Chest X-ray Classification for Tuberculosis Detection

# 1. Project Overview

Objective:  
Automated deep learning system to classify chest X-ray images into Normal and Tuberculosis (TB) classes.  
  
Motivation:  
- Early detection of TB  
- Fast and accurate classification  
- Reduce manual workload

# 2. Dataset

The dataset contains chest X-ray images of two classes: Normal and TB.  
Training/Validation/Test split: 70/15/15  
Class imbalance: ~1:5 (Normal:TB)

|  |  |
| --- | --- |
| Class | Count |
| Normal | 514 |
| TB | 2494 |

# 3. Data Preprocessing

1. Images resized to 128x128 pixels for less comutation and memory efficient  
2. Pixel values normalized with each model's preprocess\_input function  
3. Data augmentation on training set: horizontal flip, rotation, zoom, translation, brightness, contrast, Gaussian noise  
4. Class weights used to handle imbalance: Normal=3, TB=1 because, Formula = [3008/(2\*514) ~ 3 , 3008/(2\*2494) ~ 0.6]  
Validation and test sets are not augmented.

# 4. Model Selection

Pre-trained CNN models tested:  
- VGG16  
- ResNet50  
- EfficientNetB0  
  
Evaluation metrics used:  
- ROC-AUC  
- F1 Score  
- Classification Report  
- Folder-wise prediction summary

# 5. Model Training and Fine-tuning

# Summary Report

|  |  |  |
| --- | --- | --- |
| VGG16 | ResNet50 | EfficientNetB0 |
| TB Chest X-rays: - Total images: 2494 - Normal: 24 images - TB: 2470 images  Normal Chest X-rays: - Total images: 514  - Normal: 472 images - TB: 42 images | TB Chest X-rays: - Total images: 2494 - Normal: 0 images - TB: 2494 images  Normal Chest X-rays: - Total images: 514  - Normal: 249 images - TB: 265 images | TB Chest X-rays: - Total images: 2494 - Normal: 60 images - TB: 2434 images  Normal Chest X-rays: - Total images: 514  - Normal: 505 images - TB: 9 images |

# EfficientNetB0 was chosen for fine-tuning due to its lightweight architecture and good baseline performance. Fine-tuning involved unfreezing the last 40 layers, reducing learning rate(2e-5), and training additional epochs(initially 5 epochs then 20 epochs) with early\_stop and class weights.

# 6. Evaluation Results

Folder-wise prediction summaries for the fine-tuned EfficientNetB0:  
  
TB Chest X-rays:  
- Total images: 2494  
- Normal: 24 images  
- TB: 2470 images  
  
Normal Chest X-rays:  
- Total images: 514  
- Normal: 514 images  
- TB: 0 images  
  
Observation: Fine-tuned EfficientNetB0 performs excellently, correctly predicting TB and Normal images.

# 7. Deployment Decision

Chosen Model: Fine-tuned EfficientNetB0  
Reason: Highest accuracy and robust performance.  
Ready for inference and deployment.

# 8. Future Work

- Test larger EfficientNet variants  
- Evaluate on external datasets to verify generalization  
- Deploy as Streamlit web app and in AWS