomic. Future Med. Chem. 2017, 9, 381–402.

[11] Lan, W., Dang, J., Wang, Y. and Wang, S. (2018). Pedestrian detection based on Yolo network model *IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 1547–1551, IEEE, 2018.

[12] Benjumea, A., Teeti, I., Cuzzolin, F. and Bradley, A. (2021). Yolo-z: Improving small object detection in yolov5 for autonomous vehicles arXiv preprint arXiv:2112.11798, 2021.

[13 Shinde, S., Kothari, A. & Gupta, V. 2018. Yolo-based human action recognition and localization Procedia computer science, vol. 133, pp. 831–838, 2018.

[14] Ashraf, A. H., Imran, M., Qahtani, A. M., Alsufyani, A., Almutiry, O., Mahmood, A., Attique, M. and Habib, M. (2022). Weapons detection for security and video surveillance using CNN and yolo-v5s, *CMC-Comput. Mater*. Contin, vol. 70, pp. 2761–2775, 2022.

[15] Zheng, Y. and Zhang, H. (2022) Video analysis in sports by lightweight object detection network under the background of sports industry development, *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[16] Ma, H., Celik, T. and Li, H. (2021). Fer-yolo: Detection and classification based on facial expressions in Image and Graphics: 11th International Conference, ICIG 2021, Haikou, China, August 6–8, 2021, Proceedings, Part I 11, pp. 28–39, Springer,

[17] Nie, Y., Sommella, P., O'Nils, M., Liguori, C. and Lundgren, J. (2019). Automatic detection of melanoma with yolo deep convolutional neural networks in 2019 *E-Health and Bioengineering Conference (EHB)*, pp. 1–4, IEEE, 2019.

[18] Ünver, H. M. and Ayan, E. (2019). Skin lesion segmentation in dermoscopic images with a combination of yolo and grab cut algorithm, Diagnostics, vol. 9, no. 3, p. 72, 2019.

[19] Tan, L., Huangfu, T. Wu, L., and Chen, W. (2021). Comparison of retina net, SSD, and yolo v3 for real-time pill identification, *BMC medical informatics, and decision making*, vol. 21, pp. 1–11, 2021.

[20] Alsaffar, M., Alshammari, G., Alshammari, A., Aljaloud, S., Almurayziq, T. S., Hamad, A. A., Kumar, V., and Assaye B. (2021). Detection of Tuberculosis Disease Using Image Processing Technique. *Hindawi. Mobile Information Systems*, Volume 2021, Article ID 7424836, 7 pages. https://doi.org/10.1155/2021/7424836