Tuberculosis Detection Using Pre-Trained CNNs

MSc Research Project
Data Analytics

Ekansh Kothiyal
Student ID: 23222417

School of Computing
National College of Ireland

Supervisor: Shubham Subhnil

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ekansh Kothiyal
Student ID:	23222417
Programme:	Data Analytics
Year:	2024/25
Module:	MSc Research Project
Supervisor:	Shubham Subhnil
Submission Due Date:	12/12/2024
Project Title:	Tuberculosis Detection Using Pre-Trained CNNs
Word Count:	XXX
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Ekansh Kothiyal
Date:	12th December 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Tuberculosis Detection Using Pre-Trained CNNs

Ekansh Kothiyal
23222417

Abstract

Tuberculosis (TB) is a leading infectious disease globally, that continues to emphasize its high mortality rate and contagious properties, especially in low and middle-income regions. There have been many diagnostics in this area. Still, with current diagnostics methods, they face huge limitations in sensitivity, speed, and accessibility which leads to delays in early TB detection and treatment. Deep learning models are explored in this study, specifically convolutional neural networks (CNNs), which act as a tool to enhance TB diagnostics that help classify all the cell images as TB-infected or healthy cells. Two pre-trained Models MobileNetV2 and DenseNet121, have been investigated for their effectiveness in TB image classification. In this study, MobileNetV2 is selected due to its computational efficiency and DenseNet121 is chosen due to its deep feature extraction capabilities. The models were trained using transfer learning on a sputum smear microscopy dataset with TB-infected and healthy cell images. The findings in this research highlight the viability of pre-trained CNN models as an important tool in TB diagnosis.

1 Introduction

One of the top infectious killer diseases is still Tuberculosis (TB) with an estimated 10 million new cases and nearly 1.4 million deaths in 2019. Lower- and middle-income countries were majorly affected by this Chakaya et al. (2021). The cause of tuberculosis is the bacteria called Mycobacterium tuberculosis, which mainly focuses on the lungs but can sometimes affect other organs too, leading to many symptoms and diseases. There have been many cases of delay in the diagnosis of tuberculosis despite the advancement in the treatment of TB. The goal of this research is not only to enhance the early detection and treatment of TB but also to prevent long-term health impacts on the affected population with this disease. Still, nowadays the techniques for the detection of TB have limitations that slow down the process of accurate detection, especially where there is a limited resource setting. There are many challenges across the globe such as limited access to healthcare facilities and technology, high cost of diagnostic tools, and lack of trained professionals in low-resource regions Walzl et al. (2018) that highlight the urgency for new diagnostics methods and technologies that are cost-effective help in the improvement of TB detection, and also support global TB initiatives.

There have been many traditional diagnostics methods which include sputum spear samples, nucleic acid amplification tests (NAATs) Pagaduan and Altawallbeh (2023), some culture tests, and chest x-ray tests Kumar and Dhingra (2023). In highly populated countries most commonly accessible method is sputum smear microscopy, but has low sensitivity with patients with low bacterial load or extrapulmonary TB. In the case of culture tests, these are more sensitive and require several weeks for the result which usually delays the initiation of the treatment. More expensive methods are NAATs such as GeneXpert but are faster and give more accurate results. To perform NAATs specialized equipment and infrastructure may not be available in low-resource settings. These methods primarily focus on the direct detection of bacteria and may miss out on cases where bacteria is spread outside the lungs and bacteria may not be present in some sputum samples. All these limitations combined create a need for new innovative diagnostics methods that help in higher sensitivity, and speed, and help in diverse healthcare environments, mostly in areas where TB burden is at its highest.

In the detection of TB, AI plays a major role particularly deep learning which is a promising field in medical diagnostics Cao et al. (2021). These deep learning algorithms, especially convolutional neural networks (CNNs), enable computers to learn and extract important features from complex datasets through image analysis. These features of CNN are invaluable in medical imaging, for the diagnosis of diseases accurate pattern recognition is very essential. Vandana and Kannan (2021)Oncology, dermatology, and radiology are some of the fields where CNN has been successfully implemented, showing perfect accuracy in identifying disease patterns present in images containing bacteria. This study explores the potential of models in diagnosing TB with two advanced pre-trained models, Shome et al. (2024) MobileNetV2 and DenseNet121 Rochmawanti and Utaminingrum (2021), which are known for their efficiency in image recognition. These pre-trained models have become an important factor in medical diagnosis because of their adaptability to any tasks with transfer learning. MobileNetV2 and DenseNet121 were chosen here because of their computational efficiency and strong performance in medical image analysis. MobileNetV2 is known for its lightweight structure which is most useful in areas with less computational resources, while DenseNet121 is known for its deep connectivity across all the layers offering robust feature extraction which leads to enhancing the accuracy of the classification.

The main aim of this research is to assess the performance of both the models MobileNetV2 and DenseNet121 in differentiating between TB-infected positive samples and healthy cells while examining, processing speed and accuracy. Implementation of transfer learning is applied to these pre-trained models, this helps in evaluating these models' potential as more reliable alternatives than traditional TB diagnosis methods which sometimes often lack in various aspects like sensitivity, speed, or accessibility of TB control. The use of CNN in this study could support reliable and strong diagnostic processes, also reducing the delays in treatment, especially in the regions where this kind of rapid and fast diagnosis is essential in managing high cases and limited resources in healthcare infrastructure. Also, this study focuses on addressing the gaps in traditional diagnostics methods with AI-driven approaches. The development of automated diagnostics tools can potentially transform TB diagnosis enabling faster and more accurate detection in diverse healthcare contexts by the insights derived from this evaluation.

The research question of this study is "*How does the classification accuracy of pre-trained CNN models, MobileNetV2 and DenseNet121, compare in identifying tuberculosis-infected versus healthy cells using transfer learning?*" This study highlights the potential of AI-driven methods, that demonstrate their ability to surpass traditional TB diagnostic techniques. The comparative performance of the models MobileNetV2 and DenseNet121 is focused on here, this research will dig deeper into the concepts of these convolutional neural networks and whether they can deliver the precision, speed, and accessibility that are required to meet the global TB diagnostics needs. The findings that are achieved from this study could shine some light on how AI will help bridge current diagnostics gaps and offer a pathway towards high-quality TB diagnostics that could transform TB detection and management where there is limited resource setting worldwide.

2 Related Work

Tuberculosis is one the major global health issues due to its complexities which surrounds its detection and diagnosis. It can manifest in different body systems and mimic other diseases, leading to a delay in diagnosis that can worsen patient outcomes. Also, there is a rise in multidrug-resistant TB (MDR-TB) which further complicates the diagnosis and treatment landscape, leading to innovations in diagnostic technologies to improve the speed, accuracy, and accessibility. The healthcare system has faced many unresolved challenges due to reliance on traditional diagnostic tools. For example, while performing the ECG or chest X-rays that are widely used for TB screening rely on the expertise of trained radiologists, which may not always be available in resource-limited settings. There might be some challenges due to the interpretation of image results that increase the potential for diagnostic errors. These limitations are particularly present in regions with high TB burdens. There have been many advancements in digital health and artificial intelligence (AI) which have presented new opportunities for evolving the TB diagnosis. With large datasets, there are AI-driven approaches in particular with the use of convolutional neural networks (CNNs).

This study specifically focuses on the use of transfer learning with pre-trained CNN models, while exploring how these methods and architecture can contribute to improving the TB diagnosis. Minimizing the use of large datasets and training time these transfer learning model offers the ability of the existing model to do specific tasks. This approach could help address the diagnostic challenges which are posed by TB and some other mixed infections.

2.1 Challenges and Limitations in Tuberculosis Diagnosis

A comprehensive analysis was provided by (Walzl et al.; 2018) on the advancement and challenges in the development of TB diagnosis and biomarkers. There was an understanding that while significant progress has been made in understanding TB and its clinical display, these advancements are not fully helpful in diagnostic breakthroughs that can address the issues at a global level. The key highlight of this study was the dependency on traditional diagnostics methods such as sputum smear microscopy, culture tests, and nucleic acid amplification tests (NAATs). Also, G. Walzl et al. found that sputum smear microscopy lacks sensitivity, especially in individuals with extrapulmonary or paucibacillary TB. The authors have also explored the issues related to resource constraints in low- and middle-income countries, where TB is at its highest level. These diagnoses such as NAATs (e.g., GeneXpert) require high cost and specialized infrastructure. These kinds of advancements in low-income regions where TB cases are often underdiagnosed or misdiagnosed create a gap in diagnostic capabilities. There have been many other challenges that were discovered by Walzl et al. like difficulty in diagnosing latent TB infection and differentiating it from active diseases. Apart from all these challenges, this study highlights the efforts that are put in to develop new biomarkers and diagnostic tools. To become impactful in high-burden areas further validation and large-scale implementation are required but also advancements in molecular diagnosis and point-of-care testing hold promising results.

Another study was conducted by (Das et al.; 2019) with the argument for the relevance of sputum smear microscopy in TB diagnosis. Sputum smear microscopy remains a critical diagnostic tool in many low-resource regions because of its simplicity of use, affordability, and accessibility. Limitations were highlighted in this research on sputum smear microscopy which includes its low sensitivity, especially in patients with low bacterial loads and extrapulmonary tuberculosis. There have been many cases of underdiagnosis of TB or delayed treatment which impacts TB control programs negatively. Also, the preparation of sputum smears requires highly skilled personnel which leads to compromising diagnostic accuracy.

On the other hand, there was an argument that was led by Das et al. that molecular diagnoses such as NAATs which offer high sensitivity and specificity are not usually reliable in resource-constrained regions because of their high cost, requirement for specialized infrastructure, and dependency on continuous power supply. In many developing countries there is a challenge in sputum smear microscopy where the TB burden is high and healthcare resources are limited. It was discovered that sputum smear microscopy serves as a frontline diagnostic method for cases with initial stage and treatment monitoring in a healthcare setting. In contrast, molecular methods are ideal for confirming TB cases in referral centers. In conclusion of this study, Das et al. underlined the use of modern diagnostic techniques and maintained the utility of traditional methods like sputum smear microscopy

On the other hand (Andom et al.; 2023) provided valuable insights into the effectiveness of tuberculosis diagnosis and treatment in low-resource regions with focusing on the kingdom of Lesotho. Geographical and economic barriers played a critical role in this research. This research was done in the rural areas of Lesotho where patients have to cover long distances to access healthcare facilities while giving significant transportation costs. These factors lead to delays in diagnosis with an increase in disease transmission. Also, the financial burden of these diagnostic tests and other healthcare expenses often discourage individuals from seeking timely medical attention, especially in economically vulnerable regions. It was discovered that when molecular diagnostics such as GeneXpert are available, many issues such as power outages, equipment maintenance, and supply chain inefficiencies limit their abilities and utilities where there is a shortage of resources. A few strategies were discussed in this study such as decentralizing diagnostic services to bring care closer to all the communities, reducing the costs for low-income patients, and improving outcomes by integrating TB diagnostic with broader healthcare initiatives. Reducing the stigma and increasing awareness of the importance of early TB detection and treatment were discussed in this study with public health campaigns. Addressing these barriers such as systemic, economic, and social challenges in TB diagnosis requires policy-level intervention and community-based strategies usually, which were the key findings from Andom et al. in low-resource settings such as Lesotho.

(Sathitratanacheewin et al.; 2020) investigated the use of deep learning models to classify tuberculosis using chest X-rays, focusing on the dataset distribution shifts that can affect diagnostic performances. The focus of this research was addressing the critical challenges in the application of AI-driven solutions to TB diagnosis. The performance of these deep learning models often declines when tested on datasets with diverse demographic or clinical characteristics but deep learning models perform remarkably well with high accuracy with training on homogeneous datasets. There is a major limitation in developing automated diagnostic tools which is known as dataset distribution shift. For example, there might be a huge difference in the model training on chest X-rays when tested on different regions as many regions might have various imaging protocols and patient demographics. One of the challenges highlighted in this study was inconsistency in imaging equipment and techniques for TB diagnosis across healthcare facilities between high and low-resource regions. The performance of AI algorithms can also be impacted by the difference in image quality, resolutions, and noise level in the images. This acts as a risk in clinical applications where accuracy and consistency are the key factors. Clinicians often find it difficult to trust the predictions due to the deep learning models' lack of interpretability. These things lead to the inability to model decisions which complicates the adaptation in real-world scenarios where transparency and accountability are very important. To address all these challenges, Sathitratanacheewin et al. took a diverse dataset that included images from multiple regions, populations, and healthcare settings to mitigate the issues. Techniques such as domain adaptation and data augmentation were used in this study to improve the model's robustness. Also, the collaborative efforts put in between AI developers and healthcare practitioners are important for the algorithm to meet clinical standards.

Some newer diagnostic tests were discovered by (Chopra and Singh; 2020) such as molecular assays and rapid diagnostic tools, which have significantly improved the speed and specificity of TB detection. In many low-income regions advanced diagnostics like GeneXpert remains cost-prohibitive with limited access. There are restrictions on reliance on cold chain logistics and high-maintenance laboratory tools in rural areas. The limitation of point-of-care diagnostic is universally not validated for diverse TB strains and patient demographics. The need for more affordable, portable, and robust diagnostic tools in high-burden regions took the attention of the authors, their analysis highlights the gap between diagnostic innovation and practical implementation.

Few studies show the complications and challenges with post-tuberculosis lung disease (PTLD) (Allwood et al.; 2021). This study focuses on problems in identifying long-term pulmonary damage caused by TB, with inadequacy the conventional diagnostic tools, such as chest X-rays and sputum cultures. Another point raised in this study was the limited integration of lung function testing and advanced image analysis in TB diagnostic. Development of the diagnostic framework that extends beyond microbial detection and focuses on holistic lung health for addressing TB-related complications. The differentiation of latent tuberculosis infection (LTBI) from active tuberculosis (TB) was put into highlight by (Gong and Wu; 2021) which is a major challenge for effective TB control. Limitations were highlighted such as the tuberculin skin test (TST) and interferon-gamma release assays (IGRAs) which sometimes cannot differentiate between active TB and LTBI. This research also explains the lack of context-specific diagnostic frameworks which is associated in high burden regions where access to these molecular diagnostics remains limited. Their analysis helps in the necessity for research investments and global collaboration for innovative diagnostic strategies for both clinical needs and public health goals.

Some serious challenges posed by tuberculosis (TB) and HIV co-infection, in areas with high diseases show the impact on the immune system which reduces the readability of traditional diagnostic tests like tuberculin skin test (TST) and sputum smear microscopy (Letang et al.; 2020). Usually, the diagnostic process is complicated by HIV-positive individuals and often results in atypical TB presentations or extrapulmonary TB. The lack of integration between TB and HIV care services acts as a systemic barrier to timely diagnosis and treatment. An innovative approach was called in this research using diagnostics tools that are sensitive and practical for decentralized healthcare systems.

2.2 Advancements in Machine Learning and Deep Learning

Deep learning is revolutionizing the healthcare industry with its ability to handle and analyze complex datasets, keeping this in mind (Esteva et al.; 2019) conducted a study highlighting the deep learning capacity to improve diagnostic accuracy in various healthcare fields such as dermatology, radiology, and pathology, where image data is used. The performance of deep learning models was checked to its potential not only to match but also to exceed human expert performance in some specific diagnostic tasks, improving its capacity to learn complex patterns in images. The quality and quantity of the dataset used greatly influenced the model's performance, which also underlines the importance of large datasets to train deep learning models. The need for interpretability and explainability in Deep learning applications was also discussed so that clinicians can understand how these models arrive at their conclusions for integration in the clinical workflows. Black box nature is one of the challenges associated with many deep learning models which can affect the performance with adoption in sensitive healthcare areas. Future roles of the deep learning model in healthcare were pointed out by Esteva et al. with the emerging techniques in transfer learning and federated learning that have proven promising results for data privacy and security. This study focused on how the synergy between the deep learning model and clinicians can be put to use to enhance the patient's care. This partnership can lead to many better outcomes like early detection, diagnosis, and personalized treatments.

(Razzak et al.; 2018) researched how deep learning has transformed medical image processing with a focus on the automation of image-based diagnostic tasks, improving the detection, segmentation between images, and classification of any medical anomalies. CNN in particular was a highlight of this research being a cornerstone in medical image analysis because of its ability to learn the old features from raw data which reduces the reliability of manual feature extraction. Transfer learning as a solution was one of the important key highlights, enabling pre-trained models to adapt to medical imaging tasks by fine-tuning the smaller dataset. Also, advancements in generative models and unsupervised learning techniques could improve data scarcity challenges. They also suggested that combining domain knowledge with AI architecture could increase the trust and adoption of AI-driven diagnostics in healthcare or clinical environments.

(Rana and Bhushan; 2023) on the other hand discussed the potential of the machine learning model and deep learning model both in the field of medical image analysis. The important aspect of this study was the role of hybrid models that combine ML algorithms with DL framework and architecture for enhancing the adaptability of diagnostic systems across different datasets. This combination incorporates a robust feature extraction process and classification while addressing challenges such as data heterogeneity and limited labeled datasets. The importance of transfer learning was also pointed out here in medical imaging which enables pre-trained models to adapt to specific medical applications and tasks with minimal use of any computation resources. Several challenges were discussed like the need for large annotated datasets, computational costs, and patient data privacy. Despite all these challenges, the potential of ML and DL approaches helps in achieving higher accuracy and also helps in early detection, treatment process, and personalized medicines.

The images that are fed in the deep learning model are very complex and sometimes it is difficult to detect patterns or abnormalities in these images, so (Mohapatra et al.; 2021) conducted a study with medical images of radiological scans, X-rays, CT scans, MRIs, and histopathological images. CNN is used here as it has the advantage of an automatic feature extraction process which reduces the need for manual feature extraction which is usually time-consuming and prone to errors. The highlights of this study were the practical applications of CNN in diagnosing diseases through medical images. CNN works in such a way that it can detect abnormalities in images such as tumors, lesions, or other pathological changes that sometimes go unnoticed. Integration of CNN with patient demographics to create a hybrid model is also discussed. This helps in a personalized diagnostic system, where the data captured from images is combined with clinical information for better accuracy and results. They faced several issues such as class imbalance, high computational demands for training deep learning models, variations in image quality, and patient populations. Solutions such as data augmentation, model fine-tuning, and incorporating domain-specific constraints were implemented to improve the accuracy rate.

There was a concern raised by (Singh et al.; 2020) about deep learning models that it is not just important to provide accurate predictions but also to offer clear explanations for better decision-making processes. This is important in the healthcare sector where all medical practitioners must trust and understand the reasoning behind an algorithm's predictions before implementing it in a patient's care decision. The importance of explainability in deep learning models is discussed here, while deep neural networks are often regarded as black boxes because of their Complex and nonlinear nature. Many techniques such as saliency maps, attention mechanisms, and layer-wise relevance propagation are discussed here as ways to provide information about how models reach their conclusions. These techniques are helpful for clinicians and researchers to visualize which area of images has contributed most to a model's decision, making it easy to trust and identify the errors in the system. The study also addresses the challenges in image processing as images can be highly complex and can have subtle variations, which makes it difficult for the model to explain in a human-understandable manner. The need for standardization in explainable AI techniques was also discussed. They suggested that any kind of research in this area should not just focus on improving accuracy but also ensuring that they are transparent, ethical, and trustworthy.

A novel deep learning-based model was explored by (Zakareya et al.; 2023) for the diagnosis of breast cancer from medical images using CNNs to improve diagnostic accuracy. The combination of hybrid architecture with multiple deep learning layers for feature extraction and classification accuracy was introduced while addressing challenges such as class imbalance and noise in the images. Techniques like transfer learning and fine-tuning of pre-trained models showed optimizing performance in breast cancer detection. Utilization of a large dataset of medical images demonstrated the model's robustness and generalizability across different types of images like mammograms, ultrasound, and MRI scans. This study discusses the trade-offs involved in the model's complexity to make the model more transparent to clinicians. Many studies were done for the diagnosis of heart-related conditions, and an investigation of the application of the deep learning model was done by (Wong et al.; 2020) which focuses on state-of-the-art neural networks for the analysis of complex cardiovascular images, such as echocardiograms and CT angiograms for more precise detection of anomalies like coronary artery blockages, heart valve defects, and myocardial infarctions. The authors proposed an advanced deep-learning framework to incorporate multi-modal data integration which combines patient medical history with imaging data to deliver a comprehensive output. To increase the learning capabilities of the model reinforcement learning techniques can be implemented into an existing model. They also focused on the use of explainable AI methods for better visualization which helps in the decision-making process of deep learning algorithms. Certain things were proven like diagnostic insights were clinically relevant and actionable, bridging the gap between technology and healthcare practitioners.

2.3 Research Gaps in TB Diagnostics

The role of AI is one of the key factors in addressing diagnostic challenges in Tuberculosis(TB), focusing on radiology applications in one of the studies conducted by (Kulkarni and Jha; 2020). The authors of this research explored the application of deep learning in identifying the abnormalities in chest X-rays. Research gaps identified were the lack of a comprehensive dataset that represents different patient demographics and disease conditions which sometimes leads to biased AI predictions. Other issues such as the integration of AI systems into existing healthcare systems, are due to regulatory and technical barriers. The development of explainable AI models was done by the authors to improve diagnostic accuracy and transparency for clinical validation. These kinds of approaches are very important to build trust and adoption among healthcare professionals. For future research, it was suggested that the focus should be on building a scalable AI framework that would be capable of real-time diagnosis while addressing issues like data privacy and standardization.

An innovative approach was led by (Duong et al.; 2021) for tuberculosis detection using chest X-rays incorporating Vision Transformer (ViT) models and transfer learning techniques. Advanced feature extraction of ViT was used for capturing global relationships in medical images to overcome limitations in traditional machine learning algorithms. Pre-trained ViT models with transfer learning were implemented, optimizing the system for the nuances of TB detection. Again the research gap includes the need for larger, more representative datasets for the reduction of potential biases in prediction using AI. Exploration of hybrid models that combine image data with clinical information may increase diagnostic precision. This ensures that decisions made by such models are transparent by medical practitioners and professionals.

Other studies also include drug resistance patterns and good diagnostic accuracy by an integrative approach focusing on deep learning and machine learning models explored by (Liang et al.; 2022). By using the clinical and molecular datasets authors identified the potential of AI systems in identifying the complex relationships that are not identified by traditional diagnostic methods. The utilization of AI for the prediction of multi-drug resistant TB (MDR-TB) was contributed to this study. These approaches are scalable which makes them adaptable from low-resource regions to high-tech laboratories. The research gaps identified in this study include the need for harmonized datasets to train robust AI models. Another issue that was pointed out was the integration of this system into clinical practice, where the usage and interpretability by healthcare professionals, remain a significant challenge. This study also incorporates more interdisciplinary collaborations to refine these AI tools for usage in real-world applications. Other significant gaps in the diagnostic care cascade for tuberculosis were explored by (Shah et al.; 2022) in various areas including the patient, community, and healthcare systems. Challenges such as delays in diagnosis, low awareness about TB symptoms among patients, and management of the disease were discovered. These kinds of barriers often lead to failure in early treatment and contribution to disease care and transmission. A few aspects were captured from this study, focusing on the interplay between patient and systemic factors. For example, major issues in the care of the patient arise from stigma and financial constraints. Misinformation is also a key reason for delays in diagnosis among the community, patients have to face problems with limited access to diagnostic tools with under-trained professionals. For all these challenges AI has proven to be enhancing the accuracy and accessibility of diagnostic tools especially in areas with low resources. Usually, diverse and high-quality datasets are unavailable in regions with high TB cases.

A study done by (Aamir et al.; 2024) outlines how AI dependency on these data sources creates gaps in diagnosis and treatment for rural and underserved populations. Another important insight pointed out in this research was how AI is bridging the gap by implementing mobile health platforms where AI-powered smartphone applications could perform any kind of screening as per the symptoms or X-ray images, which could offer promising solutions in low-resource regions. The evolution of AI was also discussed with its being a standalone diagnostic tool for integrating the clinical decision-making process, analysis, and patent management. Due to limited diagnostic infrastructure often things are overlooked like AI could enhance drug resistance detection for TB especially. Solutions were proposed such as collaboration between AI researchers, clinicians, and policymakers that advancement in technology could be used in practical solutions for controlling TB.

3 Methodology

The methodology for this research followed a structured approach, which aimed at developing an efficient tuberculosis detection system using some deep learning techniques and algorithms. Careful documentation and execution were done in each step of this research from data collection to model evaluation for the surety of reproducibility.

Data Preparation and Management: Preparing and managing a dataset is a crucial step in any machine learning pipeline. Data was collected from one open-source platform ”Kaggle” containing TB-positive and TB-negative images. These datasets were stored in Google Drive in unified directories for a smooth streamlined process. The dataset contains images with sputum samples of patients with TB and without TB.

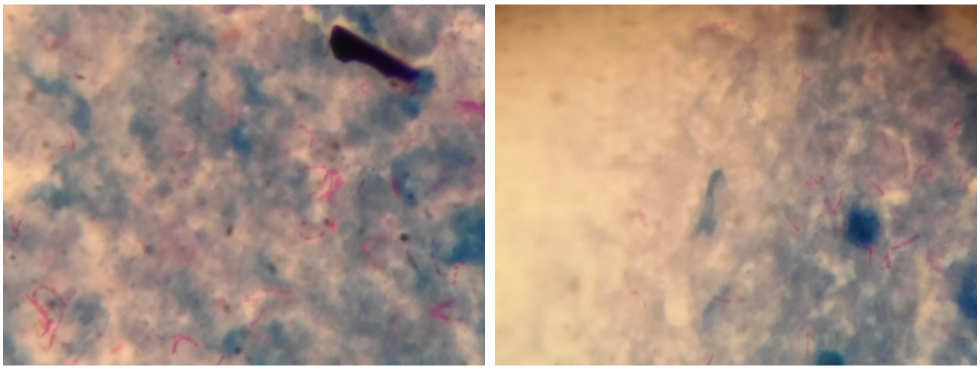


Figure 1: Positive sputum sample

As in Figure 1, the image contains a sputum sample from the patient diagnosed with bacteria called Mycobacterium tuberculosis, and in Figure 2 negative sputum samples or healthy samples. There was a class imbalance in the dataset which is a common issue in medical image analysis, which was addressed by a combination of oversampling and data augmentation techniques. With the help of oversampling it was sure that the minority class (negative class) was represented during the training of the model while data augmentation on the other hand simulated real-world scenarios to enhance the generalizability of the model. There were various augmentation techniques used in this process which were geometric transformations such as rotation, zoom, shear, and flips, as well as intensity variations. Training data was diversified, which helped the model to be well-equipped to handle previously unseen cases. After preprocessing and labeling the dataset, the data was split into training and validation subsets in an 80:20 ratio for sampling. For unbiased evaluation of the model's performance, this method was used for the preservation of the class proportions across both sets. To feed the images in the neural network advanced preprocessing strategies were implemented such as normalization and resizing.

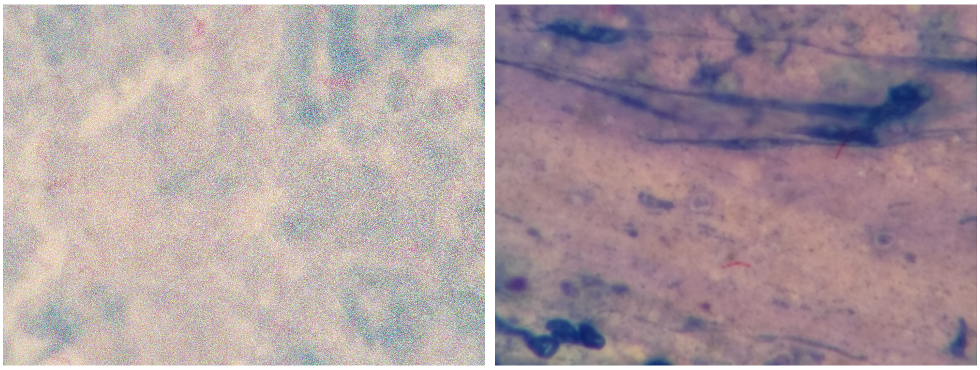


Figure 2: Negative sputum sample

Model Architecture and Training: This phase focused on the design and training of the deep learning models. The selection of pre-trained models was done based on their superior feature extraction capabilities which were MObieNetV2 and DenseNet121. The reason for the selection of MobileNetV2 was because of its lightweight architecture which would be suitable for deployment in resource-constrained devices and DenseNet was selected because of its ability to reuse features efficiently through its dense connectivity design. For encoding the knowledge in pre-trained models and using them in TB detection, transfer learning was implemented here. In the beginning feature extraction layers were frozen to retain their knowledge and a few custom layers were changed for the specific task of TB classification. To reduce the dimensionality a new layer called the GlobalAveragePooling2D layer was included, for capturing complex patterns in the images ReLU activation was implemented and for overfitting some layers were dropped out. In the end, for the production of binary classification final layers employed a sigmoid activation function. These training processes were configured in this research for robust model performance. The model was trained for 10 epochs with a batch size of 64 and 128 using the Adam optimizer for adjustment of weights. For sustaining the binary classification problems, binary cross-entropy was selected as the loss function. During the training of DenseNet121 learning rates were fine-tuned for prevention of overfitting and optimizing convergence. To further improve the training data quality techniques like oversampling and resampling were applied here and for DenseNet121 negative class were undersampled to match the positive cases for balanced training data.

Evaluation Methodology: Rigorous evaluation was done to assess the model's performance in detecting TB. Accuracy, precision, recall, F1-score, and confusion matrices were the performance metrics for providing a comprehensive view of model behavior. Accuracy was used for the model's overall correctness, while precision and recall were evaluated to correctly identify the TB positive cases among all. In dealing with imbalanced data a single measure of the model's effectiveness was measured with an F1 score balancing the precision and recall. To get deeper insights into misclassification patterns confusion matrices visualized the distribution of true positives, true negatives, false positives, and false negatives. 20 percent of the dataset was reserved for conducting validation on the model. A comparison was done with the results of MobileNetV2 and DenseNet121 for the evaluation of the model's performance. DenseNet was considered a superior model particularly classifying positive images with negative images. DenseNet model's feature extraction capabilities came into the picture in identifying which layer contribution was most significant to the classification process. These analysis processes were informed for the fine-tuning decisions, which included the unfreezing of DenseNet121's last 30 layers that allowed them to adapt to TB detection tasks. Here statistical techniques were crucial in validating the methodology. The impact of data augmentation was checked by comparing the model's performance with or without the augmentation process, which told us that recall rates for TB-positive cases improved gradually demonstrating how the augmentation is effective in enhancing the model's generalized ability. Also, the success of resampling approaches was tested by analyzing class distribution and their influence on evaluation metrics.

For the reproduction of the research, all processes were conducted on Google Collab, leveraging its GPU and TPU capabilities for a faster computation process. Libraries used in this research were TensorFlow, OpenCV, and Scikit-learn for tasks ranging from model building to pre-processing the data and evaluation of the model. Visualization tools like Matplotlib and Seaborn were used to present the results graphically including accuracy, loss, classification report, and confusion matrices. The methodology used in this research combines state-of-the-art deep learning techniques with some data pre-processing and statistical analysis to create a framework for TB diagnosis and detection. Step-by-step processes were used here for transparency and reproducibility while integrating these advanced models and handling balanced data demonstrates the potential of AI in handling and addressing critical healthcare challenges.

4 Design Specification

The implementation of detecting tuberculosis in this study is designed to focus on the architecture, framework, and techniques while detailing the requirements and functionality of the proposed design. A hybrid approach combined transfer learning with tailored preprocessing for feature extraction techniques. The framework used in this project was TensorFlow, chosen for its versatile nature in building, training, and deploying machine learning models. All the benefits of transfer learning were considered when implementing two pre-trained architectures, MobileNetV2 and DenseNet121. These two models can easily analyze medical images because of their ability to extract hierarchical features while maintaining computational efficiency. The input of the images was set to a fixed size of 224×224×3 for image preprocessing and analysis. The model's coverage during training was improved with the input requirements of MobileNetV2 and DenseNet121.

Model Architecture The primary source for detecting Tuberculosis is neural network architecture which is designed to combine transfer learning while updating the model for medical imaging analysis. For extracting the features from the images this architecture uses pre-trained CNNs, MobileNetV2 and DenseNet121. These networks are pre-trained on a diverse ImageNet dataset, which is excellent in identifying complex patterns, features, and textures in the images. A generic feature detector in these models identifies characteristics in the images like edges, shapes, and textures. Fine-tuning was done to certain layers of DenseNet121 to improve the model's ability to adapt different domain features in medical images which optimizes its performance for the detection of tuberculosis. This architecture includes custom layers that help in the refinement and interpretation of the extracted features for binary classification. To reduce the spatial dimensions of the feature maps a global average polling layer was implemented that provides a representation that minimizes overfitting tasks. For modeling the relationship between features in the images, ReLU activation functions from these dense layers are included. Drop-out layers here are added to enhance generalization by introducing randomness during the training process and also reducing dependency on specific patterns. A sigmoid-activated output layer used in the model architecture provides a probability score that indicates the chances of tuberculosis-positive diagnosis. For achieving high accuracy this architecture is scalable and its computationally efficiency is reliable in both high-resource and constrained or restricted environments.

Proposed Model Functionality Advanced image processing, feature extraction, and classification techniques can be achieved in end-to-end tuberculosis detection in sputum smear microscopy images. Starting with preprocessing the microscopic images which involves resizing to the target dimensions and scaling the pixel values across the dataset normalization was implemented. By using these preprocessing steps it was ensured that the model received consistent input with minimized variability caused by differences in the quality of the images. To detect the visual features that are relevant for tuberculosis detection this study incorporated intensity scaling for the image processing analysis. The supporting element of these models is transfer learning-based architecture with pre-trained convolutional neural networks (CNNs), specifically MobileNetV2 and DenseNet121. ImageNet served as a powerful feature extractor for the identification of patterns in sputum smear images that align with acid-fast bacilli (AFB)—the primary indicator of tuberculosis. This process took freezing-specific layers for retaining edges and texture identification mechanisms while the later layers were fine-tuned for the adaptation with the domain-specific features of sputum smear microscopy images. Certain challenges came into the picture for the imbalanced datasets which is a common issue in medical image analysis. Overfitting of the model is reduced here by applying geometric transformations like rotations flips, and scalings. Due to fewer negative samples, the oversampling technique was done to ensure enough of the samples were fed during the training phase. A comprehensive evaluation framework was assessed for the system's reliability that included metrics such as precision, recall, F1-score, and accuracy for delivering consistent and actionable results. These evaluations help the model in fine-tuning with real-world applications, where it is expected to function effortlessly in diverse environments with limited resource regions or extreme laboratory conditions. A decent amount of automation, efficiency, and accuracy was achieved by integrating these components for the classification of TB positive and negative samples. This functionality is not only limited to technical implementation but also promising in clinical settings which reduces manual efforts, accelerates the diagnosis process, and provides reliable support for managing tuberculosis diagnosis programs.

5 Implementation

The detection of tuberculosis in sputum smear microscopy was implemented using the model's robust nature in processing image classification. This section details the final stage of the implementation using the tools, framework, and methodologies for the desired outputs. This process consists of stages that are stated earlier such as preprocessed datasets, and evaluation metrics for demonstrating the system's performance. Python was the primary language, used here with libraries such as TensorFlow and Keras.NumPy and Pandas were used for data handling and manipulation, and for visualization and performance metrics, Matplotlib and Seaborn were used.

The implementation began with maintaining consistency and quality by preparing the sputum smear microscopy images dataset. The images went through resizing to standard dimensions that were suitable for the pre-trained model. This resizing process was done to make sure all the samples had uniform dimensions, eliminating the variability for better model learning. Comparing the images was an important step while maintaining the uniform range by scaling the pixel value and for the adjustment of brightness and contrast intensity scaling was applied. Also, the variations were addressed by differences in staining techniques or slide preparation. The steps undertaken here were crucial for minimizing noise from the images. MobileNetV2 and DenseNet121 showed promising results in medical imaging tasks chosen for their lightweight nature and feature-rich architecture. Pre-trained layers in these models act as feature extractors for identifying patterns in sputum images like the morphology of acid-fast bacilli (AFB), which are indicative of tuberculosis. Specific layers of DenseNet121 were fine-tuned to extract features for medical imaging. Due to this, it helped the model to capture domain-specific features that were unique to sputum smear microscopy images. Custom dense layers were added for the feature extraction phase to interpret and refine these features for binary classification into TB-positive and TB-negative categories. Implemented Adam optimizer in the training process, which is known for its efficiency in adjusting learning rates. Hyperparameters such as learning rate, batch size, and the number of epochs, are carefully fine-tuned for achieving optimal model performance. For the binary classification loss function, binary cross-entropy was used during the training. The model avoided overfitting by regularly monitoring training and validation accuracy.

Performance metrics precision, recall, F1-score, and overall accuracy provided comprehensive results of the model's ability to correctly identify TB-positive and TB-negative sample images. This model ensured the capability of processing sputum smear microscopy images, classifying them with high accuracy, and generating diagnostic probability scores.

Tools and Technologies

1. **Programming Language:** Python is the primary language used, chosen for its versatility and diverse library support for machine learning and image processing tasks.

2. **Deep Learning Libraries:** TensorFlow and Keras were used to build, train, and fine-tune the neural networks. These frameworks provide pre-trained models for simplifying the integration of MobileNetV2 and DenseNet121 into the project.

3. **Data Processing:** NumPy and Pandas were used for data manipulation, in-

cluding loading and preprocessing the image dataset. These methods help in smooth preprocessing pipelines and handling the data effectively.

4. **Visualization Tools:** Matplotlib and Seaborn were considered for visualization of exploratory data analysis, model performance evaluation, and the interpretation of results.

5. **Development Environment:** Jupyter Notebook was used for its interactive environment which allowed effortless integration of the code, documentation, and visualization.

6 Evaluation

A comprehensive analysis of this research was done by certain evaluation metrics focusing on the key points and providing rigorous training. In this section, below is the evaluation of MobileNetV2 and DenseNet121.

6.1 MobileNetV2

MobileNetV2 was trained with the dataset containing 1863 positive TB samples and 1863 negative TB samples. 10 epochs were used while training the model, during which it showed notable improvements in both training and validation performance. During the starting phase, the model achieved a training accuracy of 71.51 percent, which gradually increased to 90.81 percent by the final epoch. Validation accuracy also gradually increases from 83.11 percent to 88.74 percent, which indicates the consistent progress in the model's ability to generalize unseen data. The model's continuous learning and optimization process was seen when there was a decrease in validation loss from 0.3998 to 0.2677. The confusion matrix as seen in Figure 3 provides a more granular view of the model's performance where for negative class, 323 samples were correctly identified as negative samples, and 39 samples were misclassified as positive.

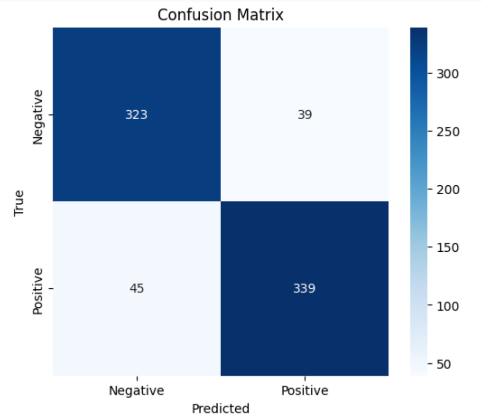
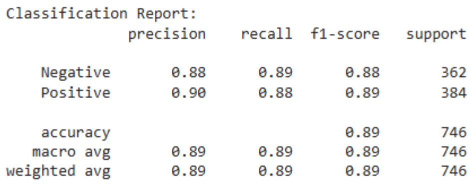


Figure 3: Confusion Matrix for MobileNetV2

For the positive class, 339 samples were correctly classified as positive, and 45 samples were misclassified as negative. These results indicate that while the model was generally accurate in classification, there were some misclassifications, particularly in the positive



```
Classification Report:
              precision    recall  f1-score   support

    Negative       0.88      0.89      0.88       362
    Positive       0.90      0.88      0.89       384

    accuracy                           0.89       746
   macro avg       0.89      0.89      0.89       746
weighted avg       0.89      0.89      0.89       746
```

Figure 4: Classification Report for MobileNetV2

class, which saw a higher number of false negatives compared to false positives in the negative class.

Evaluating further through precision, recall, and F1-score shed light on the model's balanced performance as seen in Figure 4. Precision and recall for the negative class were 0.88 and 0.89, respectively, and for the positive class precision and recall were slightly higher at 0.90 and 0.88, respectively. F1 scores for both classes were higher at 0.88 for negative and 0.89 for positive, showing the model's ability to maintain a good balance between precision and recall, which is very important in handling imbalanced datasets. These findings are proof that MObileNetV2 can effectively distinguish between positive and negative samples, with few misclassifications for positive class. The model achieved an overall accuracy of 88.74 percent, indicating MobileNetV2 is highly effective for its classification tasks. It can achieve higher precision and recall for positive class, suggesting that it could be a good fit for applications where distinguishing positive samples is critical. Also, in Figure 5 we can see the ROC curve with AUC of 0.89. The model's performance was promising in this research, but more rigorous analysis could be achieved with tools like t-tests or ANOVA, which could help in quantifying the significance of the improvements in validation accuracy. These kinds of analyses could provide a clear understanding of the increased performances, whether they are statistically significant or merely due to some random variation. The model's learning process is solid and can be understood by increasing accuracy and decreasing loss over time. This analysis could help in comparing the MobileNetV2 with other models for validating its efficacy in the context of this research.

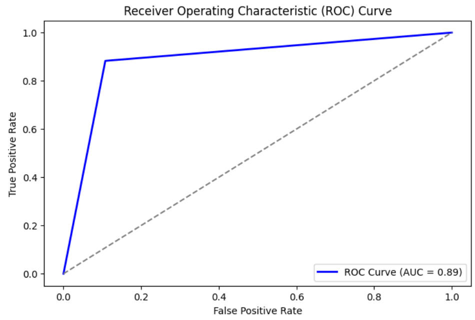


Figure 5: ROC curve

6.2 DenseNet121

DenseNet121 was trained for the second experiment on the same dataset with 1863 positive samples and 1863 negative samples for a total of 10 epochs. The validation accuracy achieved here was 85.25 percent with a validation loss of 0.3324 which indicates it performed relatively well. In the beginning, the accuracy was low at 73.89 percent but continuous improvement was seen with accuracy reaching 86.74 percent by the 10th epoch. Validation accuracy on the other hand improved from 77.48 percent to 85.25 percent reflecting the model's generalized ability to effectively work on unseen data. Although, validation accuracy was less than MobileNetV2 (88.74), DenseNet121 still performed well and showed a clear upward trend in both training and validation.

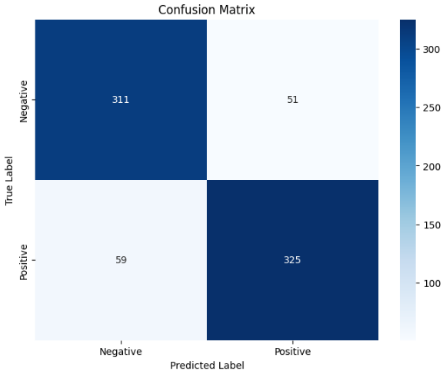
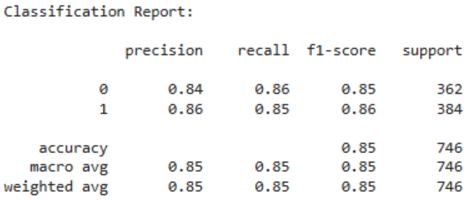


Figure 6: Confusion Matrix for DenseNet121

The confusion matrix as seen in Figure 6 provides additional insights into the model's classification behavior, for the negative class correct classification was done with 311 samples being negative while 51 samples were misclassified as positive. For positive class 325 samples were correctly classified as positive, and 59 samples were misclassified as negative. Fewer errors were shown in this evaluation with the results of a slight tendency for misclassification of positive samples that leads to a higher number of false negatives as compared to MobileNetV2. Overall, the misclassification rate remained low, with the model successfully identifying the majority of both negative and positive samples. As seen in Figure 7, the classification report consists of precision, recall, and F1-score where DenseNet121 demonstrated balanced performance. The precision and recall for the negative class were 0.84 and 0.86, respectively, while for the positive class, the precision and recall were 0.86 and 0.85, respectively. The F1 score for the negative class was 0.85 and 0.86 for the positive class which is an indication of well-rounded classification. DenseNet121 was slightly outperformed by MobileNetV2 which showed a higher precision and recall for the positive class. But DenseNet121's F1 score suggests it was still highly effective in distinguishing between the two classes.



```
Classification Report:

              precision    recall  f1-score   support

           0       0.84      0.86      0.85       362
           1       0.86      0.85      0.86       384

    accuracy                           0.85       746
   macro avg       0.85      0.85      0.85       746
weighted avg       0.85      0.85      0.85       746
```

Figure 7: Classification Report for DenseNet121

There was an observation made that DenseNet121 is similar to MobileNet121 as it exhibited steady learning throughout the epochs. We can see there was a steady decrease in the training and validation loss and accuracy gradually increased. As seen in Figure 8 there is an ROC curve of DenseNet121 with an AUC of 0.93. As compared with MobileNetV2, DenseNet121 improvements were not as promising as that of MobileNetV2. The validation accuracy of DenseNet121 reached 85.25 and 88.74 was the accuracy of MobileNetV2. For a better understanding of the differences in accuracy, some tests can be considered like a paired t-test, which would be helpful. These tests would be helpful in understanding whether MobileNetV2's superior performance is due to random variation or inherent model differences.

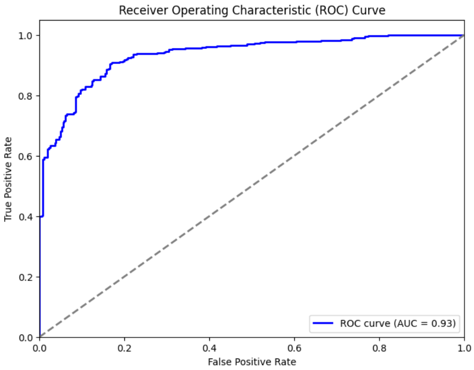


Figure 8: DenseNet121 ROC curve

6.3 Evalutaion Summary

A concise summary of the evaluation is provided by addressing key questions related to the research question, that focuses on the performance, challenges, and model significance used in tuberculosis detection.

Which model achieved higher classification accuracy in identifying tuberculosis-infected versus healthy cells?

After training both the models MobileNetV2 and DenseNet121, the prior received a higher validation accuracy of 88.74 percent outperforming the latter with 85.25 percent validation accuracy. This indicates that MobileNetV2 is more effective in distinguishing TB-positive from TB-negative samples.

How do the models compare in terms of precision, recall, and F1-score?

For TB-infected cells (Positive samples) MobileNetV2 had a precision of 0.90 with a recall of 0.88 and F1 score of 0.89 while DenseNet achieved a precision of 0.86, Recall of 0.85, and F1-Score of 0.86. For healthy cells (Negative sample) MobileNetV2 had a precision of 0.88 with a recall of 0.89 and F1 score of 0.88 while DenseNet achieved a precision of 0.84, Recall of 0.86, and F1-Score of 0.85. This indicates that MobileNetV2 demonstrated better precision and recall, especially in recognizing TB-infected samples.

What statistical approaches were used to validate the models' performance?

For handling the class imbalance and enhancing the generalizability oversampling and data augmentation techniques were implemented.ROC curve and AUC score were used to assess the model's discriminative ability. Future improvements like involving t-tests or ANOVA could be implemented to confirm the statistical significance of observed performance differences.

What are the implications of the findings for real-world applications?

MobileNetV2's high accuracy and lightweight architecture make it suitable for deployment in low-resource regions such as rural healthcare settings. DenseNet121 on the other hand was effective but may require further optimization to reduce false negatives, especially for critical TB detection tasks.

7 Conclusion and Future Work

This study is set to answer the research question: *"How does the classification accuracy of pre-trained CNN models, MobileNetV2 and DenseNet121, compare in identifying tuberculosis-infected versus healthy cells using transfer learning?"*. The results from both models demonstrated promising performance in identifying tuberculosis-infected cells. Both the models showed steady improvements in accuracy during the training period indicating the model's ability to learn and generalize effectively. Balanced precision, recall, and F1 scores were highlighted for their overall effectiveness in distinguishing between tuberculosis-infected and healthy cells. Although, a slight difference in accuracy was seen, both the models proved the capability of performing the classification task reliably and demonstrating the viability of pre-trained CNN models. In answering the research question the findings of this research suggest that both MobileNetV2 and DenseNet121, when used with transfer learning, are effective in classifying tuberculosis-infected cells. The objective of this research was met by comparing the classification accuracy of both models and showing that both of them are capable of performing well in this context. This can be proven from this research that the potential of pre-trained CNN can be used for medical applications, especially for the detection of diseases like tuberculosis.

Key Findings of this research were that both the models showed steady learning and improvement over training epochs. There was a good balance between precision, recall, and F1 scores. Despite the slight difference in performance, both models were effective in the identification of tuberculosis-infected cells. These findings could contribute to the automation of medical diagnosis helping in early detection and improving healthcare outcomes. Implications of this research were stated that larger and more diverse datasets would prove essential in various patient populations and image conditions. Integration of these models in clinical workflow could help enhance the speed and accuracy of tuberculosis diagnosis.

Future work of this research could be the use of additional pre-trained models, such as ResNet or VGG, to see if they offer improvements over MobileNetV2 and DenseNet121. Another area that could be investigated is fine-tuning these models for more specialized datasets such as different stages images of tuberculosis or varying imaging techniques that could be used for improvement in the model's robustness. Moreover, real-world testing with clinical data would provide a better understanding of their performance and reliability in real-world applications and it could be commercialized because of its potential as it could be integrated into diagnostic platforms or mobile applications for quick and accurate diagnosis.

References

Aamir, A., Khan, A. R., Malik, S. H., Tariq, H. and Mehta, R. S. (2024). Exploring the current and prospective role of artificial intelligence in disease diagnosis, *Annals of*

Medicine and Surgery **86**(2): 943.

Allwood, B. W., Byrne, A., Meghji, J., Rachow, A., van der Zalm, M. M. and Schoch, O. D. (2021). Post-tuberculosis lung disease: Clinical review of an under-recognised global challenge, *Respiration* **100**(8): 751–763.

Andom, A. T. et al. (2023). Understanding barriers to tuberculosis diagnosis and treatment completion in a low-resource setting: A mixed-methods study in the kingdom of lesotho, *PLOS ONE* **18**(5): e0285774.

Cao, X.-F., Li, Y., Xin, H.-N., Zhang, H.-R., Pai, M. and Gao, L. (2021). Application of artificial intelligence in digital chest radiography reading for pulmonary tuberculosis screening, *Chronic Diseases and Translational Medicine* **7**(1): 35–40.

Chakaya, J. et al. (2021). Global tuberculosis report 2020 – reflections on the global tb burden, treatment and prevention efforts, *International Journal of Infectious Diseases* **113**: S7–S12.

Chopra, K. K. and Singh, S. (2020). Newer diagnostic tests for tuberculosis, their utility, and their limitations, *Current Medicine Research and Practice* **10**(1): 8–11.

Das, P. K., Ganguly, S. B. and Mandal, B. (2019). Sputum smear microscopy in tuberculosis: It is still relevant in the era of molecular diagnosis when seen from the public health perspective, *Biomedical and Biotechnology Research Journal (BBRJ)* **3**(2): 77.

Duong, L. T., Le, N. H., Tran, T. B., Ngo, V. M. and Nguyen, P. T. (2021). Detection of tuberculosis from chest x-ray images: Boosting the performance with vision transformer and transfer learning, *Expert Systems with Applications* **184**: 115519.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K. Y., Cui, C., Corrado, G. S., Thrun, S. and Dean, J. (2019). A guide to deep learning in healthcare, *Nature Medicine* **25**(1): 24–29.

Gong, W. and Wu, X. (2021). Differential diagnosis of latent tuberculosis infection and active tuberculosis: A key to a successful tuberculosis control strategy, *Frontiers in Microbiology* **12**.

Kulkarni, S. and Jha, S. (2020). Artificial intelligence, radiology, and tuberculosis: A review, *Academic Radiology* **27**(1): 71–75.

Kumar, A. and Dhingra, B. (2023). Clinical and radiological diagnosis of tuberculosis, pp. 11–24.

Letang, E. et al. (2020). Tuberculosis-hiv co-infection: Progress and challenges after two decades of global antiretroviral treatment roll-out, *Archivos de Bronconeumología* **56**(7): 446–454.

Liang, S., Hu, Y., Shen, X., Zhang, X. and Zhang, Y. (2022). The application of artificial intelligence in the diagnosis and drug resistance prediction of pulmonary tuberculosis, *Frontiers in Medicine* **9**.

Mohapatra, S., Swarnkar, T. and Das, J. (2021). 2 - deep convolutional neural network in medical image processing, *Handbook of Deep Learning in Biomedical Engineering* pp. 25–60.

Pagaduan, J. V. and Altawallbeh, G. (2023). Chapter two - advances in tb testing, **115**: 33–62.

Rana, M. and Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: diagnosis to detection, *Multimed Tools Appl* **82**(17): 26731–26769.

Razzak, M. I., Naz, S. and Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future, pp. 323–350.

Rochmawanti, O. and Utaminingrum, F. (2021). Chest x-ray image to classify lung diseases in different resolution size using densenet-121 architectures, pp. 327–331.

Sathitratanacheewin, S., Sunanta, P. and Pongpirul, K. (2020). Deep learning for automated classification of tuberculosis-related chest x-ray: dataset distribution shift limits diagnostic performance generalizability, *Heliyon* **6**(8): e04614.

Shah, H. D., Patel, V. K., Dave, A., Modi, K. R. and Patel, H. C. (2022). Gaps and interventions across the diagnostic care cascade of tb patients at the level of patient, community and health system: A qualitative review of the literature, *Tropical Medicine and Infectious Disease* **7**(7).

Shome, N., Kashyap, R. and Laskar, R. H. (2024). Detection of tuberculosis using customized mobilenet and transfer learning from chest x-ray image, *Image and Vision Computing* **147**: 105063.

Singh, A., Sengupta, S. and Lakshminarayanan, V. (2020). Explainable deep learning models in medical image analysis, *Journal of Imaging* **6**(6): Art. no. 6.

Vandana, K. and Kannan, M. K. J. (2021). A method to apply convolution neural network model to detect and classify tuberculosis (tb) manifestation in x-ray images, *Journal Name* **8**(1).

Walzl, G., McNerney, R., du Plessis, A., Bateson, T., McHugh, A. G., Chegou, A. and Zumla, L. P. (2018). Tuberculosis: advances and challenges in the development of new diagnostics and biomarkers, *The Lancet Infectious Diseases* **18**(7): e199–e210.

Wong, K. K. L., Fortino, G. and Abbott, D. (2020). Deep learning-based cardiovascular image diagnosis: A promising challenge, *Future Generation Computer Systems* **110**: 802–811.

Zakareya, S., Izadkhah, H. and Karimpour, J. (2023). A new deep-learning-based model for breast cancer diagnosis from medical images, *Diagnostics* **13**(11): Art. no. 11.