Tuberculosis (TB) Chest X-ray Classification

By Chelsea Rodrigues (23200333)

Introduction

Tuberculosis (TB), caused by the bacterium Mycobacterium tuberculosis, is a contagious disease that primarily affects the lungs and remains a leading cause of death worldwide. Early diagnosis and treatment are crucial to combat this disease. Chest X-rays (CXR) are commonly used to detect TB, but there are disparities in diagnosis, particularly in high-burden countries. Computer-aided detection (CAD) systems, powered by deep learning models like Convolutional Neural Networks (CNNs), have shown promise in improving diagnostic accuracy.

Proposal

This study proposes the use of DenseNet, a deep convolutional neural network, and a Deep Forest model to classify TB in chest X-ray images. The goal is to develop a robust and accurate model that can assist radiologists in diagnosing TB, especially in areas with limited medical resources. The combination of DenseNet's deep feature extraction and Deep Forest's ensemble learning aims to enhance the classification accuracy of TB in CXR images.

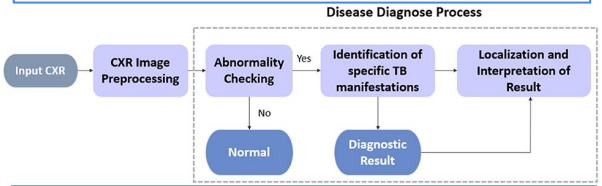
Dataset Overview

The dataset consists of 700 publicly accessible TB images and 2800 additional TB images available from the NIAID TB portal, along with 3500 standard (non-TB) images. These images were pre-processed using techniques like Contrast Limited Adaptive Histogram Equalization (CLAHE) and resized to 150 x 150 pixels for the Deep Forest model and 300 x 300 pixels for the DenseNet model.

Dataset: TB_Chest_Radiography_Database
https://www.kaggle.com/tawsifurrahman/tuberculosis-tb-chest-xray-dataset

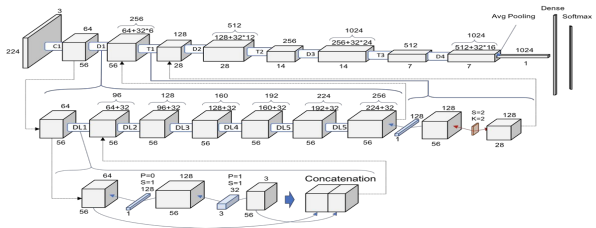
Process Flow

- Data Collection: Gather chest X-ray images of TB-positive and normal cases.
- Data Preprocessing: Enhance images using CLAHE, resize them for model input, and apply data augmentation techniques like horizontal flipping and rotation.
- Feature Extraction: Use DenseNet to extract deep features from the CXR images.
- Model Training: Train the extracted features using the Deep Forest model.
- Evaluation: Validate the models using metrics like accuracy, precision, recall, and F1-score.

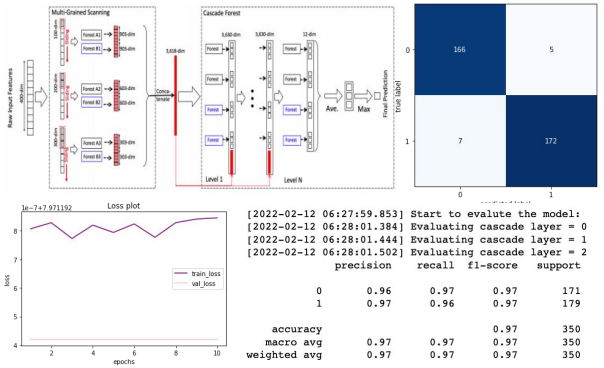


ML Methodology

DenseNet: A CNN architecture where each layer is connected to all subsequent layers, enabling efficient information flow and feature reuse. It reduces the number of parameters compared to traditional CNNs by concatenating features instead of summing them.



Deep Forest: An ensemble learning method that predicts bounding boxes for individual objects in images. It utilizes a cascaded structure and performs well in classifying CXR images without segmentation.



Results

- The Deep Forest model achieved superior performance with a classification accuracy, precision, and recall of 97% each, outperforming DenseNet, which showed an accuracy of 50%, precision of 25%, and recall of 33%.
- The results highlight the importance of using segmentation in the classification process to improve the accuracy of the DenseNet model.

Conclusion

The proposed method using a cascaded classifier with a Deep Forest model and DenseNet demonstrates the potential for automatic TB detection from chest radiographs. The study found that while DenseNet requires segmentation for optimal performance, Deep Forest performed well even without segmentation, making it a viable option for quick and accurate TB diagnosis. Future work will focus on improving DenseNet's performance by integrating image segmentation techniques.