International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Enhancing Tuberculosis Detection Through Machine Learning on Chest X-Ray Scans

G.V. Ramana[1], N.V.S Sai Ram[2], J.J.D.V Ram Kumar[3], G.Ananda Ragini[4], K.L.S Surya Kumari[4]

[1]Assistant Professor, Department of Electronics and Communcation Engineering VSM College of Engineering Ramachandrapuram, Andhra Pradesh India
[2,3,4]UG Student, Department of Electronics and Communcation Engineering VSM College of Engineering Ramachandrapuram, Andhra Pradesh India

Abstract
Tuberculosis (TB) is a serious illness disease that cause germs are a type of bacteria caused by Mycobacterium that affects the lungs, often it can also affect other parts of the body. Active disease accurately untreated lead to risk to life. Need of improve diagnosis of TB by independent of doctors by assembling leveraging machine learning methods particularly random forest , logistic regression models and embossing chest X-rays, these techniques provide more accurate, timely, and reliable diagnoses.Present machine learning utilizes technology to identify diseases .Mainly Digital image processing (DIP) can adjust and distribution of pixel intensities to enhance image contrast and Support Vector Machine (SVM) bring the effective classification tasks like disease detection from medical images. The primary objective of our project is to detect TB early using digital diagnosis of chest X - rays

Keywords: Mycobacterium Tuberculosis, Digital Image Processing (DIP), Machine Learning, Support Vector Machine (SVM) , Random forest , Logistic regression

INTRODUCTION

threat, ranking among the top ten causes of death worldwide, notably affecting developing nations, according to the World Health Organization (WHO). Conventional TB diagnostic methods, like microscopy, often prove time-consuming, leading to delays in treatment and sometimes inadequate care. In response to this challenge, machine learning algorithms have emerged as a promising avenue for enhancing TB diagnosis. Machine learning, a subset of artificial intelligence, employs algorithms and statistical models to empower systems to learn from data and make informed predictions or decisions. By scrutinizing vast datasets, machine learning algorithms can discern patterns and forecast diseases.

For this project, Support Vector Machines (SVM) have been selected due to their effectiveness in classifying data points based on their unique characteristics. SVM, a supervised learning technique, operates as a binary classification algorithm, seeking to determine the hyperplane with the greatest margin of separation between two classes within a high-dimensional feature space. It offers several advantages, including its ability to handle high-dimensional data, robust generalization capabilities, and adaptness in managing non- linearly separable data. The primary objective of this project is to develop a machine learning-based system capable of identifying TB from chest X-ray images using SVM.

The research entails employing image preprocessing techniques, such as adaptive histogram smoothing, to enhance contrast and image quality. Subsequently, SVM will be deployed to classify images accurately, distinguishing TB from other diseases presenting similar symptoms. By leveraging machine learning, particularly SVM, this system aims to advance TB diagnosis, offering a more efficient and accurate approach, particularly in resource-poor environments.

2. LITERATURE REVIEW

On our research, we analysis and discussed present ' techniques and previous projects for development these ' project. some of the existing models are

2.1 Early Detection of Tuberculosis using Chest X-Ray (CXR) with computer-aided diagnosis

Computer-aided Diagnosis (CADx) system based on image processing is proposed to assist doctors and radiologists in interpreting Chest X-rays (CXR) for early detection of lung Tuberculosis (TB).CXR can indicate lung abnormalities including TB. However, the interpretations of CXR might vary from one individual to another.The steps that were performed by the proposed system consisted of preprocessing, segmentation,feature extraction, and classification.In the preprocessing stage homomorphic filter,histogram equalization,median filter, and Contrast- Limited Adaptive Histogram Equalization (CLAHE) were applied to increase image quality. Segmentation was done by using Active Contour Model. Feature extraction was performed by analyzing the image's first first order statistical features.The last stage, classification,was based on the mean values.The results indicated that the system can increase specificity while maintaining sensitivity and accuracy of TB diagnosis.

2.2 Deep Learning-Basied Automatic Detection of Tuberculosis disease in chest x-ray images

A CNN-based model was used to detect TB from CXR images. In the current work, a CNN-based transfer learning approach using five different pre-trained models, including Inception_v3, Xception, ResNet50, VGG19, and VGG16 was utilized for the classification of CXR images to normal and TB (binary classification). The overall architecture of the proposed CNN models consists of feature extractors and a classifier (sigmoid activation function).In other words, the architecture of all pre-trained transfer models is similar and contains 4 following parts: (1) convolution layer, (2) pooling layer, (3) flattening layer, and (4) fully connected layer. Herein, the fully connected layer comprises following layers: (1) flatten, (2) dense with a size of 256 units, (3) dropout with a threshold of 0.2, and (4) a last dense with the two-element softmax activation to predict the normal and TB cases (binary classification)

3.EXISTING METHOD

TB infection is diagnosed if the person has a positive result from a TB blood test or skin test and a medical evaluation does not indicate TB disease. TB disease is diagnosed by medical history, physical examination, chest x-ray, and other laboratory tests, including culture sputum smear microscopy, sputum culture, nucleic acid amplification tests (NAATs) like Xpert MTB/RIF, drug susceptibility testing.Let know the disadvantages of these

3.1 Delays and Turnaround Times:

TB diagnosis through traditional methods can have longer turnaround times compared to molecular testing.

3.2 False Positives and Negatives Tests

Tuberculin Skin Test (TST) has limitations, including poor specificity due to BCG vaccination and non-TB mycobacterial exposure, leading to false-positive results. Interferon-Gamma Release Assay (IGRA) is a blood test, errors in collecting, transporting, or interpreting the test can decrease accuracy. Some individuals may have a negative TST result despite being infected with TB bacteria.

3.3 Challenges in Specific Populations

sensitivity of TSTs can be lower in immuno compromised individuals. IGRA and TSTs may have lower sensitivity in children. required sophisticated laboratory facilities and trained personnel, which can be challenging in resource-limited settings.

3.4 Cost and Infrastructure

Faster results and higher sensitivity, they may involve higher costs compared to conventional methods, posing challenges for implementation in resource- limited settings. They require specialized labs with expensive infrastructure and equipment, along with trained microbiologists.

3.5 Sample Collection and Handling

Errors in collecting or transporting blood specimens can decrease the accuracy of TB blood tests. cost of transport, long distances to health facilities, and lack of decentralized diagnostic services can be barriers to TB diagnosis.

4. PROPOSED SOLUTION

The four primary steps. Initially, an X-ray image is selected as input. Subsequently, the image undergoes preprocessing utilizing OpenCV's histogram equalization and Contrast Limited Adaptive Histogram Equalization (CLAHE) techniques to augment image quality. Following preprocessing, the image is classified employing a Support Vector Machine (SVM) classifier model trained on a dataset comprising X-ray images from both TB-positive and TB-negative patients. Ultimately, the output of the SVM classifier model manifests as the prediction probability, The objective of the proposed methodology is to accurately detect TB in X-ray images, prioritizing the augmentation of accuracy and reliability throughout the detection process. Implemented using Python code, the proposed model attains an accuracy rate of 97.0% with a precision of 97.0%

The proposed method can be represented visually using flowchart, the step-by-step process of the classification task .

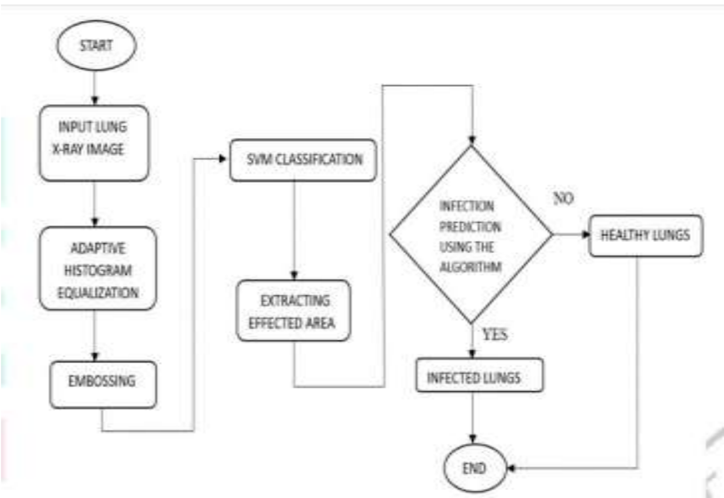

Fig A: Block Diagram

5. METHODOLOGY

Five different models in classifying chest X-ray images chest X-ray images as normal chest X-ray images

as normal Logistic Regression, Inception V3, and MobileNet. Each model underwent a series of image processing techniques, such as CLAHE, to enhance image quality.
Notably, SVM was the sole model to incorporate the equalizeHist function during the image preprocessing stage. Following comprehensive training and testing, SVM emerged as the most accurate and consistent model for this specific classification task, boasting an impressive accuracy score of 0.97. Additionally, SVM demonstrated high precision and recall scores of 0.97 each, resulting in an exceptional F1-score of 0.97.

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
SVM	0.97	0.97	0.97	0.97
RANDOM FOREST	0.95	0.95	0.95	0.95
LOGISTIC REGRESSION	0.78	0.78	0.078	0.770
INCEPTION V3	0.96	0.96	0.96	0.96
MOBILE NET	0.91	0.92	0.91	0.91

Fig B: Table of Various Model Datasets

6. SOFTWARE REQUIREMENT

6.1 HTMLM5

It is a combination of HTML, CSS and Javascript with APIs like offline storage, microdata, audio and video, drag and drop, geolocation, embedded fonts, web APIs etc.

6.2 PYTHON PROGRAMMING

You can add new classes that implement the same interface without modifying existing code.Well-defined interfaces make it easier to understand and maintain your code. Popular Python frameworks like Django and Flask provide tools and structures to simplify web development, including routing, templating, and database integration.

6.3 JUPYTER NOTE BOOK

Jupyter note is the latest web-based interactive development environment for notebooks, code, and data. Its flexible interface allows users to configure and arrange workflows in data science, scientific computing, computational journalism, and machine learning. A modular design invites extensions to expand and enrich functionality in html , image , videos and custom MIME types .

6.4 DIGITAL IMAGE PROCESSING

Using python image processing libraries to improve image quality by applying mathematical operations. Image Processing Projects involves modifying images by identification of its two dimensional signal and enhancing it by comparing with standard signal.

6.5 KAGGLE DATABASE

A team of researchers collaboration with medical doctors have created a database of chest X-ray images for Tuberculosis (TB) positive cases along with Normal images for references .

7. ADVANTAGES

7.1 ENHANCED ACCURACY AND SENSITIVITY

ML algorithms can analyze complex patterns in medical images (like chest X-rays) and other data to identify TB more accurately than traditional methods, potentially reducing misdiagnosis and false negatives.

7.2 EARLY DETECTION

By analyzing data from various sources, ML can help identify TB at an earlier stage

7.3 RAPID PROCESSING

ML models can process data quickly, enabling faster diagnoses and treatment initiation

7.4 COST LESS

Due to absence of human effort it will be at low cost 7 .

7.5 RESOURCES AND LIMITED

Use of software makes no need of external equipment

8. APPLICATIONS

8.1 Chest X-Ray Analysis

ML algorithms, particularly deep learning techniques like Convolutional Neural Networks (CNNs), can analyze chest X-rays to identify patterns indicative of TB, potentially reducing the workload on radiologists and improving diagnostic accuracy.

8.2 CAD Systems

Computer-Aided Detection (CAD) systems, powered by ML, can assist radiologists in identifying TB by highlighting suspicious areas on X-rays.

8.3 Increase X-Ray Quality

Presence of digital image Processing models doctors also prefer these process

9. EXPERIMENTAL RESULTS

The uploaded image displays the results of TB classification, indicating whether the provided chest X-ray image is positive or negative for tuberculosis.

1. X-ray images extracted from the dataset undergo histogram equalization during the pre -processing stage.

Fig C Input X-Ray Image

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Fig D X-Ray Fater CLAHE

2. The X-ray images are uploaded on the index page of the interface.

Fig E: Index Page of Interface

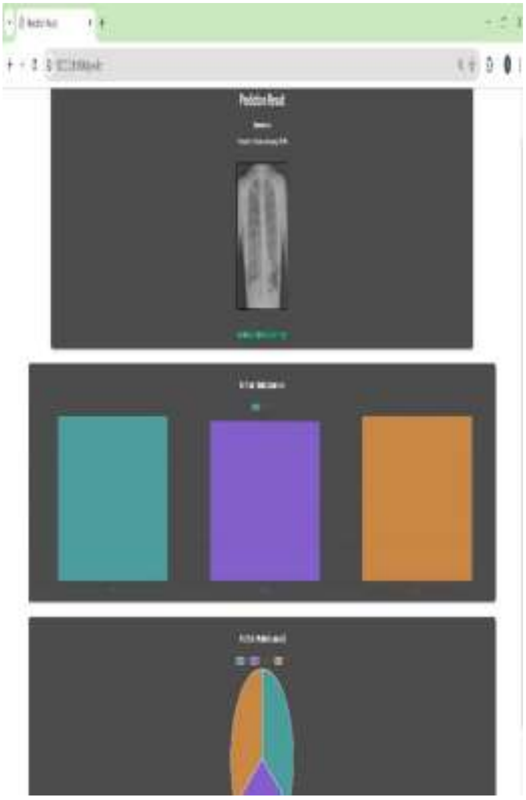

Fig F: Result for given input

10. FUTURE SCOPE

Augmenting the Dataset: While the existing dataset utilized in this project boasts substantial size, enhancing it with additional diverse and varied data holds potential to elevate the model's accuracy to even greater heights.

Incorporation with Electronic Health Records (EHRs): Embedding the model within Electronic Health Records (EHRs) offers healthcare practitioners invaluable support in decision-making processes, furnishing them with comprehensive patient histories for more informed medical diagnoses.

AI-assisted chest X-ray (CXR) imaging: A lot of mature AI software has been applied to TB diagnosis in recent years.The sensitivity of artificial neural network (ANN) was the highest at 68.5%. AI- assisted machine learning can predict the treatment outcome of TB to help doctors

11. CONCLUSION

TB detection using CXR images can help doctors and radiologists to make better interpretations of images, leading to a more accurate diagnosis. This system may increase the specificity of an early TB diagnosis. This automated TB-detection system may also act as a second opinion for doctors and radiologists and help them in making the final decision of a diagnosis Support Vector Machine (SVM) models, to enhance tuberculosis (TB) detection through differentiate between TB-positive and TB-negative cases. By implementing image processing methods like histogram equalization and Contrast Limited Adaptive Histogram Equalization (CLAHE), the project aims to improve the clarity of X-ray images, making it easier for the model to differentiate between TB-positive and TB- negative cases

12. REFERENCES

1. Heang Ping Chan et al, "Computer Aided Diagnosis: Concepts and Applications". [Online]. Available: http://www.aapm.org/meetings/04AM/pdf/14- 2366-10736.pdf. [Accessed: Nov. 25, 2017].
2. Vimala Balakrishnan, Yousra Kherabi, Ghayathri Ramanathan, Scott Arjay Paul, Chiong Kian Tiong, "Machine Learning Approaches in Diagnosing Tuberculosis Through Biomarkers", Progress in Biophysics and Molecular Biology, Volume 179, May 2023, Pages 16-25.
3. World Health Organization, "Early Detection of Tuberculosis: An Overview of Approaches, Guidelines, and Tools", WHO Press, Geneva 2011. [Online]. Available: http://apps.who.int/iris/handle/10665/70824. [Accessed: Nov 7, 2017].
4. Khairul Munadi, Kahlil Muchtar, Novi Maulina, Biswajeet Pradhan, "Image Enhancement for TB Detection Using Deep Learning", Page(s): 217897 — 217907 Electronic ISSN: 2169-3536 DOI: 10.1109/ACCESS.2020.3041867.
5. Antani Sameer P.hD, "Automated detection of lung diseases in chest x-rays, a report the Board of Scientific Counselor, April 2015, U.S. National Library of Medicine, Lister Hill National Center for Biomedical Communications, Bethesda [Online]. [Accessed: 28 May, 2018].