AI-Assisted Predictive Model for Tuberculosis Disease

Comfort Oyindamola Akanni
Department of Medicine and Surgery
Olabisi Onabanjo University
Oyindamolaakanni2@gmail.com | +2348054802423

Abstract

Tuberculosis (TB) is an infectious disease caused by the bacteria named Mycobacterium tuberculosis. It is of a major public health concern, causing millions of fatalities globally each year primarily in developing nations and among individuals with poor socioeconomic status. Effective disease management depends on early diagnosis and prompt treatment. This study is aimed to build an Artificial intelligence (AI)- assisted model that can be used for the prediction of TB based on radiological data from patients suspected to have TB with the aid of machine learning algorithms. This research also examines the potential of AI-assisted models in detecting TB at a level that will overcome the radiologist limitations. In addition, the possibility of AI-assisted predictive analysis method in improving the accuracy of TB diagnosis will be discussed. This study dataset contains 4200 Chest X-Ray (CXR) images of both TB infested and non-infected patients. In building this model, we will employ various transfer learning methods, including VGG16, VGG19, ResNet50, InceptionV3, EfficientNetB7, DenseNet201, MobileNet, Xception, AlexNet, and NASNet. These methods will be utilised to extract essential features for optimising the classification of TB within Convolutional Neural Networks (CNNs). The model was assessed based on Accuracy, precision, sensitivity, and F1-Score metrics. The result obtained indicates that the VGG16 model achieved the best result based on accuracy and recall values of 99.6% and 100% respectively.Conclusively, this study shows that AI may be used to predict TB, and it also emphasises the importance of feature selection and data preprocessing for model performance.

1. Introduction

Tuberculosis (TB) is a chronic pulmonary and systemic granulomatous infectious diseases caused most often by the bacteria named Mycobacterium tuberculosis (Natarajan et al., 2020). TB commonly affects the lungs ("pulmonary TB") and in some cases, TB affects other organs, and this is known as "extra pulmonary TB" (Olfa et al., 2022). TB is classified as symptomatic "active disease" and asymptomatic "latent infection" (Kumar et al., 2021, pp. 368). The symptoms of active TB are fever, weight loss, anorexia, drenching night sweat, chronic cough with or without haemoptysis. TB is transmitted by infected individuals through the sputum. Tuberculosis is a public health problem that is prevalent in underdeveloped countries and among people of low socioeconomic status (Elmi et al., 2014). Other risk factors of TB are immunosuppressive conditions (e.g use of steroids, HIV, Diabetes, malignancy), malnutrition, consuming unpasteurised milk, chronic alcoholism, overcrowding, extreme ages, and lack of vaccination. Furthermore, TB is ranked after HIV/AIDS as the second cause of death by infectious disease (Meraj et al., 2019). According to WHO (cited in Kumar et al., 2021, pp. 368)), in 2018, approximately 10,000,000 people are infected with TB globally with an estimation of 300,000 death in people with HIV and 1,300,000 deaths in people without HIV. The early diagnosis of Tuberculosis has been of great concern in the past because the TB diagnostic tests are either time consuming (e.g polymerase chain reaction) or they give a false positive result (e.g TB skin test) (Meraj et al., 2019). Therefore, several research have been done to overcome these short comings.

Machine learning is a subset of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computer systems to learn from and make predictions or decisions based on input data, without being explicitly programmed to do so (Olfa et al., 2022).

An AI-assisted predictive model is a machine learning model that uses artificial intelligence (AI) techniques to improve the accuracy and efficiency of its predictions. The model uses statistical algorithms and machine learning techniques to analyse large datasets and identify patterns that can be used to predict future outcomes. The "AI-assisted" aspect of the model refers to the use of artificial intelligence techniques, such as deep learning and natural language processing, to enhance the model's predictive capabilities.

In this research, Convolutional Neural Network (CNN) will serve as a backbone architecture used in building the computer-Aided Diagnostic (CAD) systems to analyse the CXR. This is because they produce a better performance in recognising the problem pattern (Murali and Ravi, 2021).

CNN and CAD are related but they serve different purposes in the context of medical image analysis and diagnosis. A CNN is a type of deep learning neural network architecture designed for analysing and processing visual data, such as images. It is specifically designed to automatically extract meaningful features and patterns from images to enable tasks like image classification, object detection, and segmentation. CAD systems, on the other hand, are software systems that assist healthcare professionals in medical diagnosis by analysing medical images using algorithms and providing decision support. CAD systems utilise various techniques, including image processing, pattern recognition, and machine learning, to detect and interpret abnormalities or specific features within medical images. Therefore, it enhances the accuracy and efficiency of the diagnosis of TB by detecting and interpreting abnormalities or specific features of interest within the images. (Shiraishi et al., 2011). The CNN will be responsible for the automated feature extraction from medical images, allowing CAD systems to detect and analyse specific patterns or abnormalities. Overall, this will assist the radiologists and doctors to detect simple and even more complicated cases of TB after analysing the patients CXR films (Vasundhara et al., 2022).

2. Methodology

DATA COLLECTION AND PREPROCESSING: The dataset used for training the model in this research is a publicly available dataset collected by Tawsifur et al., 2020. It contains 3500 CXRs of Healthy people without TB and 700 CXRs of people infected with TB of both male and female ages 18 and above. The patient's privacy was protected to meet up with the ethical considerations. This CXR images was then preprocessed to enhance their quality and extract relevant features. The CXRs were resized and normalised to standardise the images for analysis. Data resampling was done by oversampling the TB CXRs and transfer learning was done using image-net in-order to address the issue of imbalance dataset. It was then split into training, validation and test dataset at the ratio of 70:15:15.

MODEL SELECTION AND TRAINING: Deep learning models used to select and train the data are the Convolutional Neural Networks (CNNs), Visual Geometry Group 16 (VGG16), Visual Geometry Group 19 (VGG19), Residual Network 50 (ResNet50), Inception Version 3 (InceptionV3), EfficientNet Variant B3 (EfficientNetB3) Dense Convolutional Network 201 (DenseNet201), MobileNet, Xception, AlexNet and NASNet using 5 epochs and Adam optimiser.

MODEL EVALUATION: After training, the model was then evaluated using the validation set to assess its performance and identify any area that needs improvement. The metrics used to evaluate the TB prediction models include accuracy, precision, sensitivity, specificity and F1 score.

MODEL DEPLOYMENT: The trained model will then be deployed in an environment that facilitates user interaction where it can be used to predict the likelihood of TB infection on a new patient CXR. This environment will be a mobile application after choosing a framework and converting the model. The interface engine will be implemented and the performance optimised, tested, deployed and maintained while implementing encryption and and security measures to protect the user data.

TUBERCULOSIS DETECTION PROCESS: On the mobile app, the user interface will ask the users to input the symptoms they are feeling. Based on their symptoms tallying with that of TB, it then directs them to a radiologist to have a CXR done. The CXR image is then uploaded. This image is then preprocessed to align with the model's requirements. This image will then be fed into the deployed models for inference so the model will predict the likelihood of TB infection based on the learned patterns.

3. **Experimental result:** The comparative performance of all the 10 CNN models used in this AI- Assisted predictive analysis of TB is given in Fig 1 below. It is seen that VGG16 outperformed all other models by giving accuracy, precision, sensitivity and F1 Score of 99.6%, 100%, 100%, and 98.9% respectively, while AlexNet model performed poorly in the analysis.

4. **Discussion:** The successful application of AI-assisted predictive analysis using VGG16 in tuberculosis detection holds significant implications for medical diagnosis and treatment. The high accuracy, precision, sensitivity, and F1 Score achieved by VGG16 provide promising prospects for automating the screening and early detection of tuberculosis, which can aid in timely interventions and improved patient outcomes.

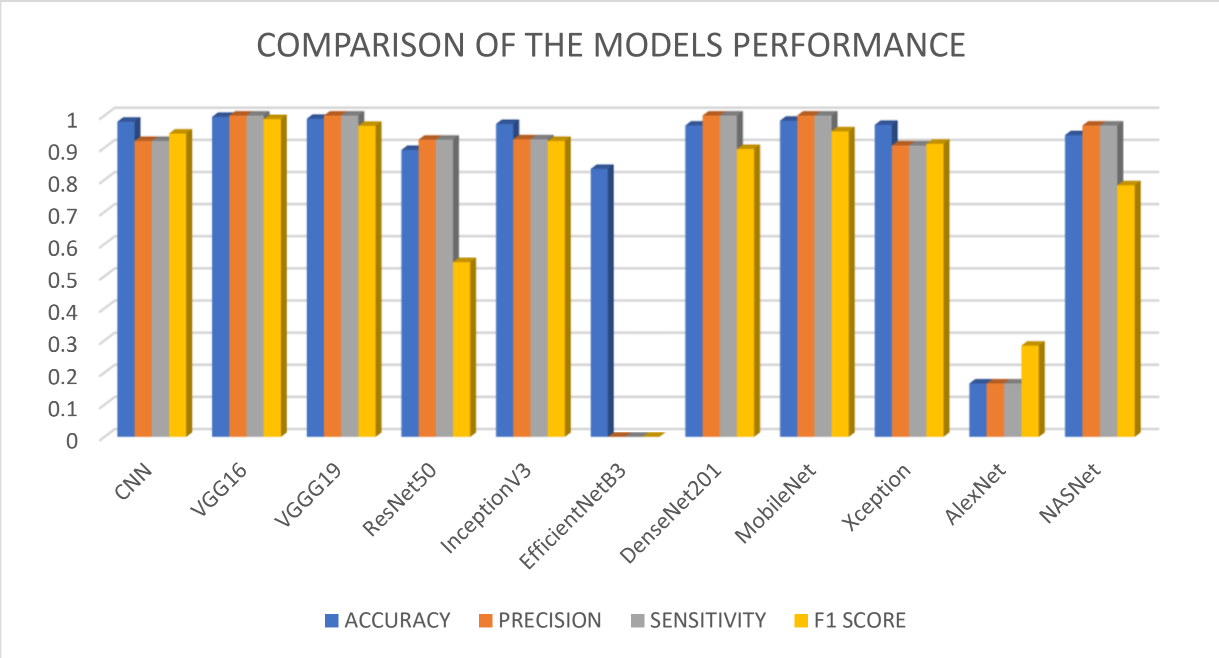

Figure 1: Chart comparing the models performance.

While the use of pre-existing datasets has its limitations, this research underscores the potential of AI-assisted approaches in medical image analysis. Further research could focus on refining the model's performance, addressing the challenges of limited data availability, and exploring alternative approaches to enhance the accuracy and efficiency of tuberculosis detection using chest X-ray images. In conclusion, this research on AI-assisted predictive analysis of tuberculosis using chest X-ray images demonstrates the effectiveness of VGG16 in accurately classifying tuberculosis cases. The findings contribute to the growing body of knowledge in the field of medical image analysis and pave the way for advancements in automated tuberculosis screening. Continued research in this area holds great promise for improving early detection and treatment outcomes in tuberculosis patients, ultimately contributing to the global efforts in combating this infectious disease.

Ethical considerations

The patient's privacy was protected by ensuring that no personal information was revealed. Also, the data was handled with utmost respect and privacy in adherence of the terms and conditions of the data provider. Informed consent is not applicable and there is no conflict of interest.

Limitations

The limitations of this research are clinical validation, generalisation and resource constraints.

Future considerations

Collaborate with a pulmonologist and a radiologist to get CXRs of patients from various healthcare settings to use for external validation of the model. Furthermore, there will be collaboration with a mobile app developer to help deploying the model on a mobile application.

Acknowledgement
The author of this paper appreciates Dr (Mrs) Sakinat Folorunso for her mentorship.

References

Akhter Y, Singh R and Vatsa M (2023) AI-based radiodiagnosis using chest X-rays: A review. Front. Big Data6:1120989. .doi: 10.3389/fdata.2023.1120989

Elmi OS, Hasan H, Abdullah S, Jeab MZM, Nadiah WA, Zilfalil BA, et al. Development and validation of a questionnaire on the knowledge of tuberculosis and the perception of tuberculosis treatment among tuberculosis patients in Malaysia. Int J Med Sci Public Health 2014;3:352-357

Kumar, V., Abbas, A. and Aster, J. (2021) Robbins and Cotran Pathologic Basis of Diseases. 10th Edition, Elsevier, Amsterdam.

Meraj, Syeda & Yaakob, Razali & Azman, Azreen & Rum, Siti & Nazri, Azree. (2019). Artificial Intelligence in Diagnosing Tuberculosis: A Review. International Journal on Advanced Science, Engineering and Information Technology. 9. 81. 10.18517/ijaspit.9.1.7567.

Murali Krishna Puttagunta and S. Ravi 2021 J. Phys.: Conf. Ser. 1767 012004

Natarajan A et al., A systemic review on tuberculosis, Indian Journal of Tuberculosis, https://doi.org/10.1016/ j.ijtb.2020.02.005

Olfa Hrizi, Karim Gasmi, Ibtihel Ben Ltaifa, Hamoud Alshammari, Hanen Karamti, Moez Krichen, Lassaad Ben Ammar, Mahmood A. Mahmood, "Tuberculosis Disease Diagnosis Based on an Optimized Machine Learning Model", Journal of Healthcare Engineering, vol. 2022, Article ID 8950243, 13 pages, 2022. https://doi.org/10.1155/2022/8950243

Shiraishi J, Li Q, Appelbaum D, Doi K. Computer-aided diagnosis and artificial intelligence in clinical imaging. Seminars in Nuclear Medicine. 2011; 41(6): 449–462. DOI: https://doi.org/10.1053/j. semnuclmed.2011.06.004

Tawsifur Rahman, Amith Khandakar, Muhammad A. Kadir, Khandaker R. Islam, Khandaker F. Islam, Zaid B. Mahbub, Mohamed Arselene Ayari, Muhammad E. H. Chowdhury. (2020) "Reliable Tuberculosis Detection using Chest X-ray with Deep Learning, Segmentation and Visualization". IEEE Access, Vol. 8, pp 191586 - 191601. DOI. 10.1109/ACCESS.2020.3031384.

Vasundhara Acharya, Gaurav Dhiman, Krishna Prakasha, Pranshu Bahadur, Ankit Choraria, Sushobhitha M, Sowjanya J, Srikanth Prabhu, Krishnaraj Chadaga, Wattana Viriyasitavat, Sandeep Kautish, "AI-Assisted Tuberculosis Detection and Classification from Chest X-Rays Using a Deep Learning Normalization-Free Network Model", Computational Intelligence and Neuroscience, vol. 2022, Article ID 2399428, 19 pages, 2022. https://doi.org/10.1155/2022/2399428

Akhter Y, Singh R, Vatsa M. AI-based radiodiagnosis using chest X-rays: A review. Front Big Data. 2023 Apr 6;6:1120989. doi: 10.3389/fdata.2023.1120989. PMID: 37091458; PMCID: PMC10116151