IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ENHANCING TUBERCULOSIS DETECTION THROUGH MACHINE LEARNING ON CHEST X-RAY SCANS

[1]Nagireddy Maheswari, [2]Bathina Pavan Kumar, [3]Konthala Hemalatha, [4]Geddam Venkata Aswith, [5]Mrs Kilari Jyothi

[1]Student(20551A0437), [2]Student(20551A0404), [3]Student(21555A0405), [4]Student(20551A0410), [5]Associate Professor

Department of Electronics and Communication Engineering,

Godavari Institute of Engineering and Technology (Autonomous), Rajamahendravaram, Andhra Pradesh, India

Abstract: Tuberculosis (TB), one of today's most deadly diseases, is caused by Mycobacterium tuberculosis and primarily affects the lungs, often exploiting weakened immune systems. TB poses a significant threat, with mortality rates escalating if left undetected. To address this challenge, various computer-assisted diagnostic methods have emerged, leveraging machine learning, particularly deep learning, in image processing. By analyzing chest X-rays, these techniques aim to provide more accurate, timely, and reliable diagnoses. Recent studies suggest that machine learning-based approaches can outperform manual diagnosis, offering superior accuracy. Notably, Digital image processing (DIP) is gaining prominence in biomedical research. Leveraging image processing, Support Vector Machine (SVM) models can effectively classify lung abnormalities indicative of TB. The primary focus of this study is to detect tuberculosis through the implementation of machine learning models trained on chest X-ray images.

Keywords - Mycobacterium Tuberculosis, Digital Image Processing (DIP), Machine Learning, Deep Learning, Support Vector Machine (SVM)

I. INTRODUCTION

Tuberculosis (TB) stands as a formidable global health 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 adeptness 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.

II. LITERATURE REVIEW

Vimala Balakrishnan et al. [1] in 2023 has proposed machine learning approaches in diagnosing tuberculosis through biomarkers. This review examines 19 studies on biomarker-based tuberculosis (TB) diagnosis using machine learning, following PRISMA guidelines. Supervised learning, particularly with Support Vector Machine and Random Forest algorithms, yielded high accuracy (97.0%), sensitivity (99.2%), and specificity (98.0%). Protein-based biomarkers were predominantly explored, with publicly available datasets commonly used. Machine learning shows promise in enhancing TB diagnosis, especially in resource-limited settings, by leveraging biomarkers for faster and more accessible detection compared to traditional methods.

Xuebin Xu et al. [2] in 2022 has proposed a tuberculosis detection method using attention and sparse R-CNN.In this project the algorithm comprises CXTCNet for TB presence judgment, employing a channel attention mechanism within DenseNet, and CXTDNet, a Sparse R-CNN-based design for TB area detection without non-maximal suppression post-processing. Preprocessing involves CLAHE for noise reduction. On TBX11K dataset, CXTCNet achieves 99.10% accuracy, outperforming existing TB classification methods, while overall detection performance yields AP of 45.35% and AP50 of 74.20%. Additionally, a new dataset of 304 chest X-rays demonstrates diagnostic accuracy comparable to radiologists, aiming to advance TB detection efforts.

Marco Beccaria et al. [3] in 2018 has proposed comprehensive gas chromatography -mass spectrometry and chemometric techniques for exhaled human breath analysis in active pulmonary TB diagnosis. This study explores using human breath analysis for diagnosing active TB, offering a non-invasive approach with machine learning algorithms achieving sensitivities of 0.82 and 1.00 and specificities of 0.92 and 0.60 in training and test data respectively, showing promise especially for patients with HIV comorbidities.

Tsung -Ting Tsai et al. [4] in 2017 has proposed diagnosis of tuberculosis using colorimetric gold nanoparticles on a paper -based analytical device. In this project a colorimetric sensing strategy utilizing gold nanoparticles and paper-based platforms was developed for TB diagnosis, leveraging surface plasmon resonance to detect changes in nanoparticle aggregation due to DNA hybridization. This label-free technique, adaptable to resource-limited settings and compatible with smartphone measurement, achieved a detection limit of 1.95×10^{-2} ng/mL for TB DNA with rapid parallel results and low reagent consumption.

Rahul Hooda et al. [5] in 2017 has proposed a potential method for TB detection using chest radiography. The paper proposes a deep learning-based method for TB detection from chest X-ray (CXR) images, achieving a 94.73% overall accuracy with the Adam optimizer, showcasing potential for early diagnosis and disease containment. Evaluation is conducted on Montgomery and Shenzhen datasets, signaling promise for computer-aided diagnosis systems in improving TB detection and treatment outcomes.

Vijaya Lakshmi Arunagiri et al. [6] in 2019 has proposed a deep learning approach to detect malaria from microscopic images. This study presents a novel approach for malaria parasite identification using a transfer learning method that combines VGG network and SVM. By unifying these models, the proposed VGG19-SVM achieves a classification accuracy of 93.1%, outperforming existing CNN models, demonstrating the potential of transfer learning in medical image analysis for malaria diagnosis.

Yan Xiong et al. [7] in 2018 has proposed a method of automatic detection of mycobacterium TB using artificial intelligence. This study developed a convolutional neural network (CNN) called TB-AI to detect acid-fast stained TB bacilli, achieving high sensitivity (97.94%) and specificity (83.65%) compared to diagnoses by human pathologists. TB-AI shows promise as a supportive tool in TB diagnosis, potentially reducing pathologists' workload and improving detection accuracy.

Ruihua Guo et al. [8] in 2020 has developed a model on TB diagnostics and localization in chest x-rays via deep learning models.The research proposes an integrated approach using convolutional neural networks (CNNs) to enhance TB diagnosis in chest X-rays, including model modification, fine-tuning, and ensemble methods, with class activation mapping for localization. This approach demonstrates superior performance in detecting lung abnormalities and diagnosing specific TB-related manifestations, offering potential support for clinicians in TB diagnosis from CXRs.

Saad I Nafisah et al. [9] in 2022 has proposed TB detection in chest radiography using CNN architecture and explainable AI. This project proposes an automatic tuberculosis (TB) detection system utilizing advanced deep learning models, incorporating sophisticated segmentation networks to extract the region of interest from chest X-rays (CXRs) and employing explainable artificial intelligence for visualization. Experimentation reveals that using segmented lung CXR images significantly enhances classification performance, with EfficientNetB3 achieving the highest accuracy of 99.1% among the tested convolutional neural network (CNN) models.

Khairul Munadi et al. [10] in 2020 has developed an imahe enhancement method for TB detection using deeplearning. This study investigates the impact of image enhancement techniques, such as Unsharp Masking, High-Frequency Emphasis Filtering, and Contrast Limited Adaptive Histogram Equalization, on deep learning-based TB diagnosis from chest x-ray images. Employing these enhancements alongside pre-trained models like ResNet and EfficientNet, the study achieved notable classification accuracy (89.92%) and AUC scores (94.8%) on the Shenzhen dataset, offering promising avenues for improving TB diagnostic performance.

III. METHODOLOGY

In this work, we explored the effectiveness of five different models in classifying chest X-ray images as normal or abnormal. These models included SVM, Random Forest, 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. This signifies SVM's remarkable ability to accurately identify abnormal chest X-rays while minimizing false positives. Although Inception V3 displayed a slightly higher precision score, its considerably lower recall score led to a diminished F1-score compared to SVM. While Random Forest, Logistic Regression, and MobileNet also exhibited respectable performances, with accuracy scores of 0.95, 0.78, and 0.91 respectively, SVM consistently outperformed these models across all metrics, including accuracy, precision, recall, and F1-score. Overall, our experimental findings underscore SVM as the most reliable model for classifying chest X-ray images as normal or abnormal. Its superior performance can be attributed to the combined utilization of CLAHE and equalizeHist during image preprocessing, facilitating enhanced contrast and detail within the images.

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

Table 1: Analysis of various models on the dataset

IV. PROPOSED METHOD

The proposed methodology for this study encompasses 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, denoting the likelihood of the input image being TB-positive or TB-negative, with the final result presented to the user. CLAHE, an abbreviation for Contrast Limited Adaptive Histogram Equalization, stands as a variant of histogram equalization designed to bolster local contrast while curbing noise amplification within an image. By segmenting the image into small, overlapping tiles and applying histogram equalization individually to each tile, CLAHE effectively enhances the contrast of local features without impacting the global contrast of the image. The outcome is an image exhibiting enhanced local contrast and reduced noise amplification, facilitating easier visualization and analysis of intricate details within the image. 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 a flowchart, delineating the step-by-step process of the image classification task. It begins with an input X-ray image, proceeds to the pre-processing stage involving CV2 histogram equalization and CLAHE, followed by SVM classification. Finally, the flowchart concludes with the output of prediction probability and result.

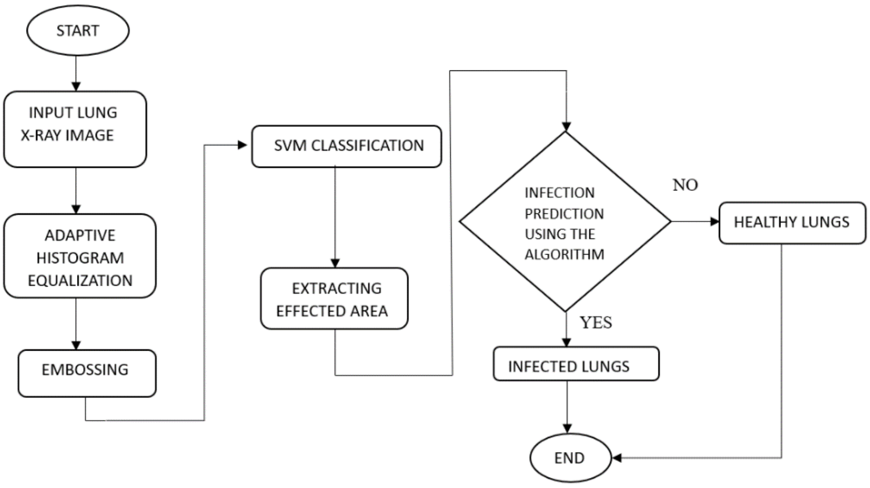

Fig 1: Flow Chart of TB Detection through Machine Learning

V. RESULTS AND DISCUSSION

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 2: Input X-Ray Image

Fig 3: X-Ray after CLAHE

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

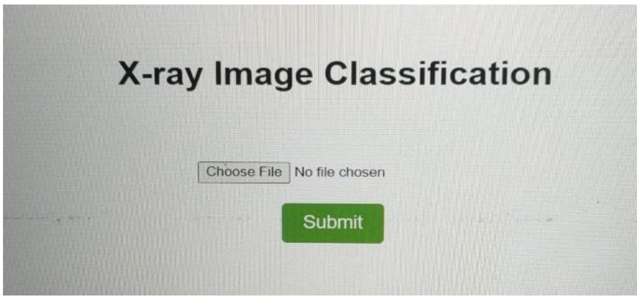

Fig 4: Index Page of Interface

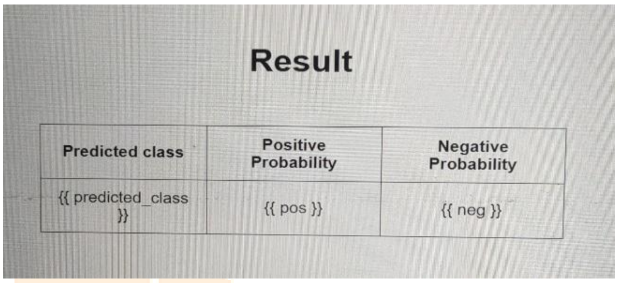

Fig 5: Result

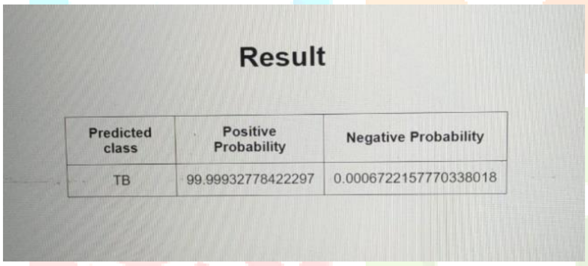

Fig 6: Positive Case of TB

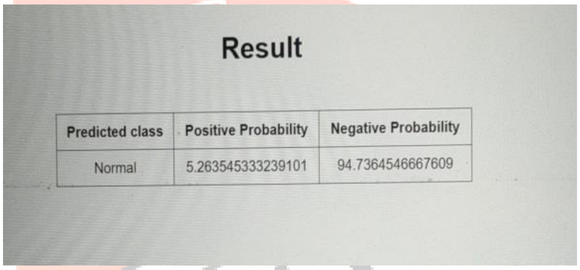

Fig 7: Negative Case of TB

VI. CONCLUSION

This paper underscores the transformative role of machine learning in facilitating early detection and diagnosis of tuberculosis by analyzing chest X-ray images. Through the development and training of machine learning models on a comprehensive dataset comprising TB and non-TB chest X-ray images, the system adeptly classifies new images as positive or negative cases of TB with precision. Particularly, SVM emerges as a fitting and efficient machine learning algorithm for accurately classifying and diagnosing tuberculosis from chest X-ray images, presenting a promising avenue for future endeavors in this field.

VII. FUTURE SCOPE

1) 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.
2) 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.

VIII. REFERENCES

1. 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.
2. Xuebin Xu, Jiada Zhang, Xiaorui Cheng, Longbin Lu, Yuqing Zhao, Zongyu Xu, Zhuangzhuang Gu, "A Tuberculosis Detection Method Using Attention and Sparse R-CNN", Vol. 16, No. 7, July 31, 2022 doi>10.3837/tiis.2022.07.001.
3. Marco Beccaria , Carly Bobak, Boitumelo Maitshotlo, Theodore R Mellors, Giorgia Purcaro, Flavio A Franchina, Christiaan A Rees, Mavra Nasir, Wendy S Stevens, Lesley E Scott, Andrew Black, Jane E Hill, "Exhaled human breath analysis in active pulmonary TB diagnosis by comprehensive gas chromatography -mass spectrometry and chemometric techniques", PMID: 30394364 PMCID: PMC6537098 DOI: 10.1088/1752-7163/aae80e.
4. Tsung-Ting Tsai , Chia-Yu Huang , Chung-An Chen , Shu-Wei Shen , Mei-Chia Wang , Chao-Min Cheng , Chien-Fu Chen, "Diagnosis of Tuberculosis Using Colorimetric Gold Nanoparticles on a Paper-Based Analytical Device", PMID: 28901134 DOI: 10.1021/acssensors.7b00450.
5. Rahul Hooda; Sanjeev Sofat; Simranpreet Kaur; Ajay Mittal; Fabrice Meriaudeau, " A Potential Method for TB Detection using Chest Radiography", 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), DOI: 10.1109/ICSIPA.2017.8120663.
6. Fei Deng, Jibing Huang, Xiaoling Yuan, Chao Cheng, Lanjing Zhang "Performance and Efficiency of Machine Learning Algorithms for Analyzing rectangular biomedical data", Laboratory Investigation volume 101, pages430–441 (2021)
7. Yan Xiong, Xiaojun Ba, Ao Hou, Kaiwen Zhang, Longsen Chen, Ting Li , "Automatic Detection of Mycobacterium TB using Artificial Intelligence", PMCID: PMC5906344 PMID: 29707349 doi: 10.21037/jtd.2018.01.91.
8. Ruihua Guo, Kalpdrum Passi, Chakresh Kumar Jain, "TB Diagnostics and Localization in Chest X-Rays Via Deep Learning Models", PMCID: PMC7861240 PMID: 33733221 doi: 10.3389/frai.2020.583427
9. Saad I Nafisah , Ghulam Muhammad , "TB detection in chest radiography using CNN architecture and explainable AI", PMID: 35462630 PMCID: PMC9016694 DOI: 10.1007/s00521-022-07258-6.
10. 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.