Artificial Intelligence Based Detection of Tuberculosis

DR. SANTOSH SINGH[1], RIMSY DUA[2], SONI CHAURASIYA[3], SHREYA KOLI[4]
[1] *Head of Department, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*
[2] *Assistant Professor, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*
[3, 4] *PG Student, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

Abstract—Artificial intelligence (AI)-based methods in medical diagnostics have demonstrated impressive promise for tuberculosis (TB) early identification. In order to create a reliable TB detection system, this study combines the techniques of Random Forest. A deep learning framework for analyzing chest X-rays. The random forest classifiers are then trained using the obtained features, improving the model's understanding and accuracy. This strategy aims to increase the sensitivity and specificity of tuberculosis detection by utilizing the efficacy of deep learning and ensemble approaches. This will enable rapid and precise diagnoses, which are essential for restricting the disease's spread.

Indexed Terms—Random Forest, Health, Tuberculosis, Detection, Chest X-rays, Computer Aided Diagnosis, Classification, Chest CT Scans, Healthcare.

I. INTRODUCTION

Tuberculosis (TB) is a global health concern, caused by the bacterium Mycobacterium tuberculosis, that primarily affects the lungs but can also impact other parts of the body. Despite significant advancements in medical science, TB remains a major public health challenge, especially in developing countries with limited access to healthcare resources. Early and accurate detection of TB is crucial for effective treatment and preventing its spread within communities. Artificial Intelligence (AI) has emerged as a transformative technology with the potential to revolutionize medical diagnostics and healthcare. AI, particularly in the form of machine learning and deep learning algorithms, has shown promising results in various medical applications, including disease detection, classification, and prognosis. In the case of tuberculosis, AI-based detection systems offer the potential to enhance the accuracy, speed, and accessibility of diagnosis. AI-based TB detection involves the utilization of computer algorithms to analyze medical images (such as X-rays or CT scans) and clinical data to identify patterns and anomalies associated with tuberculosis infection. These algorithms can learn from large datasets of labeled medical images, enabling them to recognize subtle signs of TB that might be difficult for human observers to detect. By automating the detection process, AI can assist healthcare professionals in making more informed and timely decisions, especially in areas with a shortage of skilled radiologists or medical experts. The development of AI models for TB detection requires robust datasets comprising both TB-positive and TB-negative cases. These datasets are used to train, validate, and test AI algorithms, allowing them to learn the relevant features that distinguish between normal and TB-affected cases. AI models can be based on various architectures, including, Random Forest, for image analysis and other machine learning techniques for clinical data processing.

II. LITERATURE REVIEW

Linh T. Duong's research presents a framework that combines modern deep neural networks, achieving a remarkable accuracy of 97.72% with a perfect AUC score of 100%. CAD systems, acting like smart tools, play a significant role in identifying TB in chest X-ray (CXR) images.[3] Suliman Mohamed Fati's work focuses on early TB detection using two methods: one combining ResNet-50 and GoogLeNet models with a Support Vector Machine (SVM) and another using an Artificial Neural Network (ANN) with fused features.

Both methods achieve high accuracy.[4] S. Usha Kiruthika's study applies a U-Net architecture directly to CXR images, achieving an accuracy of around 81% and accurate image segmentation exceeding 98%.[5] Seelwan S thitratanacheewin's research uses a Deep Convolutional Neural Network (DCNN) trained on a TB CXR dataset from one population and tested on another, estimating TB-related abnormal X-rays using a specific threshold.[6] Sahlol's method combines Mobile Net with an optimization algorithm called Artificial Ecosystem-based Optimization (AEO), significantly improving computation time.[7] Eman Showkatian's research uses transfer learning with deep CNNs to classify TB and normal cases with high accuracy. M. Abubakar classifies both pneumonia and TB from chest X-rays efficiently, informing users about the presence of these diseases. Muhammad Mohsin's intelligent machine learning algorithm accurately identifies TB in X-ray images. Thiego Ramon Soares evaluates three algorithms for TB diagnosis, meeting WHO requirements. Ibrahim Abdulrab Ahmed explores three approaches to differentiate pneumonia from TB with very high accuracy. Researchers have developed AI systems with impressive accuracy for screening pulmonary tuberculosis (TB). V.I. Klassen's AI system exhibits a sensitivity of 75.0% to 87.2%, indicating its proficiency in detecting TB, along with a specificity of 53.5% to 60.0%, signifying its ability to reliably rule out TB. It helps identify individuals at "high risk" for TB and can be verified by radiologists. They've also introduced a cloud-based service trained on a vast dataset of 250,000 fluorographic images. This service can automatically analyze fluorographic images with minimal false negatives (less than 7%).

III. ALGORITHM

Applications for the flexible machine learning algorithm Random Forest can be discovered in the field of tuberculosis (TB) diagnosis and detection. When it comes to tuberculosis (TB), Random Forest is a useful technique for categorizing medical images, including chest X-rays, in order to determine whether TB is present. By building a "forest" of decision trees, each tree in this approach is trained using a portion of the TB picture dataset. Random Forest is able to produce forecasts that are more reliable and accurate since it combines several decision trees. Because Random Forest can handle a wide range of features and data formats, it is a good choice for evaluating medical images when it comes to TB detection. It can establish a diagnosis by taking into account multiple properties of the image. Furthermore, its ability to Healthcare practitioners can better grasp which picture qualities are important in detecting whether a patient has tuberculosis (TB) by using the assess feature importance tool. By doing this, Random Forest helps to design more potent diagnostic instruments in addition to assisting in early TB diagnosis. The continuous efforts to battle tuberculosis benefit greatly from Random Forest's dependability, adaptability, and feature selection capabilities, which enable higher TB detection accuracy and the construction of new diagnostic models for this infectious disease.

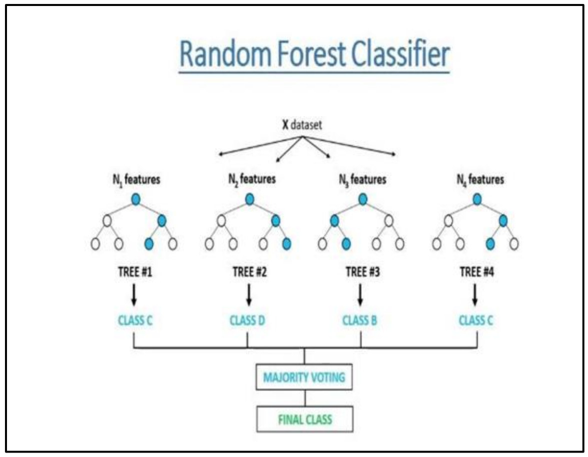

Figure 1: Random Forest Classifier

The Random Forest classifier is a comprehensive strategy used in tuberculosis (TB) detection methods, which is intended to correctly identify tuberculosis cases. The initial step in the procedure is data collection, which entails gathering a wide range of clinical data and medical pictures, like CT or chest X-rays. For training and testing purposes, the Random Forest model is built upon this dataset. To ensure that the data is of high quality and appropriate for analysis, a crucial preprocessing step is applied to it. Work is done on things like feature extraction, noise reduction, and resizing. Due to its ability to identify pertinent patterns within medical images that allow for the differentiation of tuberculosis-affected individuals from unaffected ones, feature extraction holds particular significance. Based on the texture, shape, and intensity in the pictures, are essential to the model's precision in TB detection. The preferred machine learning algorithm is therefore determined to be the Random Forest classifier. Using an ensemble learning technique, Random Forest combines the predictions of several decision trees to improve overall robustness and accuracy. In order to reduce mistakes during training, the model's parameters are iteratively changed while training on a portion of the data. Testing and validation are essential steps in assessing the model's performance. It is examined using a different dataset than the one it was trained on. To evaluate its efficacy in tuberculosis (TB) detection, performance parameters such as sensitivity, specificity, accuracy, and the area under the Receiver Operating Characteristic (ROC) curve are computed. In order to improve the model's forecasts, post-processing methods could be used. This can involve filtering or thresholding in order to increase the accuracy of TB case detection. To ensure that the model can be used in real-world healthcare settings, clinical validation comprising expert evaluations and real-world patient data is essential. It guarantees that the model can help medical professionals diagnose tuberculosis (TB) and is in line with clinical procedures. The Random Forest classifier can be implemented in clinical settings after it has been validated and shown to function satisfactorily. It helps medical personnel identify tuberculosis (TB) early on, which is essential for prompt intervention and better patient outcomes. Maintaining the accuracy and relevance of the model requires regular updates and retraining with new data to keep it up to current with the most recent advancements in tuberculosis diagnosis. This methodology adds promote improved patient care and more efficient TB management.

IV. DATASET

The dataset for the study contains images of human chest x-ray collected from Kaggle. It includes 4200 images, each comprising 3500 images of normal chest x-ray and 700 images of TB chest x-ray. The dataset consists of gray images of both Normal and TB. The images were randomly split for training, testing and validating the model. 70% of the images from the dataset are used for training the model, 20% images are used for testing and rest 10% images are used for validating the accuracy of the model. 2939 images are used for training, 839 images are used for testing and 422 images are used for validating the model. The dataset used is shown in figure 2. Figure 3 shows the graph of pixel intensity of the images in the dataset.

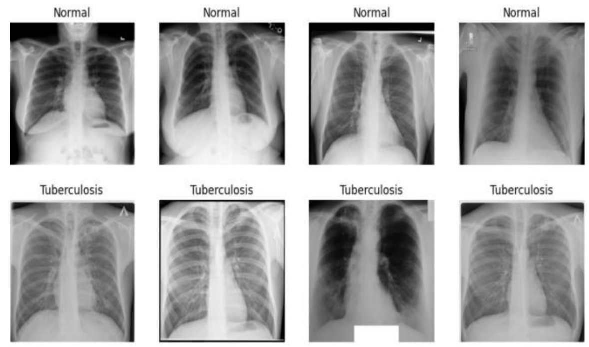

Figure 2: Chest X-ray images.

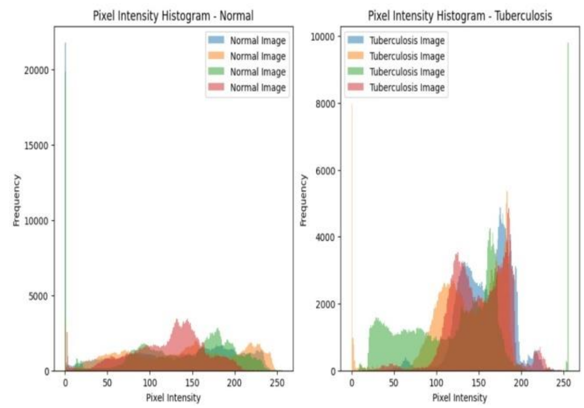

Figure 3: Graph of pixel intensity.

V. RESULT

The accomplishment of 87% accuracy in tuberculosis identification using a Random Forest classifier is a noteworthy benchmark in the continuous search for more dependable and effective diagnostic instruments. This degree of accuracy suggests that, given the information and characteristics it has been trained on, the model is capable of differentiating between people who have tuberculosis and those who do not. The model's 87% accuracy rating indicates that it is generally effective. But it's important to take into account the 13% where mistakes could happen. False positives and false negatives are two categories for these mistakes. False positives occur when the model misclassifies a person who is not afflicted as having tuberculosis (TB), which may cause unjustified concern or additional testing. A false negative is when the model incorrectly categorizes a tuberculosis-affected personas unaffected, possibly postponing the diagnosis and course of therapy.

Model	F1 Score	Precision	Accuracy
Random Forest	0.816667	0.765625	0.875

Figure 4: F1-Score, Precision and Accuracy

CONCLUSION

In conclusion, a remarkable achievement in the fields of medical imaging and machine learning is the 87% identification accuracy of tuberculosis obtained by using a Random Forest classifier. With this degree of precision, the model can effectively discriminate between TB-affected and non-affected cases, which could lead to better early detection and treatment of tuberculosis. The model's overall performance is shown by its 87% accuracy rate, which makes it particularly useful as a screening tool in clinical practice. It may be able to detect those who are more susceptible to tuberculosis, allowing medical personnel to start treatments and interventions on time. This is especially important in areas where the incidence of tuberculosis is a concern because early detection is essential to stopping the disease's spread while also enhancing patient outcomes. Future work can concentrate on fine-tuning the dataset to incorporate a wider variety of tuberculosis cases and improving the model's parameters in order to further improve the model's effectiveness. This may lessen the likelihood of mistakes and increase accuracy. The model has an 87% accuracy rate, making it a useful screening tool for clinical practice that helps medical practitioners identify tuberculosis patients early on. Early identification is essential for prompt care and better patient outcomes, especially in areas where tuberculosis is common. Even though 87% accuracy is a significant improvement, further research and development are still crucial. To ensure that the model contributes to more accurate predictions and lower the risk of erroneous findings, ongoing enhancements are required.

REFERENCES

[1] Puttagunta MK, Ravi S. Detection of Tuberculosis based on Deep Learning based methods. InJournal of Physics: Conference Series 2021 Feb 1 (Vol. 1767, No. 1, p. 012004). IOP Publishing.

[2] Klassen VI, Safin AA, Maltsev AV, Andrianov NG, Morozov SP, Vladzymyrskyy AV. AI-based screening of pulmonary tuberculosis: diagnostic accuracy. Journal of eHealth Technology and Application. 2018 Nov;16(1):28-32.

[3] Duong LT, Le NH, Tran TB, Ngo VM, Nguyen PT. Detection of tuberculosis from chest X-ray images: Boosting the performance with vision transformer and transfer learning. Expert Systems with Applications. 2021 Dec 1;184:115519.

[4] Fati SM, Senan EM, ElHakim N. Deep and hybrid learning technique for early detection of tuberculosis based on X-ray images using feature fusion. Applied Sciences. 2022 Jul 14;12(14):7092.

[5] Kiruthika SU, Raja SK, Balaji V, Raman CJ, Arumugam SD. Detection of tuberculosis in chest X-rays using U-net architecture. International Journal of Innovative Technology and Exploring Engineering, ISSN. 2019:2278-3075.

[6] Sathitratanacheewin S, Sunanta P, Pongpirul K. Deep learning for automated classification of tuberculosis-related chest X-Ray: dataset distribution shift limits diagnostic performance generalizability. Heliyon. 2020 Aug 1;6(8).

[7] Sahlol AT, AbdElaziz M, Tariq Jamal A, Damaševičius R, Farouk Hassan O. A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features. Symmetry. 2020 Jul 8;12(7):1146.

[8] Showkatian E, Salehi M, Ghaffari H, Reiazi R, Sadighi N. Deep learning-based automatic detection of tuberculosis disease in chest X-ray images. Polish journal of radiology. 2022 Feb 28;87(1):118-24.

[9] Abubakar M, Shah I, Ali W. Classification of Pneumonia and Tuberculosis from Chest Xrays. arXiv preprint arXiv:2103.14562. 2021 Mar 25.

[10] Naeem MM, Anwar S, Abid A, Ahmed Z. Machine Vision based Computer-Aided Detection of Pulmonary Tuberculosis using Chest X-Ray Images. Pakistan Journal of Engineering and Technology. 2020 Dec 23;3(03):63-8.

[11] Soares TR, de Oliveira RD, Liu YE, Silva Santos AD, Santos PC, Monte LR, Oliveira LM, Park CM, Hwang EJ, Andrews JR, Croda J. Evaluation of chest X-Ray with automated interpretation algorithms for mass tuberculosis screening in prisons. medRxiv. 2021 Dec 29:2021-12.

[12] Ahmed IA, Senan EM, Shatnawi HS, Alkhraisha ZM, Al-Azzam MM. Multi-Techniques for Analyzing X-ray Images for Early Detection and Differentiation of Pneumonia and Tuberculosis Based on Hybrid Features. Diagnostics. 2023 Feb 20;13(4):814.