iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50382

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Pulmonary Tuberculosis Detection from Chest X-Ray Images Using Machine Learning

Asst. Prof. Ms. K Sowjanya[1], G. Poojitha[2], Ch. Krishna Saran[3], B. Priyanka[4], D. Ahalya[5]
[1, 2, 3, 4, 5]*Department of the Electronics and Communication Engineering, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru*

Abstract: ***One of today's deadliest diseases, tuberculosis (TB), is caused by "mycobacterium tuberculosis", which usually targets the lungs, due to weakening of immune system. Tuberculosis is very common, and if it is not detected, the patient's risk of death increases over time. Several computer diagnostic methods have been proposed to diagnose tuberculosis based on chest X-rays, as machine learning has been widely used in the field of image processing, especially deep learning. Adapting these machine learning techniques can provide more accurate, timely and reliable diagnostic results. Current research shows that manual diagnosis can be replaced by machine learning-based diagnosis with properly trained models, which provide more accurate results. DIP is becoming more prominent in the field of biomedicine. With image processing, a Support Vector Machine (SVM) learning model can be used to classify disabled lungs. The main objective of this paper is, to diagnose tuberculosis using machine learning trained models with chest x-ray images.***
Keywords: DIP, SVM, Machine Learning, TB.

I. INTRODUCTION

According to the World Health Organization (WHO), Tuberculosis (TB) is a deadly infectious disease and one of the top ten causes of death worldwide, particularly in developing countries. Traditional methods for diagnosing TB, such as microscopy, can be time-consuming leading to delayed or inadequate treatment. To address this issue, machine learning algorithms have emerged as a promising approach to improve TB diagnosis. Machine learning is a subset of artificial intelligence that involves the use of algorithms and statistical models to enable systems to learn and make predictions or decisions based on data. By analyzing large amounts of data, machine learning algorithms can identify trends and predict diseases. Support Vector Machines (SVM) were chosen for this project due to their effectiveness in classifying data points into different groups based on their characteristics. SVM is a supervised learning technique that is a binary classification algorithm that finds the hyperplane with the largest margin of separation between two classes in a high-dimensional feature space. SVM has several advantages, such as being able to handle high-dimensional data, having a robust generalization ability, and being able to handle non-linearly separable data. This project aims to develop a machine learning-based system to recognize TB from chest X-rays using SVM. The research involves image preprocessing techniques, such as adaptive histogram smoothing, to improve contrast and image quality. SVM will then be used to classify the images and accurately differentiate TB from other diseases with similar symptoms. By using machine learning (SVM), this system can improve TB diagnosis and provide a more efficient and accurate approach for diagnosing TB, particularly in resource-poor environments.

II. LITERATURE SURVEY

In four separate studies, the authors have used different imaging sources such as X-Ray, CT scan, and microscopic images, and different machine learning algorithms such as CNN, KNN, and SVM to detect different diseases. For instance, Stefanus Kieu Tao Hwa et al. [1] used CNN algorithm on canny edge detected images resulting in 93.59% accuracy. Satheeshkumar et al. [2] used a CNN algorithm on X-Ray images to detect lung cancer, resulting in a 94% accuracy rate. Similarly, P. PrasannaKumari et al. [3] used a KNN algorithm on CT scan images to detect COVID-19, achieving an 82.5% accuracy rate. Finally, R. Dinesh Jackson Samuel et al. [4] employed an SVM algorithm on microscopic images to detect malaria, obtaining an accuracy rate of 95.05%.

All the studies have reported high accuracy rates in detecting diseases using machine learning algorithms. However, there are some differences in the imaging techniques and algorithms used. Stefanus Kieu Tao Hwa et al. and Satheeshkumar et al. used a CNN algorithm, which is a deep learning algorithm known for its ability to learn complex patterns in data. P. PrasannaKumari et al. used a KNN algorithm, which is a non-parametric algorithm used for classification tasks. R. Dinesh Jackson Samuel et al. used an SVM algorithm, which is a supervised learning algorithm used for classification and regression analysis.

©IJRASET: All Rights are Reserved

III. METHODOLOGY

During the course of the project, four different models were used to classify chest X-ray images as either normal or abnormal. The four models were *SVM1, SVM2, ResNet50, and VGG16. All of the models utilized a combination of image processing techniques such as CLAHE to enhance image quality. SVM1, however, was the only model that also used the equalizeHist function in the image pre-processing step. After training and testing all four models, we found that SVM1 was the most accurate and consistent model for this particular classification task, with an accuracy score of 0.953. SVM1 also had high precision and recall scores of 0.953 and 0.950, respectively, resulting in an F1-score of 0.949. This means that SVM1 was able to correctly identify abnormal chest X-rays with high accuracy, while also minimizing the number of false positives. While SVM2 had a slightly higher precision score, its recall score was significantly lower, resulting in a lower F1-score compared to SVM1. ResNet50 and VGG16 also performed relatively well, with accuracy scores of 0.900 and 0.925, respectively. However, SVM1 still outperformed both models in terms of accuracy, precision, recall, and F1-score. Overall, the results of my experiments indicate that SVM1 is the most reliable model for classifying chest X-ray images as normal or abnormal. Its superior performance can be attributed to the use of both CLAHE and equalizeHist in the image pre-processing step, which allowed for better contrast and detail enhancement in the images.

(*SVM1- SVM classifier using normal histogram equalization and CLAHE together in pre-processing SVM2- Only CLAHE in pre-processing)

TABLE I
Analysis of various models on the dataset

Model	Accuracy	Precision	Recall	F-1
SVM1	0.950	0.963	0.93	0.95
SVM2	0.914	0.953	0.87	0.91
ResNet50	0.900	0.84	0.98	0.90
VGG 16	0.925	0.93	0.93	0.92

1) *Precision:* measures the proportion of true positives (correctly predicted positive instances) among all predicted positive instances. A high precision score indicates that the model has a low rate of false positives.
2) *Accuracy:* measures the proportion of correct predictions over the total number of predictions made. It provides an overall measure of the model's performance, taking into account both true positives and true negatives.
3) *Recall:* A recall score is a metric that determines the ratio of correctly identified positive instances out of all actual positive instances. A high recall score implies that the model has a minimal occurrence of falsely labelled negative instances.
4) *F1 score:* is the harmonic mean of precision and recall, and is often used as a balanced measure that takes into account both precision and recall. A high F1 score indicates that the model is performing well in terms of both precision and recall.
5) *Dataset:* A group of scientists, including researchers from Qatar University and the University of Dhaka, Bangladesh, and their partners from Malaysia, have collaborated with medical professionals from Hamad Medical Corporation and Bangladesh to construct a database of chest X-ray images comprising TB-positive cases as well as normal images.[5][9]

IV. PROPOSED METHOD

The proposed method for this study involves four main steps. First, an X-ray image is chosen as input. Second, the image undergoes pre-processing using OpenCV's histogram equalization and Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance image quality. Third, the pre-processed image is classified using a Support Vector Machine (SVM) classifier model, which has been trained on a dataset of X-ray images of both TB positive and negative patients. Finally, the output of the SVM classifier model is the prediction probability, which indicates the likelihood of the input image being TB positive or negative, and the final result is displayed to the user.

©IJRASET: All Rights are Reserved

CLAHE: Contrast Limited Adaptive Histogram Equalization, is a variant of histogram equalization that improves local contrast while limiting the amplification of noise in an image. CLAHE divides the image into small, overlapping tiles and applies histogram equalization to each tile separately. This allows CLAHE to enhance the contrast of small, local features without affecting the global contrast of the image The result is an image with improved local contrast and reduced noise amplification, making it easier to visualize and analyse small details in the image. The proposed method aims to accurately detect TB in X-ray images, with a focus on enhancing the accuracy and reliability of the detection process. It was implemented using python code. The proposed model produces an accuracy of 95.0% with 96.3% precision.

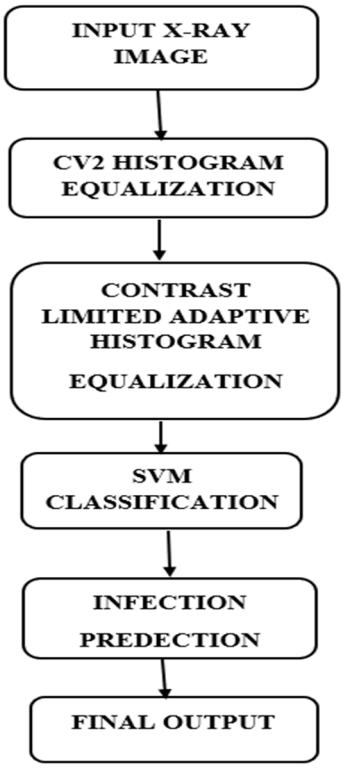


Fig 1: Flowchart

The proposed method can be visualized through a flowchart that outlines the step-by-step process of the image classification task. The flowchart starts with the input of an X-ray image, followed by the pre-processing stage(cv2 histogram equalization and CLAHE), SVM classification, and finally, output of the prediction probability and result.

V. RESULTS

The uploaded image with the TB classification results shows whether the given chest X-ray image is positive or negative for TB.

A. X-ray Images from the dataset are taken and histogram equalization is performed in pre-processing stage



Fig 2**:** Input X-ray



Fig 3**:** X-ray after CLAHE

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. The X-ray images are uploaded in the index page of interface

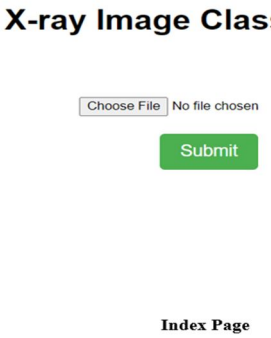


Fig 2: Index page of Interface

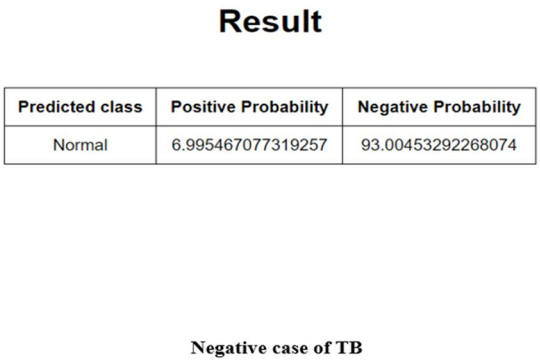


Fig 3: Negative Case Result

Additionally, the probabilities for both classes (positive and negative) are also displayed. These results provide important information for medical professionals to make informed decisions about the diagnosis and treatment of TB. The probabilities can also help in determining the confidence level of the classification model's prediction.

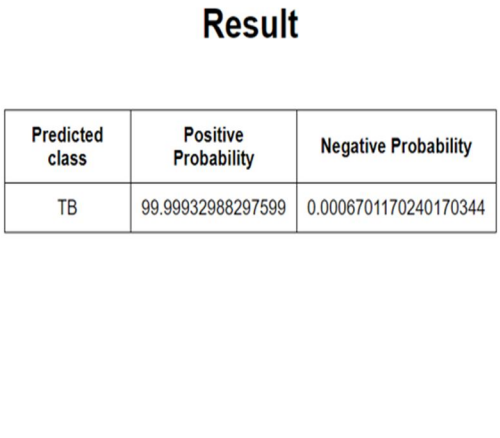


Fig 4: Positive Case Result

VI. CONCLUSION

In conclusion, this project has demonstrated the potential of machine learning to aid in the early detection and diagnosis of tuberculosis through the analysis of chest X-ray images. By developing and training machine learning models on a dataset of TB and non-TB chest X-ray images, the system is able to classify new images as either positive or negative cases of TB accurately. SVM has shown to be a suitable and effective machine learning algorithm for accurately classifying and diagnosing tuberculosis from chest X-ray images, making it a promising approach for this project.

VII. FUTURE SCOPE

1) *Expand The Dataset:* Although the current dataset used in this project is quite large, adding more diverse and varied data can improve the accuracy of the model even further.
2) *Integration with Electronic Health Records (EHRs):* Integrating the model with EHRs can help healthcare professionals in decision-making and can provide a more comprehensive patient history for medical diagnosis.

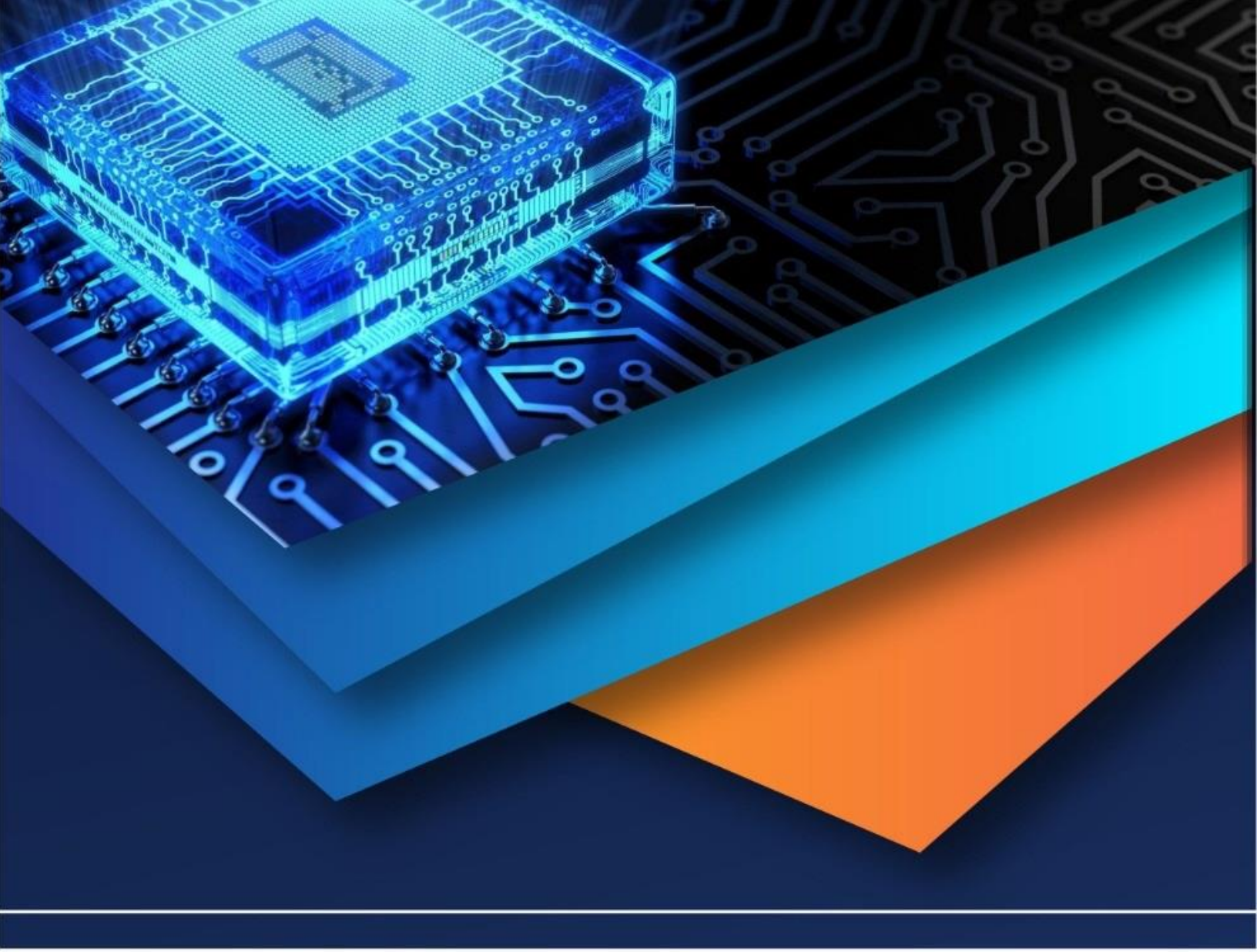
VIII. ACKNOWLEDGMENT

We extend our heartfelt gratitude and appreciation to Dr. Y. Ramakrishna, Professor and Head of the Department of Electronics and Communication Engineering, for his valuable support and encouragement throughout the successful completion of our project.

REFERENCES

[1] Hwa, Stefanus & Bade, Abdullah & Hijazi, Mohd & Jeffree, Mohammad. (2020). “Tuberculosis detection using deep learning & X-Ray images”. International Journal of Artificial Intelligence (IJ-AI). 9
[2] K. G. Satheeshkumar, V Arunachalam, “Automatic Detection of Tuberculosis from Chest X-Rays using Convolutionary Neural Networks,” 2020 International Journal of Engineering & Advanced Technologies (IJEAT), vol 9, pp. 72-77. 2020. doi:10.35940/ijeat.E9292.069520.
[3] P. PrasannaKumari, B. Prabhakara Rao, “Abnormality Detection of TB using Instance Learned Classifier on Lung CT Images,” 2019 International Journal of Recent Technology and Engineering (IJRTE), vol 8, pp. 979-982. doi:10.35940/ijrte.B1163.0982S1119.
[4] R. Dinesh Jackson Samuel, “Tuberculosis detection system using deep neural networks”. Neural Computing & Applications, 2019, vol 31 Issue 5, p1533–1545.
[5] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, " Two chest Xray image datasets for computer supported screening of pulmonary diseases," Quantitative imaging in medicine & surgery, vol. 4 (6), p. 475(2014)
[6] B. P. Health. (2020). BELARUS TUBERCULOSIS PORTAL http://tuberculosis.by/.
[7] NIAID TB portal program dataset https://data.tbportals.niaid.nih.gov/.
[8] kaggle. RSNA Pneumonia Detection Challenge https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data.
[9] Tuberculosis (TB) Chest X-ray Database https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)