**Submission date**: 11-Dec-2023

**USN:** ENG21CY0012

**Topic:** Tuberculosis Detection on X-Ray Images with CNN

**Word count**: 1552

**Character count:** 9881

C

PROPOSAL BY

CHUNDURU NARASIMHA SATWIK

**Abstract:** Tuberculosis (TB) remains a major global health challenge, requiring innovative strategies for early detection and prediction**.** Globally, Tuberculosis is a serious health issue for which an early diagnosis is crucial to an effective course of therapy. Machine learning techniques have been used to predict tuberculosis with a high degree of accuracy. This study presents an integrated approach using machine learning techniques to improve the accuracy and efficiency of tuberculosis prediction. A comprehensive dataset containing clinical, demographic, and laboratory parameters along with X-ray scans of various TB patients was used to train and validate a predictive model by using multiple Machine Learning (ML) models. Various operations are done on the dataset such as data preprocessing, splitting the dataset, modelling, evaluation, model prediction, and lung segmentation.

**1. Introduction**

Tuberculosis is a disease which is caused by a contagious bacterial infection called Mycobacterium tuberculosis which primarily affecting the lungs. Effective disease management and preventing the spread of tuberculosis (TB) depend heavily on early detection and precise prediction of the illness. The field of tuberculosis detection and prediction now has more options because to recent developments in machine learning. Large databases of medical data, such as chest X-rays, CT scans, and patient demographics, can be analysed by machine learning algorithms to find trends linked to tuberculosis infection. Next, using this data, models that can accurately predict the presence of tuberculosis can be created.

The medical field has seen the rise of diagnostics after it’s integration with the field of machine learning and in this project multiple machine learning projects such as Convolutional Neural Networks (CNN), Decision Trees, Random Forest, k-Nearest Neighbours (KNN), and Support Vector Machines (SVM). Conventional methods of tuberculosis (TB) detection have a number of drawbacks that prevent prompt diagnosis and efficient treatment. These are a few of the main obstacles:

* **Low Sensitivity of Microscopy**
* Slow turnaround time
* **Inability to distinguish active from latent TB**
* **Difficulty in diagnosing extrapulmonary TB**

**2. Problem Statement**

Tuberculosis (TB) remains a significant global health issue, affecting millions of people worldwide. Timely and accurate diagnosis is crucial for effective management and control of the disease. Chest X-ray (CXR) imaging is a common diagnostic tool for TB, providing detailed insights into the condition of the lungs. Convolutional Neural Networks (CNNs) have demonstrated remarkable capabilities in image classification tasks, making them a promising technology for automated TB detection.

The dataset consists of:

* Amount: 700 Tuberculosis images / 3500 normal images. (Class imbalance)
* File type .PNG
* Image size: 512x512
* Channel: 3 channels (RGB)

**3. Literature Survey**

This project contains various techniques which helps in early prediction and detection of Tuberculosis and enhances the accuracy of the detection in the current scenario. The main learning technique was through Convolutional Neural Network (CNN). This project also compares the accuracy of other machine learning and deep learning algorithms with accuracy and precision as their parameters. Here are some the research papers from where the project has references:

1. AI-based prediction for early detection of Tuberculosis in India based on environmental factors

Authors : [Nupur Giri](https://ieeexplore.ieee.org/author/37411022900); [Richard Joseph](https://ieeexplore.ieee.org/author/37086357164); [Sanika Chavan](https://ieeexplore.ieee.org/author/37088413379); [Raghav Heda](https://ieeexplore.ieee.org/author/37088414407); [Reema Israni](https://ieeexplore.ieee.org/author/37088413709); [Ritika Sethiya](https://ieeexplore.ieee.org/author/37088412753)

1. Analysis and Prediction of Tuberculosis using Machine Learning Classifiers

Authors : [M. Senthilmurugan](https://ieeexplore.ieee.org/author/37088843398); [M. Latha](https://ieeexplore.ieee.org/author/37086172598); [R. Chinnaiyan](https://ieeexplore.ieee.org/author/37086289002)

1. Tuberculosis Disease Diagnosis Based on an Optimized Machine Learning Model

Authors: Olfa Hrizi,1**Karim Gasmi,** Ibtihel Ben Ltaifa, Hamoud Alshammari, Hanen Karamti, Moez Krichen, Lassaad Ben Ammar,and Mahmood A. Mahmood

1. Prediction of tuberculosis using an automated machine learning platform for models trained on synthetic data

Authors : [Hooman H. Rashidi](https://pubmed.ncbi.nlm.nih.gov/?term=Rashidi%20HH%5BAuthor%5D), [Imran H. Khan](https://pubmed.ncbi.nlm.nih.gov/?term=Khan%20IH%5BAuthor%5D), [Luke T. Dang](https://pubmed.ncbi.nlm.nih.gov/?term=Dang%20LT%5BAuthor%5D), [Samer Albahra](https://pubmed.ncbi.nlm.nih.gov/?term=Albahra%20S%5BAuthor%5D), [Ujjwal Ratan](https://pubmed.ncbi.nlm.nih.gov/?term=Ratan%20U%5BAuthor%5D), [Nihir Chadderwala](https://pubmed.ncbi.nlm.nih.gov/?term=Chadderwala%20N%5BAuthor%5D), [Wilson To](https://pubmed.ncbi.nlm.nih.gov/?term=To%20W%5BAuthor%5D), [Prathima Srinivas](https://pubmed.ncbi.nlm.nih.gov/?term=Srinivas%20P%5BAuthor%5D), [Jeffery Wajda](https://pubmed.ncbi.nlm.nih.gov/?term=Wajda%20J%5BAuthor%5D), and [Nam K. Tran](https://pubmed.ncbi.nlm.nih.gov/?term=Tran%20NK%5BAuthor%5D).

1. Machine Learning Techniques for Tuberculosis Prediction

Authors: Akshitha Tiwari, Srabanti Maji

This literature survey has illustrated that all these 5 papers aim to improve and better the prediction of the machine learning model to predict TB accurately. Numerous machine learning techniques, such as decision trees, random forests, support vector machines, and neural networks, are employed in these studies. Genomic and metabolomic data, as well as clinical and demographic information, are among the features that are used for prediction.

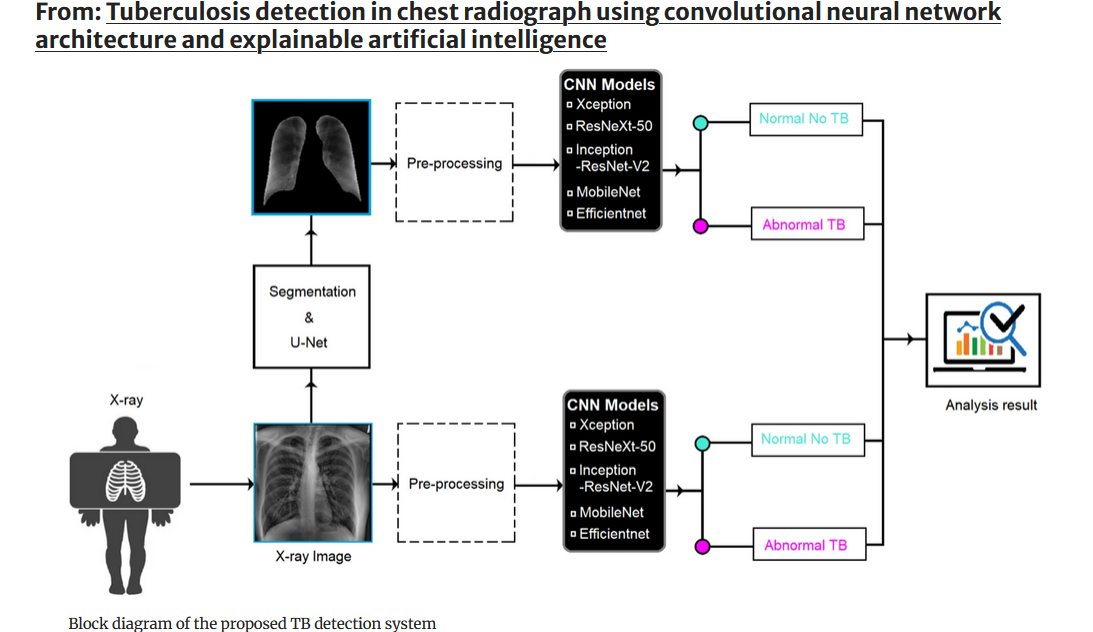
The dataset used for all these algorithms is available in Kaggle. This project also works with LIME (Local Interpretable Model-Agnostic Explanations) which is used in deep learning models.

**4. Methodology**

For this project multiple ML algorithms have been used such as :

* Convolutional Neural Networks
* Decision Tree Classifier
* Random forest Classifier
* K- Nearest Neighbours
* Support Vector Machines

**Convolutional Neural Networks:**



* + Data Preprocessing:

Converting PNG Images to NumPy Arrays.

**Pros:**

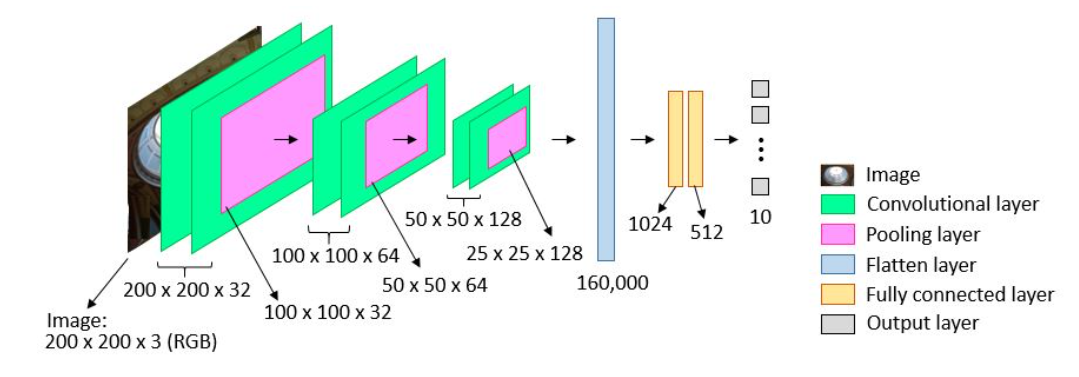
**Memory Efficiency:** Because NumPy arrays use less memory than raw image files, they are ideal for large datasets.

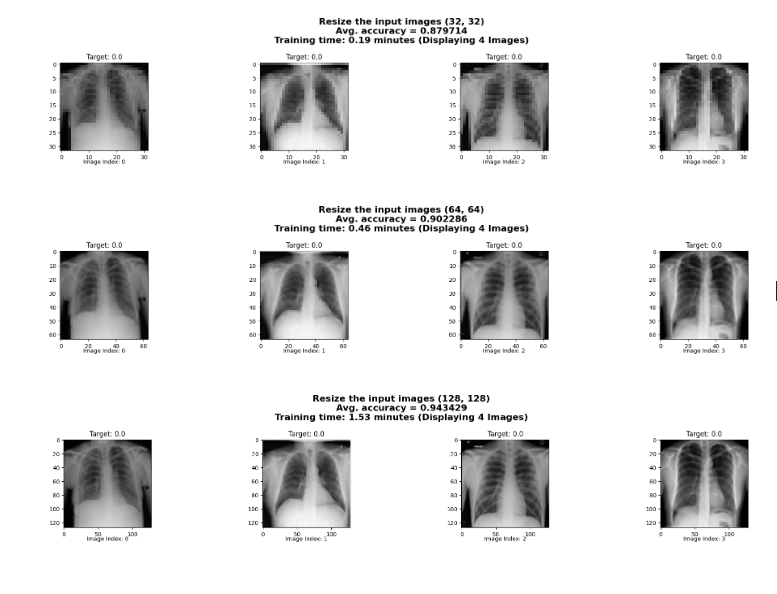
NumPy array operations are optimized and faster in terms of computational efficiency.

**Cons:**

**Lossy Compression:** Depending on the image format and compression settings used, converting images to NumPy arrays may involve lossy compression, resulting in some image quality loss.

**Storage and Disk Space:** While NumPy arrays are more memory-efficient, they can still take up a lot of disk space, especially if you have a lot of images or a lot of them.

Convolutional Neural Networks are a type of deep learning algorithm that is designed mainly for image processing, pattern recognition, and computer vision tasks. CNN are very prominent in the field of Deep Learning. They have revolutionized the field of computer vision and achieved cutting-edge results in a variety of image-related applications such as object detection, image classification, segmentation, and others. CNNs are designed to perform tasks on 2-D images. In the context of binary classification for TB detection, we can customize the CNN's output layer to have a single unit that will classify the input as TB or normal.

****

**Training Process:**

* Forward propagation
  + Receives the input images in a variable (X for example) and applies a series of linear transformations through hidden layers until the **final output** layer produces predictions for the given input.
* Backward propagation
  + Randomly initialized the weights, biases during this step to minimize the Error function (**Gradient Optimization**) and improve the model's performance.

CNNs Learning: Detecting Features in Images

* Convolutional Layers: ability to learn and detect the entire image and patterns and features.
* Activation Functions: each layer applies an activation function, input images and calculated by activation function transform to the output of each layers.
* Combining Layers: retaining the most crucial information while reducing the image and dataset complexity.

Summary using a Convolution or Pooling layer we reduce the dimensions of the input image of dimensions

**At the end it combines the learned features from previous layers and making final predictions in a neural network.**

**Splitting the Dataset:**

* 80% for training,
* 10% for validation,
* and 10% for testing. Ensure that the classes are balanced in each split.

**Decision Tree Classifier:**

**1. Feature extraction:**

Rather than using pixel values directly, images are first converted into a set of features that capture important information such as color, texture, and shape.

**2. Decision tree construction:**

The decision tree is constructed recursively by dividing the data into smaller subsets based on the feature values.

**3. Classification**

The features of a new image are extracted and compared to the decision rules at each node of the tree to classify it.

**Random Forest:**

Random forests are a combination of multiple decision trees which are clubbed together to improve the accuracy of decision making. While deep learning models dominate the field of image recognition, random forests can still be a useful tool in some cases, particularly when dealing with limited data or specific requirements.

**5. Results**

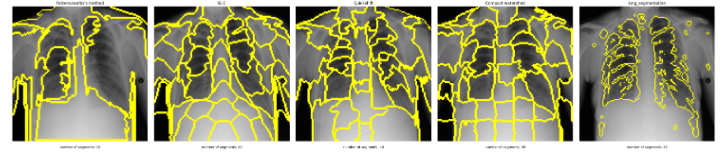
****

Figure 1 Lung Segmentation using LIME.

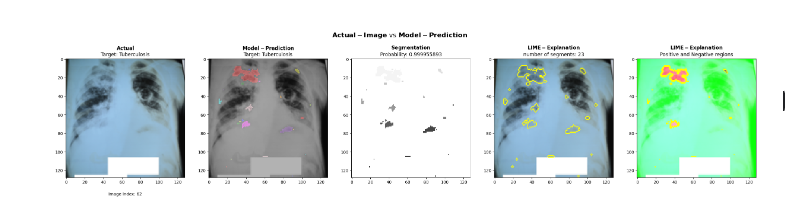
****

Figure 2 Sample output

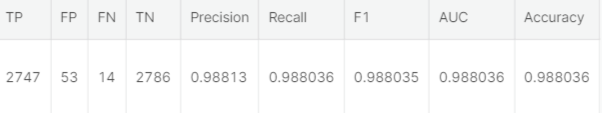
**Accuracy provided by different ML Algorithms :**

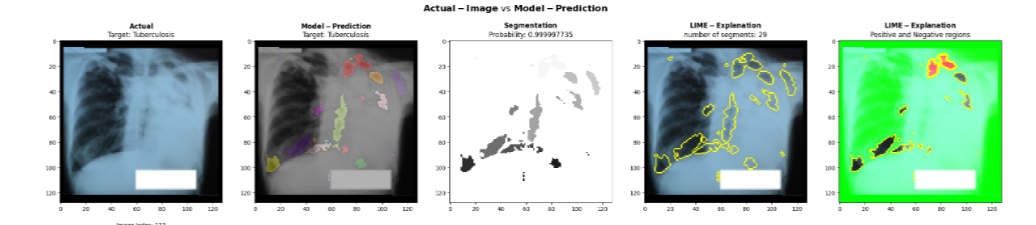
****

****

****

CNN:





Here is a sample output for the input which is provided. As you can the image has been divided into multiple segments by LIME and the model has predicted that this image is belonging to the “non tuberculosis” category.

**6. References**

[1] Jaeger S, Karargyris A, Candemir S, et al. "Automated Tuberculosis Screening Using Chest X-rays" IEEE Transactions on Medical Imaging, 33(2), 233-245. (2014)

[2] Lopes FM, de Paiva AC, Gonçalves LM. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics, and Transfer Learning" Journal of Applied Mathematics, vol. 2017, Article ID 6514793. (2017).

[3] Das DK, Ahmed MS, Rahman MM "Tuberculosis Screening and Classifying Using Convolutional Neural Networks" 2017 IEEE Region 10 Symposium (TENSYMP), Cochin, India. (2017)

[4] Rahman MT, Antani SK, Thoma GR. "Deep Learning for Tuberculosis Detection in Chest Radiographs: A Systematic Review and Meta-analysis" Journal of Imaging, 4(12), 189. (2018)

[5] Muktabh Mayank Srivastava, Towards Automated Tuberculosis detection using Deep Learning IEEE Conference.

[6] Fabrice Meriaudeau, Deep-learning: A potential method for tuberculosis detection using chest radiography. 2017 IEEE International Conference on Signal and Image Processing Applications.

[7] Chang Liu , Yu Cao , Marlon Alcantara , Benyuan Liu , Maria Brunette , Jesus Peinado , Walter Curioso, TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural networks

Proceedings of the 2017 IEEE International Conference on Image Processing , p. 2314 - 2318 Posted: 2017.

[8] M Srinivas , D Roy , C K Mohan, Discriminative feature extraction from x-ray images using deep convolutional neural networks. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing.