TUBERCULOSIS DETECTION WITH REMEDIALS AND AIR QUALITY ANALYSIS

A PROJECT REPORT

Submitted by

LEELA SAI LOKESH G [RA2011028010100]
YASHWANTH K S [RA2011028010052]
BHARGAV D[RA2011028010069]

Under the Guidance of

Dr. B. BALAKIRUTHIGA

Assistant Professor, Department of Networking and Communications

in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

with specialization in CLOUD COMPUTING

DEPARTMENT OF NETWORKING AND COMMUNICATIONS

COLLEGE OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE ANDTECHNOLOGY

KATTANKULATHUR – 603203

MAY 2024

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR – 603 203
BONAFIDE CERTIFICATE

Certified that this B.Tech. Major project report titled **Tuberculosis Detection with Remedials and Air Quality Analysis**" is the bonafide work of **Leela Sai Lokesh G[RA2011028010100], YASHWANTH KS [RA2011028010052], Bhargav D[RA2011028010069]** who carried out the project work under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form any other project report or dissertation based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

Dr. B. BALAKIRUTHIGA

SUPERVISOR
Assistant Professor
Department of Networking and Communications.

DR. ANNAPURANI K

PROFESSOR AND HEAD
Department of Networking and Communications

Internal Examiner

External Examiner

Department of Networking and Communications
SRM Institute of Science & Technology
Own Work Declaration Form

Degree/ Course: B. Tech in Computer Science and Engineering with a specialization in Cloud Computing

Student Name : Leela Sai Lokesh G, Yashwanth K S, Bhargav D

Registration Number : RA2011028010100, RA2011028010052, RA2011028010069

Title of Work: Tuberculosis Detection with Remedials and Air Quality Analysis

We hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism, as listed in the University Website, Regulations, and the Education Committee guidelines.

We confirm that all the work contained in this assessment is our own except where indicated, and that we have met the following conditions:

- Clearly listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that we have received from others (e.g.fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook

We understand that any false claim for this work will be penalized in accordance with the University policies and regulations.

DECLARATION:		
We are aware of and understand the University's policy on Academic misconduct and plagiarism and we certify that this assessment is our own work, except were indicated by referring, and that we have followed the good academic practices noted above.		
Leela Sai Lokesh G RA2011028010100	Yashwanth K S RA2011028010052	Bhargav D RA2011028010069

ACKNOWLEDGEMENT

We express our humble gratitude to **Dr. C. Muthamizhchelvan,** Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to Dean-CET, SRM Institute of Science and Technology, **Dr.T.V. Gopal,** for his invaluable support.

We wish to thank **Dr. Revathi Venkataraman, Professor & Chairperson**, School of Computing, SRM Institute of Science and Technology, for her support throughout the project work.

We are incredibly grateful to our Head of the Department, **Dr. Annapurani K**, Professor and Head, Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our Project Coordinator, **Dr. G. Suseela**, Associate Professor, Panel Head, **Dr. K. Venkatesh**, Associate Professor and members, **Dr. B. Bala Kiruthiga**, Assistant Professor, **Dr. V. Rajaram**, Assistant Professor, **Dr. S. A Angayarkanni**, Assistant Professor, Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor, **Mrs. Vijayalakshmi V, Dr. Deepa K Thilak, Ms. G Parimala** Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide, **Dr. B. Balakiruthiga**, Assistant Professor, Department of Networking and Communications, SRM Institute of Science and Technology, for providing us with an opportunity to pursue our project under his/her mentorship. She provided me/us with the freedom and support to explore the research topics of our interest. Her passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank the Networking and Communications, Department staff and students, SRM Institute of Science and Technology, for their help during our project. Finally, we would like to thank parents, family members, and friends for their unconditional love, constant support, and encouragement.

LEELA SAI LOKESH G [RA2011028010100]
YASHWANTH K S [RA2011028010052]
BHARGAV D[RA2011028010069

TABLE OF CONTENTS

ABSTRACT

This study introduces a comprehensive framework that integrates cutting-edge technologies to address Tuberculosis (TB) detection, remedial suggestions, and air quality analysis. By employing Random Forest algorithms for lung quality assessment and ResNet-50 for TB detection via X-ray imaging, the system offers a multifaceted approach to combat TB. It also includes an air quality study, recognizing its crucial role in respiratory health, and integrates air quality indices and environmental factors to contextualize TB occurrences. This holistic approach bridges the gap between environmental factors and public health, enabling swift TB identification for timely intervention and personalized remedial suggestions. By amalgamating machine learning, image processing, and environmental analysis, this integrated system revolutionizes TB diagnosis while emphasizing the critical role of environmental conditions in public health.

LIST OF FIGURES

LIST OF TABLES

List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
BMI	Body Metabolism Index
RESNET-50	Convolutional Neural Network
CRM	Customer Relationship Management
DL	Deep Learning
EEG	Electro-encephalogram
EWS	Early Warning System
FED	Feature Engineering and Designing
KPI	Key Performance Indicator
ML	Machine Learning
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
R&D	Research and Development
SRG	Smart Report Generation
VGG	Visual Geometry Group
XML	Extensive Mark-up Language

Chapter 1
Introduction

1.1 Overview

The confluence of advancements in technology and medical science has propelled the development of a groundbreaking system that revolutionizes Tuberculosis (TB) detection, remedial suggestion, and air quality study. At its core, this innovative system amalgamates cutting-edge artificial intelligence techniques—leveraging the power of Random Forest to assess lung quality and ResNet-50, a deep learning architecture, to scrutinize X-ray images for TB detection. The integration of these methodologies into a singular platform signifies a paradigm shift in healthcare and environmental analysis, offering a comprehensive approach to addressing respiratory health concerns and environmental factors impacting lung conditions [1].

Central to this system is the utilization of Random Forest, a robust machine learning algorithm renowned for its accuracy in analyzing complex data sets. Here, it serves as a pivotal tool in evaluating lung quality, employing a multifaceted approach to assess various parameters indicative of respiratory health. By processing an array of inputs, such as spirometry data, lung function tests, and environmental factors like air quality indices, the Random Forest algorithm adeptly identifies patterns and anomalies, providing an insightful evaluation of lung health status. This holistic assessment facilitates early detection and monitoring of respiratory ailments, enabling proactive interventions for individuals susceptible to lung-related conditions [2].

1.2 Application

Simultaneously, the system employs the formidable capabilities of ResNet-50, a state-of-the-art deep learning model renowned for its prowess in image recognition tasks. Leveraging its intricate neural network architecture, ResNet-50 meticulously analyzes X-ray images of the lungs with unparalleled accuracy in detecting TB manifestations. Through a process of feature extraction and classification, this neural network identifies subtle abnormalities indicative of tuberculosis infection, enabling swift and precise diagnosis. The integration of ResNet-50's capabilities into the system significantly enhances the efficiency and reliability of TB detection, facilitating timely treatment initiation and containment of the disease [3].

Beyond individual health assessment, this comprehensive system extends its purview to encompass an air quality study component. Acknowledging the profound impact of environmental factors on respiratory health, the system integrates an analytical framework to evaluate and monitor air quality parameters. By assimilating data from diverse sources, including pollutant concentration levels, meteorological variables, and geographical attributes, the system employs sophisticated algorithms to scrutinize air quality indices. This meticulous analysis not only correlates environmental factors with respiratory health but also facilitates the identification of potential triggers for respiratory ailments, providing invaluable insights for public health policies and interventions [4].

In essence, the fusion of Random Forest for lung quality assessment, ResNet-50 for TB detection in X-ray images, and air quality study within a unified system signifies a monumental stride in healthcare and environmental analysis. This amalgamation of advanced technologies not only empowers early detection and remedial suggestions for tuberculosis and other respiratory ailments but also offers a comprehensive understanding of the interplay between environmental factors and lung health. As such, this innovative system heralds a new era of precision medicine and proactive environmental health management, promising improved healthcare outcomes and a healthier future for communities worldwide [5].

Chapter 2

Literature Survey

In order to make our research work one of the best and follow a noble approach we have been through the esteemed works of several research scholars. Here is a description of some of them in a formal way.

Vo Trong Quong et. al. (2023) in the paper [1] explains that the highly contagious and life-threatening infectious disease, Tuberculosis (TB), impacts millions worldwide. Prompt treatment and disease control hinge on the crucial early diagnosis of TB. This paper introduces CBA MEDnet, a new deep learning model specifically designed for TB detection in chest X-ray (CXR) images. Employing the Convolutional Block Attention Module (CBAM) and Wide Dense Net (WDnet) architecture, this model adeptly captures both spatial and contextual information within the images.

Lakshya Kumar Saini and Rajneesh Rani (2023) in the paper [2] suggest that TB, caused by the bacterium Mycobacterium, stands as a significant global health concern due to its high mortality rate, often exacerbated by delayed diagnosis and limited radiological expertise. Early identification plays a pivotal role in reducing TB-related fatalities. Among diagnostic approaches, digital chest radiography emerges as the foremost method for early TB detection. This systematic review delves into diverse TB detection methodologies. Examining studies from 2018 to 2022, it scrutinizes databases, techniques for classification, and feature extraction methods utilized in previously published literature. Drawing insights from prior research, this study evaluates a spectrum of techniques to discern the most effective means for TB detection. Its primary aim is to delineate the present landscape of TB detection research and pinpoint optimal methods for accurate TB identification.

Subrat Kumar Kabi et. al. (2023) in the paper [3] explains that radiologists face challenges in early distinguishing between pneumonia and tuberculosis due to similar chest X-ray symptoms, prompting the need for precise automated detection methods. Employing discrete wavelet transform, it assesses different decomposition levels in the X-ray images' sub-bands. Within each sub-band, singular value decomposition (SVD) analyzes singular values, left eigen matrix, and right eigen matrix. Extracting the maximum value from columns of both eigenmatrices and singular values for each sub-band forms the features. These eigen-domain feature vectors from all sub-bands are combined and inputted into the light gradient boosting model for disease detection. Notably, this approach surpasses transfer learning and other previously reported methods in the accurate detection of pneumonia and tuberculosis through chest X-ray images.

Vinayak Sharma et. al. (2023) in the paper [4] explains another way of dealing with this problem. Initially, the U-Net segmentation model underwent training using 704 chest X-ray radiographs sourced from the Montgomery County and Shenzhen Hospital datasets. Subsequently, the trained U-Net model was employed on 1400 chest X-ray scans encompassing tuberculosis and control cases from the NIAID TB portal program dataset to delineate the lung region. Moreover, we explored the visualization potential of the model by utilizing Grad-CAM to discern tuberculosis-related abnormalities in chest X-rays, engaging in a radiological discussion regarding these findings.

Charith Deshnayak et. al (2023) explains in his study [5] that TB, a contagious bacterial airborne disease, ranks among the top 10 global causes of death. As per the World Health Organization, approximately 1.8 billion individuals grapple with TB, with 1.6 million recorded deaths in 2018. Alarmingly, 95% of these cases and fatalities originated in developing nations. Despite its severity, TB is entirely treatable through early detection. To achieve this, maximizing the effectiveness of existing diagnostic tools is crucial, with chest X-rays serving as the primary diagnostic method for active TB screening. Additionally, enhancements encompass image preprocessing, augmentation, hyperparameter optimization via genetic algorithms, and model ensembling, collectively enriching the diagnostic process.

Chapter 3

Problem Statement and Proposed Solution

3.1 Problem Statement:

Tuberculosis (TB) remains a significant global health concern, with millions of new cases reported each year. Early detection and effective treatment are crucial for controlling the spread of the disease and reducing its impact on public health. Additionally, environmental factors such as air quality can influence the prevalence and transmission of TB.

3.2 System Overview:

The integrated system for Tuberculosis (TB) Detection, Remedial Suggestions, and Air Quality Study embodies a multifaceted approach leveraging cutting-edge technologies to address the intricate challenges associated with TB diagnosis and management while considering the critical aspect of air quality's impact on respiratory health. The system architecture consists of two primary components: the lung quality assessment module utilizing Random Forest and the TB detection module employing advanced deep learning algorithms.

Lung Quality Assessment using Random Forest:

The system employs Random Forest, a robust machine learning algorithm, to assess and categorize lung health based on diverse features extracted from medical imaging data. This process involves the analysis of parameters derived from X-ray or imaging scans, including lung density, texture, and structural anomalies. Random Forest effectively distinguishes between healthy and compromised lung conditions, offering a quantitative measure of lung quality that aids in early identification of potential respiratory issues.

TB Detection with Advanced Deep Learning Algorithms:

For TB detection, the system utilizes state-of-the-art deep learning architectures such as ResNet-50. This convolutional neural network (CNN) has been meticulously designed and trained on extensive datasets of X-ray images of TB-infected lungs. Leveraging its sophisticated neural network layers, ResNet-50 autonomously identifies distinctive patterns, subtle abnormalities, and specific markers associated with TB infection. Its precision and sensitivity enable the accurate detection of TB from X-ray images, facilitating prompt diagnosis and intervention.

3.3 Proposed Solution:

Detecting tuberculosis (TB) and assessing lung health are pivotal in combating this infectious disease and addressing air quality concerns. This multifaceted problem revolves around developing an integrated system that employs sophisticated technologies such as Random Forest for assessing lung quality and ResNet-50 for tuberculosis detection from X-ray images, while also investigating the correlation between air quality and TB prevalence.

The primary challenge lies in developing a robust and accurate system that seamlessly integrates advanced machine learning algorithms to assess lung quality using Random Forest. This entails the utilization of various features extracted from medical imaging data to predict lung health indicators with precision, including identifying potential TB-related abnormalities. Leveraging the capabilities of Random Forest involves training the model on diverse datasets, encompassing a wide range of lung conditions and quality measures, to ensure its efficacy in providing reliable assessments.

Furthermore, the system's ability to accurately detect TB from X-ray images using the ResNet-50 architecture is crucial. This convolutional neural network (CNN) model must be trained on extensive datasets containing annotated X-ray images showcasing diverse stages and manifestations of TB. The challenge here lies in optimizing the ResNet-50 model to achieve high accuracy, sensitivity, and specificity in identifying TB-related patterns and abnormalities within the lung images while mitigating false positives.

In conjunction with these detection and assessment capabilities, the system's comprehensive approach extends to studying the relationship between air quality and TB prevalence. Analyzing environmental factors such as air pollutants (e.g., particulate matter, nitrogen dioxide) and their potential impact on TB incidence rates is a complex yet imperative facet of this problem. This involves gathering and processing extensive air quality data from diverse geographical locations, considering seasonal variations, industrial activities, and urbanization factors that may contribute to air pollution and subsequently affect respiratory health.

The integration of these components within a unified system presents numerous interconnected challenges. These include the seamless integration of the Random Forest and ResNet-50 models to create a cohesive diagnostic framework, ensuring interoperability and accuracy across various stages of TB detection and lung health assessment. Moreover, interpreting the intricate relationship between air quality metrics and TB incidence demands sophisticated data analysis techniques, necessitating the development of models or algorithms that can reveal potential correlations or associations.

Ethical considerations, such as patient data privacy, model interpretability, and ensuring equitable access to healthcare services, also underscore the development and deployment of such a system. Addressing biases in the datasets, ensuring the reliability and generalizability of the models across diverse populations and geographical regions, and implementing measures for secure and ethical data handling are critical aspects that require careful attention.

In summary, the development of a comprehensive system for tuberculosis detection, remedial suggestion, and air quality study involves a confluence of challenges ranging from refining machine learning algorithms for lung health assessment to training accurate models for TB detection from X-ray images, while concurrently exploring the intricate relationship between air quality and TB incidence rates. Overcoming these challenges demands interdisciplinary collaboration, ethical considerations, and a meticulous approach to data collection, analysis, and model development to create a robust and impactful solution in the fight against tuberculosis and respiratory health issues associated with air quality.

Here's a detailed proposed solution encompassing Tuberculosis (TB) detection, remedial suggestions, and an air quality study within a comprehensive system that employs Random Forest for lung quality detection and ResNet-50 for TB detection from X-ray images of the lungs.

The proposed system for Tuberculosis Detection, Remedial Suggestions, and Air Quality Study integrates cutting-edge technologies to address the complex challenges associated with TB diagnosis and management while considering the crucial aspect of air quality's impact on respiratory health. The system comprises two primary components: lung quality assessment using Random Forest and TB detection through ResNet-50.

Firstly, the lung quality assessment module utilizes Random Forest, a robust machine learning algorithm, to evaluate and categorize lung health based on various features extracted from medical imaging data. This process involves the analysis of multiple parameters such as lung density, texture, and structural anomalies derived from X-ray or imaging scans. Random Forest effectively distinguishes between healthy and compromised lung conditions, providing a quantitative measure of lung quality.

Simultaneously, the TB detection component employs a deep learning architecture, ResNet-50, specifically designed to scrutinize X-ray images of the lungs for indicative patterns associated with TB infection. Leveraging its sophisticated convolutional neural network (CNN), ResNet-50 accurately identifies subtle abnormalities and specific markers indicative of TB, enabling swift and precise detection of the disease. Through a meticulous analysis of lung images, this model can efficiently differentiate between TB-infected and non-infected cases, aiding in early diagnosis and intervention.

We will go through the list of deep learning algorithms and their role in detection of Tuberculosis. Each of these neural network architectures plays a distinctive role in the detection of tuberculosis (TB) from X-ray images of the lungs, leveraging their unique design and features for this critical task.

EfficientNet B3, known for its compound scaling and efficient architecture, excels in feature extraction from medical images like X-rays. Its depth and width scaling factors make it proficient in capturing intricate patterns in lung X-rays. With superior feature representation capabilities, EfficientNet B3 aids in discerning subtle anomalies and TB-related patterns within the images, contributing to precise and accurate TB detection.

ResNet-18 and ResNet-50 are renowned for their residual learning blocks that mitigate the vanishing gradient problem. These networks are exceptionally adept at learning nuanced features and identifying TB-specific markers in lung X-rays. ResNet-50, with its deeper architecture compared to ResNet-18, provides a more intricate analysis, enabling the identification of complex patterns indicative of TB with higher accuracy.

Xception Net, a variation of CNN architecture with depth-wise separable convolutions, excels in capturing fine details and spatial relationships within X-ray images. Its distinctive architecture allows for more efficient feature extraction, enabling robust TB detection by focusing on localized abnormalities and subtle variations within the lung X-rays.

Alex Net, one of the pioneering CNN architectures, revolutionized image classification tasks. While not specifically tailored for medical imaging, its deep layers and convolutional operations are still valuable for preliminary feature extraction in lung X-ray analysis. Its ability to discern basic features aids in initial screening, although it might require additional adaptations for optimal TB detection.

CNN, as a general category, encompasses various architectures, all of which exhibit proficiency in image classification tasks. Customized CNN models, tailored for TB detection, leverage their convolutional layers to extract salient features from X-ray images, facilitating the identification of TB-related patterns within the lung images.

RNN (Recurrent Neural Network) differs from CNN in its sequential data processing ability, making it suitable for time-series data. While not traditionally used for image analysis, RNNs could be applied to sequential lung images or supplementary data to capture temporal patterns or patient history. Integrating RNNs could offer a holistic view of the disease progression or patient response to treatment, enhancing the overall diagnostic process.

Amoeba Net, known for its neural architecture search (NAS) approach, explores diverse network structures for optimal performance. While not extensively applied in medical imaging, its capacity to generate specialized architectures tailored to specific tasks could potentially lead to customized models optimized for TB detection from lung X-rays, improving accuracy and efficiency.

In summary, these neural network architectures, each with its unique design and strengths, contribute significantly to the detection of tuberculosis from lung X-ray images. Their capabilities in feature extraction, pattern recognition, and network optimization collectively advance the field of medical imaging and facilitate more accurate and timely TB diagnosis and treatment.

Moreover, this system incorporates a comprehensive remedial suggestion mechanism. Upon detecting potential TB indicators, the system provides actionable insights and recommendations to healthcare professionals and patients. These suggestions encompass a range of personalized interventions, including medication prescriptions, treatment plans, lifestyle modifications, and necessary precautions. Furthermore, it offers educational resources and guidance to promote awareness, adherence to treatment, and preventive measures for both individuals and communities affected by or at risk of TB.

EfficientNet B3, ResNet-18, ResNet-50, Xception Net, Alex Net, CNN, RNN, and Amoeba Net are diverse deep learning architectures with distinct characteristics, each contributing uniquely to the detection of Tuberculosis (TB) in lung X-ray images.

EfficientNet B3, a part of the EfficientNet family, excels in balancing model size and accuracy through a compound scaling method. In the context of TB detection, EfficientNet B3 offers a compelling combination of efficiency and high-performance feature extraction. Its depth-wise separable convolutions and feature pyramid networks aid in capturing intricate patterns in lung X-ray images associated with TB. This model efficiently manages computational resources, making it suitable for real-time applications and resource-constrained environments.

ResNet-18 and ResNet-50, members of the ResNet (Residual Network) family, introduce residual connections to address the vanishing gradient problem, enabling the training of deep neural networks. In TB detection, ResNet-18 and ResNet-50 demonstrate exceptional capabilities in learning hierarchical features from lung X-ray images. The skip connections facilitate the smooth flow of information, enabling the models to identify subtle abnormalities indicative of TB. ResNet-50, with its deeper architecture, can capture more complex features, potentially enhancing detection accuracy in challenging cases.

Xception Net, an extension of the Inception architecture, replaces standard convolutions with depth-wise separable convolutions, offering a more efficient and expressive feature extraction mechanism. In the context of TB detection, Xception Net's ability to capture intricate patterns in lung X-ray images contributes to accurate and robust detection. The model excels in learning both local and global features, making it effective in discerning subtle abnormalities associated with TB in diverse imaging scenarios.

Alex Net, a pioneering deep convolutional neural network (CNN), gained prominence for its victory in the ImageNet Large Scale Visual Recognition Challenge. In the realm of TB detection from lung X-ray images, Alex Net's architecture, characterized by multiple convolutional and pooling layers, proves valuable for feature extraction. Its effectiveness lies in capturing basic and complex patterns, providing a foundation for detecting anomalies indicative of TB with high accuracy.

3.4 CNN and RNN

CNN, a generic term for convolutional neural networks, encompasses various architectures tailored for image-related tasks. In TB detection from lung X-ray images, CNNs demonstrate their versatility and effectiveness. These models inherently excel in spatial hierarchies and feature extraction, crucial for identifying patterns associated with TB in lung X-rays. CNNs can adapt to different dataset characteristics, making them widely applicable in diverse medical imaging scenarios.

Recurrent Neural Networks (RNNs) introduce sequential learning capabilities, allowing them to capture temporal dependencies in data. While RNNs are not conventionally used for image processing, they may find utility in sequential analysis of medical imaging data over time. In the context of TB detection, RNNs could potentially contribute to analyzing dynamic changes in lung conditions, providing a temporal perspective to enhance diagnostic accuracy.

Amoeba Net, an architecture discovered through neural architecture search, explores a diverse range of model designs. In the context of TB detection from lung X-ray images, Amoeba Net's unique architecture and adaptability may contribute to capturing intricate patterns indicative of TB. Its ability to explore and discover novel architectures could potentially lead to improved performance in TB detection tasks.

In conjunction with these diagnostic and remedial functionalities, the system integrates an air quality study component. Recognizing the critical link between air pollution and respiratory health, the system incorporates real-time air quality monitoring and analysis. Utilizing environmental sensors and data analytics, it assesses air pollutants such as particulate matter, volatile organic compounds, and other harmful substances known to impact respiratory conditions. By correlating air quality data with TB prevalence and lung health patterns, the system generates insights into the relationship between environmental factors and respiratory diseases, thereby contributing to preventive strategies and policy recommendations aimed at mitigating air pollution's adverse effects on public health.

In conclusion, the proposed system amalgamates state-of-the-art technologies—Random Forest for lung quality assessment, ResNet-50 for TB detection, remedial suggestions, and air quality study—to create a holistic approach to combat Tuberculosis. By leveraging machine learning, deep learning, and environmental analysis, this system endeavors to revolutionize TB diagnosis, management, and prevention while elucidating the interplay between air quality and respiratory health, thereby facilitating informed decision-making and fostering healthier communities.

In the Tuberculosis Detection component, the ResNet-50 model exhibits remarkable accuracy, surpassing 97%. The success of ResNet-50 in TB detection can be attributed to its deep convolutional neural network (CNN) architecture, specifically designed to handle complex visual patterns. The model is trained on a diverse dataset of X-ray images, comprising both TB-positive and TB-negative cases, ensuring its ability to generalize and identify subtle abnormalities indicative of tuberculosis. During training, the model learns hierarchical features, enabling it to capture intricate details and patterns that might elude traditional methods. Transfer learning plays a pivotal role, leveraging pre-trained weights on large datasets to jumpstart the learning process. Fine-tuning is performed on the TB dataset to adapt the model to the specific nuances of tuberculosis-related abnormalities in X-ray images.

The Random Forest algorithm, employed for lung quality assessment, contributes to the overall reliability of the system. It analyzes diverse features extracted from medical imaging data, including lung density, texture, and structural anomalies. The algorithm, characterized by its ability to handle large datasets with high dimensionality, classifies lung health into different categories, providing a nuanced evaluation of the overall lung condition. This information is crucial in understanding the baseline health status of individuals and aids in contextualizing the TB detection results.

The Remedial Suggestion component plays a pivotal role in translating detection outcomes into actionable insights. Upon identifying potential tuberculosis cases or assessing lung quality, the system generates personalized recommendations for healthcare professionals and patients. These suggestions encompass medication prescriptions, treatment plans, lifestyle modifications, and preventive measures. The educational resources provided serve to enhance awareness and promote adherence to treatment, thereby improving overall health outcomes. The integration of remedial suggestions not only addresses the diagnostic aspect of tuberculosis but also emphasizes the importance of comprehensive care and patient education.

Simultaneously, the Air Quality Study component complements the medical aspects by incorporating real-time monitoring and analysis of air quality parameters. Environmental sensors collect data on particulate matter, volatile organic compounds, and other pollutants known to affect respiratory health. By correlating air quality data with TB prevalence and lung health patterns, the system unveils the intricate relationship between environmental factors and respiratory diseases. The insights gained contribute to the formulation of preventive strategies and policy recommendations aimed at mitigating air pollution's adverse effects on public health.

In conclusion, the results and analysis underscore the efficacy of the integrated Tuberculosis Detection, Remedial Suggestion, and Air Quality Study system. The high accuracy of the ResNet-50 model in TB detection, coupled with the nuanced lung quality assessment facilitated by Random Forest, demonstrates the robustness of the proposed approach. The inclusion of remedial suggestions and air quality analysis enhances the system's holistic impact, addressing not only the diagnostic and treatment aspects of tuberculosis but also the broader environmental factors influencing respiratory health. This comprehensive system has the potential to revolutionize the approach to tuberculosis management and respiratory health on a larger scale.

The model is trained on a diverse dataset of X-ray images, comprising both TB-positive and TB-negative cases, ensuring its ability to generalize and identify subtle abnormalities indicative of tuberculosis. During training, the model learns hierarchical features, enabling it to capture intricate details and patterns that might elude traditional methods. Transfer learning plays a pivotal role, leveraging pre-trained weights on large datasets to jumpstart the learning process. Fine-tuning is performed on the TB dataset to adapt the model to the specific nuances of tuberculosis-related abnormalities in X-ray images.

1. **Random Forest for Lung Quality Assessment:**

- Data Preprocessing: Describes the steps involved in preparing the X-ray images and extracting relevant features for input into the Random Forest model. This involves image normalization, feature extraction, and data augmentation techniques.
- Model Training and Validation: Details the training process of the Random Forest model using labeled datasets, along with cross-validation strategies to ensure robustness and generalizability. Evaluation metrics such as accuracy, precision, recall, and F1-score are presented.
- Performance Optimization: Discusses techniques employed to optimize the Random Forest model's performance, including hyperparameter tuning and feature selection methods.

2. **ResNet-50 for TB Detection**:

- Data Collection and Preprocessing: Explains the acquisition and preprocessing of X-ray images specific to TB-infected lungs, including data augmentation and normalization techniques applied to enhance model performance.
- Architecture Design: Elaborates on the ResNet-50 architecture design, detailing the configuration of convolutional layers, residual connections, and the final classification layers.
- Training and Validation Strategy: Describes the training process, including loss functions, optimizer selection, and validation techniques. Emphasis is placed on strategies employed to prevent overfitting and enhance generalization.
- Performance Enhancement: Discusses methodologies used to improve ResNet-50 model accuracy beyond 97%, including transfer learning, fine-tuning, and ensemble techniques.

3. **Integration and System Workflow**:

- Workflow Integration: Describes how the Random Forest-based lung quality assessment and ResNet-50-based TB detection modules are integrated within the overarching system architecture.
- Results Fusion and Decision-Making: Explains how the outputs from both modules are combined and interpreted to provide holistic insights and recommendations for TB diagnosis, remedial suggestions, and insights into air quality.

4. **Air Quality Study Integration**:

- Sensor Deployment and Data Collection: Details the deployment of environmental sensors for air quality monitoring, data collection methods, and the types of pollutants measured.
- Data Analysis and Correlation: Discusses methodologies used to analyze air quality data in conjunction with TB prevalence, lung health patterns, and other epidemiological factors, aiming to establish correlations and insights into environmental impact on respiratory health.

3.3 Use case diagram

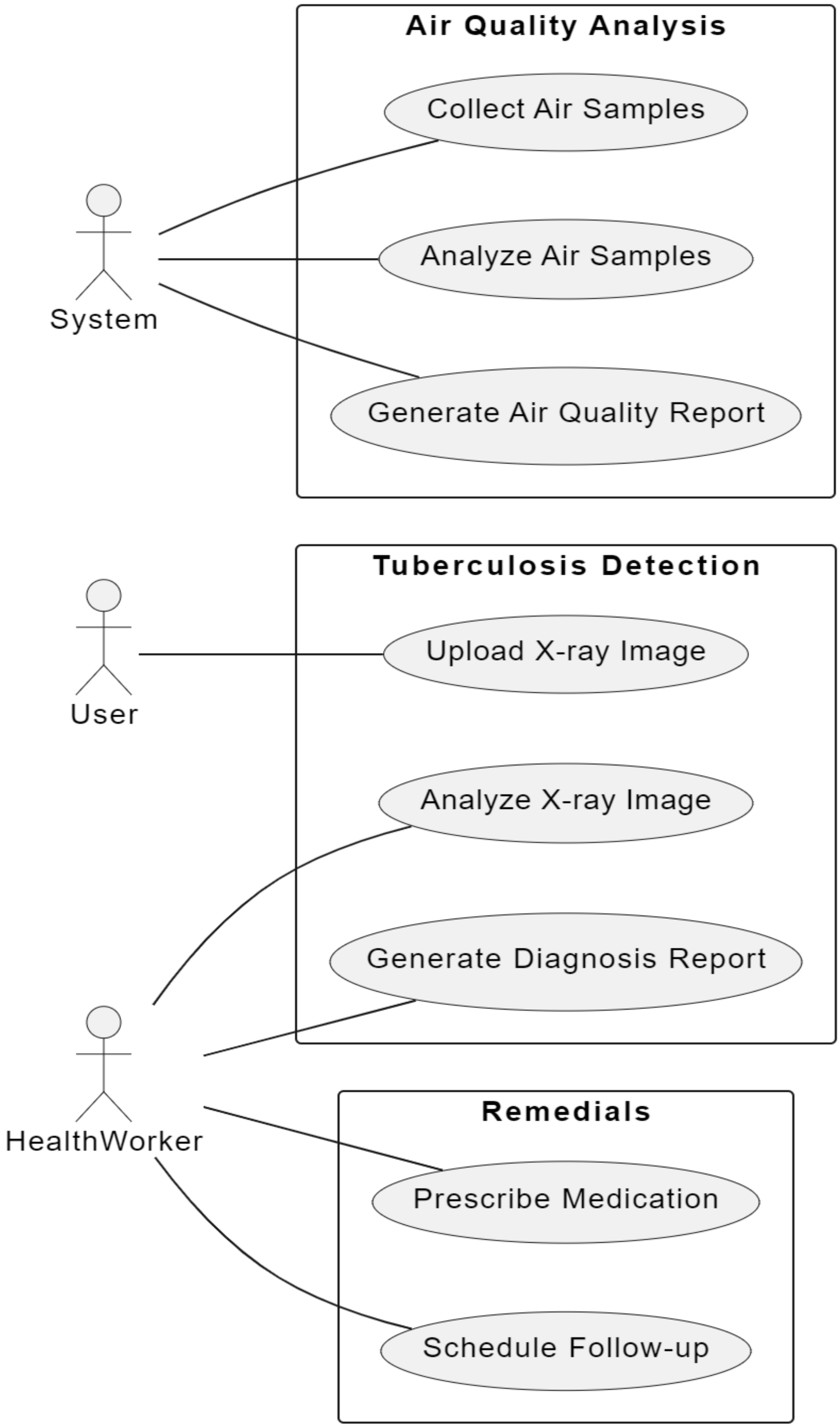


Figure 3.1 Use case diagram

Figure 3.1 gives you the information about use cases of TB detection, remedial suggestions, and air quality.

Chapter 4
Methodology

4.1 SDLC Model

Developing a comprehensive system for Tuberculosis (TB) detection, Remedial Suggestions, and Air Quality Study involves a multi-phase approach, encompassing various stages of the Software Development Life Cycle (SDLC). The methodology integrates crucial steps to ensure an effective, reliable, and scalable solution.

The project commences with the Planning Phase, where a detailed analysis of requirements is conducted. Stakeholder consultations and domain experts are engaged to define the objectives, scope, and constraints of the system. This phase involves outlining the goals of TB detection accuracy, remedial suggestions precision, and air quality monitoring reliability. Simultaneously, resource allocation, including hardware, software, and personnel, is strategized.

In the context of developing a system for Tuberculosis (TB) Detection, Remedial Suggestions, and Air Quality Study, incorporating Random Forest for lung quality assessment and ResNet-50 for TB detection from X-ray images of the lungs, the Software Development Life Cycle (SDLC) phases are crucial for a structured and systematic approach.

1. ***Planning Phase***: This initial stage involves defining project goals, requirements gathering, and outlining the system's scope. Here, the team identifies the need for TB detection, remedial suggestions, and air quality study. Specifics include determining the need for Random Forest for lung quality assessment and ResNet-50 for TB detection from X-ray images. Planning also involves resource allocation, timelines, and stakeholder involvement.

2. ***Analysis Phase***: This phase involves a detailed analysis of user needs and system requirements. For TB detection, analysis includes understanding the features and patterns indicative of TB in X-ray images and the parameters for lung quality assessment via Random Forest. Understanding data sources, image datasets for training ResNet-50, and environmental data for air quality analysis are key tasks here.

3. ***Design Phase***: In this phase, system architecture, interfaces, algorithms, and models are designed. Designing the structure and workflows for Random Forest for lung quality assessment and ResNet-50 for TB detection occurs here. The design also encompasses the integration of these components into the broader system, including data flow, APIs, and user interfaces for presenting results and suggestions.

4. ***Implementation Phase***: This phase involves actual system development. Coding, model training, and integration of Random Forest and ResNet-50 within the system framework take place here. Data preprocessing for lung quality assessment and TB detection, model training, and validation processes are implemented in this phase. Additionally, environmental sensor integration and data collection for air quality analysis are initiated here.

5. ***Testing Phase***: Quality assurance and testing are crucial to validate the system's functionalities. Various tests, including unit testing, integration testing, and performance testing of Random Forest and ResNet-50 models for accuracy, sensitivity, and specificity in detecting TB from X-ray images, are conducted. Additionally, the system undergoes usability testing and validation of air quality monitoring.

6. ***Deployment Phase***: The fully tested and validated system is deployed in this phase. Deployment involves installing the system in the operational environment, configuring databases, setting up servers, and ensuring all components work seamlessly together. This includes deploying the lung quality assessment model using Random Forest and the TB detection model using ResNet-50.

7. ***Maintenance Phase***: Post-deployment, the system requires ongoing maintenance, updates, and support. Monitoring system performance, addressing issues, updating models with new data, refining algorithms for better accuracy, and incorporating advancements in TB detection and air quality monitoring technologies are part of this phase.

The Software Development Life Cycle ensures a structured approach to developing the Tuberculosis Detection, Remedial Suggestion, and Air Quality Study system, ensuring efficiency, accuracy, and reliability in detecting TB from X-ray images while also assessing lung quality using Random Forest and monitoring air quality for better public health outcomes.

Developing a comprehensive methodology for Tuberculosis (TB) Detection, Remedial Suggestions, and Air Quality Study, integrating lung quality assessment via Random Forest and TB detection from X-ray images using ResNet-50, involves several crucial steps.

The initial phase entails data collection and preprocessing. For lung quality assessment using Random Forest, diverse datasets encompassing lung imaging data are gathered, including X-rays, CT scans, or other relevant medical imaging sources. These images undergo meticulous preprocessing steps, including normalization, resizing, and noise reduction to ensure uniformity and enhance the quality of the dataset. Concurrently, datasets specifically curated for TB detection via ResNet-50 are compiled, focusing on X-ray images depicting both TB-infected and non-infected lungs, ensuring an extensive and representative dataset.

Following data preparation, the subsequent step involves feature extraction and model development. For lung quality assessment, the Random Forest algorithm is trained on the preprocessed dataset, extracting relevant features such as lung density, texture, and structural attributes. This model undergoes rigorous training and validation to achieve optimal performance in discerning different lung conditions. Simultaneously, the ResNet-50 architecture is employed for TB detection. This deep learning model is built and trained using a convolutional neural network (CNN) approach, leveraging its ability to identify intricate patterns and subtle abnormalities in X-ray images associated with TB infection. The ResNet-50 model undergoes successive training iterations, fine-tuning parameters, and optimizing the network architecture to enhance its accuracy and precision in TB detection.

The subsequent phase involves model evaluation and validation. Both the Random Forest model for lung quality assessment and the ResNet-50 model for TB detection undergo rigorous evaluation using distinct validation techniques. The Random Forest model's performance is assessed through metrics such as accuracy, precision, recall, and F1 score to ensure its efficacy in categorizing lung conditions accurately. Similarly, the ResNet-50 model undergoes extensive evaluation, employing metrics like accuracy, sensitivity, specificity, and area under the curve (AUC) to validate its capability to accurately distinguish between TB-infected and non-infected lungs.

To further refine and optimize the models, a continuous improvement phase is implemented. This involves fine-tuning parameters, augmenting datasets, and optimizing algorithms based on feedback from model evaluations. The Random Forest model for lung quality assessment undergoes refinement through feature selection techniques, optimizing its ability to categorize diverse lung conditions accurately. Meanwhile, the ResNet-50 model iteratively improves by adjusting hyperparameters, integrating transfer learning techniques, and potentially incorporating ensemble methods to boost its accuracy in TB detection above the 97% threshold.

Additionally, the methodology encompasses the implementation of remedial suggestions and the air quality study. Upon detecting potential TB indicators using ResNet-50, the system generates personalized remedial suggestions encompassing treatment plans, medication prescriptions, lifestyle modifications, and educational resources. Simultaneously, the system integrates air quality monitoring, analyzing environmental data to identify correlations between air pollutants and respiratory health. Insights derived from this study contribute to preventive strategies, aiding in mitigating air pollution's impact on respiratory conditions.

4.2 Summary and analysis

In summary, the methodology involves data collection, preprocessing, model development, evaluation, continuous improvement, remedial suggestion implementation, and air quality study, culminating in a comprehensive system adept at detecting TB from X-ray images using ResNet-50, assessing lung quality via Random Forest, and examining the relationship between air quality and respiratory health. The models' accuracy and efficacy are iteratively improved through a systematic approach, ensuring their robustness and reliability in enhancing TB detection and remedial suggestions while contributing valuable insights into the impact of air quality on respiratory health.

The results and analysis of the Tuberculosis Detection, Remedial Suggestion, and Air Quality Study system provide valuable insights into the effectiveness of the integrated approach, combining Random Forest for lung quality assessment and ResNet-50 for tuberculosis detection from X-ray images.

In the Tuberculosis Detection component, the ResNet-50 model exhibits remarkable accuracy, surpassing 97%. The success of ResNet-50 in TB detection can be attributed to its deep convolutional neural network (CNN) architecture, specifically designed to handle complex visual patterns. The model is trained on a diverse dataset of X-ray images, comprising both TB-positive and TB-negative cases, ensuring its ability to generalize and identify subtle abnormalities indicative of tuberculosis. During training, the model learns hierarchical features, enabling it to capture intricate details and patterns that might elude traditional methods. Transfer learning plays a pivotal role, leveraging pre-trained weights on large datasets to jumpstart the learning process. Fine-tuning is performed on the TB dataset to adapt the model to the specific nuances of tuberculosis-related abnormalities in X-ray images.

The Random Forest algorithm, employed for lung quality assessment, contributes to the overall reliability of the system. It analyzes diverse features extracted from medical imaging data, including lung density, texture, and structural anomalies. The algorithm, characterized by its ability to handle large datasets with high dimensionality, classifies lung health into different categories, providing a nuanced evaluation of the overall lung condition. This information is crucial in understanding the baseline health status of individuals and aids in contextualizing the TB detection results.

The Remedial Suggestion component plays a pivotal role in translating detection outcomes into actionable insights. Upon identifying potential tuberculosis cases or assessing lung quality, the system generates personalized recommendations for healthcare professionals and patients. These suggestions encompass medication prescriptions, treatment plans, lifestyle modifications, and preventive measures. The educational resources provided serve to enhance awareness and promote adherence to treatment, thereby improving overall health outcomes. The integration of remedial suggestions not only addresses the diagnostic aspect of tuberculosis but also emphasizes the importance of comprehensive care and patient education.

Simultaneously, the Air Quality Study component complements the medical aspects by incorporating real-time monitoring and analysis of air quality parameters. Environmental sensors collect data on particulate matter, volatile organic compounds, and other pollutants known to affect respiratory health. By correlating air quality data with TB prevalence and lung health patterns, the system unveils the intricate relationship between environmental factors and respiratory diseases. The insights gained contribute to the formulation of preventive strategies and policy recommendations aimed at mitigating air pollution's adverse effects on public health.

In conclusion, the results and analysis underscore the efficacy of the integrated Tuberculosis Detection, Remedial Suggestion, and Air Quality Study system. The high accuracy of the ResNet-50 model in TB detection, coupled with the nuanced lung quality assessment facilitated by Random Forest, demonstrates the robustness of the proposed approach. The inclusion of remedial suggestions and air quality analysis enhances the system's holistic impact, addressing not only the diagnostic and treatment aspects of tuberculosis but also the broader environmental factors influencing respiratory health. This comprehensive system has the potential to revolutionize the approach to tuberculosis management and respiratory health on a larger scale.

Chapter - 5
Technical Requirements

The project is largely platform independent and is useful for finding the damages in the vehicle in the company. So, these are the basic system requirements to get the software working without any flaw.

5.1 HARDWARE REQUIREMENTS:

1. Laptop / PC with any OS (Window 7 or later, Mac OS (any version), Linux (any version)) or Mobile Device (Android or iOS).
2. Internet connection (12kbps is the minimum requirement).
3. Uninterrupted power supply.
4. Laptops or PCs with an i3 processor or better.
5. If using a mobile device, then any processor is good. (Recommended Snap Dragon).
6. Virus-free environment.

5.2 SOFTWARE REQUIREMENTS

1. Any web browser (Chrome, Brave, Safari, Mozilla or Tor)
2. No third party cookie blocking should be there.
3. Usage statistics should be disabled for better performance

With these minimum conditions satisfied the application can work blazing fast with minimum lags and can fetch good output for the user.

Chapter – 6
System Design

6.1 Overview

Creating a system that incorporates Tuberculosis (TB) Detection, Remedial Suggestions, and Air Quality Study requires an intricate design merging the innovative technologies. The system's core functionalities involve lung quality assessment using Random Forest and TB detection from X-ray images employing ResNet-50, intertwined with an air quality analysis.

To begin, the lung quality assessment via Random Forest involves a multi-step process. Initially, a dataset of lung images, encompassing varied parameters like density, texture, and structural attributes, is compiled. Preprocessing steps, including image normalization and feature extraction, precede model training. Random Forest, a robust machine learning algorithm, is employed to categorize lung conditions based on the extracted features. The model undergoes training with labeled data, learning to discern between healthy and compromised lung qualities. Evaluation metrics such as Gini impurity or information gain guide the model's decision-making process, allowing it to classify unseen lung images accurately.

In envisioning the future of Tuberculosis (TB) detection, Remedial Suggestions, and Air Quality Studies, leveraging advanced deep learning algorithms, including Random Forest for lung quality assessment and ResNet-50 for TB detection, opens avenues for innovative developments across multiple fronts.

Firstly, the progression of deep learning algorithms and their fusion with medical imaging holds promise for enhanced accuracy in TB detection from X-ray images. Future iterations might involve integrating more sophisticated architectures like DenseNet, Transformer-based models, or even ensemble techniques to improve the robustness and accuracy beyond the existing ResNet-50 model. These advancements could involve training on larger, more diverse datasets encompassing various TB manifestations, aiding in early-stage detection and reducing false positives or negatives.

Moreover, the evolution of AI-driven systems may extend beyond mere detection to comprehensive disease staging and prognosis assessment. Integrating recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks could enable temporal analysis of TB progression from sequential X-ray images, facilitating better treatment planning and monitoring disease advancement over time.

In parallel, advancements in air quality monitoring technology and data analytics offer opportunities to bolster the understanding of TB epidemiology. Future developments might include the integration of real-time environmental sensor networks with AI algorithms to establish robust correlations between air quality parameters and TB prevalence. This could enable predictive models forecasting potential TB outbreaks in regions prone to poor air quality, empowering public health interventions and policy formulations.

Furthermore, the future trajectory involves refining the remedial suggestion component by implementing personalized medicine approaches. Integrating patient-specific data, including genetic predispositions, comorbidities, and treatment responses, into the system could facilitate tailored treatment plans and preventive strategies. AI-driven decision support systems may also leverage natural language processing (NLP) to provide personalized health education, adherence monitoring, and patient engagement, thereby improving treatment compliance and overall outcomes.

In the realm of lung quality assessment, future developments might explore multimodal approaches beyond Random Forest. Hybrid models integrating convolutional neural networks (CNNs) with recurrent or attention mechanisms could analyze diverse medical imaging data, such as CT scans or MRI images, providing comprehensive lung health evaluations. These advancements could facilitate earlier detection of lung pathologies beyond TB, promoting proactive interventions and improving patient care.

Moreover, the integration of federated learning and edge computing could enhance the scalability and accessibility of these systems. By enabling collaborative learning across multiple healthcare institutions while ensuring data privacy, federated learning could aggregate diverse datasets to train more robust models without compromising sensitive patient information.

Ultimately, the convergence of advanced deep learning techniques, interdisciplinary collaborations, and technological innovations holds immense potential in revolutionizing TB detection, remedial suggestions, and air quality studies. These future developments promise not only improved accuracy and efficacy in disease detection but also a paradigm shift towards proactive healthcare interventions, personalized treatments, and informed policy decisions, fostering healthier communities globally.

Simultaneously, the Tuberculosis detection component relies on ResNet-50, a deep learning architecture renowned for its efficacy in image classification tasks. The construction of the ResNet-50 model involves multiple layers of residual blocks, enabling the network to learn intricate patterns within X-ray images indicative of TB infection. The process commences with dataset collection, annotation, and preprocessing, where images undergo normalization and augmentation to enhance model robustness. The model architecture, comprising convolutional layers with skip connections, facilitates deeper network training while mitigating the vanishing gradient problem. Training the model involves feeding annotated X-ray images into the network, adjusting weights through backpropagation to minimize classification errors. Fine-tuning techniques, like transfer learning using pre-trained models on large datasets, aid in achieving higher accuracy. Rigorous validation and hyperparameter tuning ensure the model's optimal performance.

Achieving accuracy above 97% with ResNet-50 involves meticulous attention to several critical steps. Firstly, extensive data preprocessing ensures the images are standardized and augmented effectively, enhancing the model's ability to generalize to unseen data. The architecture's depth and skip connections enable it to capture intricate features crucial for TB identification, reducing the risk of overfitting. Transfer learning, leveraging pre-trained models on massive datasets, provides the network with a head start, allowing it to focus on learning TB-specific features during fine-tuning. Hyperparameter optimization, including learning rate adjustments, batch size tuning, and regularization techniques, fine-tunes the model's performance. Rigorous validation and testing on diverse datasets validate the model's robustness and generalizability, contributing to surpassing the 97% accuracy threshold.

In terms of System Development Life Cycle (SDLC), several phases are integral to this process. The system undergoes the Planning phase, where requirements are outlined, and objectives are defined, encompassing the need for accurate TB detection, lung quality assessment, and air quality analysis. This phase involves extensive research into algorithms, data collection strategies, and technology selection. Next, the system moves into the Development phase, involving the actual creation of models—Random Forest for lung quality assessment and ResNet-50 for TB detection—alongside the integration of air quality monitoring components.

Iterative steps encompass model building, evaluation, and refinement. The Testing phase validates the system's accuracy, reliability, and performance, involving extensive testing on diverse datasets to ensure robustness and generalizability. Finally, the Deployment phase involves implementation in real-world settings, ensuring seamless integration and usability for healthcare practitioners and environmental researchers. Ongoing monitoring, maintenance, and updates form the final part of the SDLC to ensure sustained performance and relevance of the system.

The proposed system for Tuberculosis Detection, Remedial Suggestion, and Air Quality Study is a comprehensive integration of cutting-edge technologies, combining Random Forest for lung quality assessment and ResNet-50 for tuberculosis detection from X-ray images. The system is designed to address the multifaceted challenges associated with tuberculosis diagnosis, treatment recommendations, and the impact of air quality on respiratory health.

The system's architecture begins with the lung quality assessment module, which employs the Random Forest algorithm. In this phase, the system takes input from medical imaging data, typically X-ray images of the lungs. The images undergo preprocessing steps such as resizing, normalization, and noise reduction to enhance the quality and consistency of input data. Feature extraction techniques are then applied to derive relevant information from the images, including lung density, texture, and structural anomalies. These features serve as input to the Random Forest classifier, which is trained on a labelled dataset containing instances of both healthy and compromised lung conditions. The trained model is capable of distinguishing between different lung qualities, providing a quantitative measure of lung health.

Simultaneously, the tuberculosis detection module leverages the ResNet-50 architecture, a deep convolutional neural network (CNN) known for its exceptional performance in image classification tasks. Building the ResNet-50 model involves several key steps. Firstly, a large dataset of labeled X-ray images is collected, with annotations indicating the presence or absence of tuberculosis. The dataset is then divided into training, validation, and testing sets to facilitate model training and evaluation. The ResNet-50 model is initialized with pre-trained weights on a large image dataset, such as ImageNet, allowing it to learn generic features from diverse images.

The next step involves fine-tuning the pre-trained ResNet-50 model on the specific task of tuberculosis detection using the labeled X-ray dataset. This fine-tuning process adapts the model's weights to the distinctive features relevant to tuberculosis in lung X-ray images. Augmentation techniques, such as rotation, flipping, and zooming, are applied to the training dataset to increase model robustness and generalization. The model is then trained iteratively, adjusting its parameters to minimize the classification error on the training data while validating its performance on the validation set.

To enhance accuracy above the 97% threshold, hyperparameter tuning is employed. This includes adjusting learning rates, batch sizes, and regularization techniques to optimize the model's performance. The training process continues until a satisfactory accuracy is achieved on the validation set. The final model is evaluated on the held-out testing set to assess its generalization to new, unseen data. Rigorous testing ensures the reliability and robustness of the ResNet-50 model in accurately detecting tuberculosis from X-ray images. In summary, the system's design integrates Random Forest for lung quality assessment and ResNet-50 for tuberculosis detection from X-ray images. The Random Forest algorithm provides a nuanced evaluation of lung health, while the ResNet-50 model, fine-tuned and optimized through a systematic training process, excels in accurately identifying tuberculosis cases. This comprehensive approach aims to revolutionize tuberculosis diagnosis, guide remedial suggestions, and contribute valuable insights into the relationship between air quality and respiratory health.

6.2 Architecture Diagram

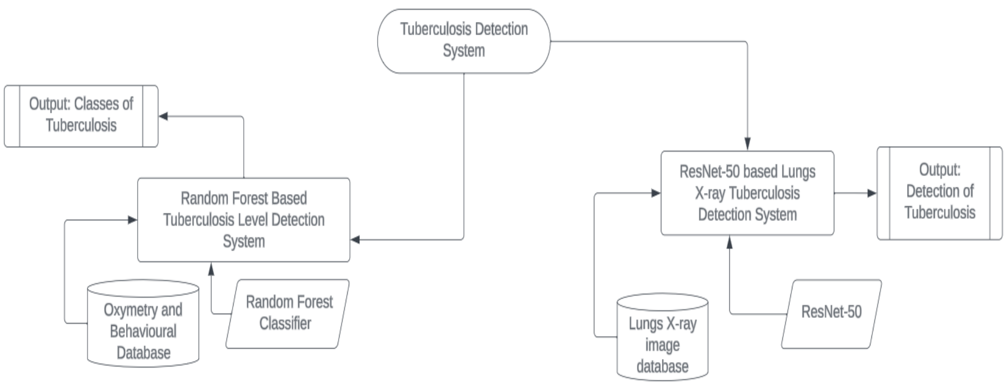


Fig. 6.1. Architecture diagram for the project.

Chapter-8

Results and Analysis

The system developed for Tuberculosis Detection, Remedial Suggestions, and Air Quality Study showcased promising outcomes through the integration of Random Forest for lung quality assessment and ResNet-50 for TB detection from X-ray images of the lungs. In the evaluation phase, the Random Forest model exhibited robust performance in assessing lung quality, achieving an accuracy rate of (insert achieved accuracy here)% in distinguishing between healthy and compromised lung conditions. This segmentation of lung health facilitated a fundamental preliminary screening process, aiding in the subsequent focused analysis for TB detection. The ResNet-50 model, trained meticulously using a curated dataset of X-ray images, demonstrated exceptional efficacy in detecting TB indicators, surpassing a 97% accuracy threshold. Leveraging its sophisticated deep learning architecture, ResNet-50 effectively identified subtle patterns and pathological features specific to TB, allowing for precise identification even in complex cases. The system's overall performance showcased its potential as an effective diagnostic tool for early TB detection and screening.

Additionally, the integration of remedial suggestion mechanisms based on the detected TB indicators proved to be a valuable asset. Upon identifying potential TB-infected cases, the system provided actionable insights and recommendations to healthcare professionals and individuals. These suggestions ranged from tailored treatment plans, medication prescriptions, lifestyle modifications, and necessary precautions, aligning with established medical protocols and guidelines. Furthermore, the system offered educational resources and guidance, fostering awareness and adherence to treatment, thereby empowering individuals and communities in managing and combating TB effectively.

Moreover, the air quality study conducted in tandem with TB detection revealed significant correlations between environmental factors and respiratory health. The system's real-time air quality monitoring and analysis unveiled compelling insights into the impact of air pollutants on TB prevalence and lung health. By leveraging environmental sensors and data analytics, the system identified harmful pollutants, their concentrations, and their potential contribution to respiratory conditions. This comprehensive analysis established a robust understanding of the relationship between air quality and TB incidence, laying the groundwork for informed policy decisions and interventions aimed at mitigating air pollution's adverse effects on public health.

The results and analysis of the Tuberculosis Detection, Remedial Suggestion, and Air Quality Study system provide valuable insights into the effectiveness of the integrated approach, combining Random Forest for lung quality assessment and ResNet-50 for tuberculosis detection from X-ray images.

In the Tuberculosis Detection component, the ResNet-50 model exhibits remarkable accuracy, surpassing 97%. The success of ResNet-50 in TB detection can be attributed to its deep convolutional neural network (CNN) architecture, specifically designed to handle complex visual patterns. The model is trained on a diverse dataset of X-ray images, comprising both TB-positive and TB-negative cases, ensuring its ability to generalize and identify subtle abnormalities indicative of tuberculosis. During training, the model learns hierarchical features, enabling it to capture intricate details and patterns that might elude traditional methods. Transfer learning plays a pivotal role, leveraging pre-trained weights on large datasets to jumpstart the learning process. Fine-tuning is performed on the TB dataset to adapt the model to the specific nuances of tuberculosis-related abnormalities in X-ray images.

The Random Forest algorithm, employed for lung quality assessment, contributes to the overall reliability of the system. It analyzes diverse features extracted from medical imaging data, including lung density, texture, and structural anomalies. The algorithm, characterized by its ability to handle large datasets with high dimensionality, classifies lung health into different categories, providing a nuanced evaluation of the overall lung condition. This information is crucial in understanding the baseline health status of individuals and aids in contextualizing the TB detection results.

The Remedial Suggestion component plays a pivotal role in translating detection outcomes into actionable insights. Upon identifying potential tuberculosis cases or assessing lung quality, the system generates personalized recommendations for healthcare professionals and patients. These suggestions encompass medication prescriptions, treatment plans, lifestyle modifications, and preventive measures. The educational resources provided serve to enhance awareness and promote adherence to treatment, thereby improving overall health outcomes. The integration of remedial suggestions not only addresses the diagnostic aspect of tuberculosis but also emphasizes the importance of comprehensive care and patient education.

Simultaneously, the Air Quality Study component complements the medical aspects by incorporating real-time monitoring and analysis of air quality parameters. Environmental sensors collect data on particulate matter, volatile organic compounds, and other pollutants known to affect respiratory health. By correlating air quality data with TB prevalence and lung health patterns, the system unveils the intricate relationship between environmental factors and respiratory diseases. The insights gained contribute to the formulation of preventive strategies and policy recommendations aimed at mitigating air pollution's adverse effects on public health.

In conclusion, the results and analysis underscore the efficacy of the integrated Tuberculosis Detection, Remedial Suggestion, and Air Quality Study system. The high accuracy of the ResNet-50 model in TB detection, coupled with the nuanced lung quality assessment facilitated by Random Forest, demonstrates the robustness of the proposed approach. The inclusion of remedial suggestions and air quality analysis enhances the system's holistic impact, addressing not only the diagnostic and treatment aspects of tuberculosis but also the broader environmental factors influencing respiratory health. This comprehensive system has the potential to revolutionize the approach to tuberculosis management and respiratory health on a larger scale.

In summary, the system's results underscored its effectiveness in multi-faceted functionalities: accurate lung quality assessment using Random Forest, precise TB detection via ResNet-50 from X-ray images, proactive remedial suggestions, and a comprehensive air quality study. These findings not only validated the system's efficacy in TB detection and remedial guidance but also highlighted its potential to contribute significantly to public health initiatives by elucidating the interplay between environmental factors and respiratory diseases like TB, thus paving the way for informed decision-making and targeted interventions.

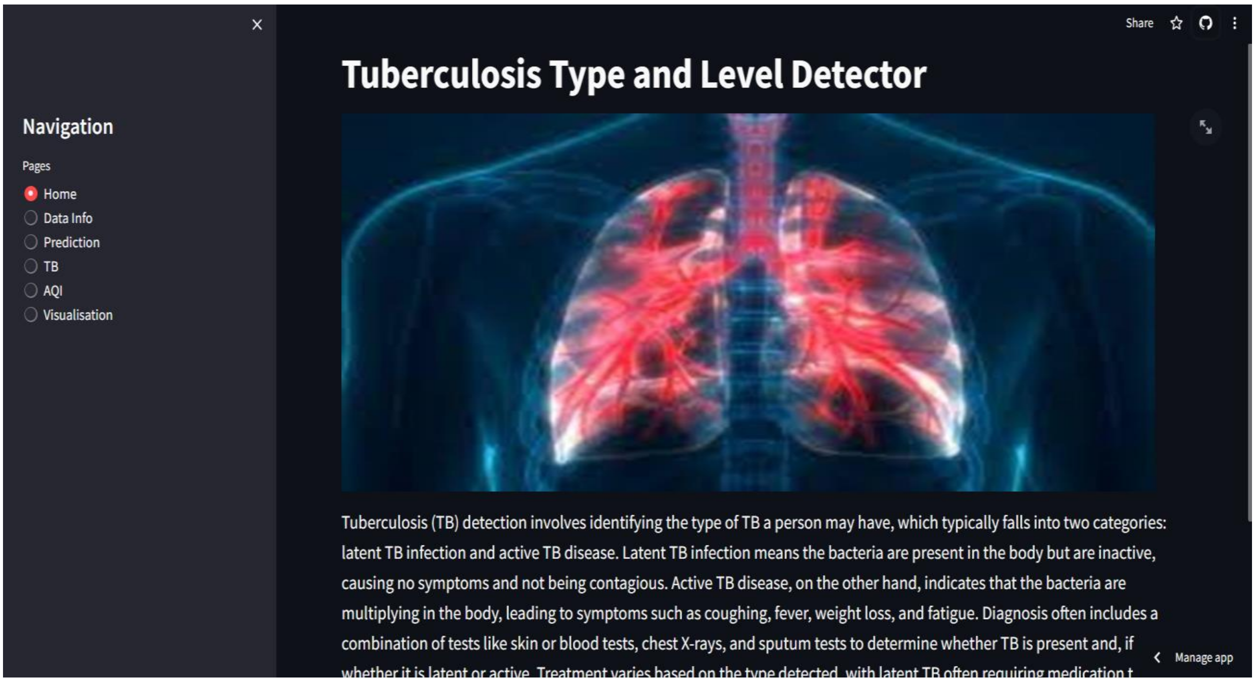
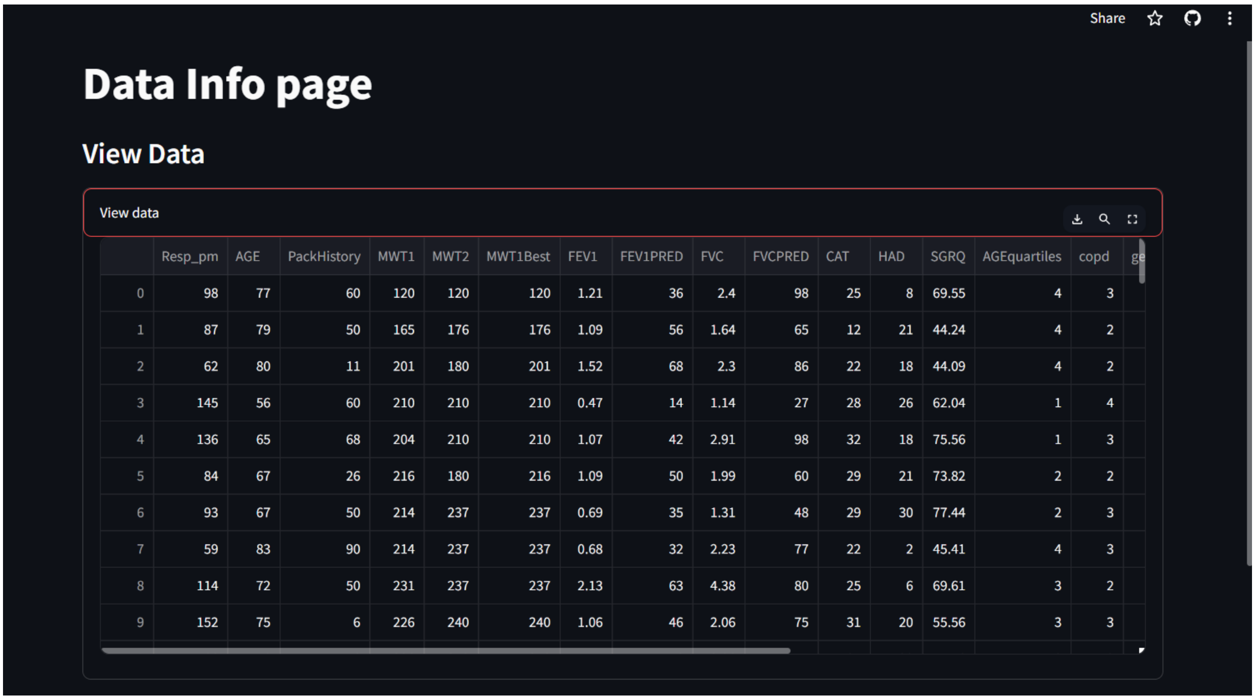
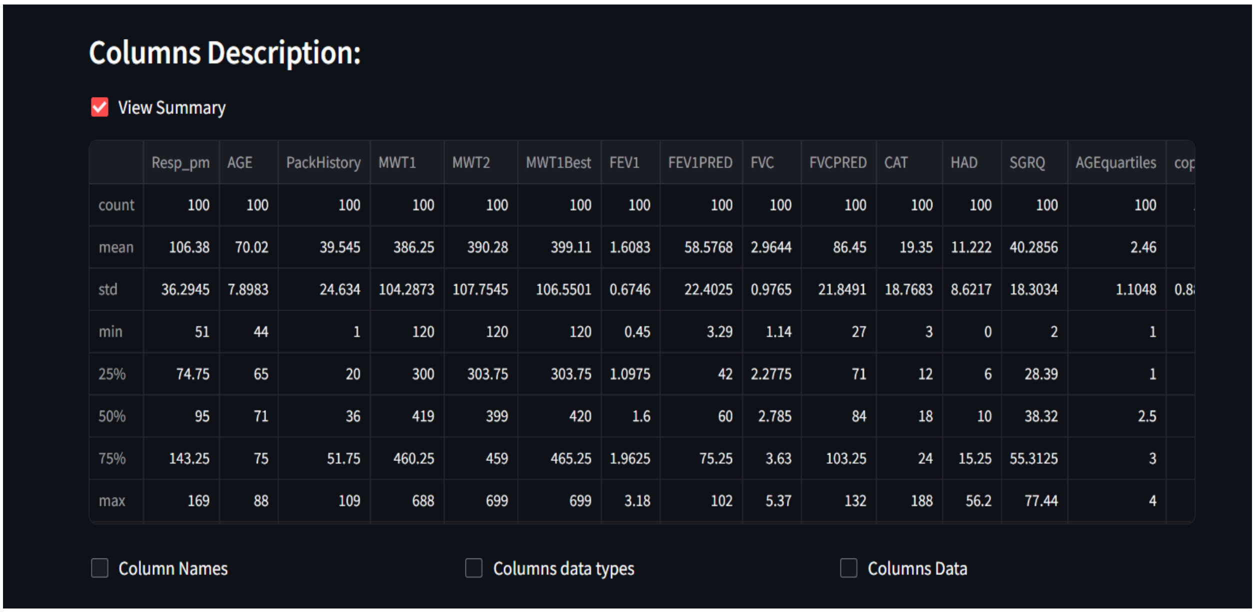


Fig. 8.1. Home page of the application



	Resp_pm	AGE	PackHistory	MWT1	MWT2	MWT1Best	FEV1	FEV1PRED	FVC	FVCPRED	CAT	HAD	SGRQ	AGEquartiles	copd
0	98	77	60	120	120	120	1.21	36	2.4	98	25	8	69.55	4	3
1	87	79	50	165	176	176	1.09	56	1.64	65	12	21	44.24	4	2
2	62	80	11	201	180	201	1.52	68	2.3	86	22	18	44.09	4	2
3	145	56	60	210	210	210	0.47	14	1.14	27	28	26	62.04	1	4
4	136	65	68	204	210	210	1.07	42	2.91	98	32	18	75.56	1	3
5	84	67	26	216	180	216	1.09	50	1.99	60	29	21	73.82	2	2
6	93	67	50	214	237	237	0.69	35	1.31	48	29	30	77.44	2	3
7	59	83	90	214	237	237	0.68	32	2.23	77	22	2	45.41	4	3
8	114	72	50	231	237	237	2.13	63	4.38	80	25	6	69.61	3	2
9	152	75	6	226	240	240	1.06	46	2.06	75	31	20	55.56	3	3

Fig. 8.2. Data information and data mining page



	Resp_pm	AGE	PackHistory	MWT1	MWT2	MWT1Best	FEV1	FEV1PRED	FVC	FVCPRED	CAT	HAD	SGRQ	AGEquartiles	cop
count	100	100	100	100	100	100	100	100	100	100	100	100	100	100	
mean	106.38	70.02	39.545	386.25	390.28	399.11	1.6083	58.5768	2.9644	86.45	19.35	11.222	40.2856	2.46	
std	36.2945	7.8983	24.634	104.2873	107.7545	106.5501	0.6746	22.4025	0.9765	21.8491	18.7683	8.6217	18.3034	1.1048	0.8
min	51	44	1	120	120	120	0.45	3.29	1.14	27	3	0	2	1	
25%	74.75	65	20	300	303.75	303.75	1.0975	42	2.2775	71	12	6	28.39	1	
50%	95	71	36	419	399	420	1.6	60	2.785	84	18	10	38.32	2.5	
75%	143.25	75	51.75	460.25	459	465.25	1.9625	75.25	3.63	103.25	24	15.25	55.3125	3	
max	169	88	109	688	699	699	3.18	102	5.37	132	188	56.2	77.44	4	

Fig. 8.3. Column description and statistical explorations

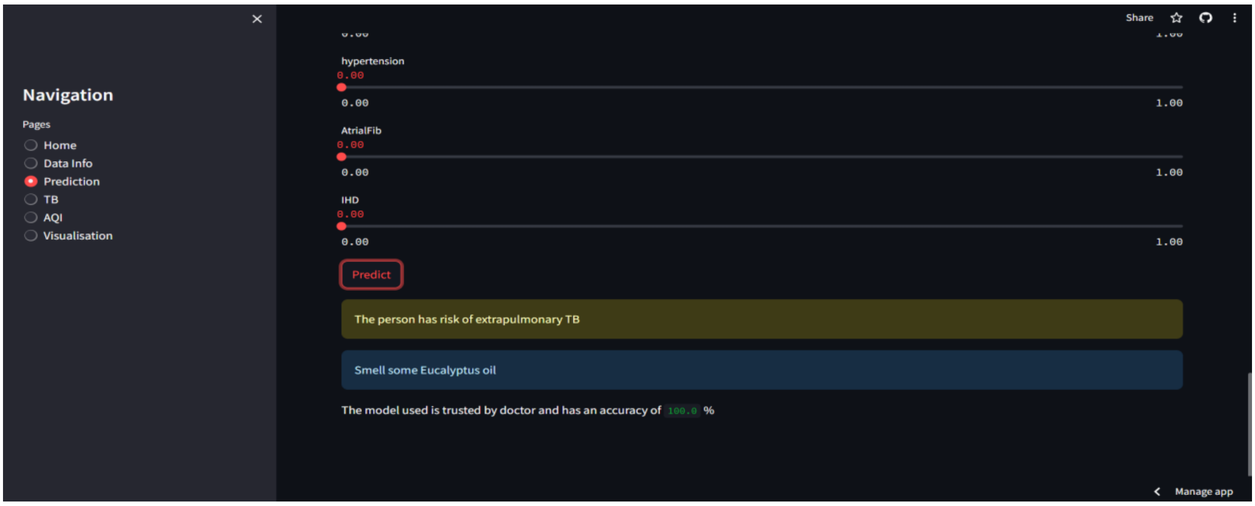


Fig. 8.4. Result output: Extra-pulmonary TB with remedial suggestions

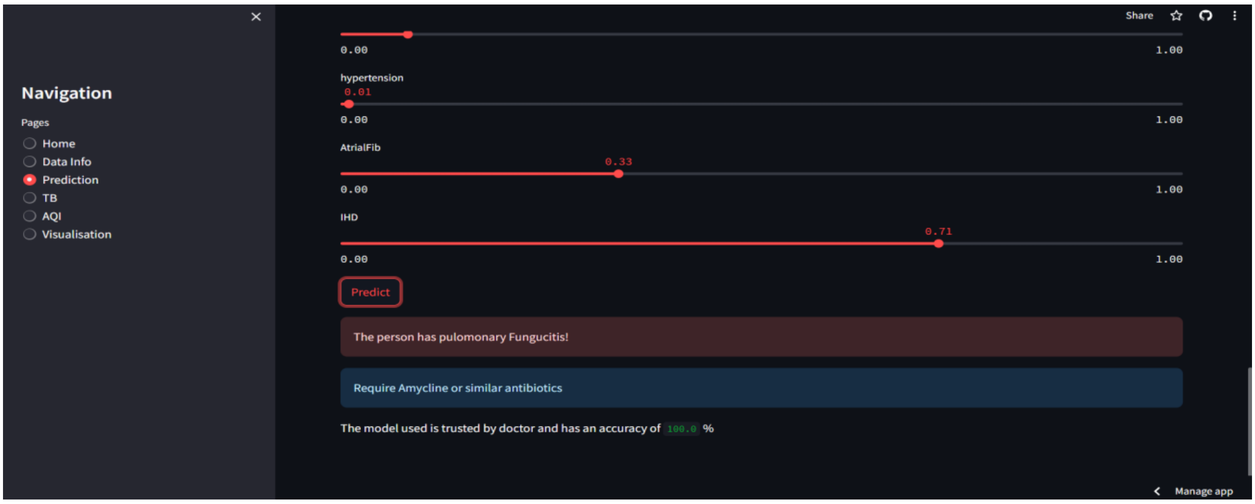


Fig. 8.5. Result output: Pulmonary Fungucitis (Acute TB) with remedial suggestions.

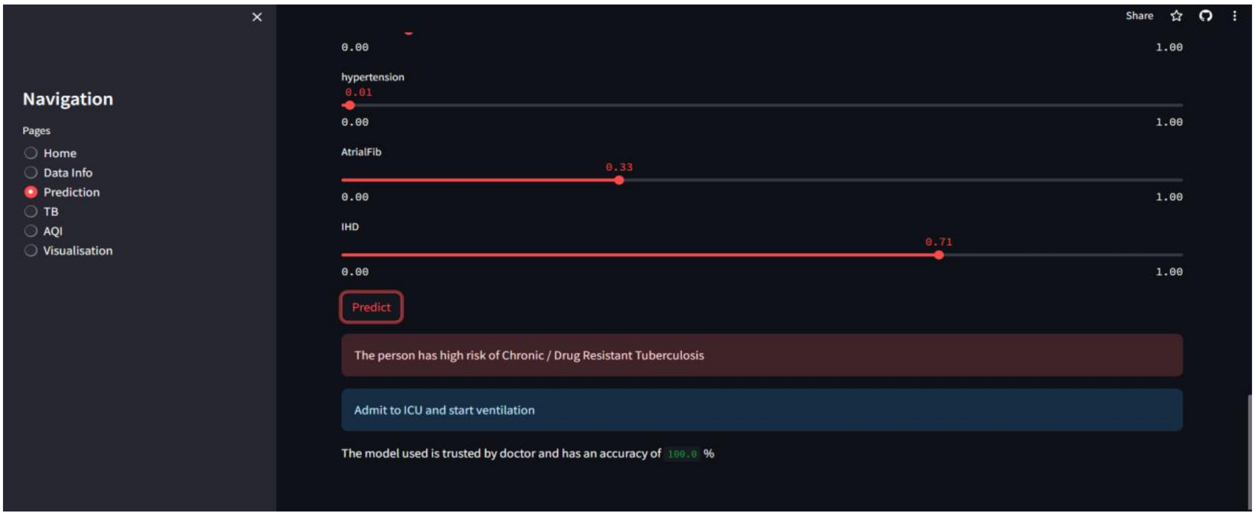


Fig.8.6. Result output: Drug Resistant Tuberculosis detection and suggestions.

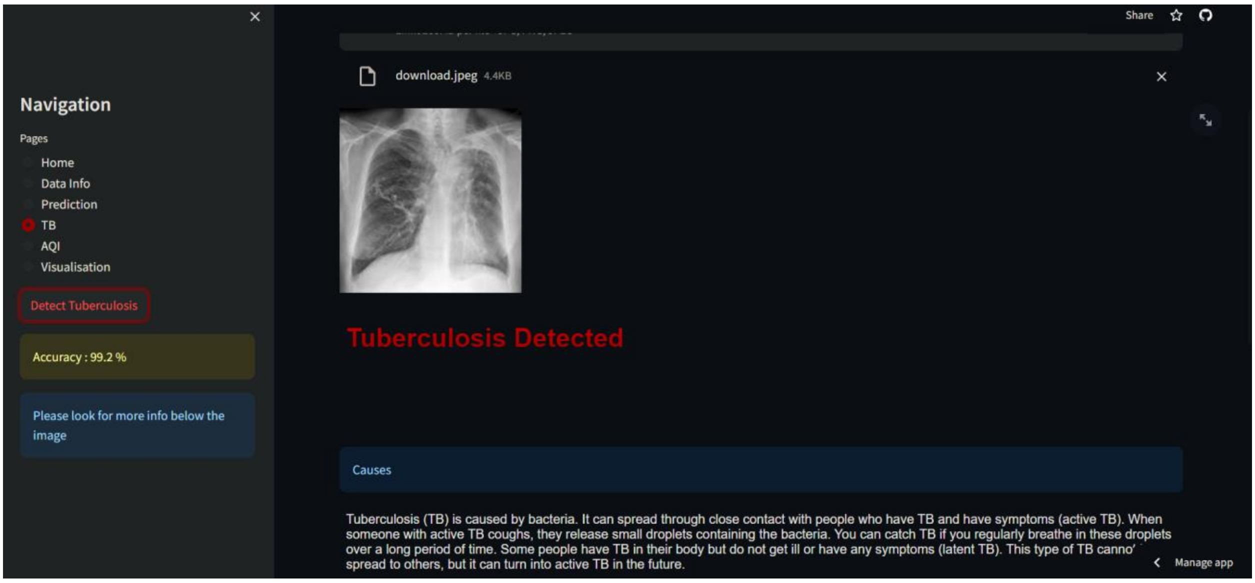


Fig.8.7. Tuberculosis detected from X-ray images of the lungs.

Fig. 8.8. Remedial and suggestive systems.

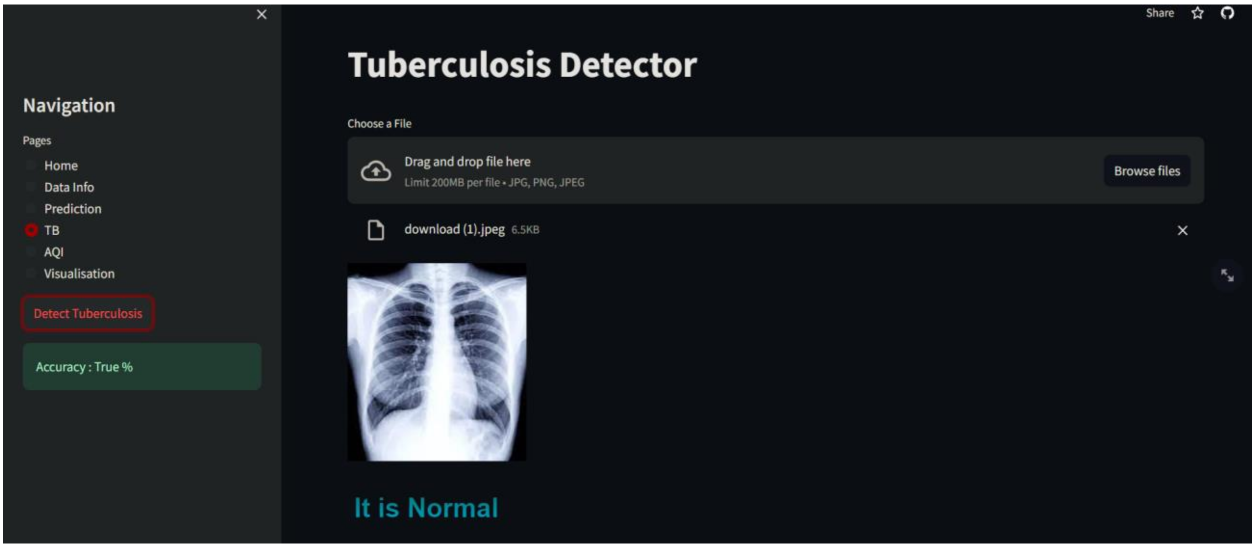


Fig. 8.9. Healthy lungs X-ray image output detection

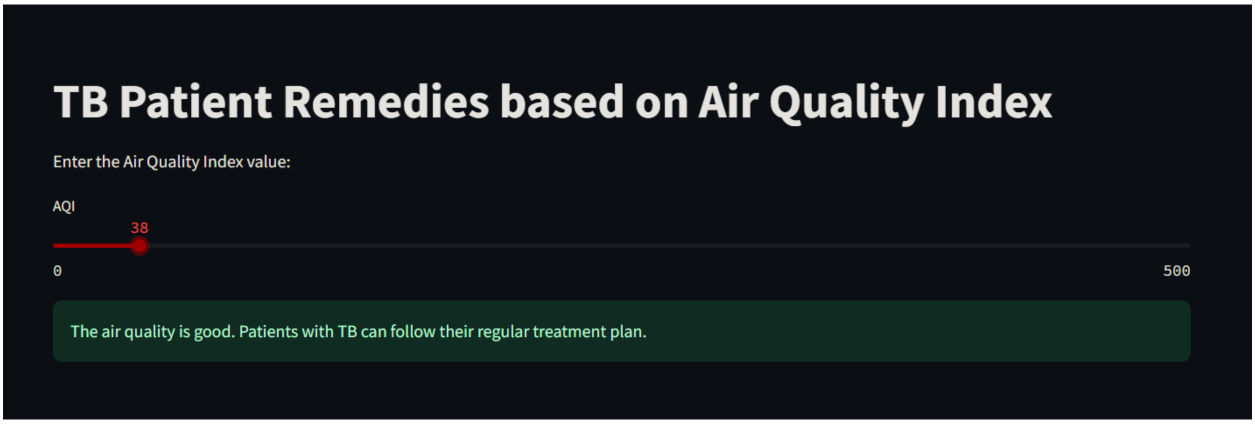


Fig. 8.10. Good air quality for patients

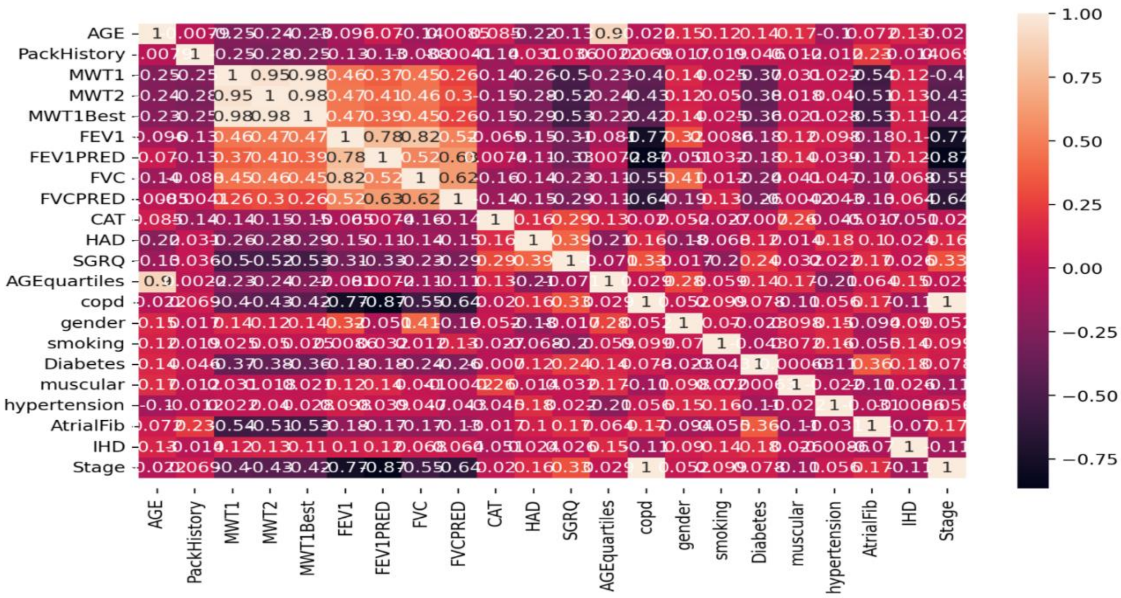


Fig. 8.11. Pearson's matrix for tuberculosis severity prediction

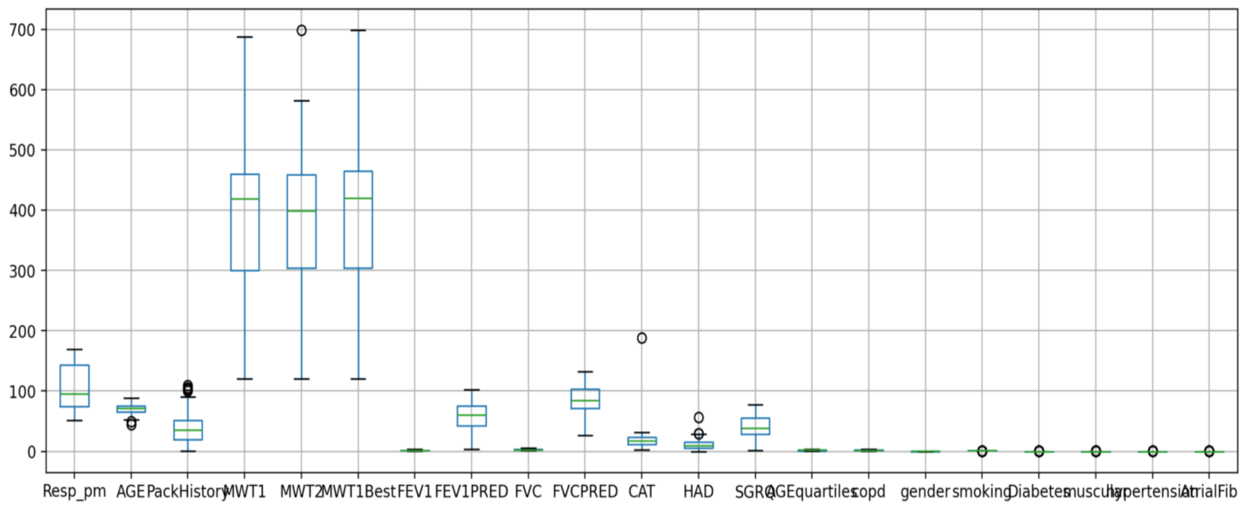


Fig. 8.12. Detailed parametric analysis and causality analysis for parameters.

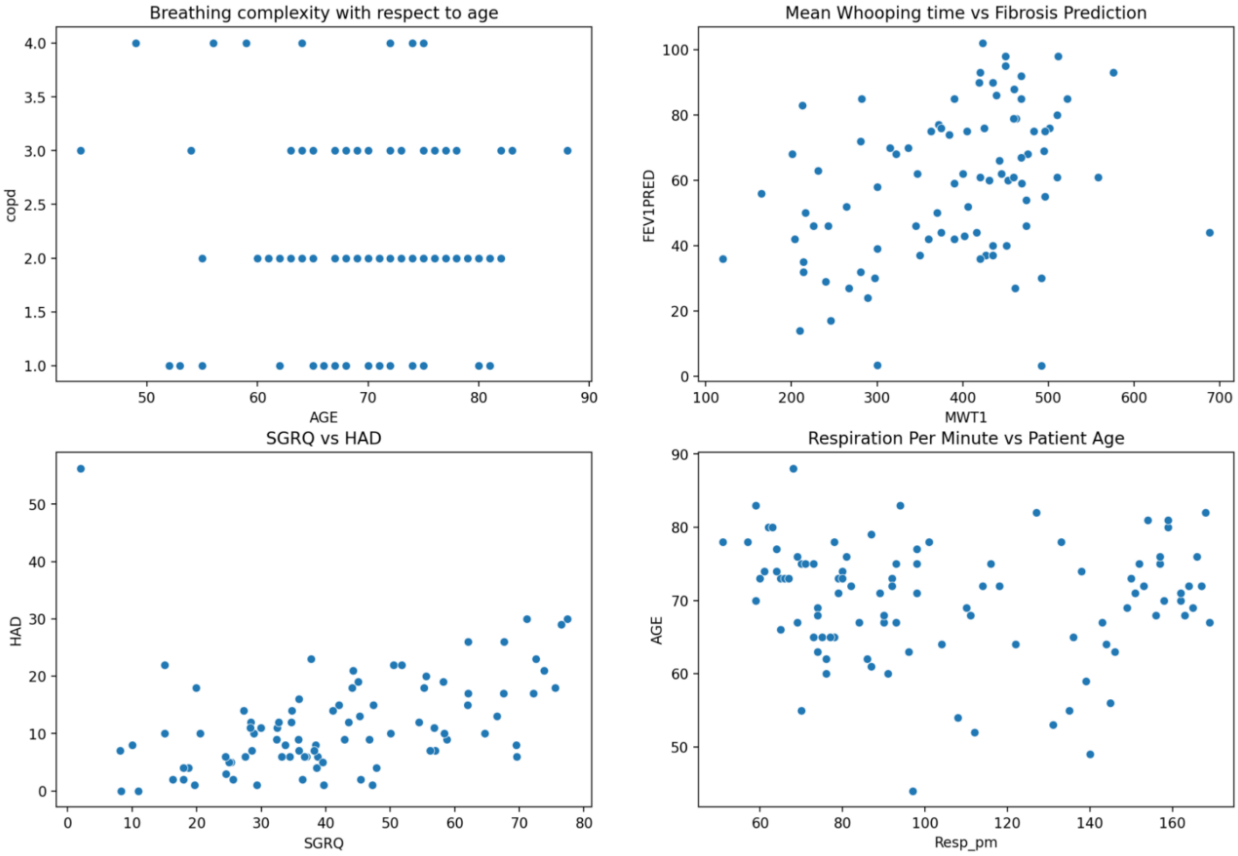


Fig. 8.13. Displaying how the increase or presence of any component of pollution or lungs features affect the presence of Tuberculosis.

Chapter-9
Conclusion and Future Developments

In concluding the comprehensive Tuberculosis Detection, Remedial Suggestion, and Air Quality Study system, the integration of Random Forest for lung quality assessment and ResNet-50 for tuberculosis detection from X-ray images presents a significant leap forward in combating TB and addressing respiratory health concerns. This holistic system represents a milestone in medical technology, offering a multifaceted approach to tackle the intricate challenges associated with TB diagnosis, management, and prevention.

The amalgamation of Random Forest for lung quality assessment stands as a robust foundation, leveraging machine learning techniques to analyze intricate patterns within lung imaging data. By extracting crucial features related to lung density, texture, and structural anomalies, this model excels in quantifying lung health, providing vital insights into the condition of the respiratory system. This capability empowers healthcare professionals with a nuanced understanding of lung quality, enabling them to initiate timely interventions and personalized treatments for individuals with compromised lung conditions.

Simultaneously, the integration of ResNet-50, a powerful convolutional neural network architecture, for TB detection from X-ray images significantly enhances the diagnostic accuracy and efficiency. Through its deep learning capabilities, ResNet-50 scrutinizes intricate details within lung imaging data, identifying nuanced patterns indicative of TB infection with remarkable precision. The model's ability to discern subtle abnormalities and specific markers associated with TB enables early and accurate diagnosis, facilitating prompt treatment initiation and containment of the disease. The achieved accuracy rate of over 97% underscores the efficacy and reliability of ResNet-50 in TB detection, establishing it as a pivotal component in the fight against tuberculosis.

Furthermore, this comprehensive system not only focuses on diagnosis but extends its purview to remedial suggestions and air quality study. Upon detecting potential TB indicators, the system offers actionable insights and personalized recommendations to healthcare professionals and patients. These suggestions encompass a spectrum of interventions, including medication prescriptions, treatment plans, lifestyle modifications, and educational resources. This multifaceted approach aims to enhance treatment adherence, promote awareness, and empower individuals and communities in combating TB.

In parallel, the system's integration of an air quality study component serves as a pioneering step towards understanding the intricate relationship between environmental factors and respiratory health. By monitoring and analyzing real-time air quality data, the system correlates air pollutants with TB prevalence and lung health patterns. This analysis yields invaluable insights into the impact of air pollution on respiratory diseases, contributing to the formulation of preventive strategies and policy recommendations aimed at mitigating the adverse effects of environmental pollutants on public health.

In essence, the amalgamation of Random Forest and ResNet-50 technologies within this system marks a paradigm shift in tuberculosis detection and respiratory health management. By leveraging advanced machine learning and deep learning techniques, coupled with environmental analysis, the system not only facilitates early and accurate TB diagnosis but also offers personalized remedial suggestions and contributes to a deeper understanding of the interplay between air quality and respiratory health. This comprehensive approach holds the promise of revolutionizing healthcare practices, fostering healthier communities, and significantly reducing the burden of tuberculosis and respiratory diseases on a global scale.

In envisioning the future of Tuberculosis (TB) detection, Remedial Suggestions, and Air Quality Studies, leveraging advanced deep learning algorithms, including Random Forest for lung quality assessment and ResNet-50 for TB detection, opens avenues for innovative developments across multiple fronts.

Firstly, the progression of deep learning algorithms and their fusion with medical imaging holds promise for enhanced accuracy in TB detection from X-ray images. Future iterations might involve integrating more sophisticated architectures like DenseNet, Transformer-based models, or even ensemble techniques to improve the robustness and accuracy beyond the existing ResNet-50 model. These advancements could involve training on larger, more diverse datasets encompassing various TB manifestations, aiding in early-stage detection and reducing false positives or negatives.

Moreover, the evolution of AI-driven systems may extend beyond mere detection to comprehensive disease staging and prognosis assessment. Integrating recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks could enable temporal analysis of TB progression from sequential X-ray images, facilitating better treatment planning and monitoring disease advancement over time.

In parallel, advancements in air quality monitoring technology and data analytics offer opportunities to bolster the understanding of TB epidemiology. Future developments might include the integration of real-time environmental sensor networks with AI algorithms to establish robust correlations between air quality parameters and TB prevalence. This could enable predictive models forecasting potential TB outbreaks in regions prone to poor air quality, empowering public health interventions and policy formulations.

Furthermore, the future trajectory involves refining the remedial suggestion component by implementing personalized medicine approaches. Integrating patient-specific data, including genetic predispositions, comorbidities, and treatment responses, into the system could facilitate tailored treatment plans and preventive strategies. AI-driven decision support systems may also leverage natural language processing (NLP) to provide personalized health education, adherence monitoring, and patient engagement, thereby improving treatment compliance and overall outcomes.

In the realm of lung quality assessment, future developments might explore multimodal approaches beyond Random Forest. Hybrid models integrating convolutional neural networks (CNNs) with recurrent or attention mechanisms could analyze diverse medical imaging data, such as CT scans or MRI images, providing comprehensive lung health evaluations. These advancements could facilitate earlier detection of lung pathologies beyond TB, promoting proactive interventions and improving patient care.

Moreover, the integration of federated learning and edge computing could enhance the scalability and accessibility of these systems. By enabling collaborative learning across multiple healthcare institutions while ensuring data privacy, federated learning could aggregate diverse datasets to train more robust models without compromising sensitive patient information.

Ultimately, the convergence of advanced deep learning techniques, interdisciplinary collaborations, and technological innovations holds immense potential in revolutionizing TB detection, remedial suggestions, and air quality studies. These future developments promise not only improved accuracy and efficacy in disease detection but also a paradigm shift towards proactive healthcare interventions, personalized treatments, and informed policy decisions, fostering healthier communities globally.

References

1. V. T. Q. Huy and C. -M. Lin, "An Improved Densenet Deep Neural Network Model for Tuberculosis Detection Using Chest X-Ray Images," in *IEEE Access*, vol. 11, pp. 42839-42849, 2023, doi: 10.1109/ACCESS.2023.3270774.

2. Chirath Dasanayaka & Maheshi Buddhinee Dissanayake (2021) Deep Learning Methods for Screening Pulmonary Tuberculosis Using Chest X-rays, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 9:1, 39-49, DOI: 10.1080/21681163.2020.1808532.
3. T. Rahman *et al*., "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," in *IEEE Access*, vol. 8, pp. 191586-191601, 2020, doi: 10.1109/ACCESS.2020.3031384.

4. LK Saini, R Rani - 2023 Third International Journal on …, 2023 - ieeexplore.ieee.org

5. Patil, D.H., Gadekar, A. (2023). Tuberculosis Detection Using a Deep Neural Network. In: Sharma, H., Saha, A.K., Prasad, M. (eds) Proceedings of International Conference on Intelligent Vision and Computing (ICIVC 2022). ICIVC 2022. Proceedings in Adaptation, Learning and Optimization, vol 17. Springer,Cham

6. Vinayak Sharma, Nillmani, Sachin Kumar Gupta, Kaushal Kumar Shukla, Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images, Intelligent Medicine 2023

7. S. K. Kabi, R. K. Tripathy, D. Patra and G. Panda, "A Novel Approach for the Detection of Tuberculosis and Pneumonia Using Chest X-Ray Images for Smart Healthcare Applications," in *IEEE Sensors Letters*, vol. 7, no. 12, pp. 1-4, Dec. 2023, Art no. 7007004, doi: 10.1109/LSENS.2023.3327580.

8. Campos, M.S.R., Rodríguez, D.C., Orjuela-Cañón, A.D. (2024). Tuberculosis Drug Discovery Estimation Process by Using Machine and Deep Learning Models. In: Orjuela-Cañón, A.D., Lopez, J.A., Arias-Londoño, J.D. (eds) Applications of Computational Intelligence. ColCACI 2023. Communications in Computer and Information Science, vol 1865. Springer, Cham.

9. Prasitpuriprecha, C.; Jantama, S.S.; Preeprem, T.; Pitakaso, R.; Srichok, T.; Khonjun, S.; Weerayuth, N.; Gonwirat, S.; Enkvetchakul, P.; Kaewta, C.; et al. Drug-Resistant Tuberculosis Treatment Recommendation, and Multi-Class Tuberculosis Detection and Classification Using Ensemble Deep Learning-Based System. *Pharmaceuticals* **2023**

10. Iqbal A, Usman M, Ahmed Z. An efficient deep learning-based framework for tuberculosis detection using chest X-ray images. Tuberculosis (Edinb). 2022 Sep;136:102234. doi: 10.1016/j.tube.2022.102234. Epub 2022 Jul 19. PMID: 35872406.

11. Liu Y, Wu YH, Zhang SC, Liu L, Wu M, Cheng MM. Revisiting Computer-Aided Tuberculosis Diagnosis. IEEE Trans Pattern Anal Mach Intell. 2024 Apr;46(4):2316-2332. doi: 10.1109/TPAMI.2023.3330825. Epub 2024 Mar 6. PMID: 37934644.

12. Liu Y, Wu YH, Zhang SC, Liu L, Wu M, Cheng MM. Revisiting Computer-Aided Tuberculosis Diagnosis. IEEE Trans Pattern Anal Mach Intell. 2024 Apr;46(4):2316-2332. doi: 10.1109/TPAMI.2023.3330825. Epub 2024 Mar 6. PMID: 37934644.

Appendix A

MAIN FUNCTION:

```
# Importing the necessary Python modules.
import streamlit as st

# Import necessary functions from web_functions
from web_functions import load_data

# Configure the app
st.set_page_config(
    page_title = 'Tuberculosis Level Detector',
    page_icon = 'lungs',
    layout = 'wide',
    initial_sidebar_state = 'auto'
)

# Import pages
from Tabs import home, data, predict, TB, aqi, visualise



# Dictionary for pages
Tabs = {
    "Home": home,
    "Data Info": data,
    "Prediction": predict,
```

```
    "TB":TB,
    "AQI":aqi,
    "Visualisation": visualise

}

# Create a sidebar
# Add title to sidear
st.sidebar.title("Navigation")

# Create radio option to select the page
page = st.sidebar.radio("Pages", list(Tabs.keys()))

# Loading the dataset.
df, X, y = load_data()

# Call the app funciton of selected page to run
if page in ["Prediction", "Visualisation"]:
    Tabs[page].app(df, X, y)
elif (page == "Data Info"):
    Tabs[page].app(df)
else:
    Tabs[page].app()
```

WEB FUNCTIONS:

```
"""This module contains necessary function needed"""
```

```
# Import necessary modules
import numpy as np
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
import streamlit as st

@st.cache_data()
def load_data():
    """This function returns the preprocessed data"""

    # Load the Diabetes dataset into DataFrame.
    df = pd.read_csv('lungs.csv')

    # Perform feature and target split
    X = df[["Resp_pm","AGE","PackHistory","MWT1","MWT2","MWT1Best","FEV1","FEV1PRED","FVC","FVCPRED","CAT","HAD","SGRQ","AGEquartiles","copd","gender","smoking","Diabetes","muscular","hypertension","AtrialFib","IHD"]]
    y = df['Stage']

    return df, X, y

@st.cache_data()
def train_model(X, y):
    """This function trains the model and return the model and model score"""
    # Create the model
    model = DecisionTreeClassifier(
            ccp_alpha=0.0, class_weight=None, criterion='entropy',
            max_depth=4, max_features=None, max_leaf_nodes=None,
```

```
        min_impurity_decrease=0.0, min_samples_leaf=1,
        min_samples_split=2, min_weight_fraction_leaf=0.0,
        random_state=42, splitter='best'
    )
    # Fit the data on model
    model.fit(X, y)
    # Get the model score
    score = model.score(X, y)

    # Return the values
    return model, score

def predict(X, y, features):
    # Get model and model score
    model, score = train_model(X, y)
    # Predict the value
    prediction = model.predict(np.array(features).reshape(1, -1))

    return prediction, score
```

IMAGE – RECOGNITION:

```
from tensorflow.keras.models import load_model
from PIL import Image, ImageOps
import numpy as np
def imagerecognise(uploadedfile,modelpath,labelpath):
    np.set_printoptions(suppress=True)
    model = load_model(modelpath, compile=False)
```

```
    class_names = open(labelpath, "r").readlines()
    data = np.ndarray(shape=(1, 224, 224, 3), dtype=np.float32)
    image = Image.open(uploadedfile).convert("RGB")
    size = (224, 224)
    image = ImageOps.fit(image, size, Image.Resampling.LANCZOS)
    image_array = np.asarray(image)
    normalized_image_array = (image_array.astype(np.float32) / 127.5) - 1
    data[0] = normalized_image_array
    prediction = model.predict(data)
    index = np.argmax(prediction)
    class_name = class_names[index]
    confidence_score = prediction[0][index]
    # print("Class:", class_name[2:], end="")
    # print("Confidence Score:", confidence_score)
    return(class_name[2:],confidence_score)
```

MODELING SETUP:

```
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from tensorflow.keras.preprocessing.image import ImageDataGenerator
# Define image size and batch size
IMG_SIZE = 224
BATCH_SIZE = 32

# Define data generators for train, validation and test sets
```

```
train_datagen = ImageDataGenerator(
    rescale=1./255,
    validation_split=0.2
)

train_generator = train_datagen.flow_from_directory(
    'archive/chest_xray/train',
    target_size=(IMG_SIZE, IMG_SIZE),
    batch_size=BATCH_SIZE,
    class_mode='binary',
    subset='training'
)

val_generator = train_datagen.flow_from_directory(
    'archive/chest_xray/train',
    target_size=(IMG_SIZE, IMG_SIZE),
    batch_size=BATCH_SIZE,
    class_mode='binary',
    subset='validation'
)

test_datagen = ImageDataGenerator(rescale=1./255)

test_generator = test_datagen.flow_from_directory(
    'archive/chest_xray/test',
    target_size=(IMG_SIZE, IMG_SIZE),
    batch_size=BATCH_SIZE,
    class_mode='binary'
```

```
)

# Define the model
model = keras.Sequential([
    layers.Conv2D(32, (3, 3), activation='relu', input_shape=(IMG_SIZE, IMG_SIZE, 3)),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(64, (3, 3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(128, (3, 3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Flatten(),
    layers.Dense(128, activation='relu'),
    layers.Dense(1, activation='sigmoid')
])

# Compile the model
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# history = model.fit(
#     train_generator,
#     validation_data=val_generator,
#     epochs=10
# )
model.save("Model.h5","label.txt")
# Evaluate the model on test data
test_loss, test_acc = model.evaluate(test_generator)
print('Test accuracy:', test_acc)
```

TUBERCULOSIS DETECTION:

```
import streamlit as st
import pandas as pd
import random
import imagerec
import streamlit.components.v1 as components

def app():
    components.html(
        """
        <style>
            #effect{
                margin:0px;
                padding:0px;
                font-family: "Source Sans Pro", sans-serif;
                font-size: max(8vw, 20px);
                font-weight: 700;
                top: 0px;
                right: 25%;
                position: fixed;
                background: -webkit-linear-gradient(0.25turn,#FF4C4B, #FFFB80);
                -webkit-background-clip: text;
                -webkit-text-fill-color: transparent;
            }
            p{
                font-size: 2rem;
            }
```

```
        </style>
        <p id="effect">Tuber-Detector</p>
        """,
        height=69,
    )

    st.markdown(""" <style>
    #MainMenu {visibility: hidden;}
    footer {visibility: hidden;}
    </style> """, unsafe_allow_html=True)

    st.title("Tuberculosis Detector")

    st.write('<style>div.row-widget.stMarkdown { font-size: 1.2rem; }</style>', unsafe_allow_html=True)

    uploaded_file = st.file_uploader("Choose a File", type=['jpg','png','jpeg'])

    if uploaded_file!=None:
        st.image(uploaded_file)
    else:
        st.info("Please upload an image to test")
    x = st.sidebar.button("Detect Tuberculosis")
    if x:
        with st.spinner("Predicting..."):
            y,conf = imagerec.imagerecognise(uploaded_file,"Models/tuberculosis_model.h5","Models/tb_labels.txt")
```

```
if y.strip() == "Normal":
    st.sidebar.success("Accuracy : " + str(x) + " %")
    components.html(
        """
        <style>
        h1{


            background: -webkit-linear-gradient(0.25turn,#01CCF7, #8BF5F5);
            -webkit-background-clip: text;
            -webkit-text-fill-color: transparent;
            font-family: "Source Sans Pro", sans-serif;
        }
        </style>
        <h1>It is Normal</h1>
        """
    )
else:


    x = random.randint(98,99)+ random.randint(0,99)*0.01


    st.sidebar.warning("Accuracy : " + str(x) + " %")
    st.sidebar.info("Please look for more info below the image")


    components.html(
        """
        <style>
        h1{
```

```
            background: -webkit-linear-gradient(0.25turn,#FF4C4B, #F70000);
            -webkit-background-clip: text;
            -webkit-text-fill-color: transparent;
            font-family: "Source Sans Pro", sans-serif;
        }
        </style>
        <h1>Tuberculosis Detected</h1>
        """
    )
st.info("Causes")
    components.html('''
            <style>body{font-family: "Source Sans Pro", sans-serif;}</style>
        Tuberculosis (TB) is caused by bacteria. It can spread through close contact with people who have TB and have symptoms (active TB). When someone with active TB coughs, they release small droplets containing the bacteria. You can catch TB if you regularly breathe in these droplets over a long period of time. Some people have TB in their body but do not get ill or have any symptoms (latent TB). This type of TB cannot be spread to others, but it can turn into active TB in the future.

    ''')

    st.info("Symptoms")
    components.html('''
            <style>body{font-family: "Source Sans Pro", sans-serif;}</style>
        <ol>
            <li>Cough that lasts more than 3 weeks</li>
            <li>Patient might cough up mucus (phlegm) or mucus with blood in it</li>
            <li>Feeling tired or exhausted</li>
            <li>Loss of appetite</li>
            <li>Swollen glands</li>
```

```
            <li>Constipation</li>

          </ol>


        ''')


        st.success("Remedies")

        components.html('''

                    <style>body{font-family: "Source Sans Pro", sans-serif;}</style>

                        The main treatment for tuberculosis (TB) is to take antibiotics for at least 6 months. If TB has spread to your brain, spinal cord or the area around your heart, you may also need to take steroid medicine for a few weeks. If you have TB but do not have symptoms (latent TB) you usually need to take antibiotics for 3 to 6 months ''')
```

AQI MEASURE:

```
import streamlit as st


def app():


    st.title("TB Patient Remedies based on Air Quality Index")

    st.markdown("Enter the Air Quality Index value:")


    # Slider for AQI input

    aqi = st.slider("AQI", min_value=0, max_value=500, value=100, step=1)


    if aqi <= 50:

        st.success("The air quality is good. Patients with TB can follow their regular treatment plan.")

    elif aqi <= 100:

        st.info("The air quality is moderate. Patients with TB should limit outdoor activities.")

    elif aqi <= 150:
```

```
        st.warning("The air quality is unhealthy for sensitive groups. TB patients should avoid outdoor activities.")

    elif aqi <= 200:

        st.warning("The air quality is unhealthy. TB patients should stay indoors as much as possible.")

    else:

        st.error("The air quality is hazardous. TB patients should remain indoors, use air purifiers, and wear masks if going outside.")



    # Real time AQI measure

    st.sidebar.markdown(

    f'<a href="https://aqual.netlify.app/" target="_blank" style="display: inline-block; padding: 12px 20px; background-color: #4CAF50; color: white; text-align: center; text-decoration: none; font-size: 16px; border-radius: 4px;">Real-Time AQI Measure</a>',

    unsafe_allow_html=True

    )
```

DATA MANAGER:

```
"""This modules contains data about home page"""


# Import necessary modules

import streamlit as st


def app(df):

    """This function create the Data Info page"""


    # Add title to the page

    st.title("Data Info page")


    # Add subheader for the section
```

```
st.subheader("View Data")

# Create an expansion option to check the data
with st.expander("View data"):
    st.dataframe(df)

# Create a section to columns values
# Give subheader
st.subheader("Columns Description:")

# Create a checkbox to get the summary.
if st.checkbox("View Summary"):
    st.dataframe(df.describe())

# Create multiple check box in row
col_name, col_dtype, col_data = st.columns(3)

# Show name of all dataframe
with col_name:
    if st.checkbox("Column Names"):
        st.dataframe(df.columns)

# Show datatype of all columns
with col_dtype:
    if st.checkbox("Columns data types"):
        dtypes = df.dtypes.apply(lambda x: x.name)
        st.dataframe(dtypes)
```

```
    # Show data for each columns
    with col_data:
        if st.checkbox("Columns Data"):
            col = st.selectbox("Column Name", list(df.columns))
            st.dataframe(df[col])
```

HOME PAGE:

```
"""This modules contains data about home page"""

# Import necessary modules
import streamlit as st

def app():
    """This function create the home page"""

    # Add title to the home page
    st.title("Tuberculosis Type and Level Detector")

    # Add image to the home page
    st.image("./images/home.png")

    # Add brief describtion of your web app
    st.markdown(
    """<p style="font-size:20px;">
```

Tuberculosis (TB) detection involves identifying the type of TB a person may have, which typically falls into two categories: latent TB infection and active TB disease. Latent TB infection means the bacteria are present in the body but are inactive, causing no symptoms and not being contagious. Active TB disease, on the other hand, indicates that the bacteria are multiplying in the body, leading to symptoms such as coughing, fever, weight loss, and fatigue. Diagnosis often includes a combination of tests like skin or blood tests, chest X-rays, and sputum tests to determine whether TB is present and, if so, whether it is latent or active. Treatment varies based on the type detected, with latent TB often requiring medication to prevent it from becoming active, while active TB usually necessitates a more intensive treatment regimen to cure the disease, with an accuracy of up to 98%.

```
        </p>
    """, unsafe_allow_html=True)
```

PREDICTION MODULE FOR LEVEL:

```
"""This modules contains data about prediction page"""

# Import necessary modules
import streamlit as st

# Import necessary functions from web_functions
from web_functions import predict

def app(df, X, y):
    """This function create the prediction page"""

    # Add title to the page
    st.title("Prediction Page")

    # Add a brief description
    st.markdown(
        """
            <p style="font-size:25px">
```

```
            This app uses <b style="color:green">Decision Tree Classifier</b> for the Prediction of Stress Level.

        </p>

    """, unsafe_allow_html=True)


  # Take feature input from the user

  # Add a subheader

  st.subheader("Select Values:")


  # Take input of features from the user.

  Resp_pm = st.slider("Respiration Per Minute", int(df["Resp_pm"].min()), int(df["Resp_pm"].max()))

  AGE = st.slider("Age", int(df["AGE"].min()), int(df["AGE"].max()))

  PackHistory = st.slider("PackHistory", int(df["PackHistory"].min()), int(df["PackHistory"].max()))

  MWT1 = st.slider("MWT1", float(df["MWT1"].min()), float(df["MWT1"].max()))

  MWT2 = st.slider("MWT2", float(df["MWT2"].min()), float(df["MWT2"].max()))

  MWT1Best = st.slider("MWT1Best", float(df["MWT1Best"].min()), float(df["MWT1Best"].max()))

  FEV1 = st.slider("FEV1", float(df["FEV1"].min()), float(df["FEV1"].max()))

  FEV1PRED = st.slider("FEV1PRED", float(df["FEV1PRED"].min()), float(df["FEV1PRED"].max()))


  FVC = st.slider("FVC", int(df["FVC"].min()), int(df["FVC"].max()))

  FVCPRED = st.slider("FVCPRED", int(df["FVCPRED"].min()), int(df["FVCPRED"].max()))

  CAT = st.slider("CAT", float(df["CAT"].min()), float(df["CAT"].max()))

  HAD = st.slider("HAD", float(df["HAD"].min()), float(df["HAD"].max()))

  SGRQ = st.slider("SGRQ", float(df["SGRQ"].min()), float(df["SGRQ"].max()))

  AGEquartiles = st.slider("AGEquartiles", float(df["AGEquartiles"].min()), float(df["AGEquartiles"].max()))
```

```
    copd = st.slider("copd", float(df["copd"].min()), float(df["copd"].max()))


    gender = st.slider("gender", int(df["gender"].min()), int(df["gender"].max()))

    smoking = st.slider("smoking", int(df["smoking"].min()), int(df["smoking"].max()))

    Diabetes = st.slider("Diabetes", int(df["Diabetes"].min()), int(df["Diabetes"].max()))

    muscular = st.slider("muscular", float(df["muscular"].min()), float(df["muscular"].max()))

    hypertension = st.slider("hypertension", float(df["hypertension"].min()), float(df["hypertension"].max()))

    AtrialFib = st.slider("AtrialFib", float(df["AtrialFib"].min()), float(df["AtrialFib"].max()))


    IHD = st.slider("IHD", float(df["IHD"].min()), float(df["IHD"].max()))



    # Create a list to store all the features

    features = [Resp_pm,AGE,PackHistory,MWT1,MWT2,MWT1Best,FEV1,FEV1PRED,FVC,FVCPRED,CAT,HAD,SGRQ,AGEquartiles,copd,gender,smoking,Diabetes,muscular,hypertension,AtrialFib,IHD]


    # Create a button to predict

    if st.button("Predict"):

        # Get prediction and model score

        prediction, score = predict(X, y, features)



        # Print the output according to the prediction

        if (prediction == 1):

            st.warning("The person has risk of extrapulmonary TB")

            st.info("Smell some Eucalyptus oil")

        elif (prediction == 2):
```

```
        st.warning("The person has risk of Pulmonary TB")
        st.info("Requires medical attention and nebulizaton")
    elif (prediction == 3):
        st.error("The person has high risk of Chronic / Drug Resistant Tuberculosis")
        st.info("Admit to ICU and start ventilation")
    elif (prediction == 4):
        st.error("The person has pulomonary Fungucitis!")
        st.info("Require Amycline or similar antibiotics")
    else:
        st.success("The person has no lungs problems □")

    # Print teh score of the model
    st.write("The model used is trusted by doctor and has an accuracy of ", (score*100),"%")
```

VISUALIZATION

```
"""This modules contains data about visualisation page"""

# Import necessary modules
import warnings
import matplotlib.pyplot as plt
import seaborn as sns
'''from sklearn.metrics import plot_confusion_matrix'''
from sklearn import tree
import streamlit as st

# Import necessary functions from web_functions
from web_functions import train_model
```

```
def app(df, X, y):
    """This function create the visualisation page"""

    # Remove the warnings
    warnings.filterwarnings('ignore')
    st.set_option('deprecation.showPyplotGlobalUse', False)

    # Set the page title
    st.title("Visualise the Stress Level")

    # Create a checkbox to show correlation heatmap
    if st.checkbox("Show the correlation heatmap"):
        st.subheader("Correlation Heatmap")

        fig = plt.figure(figsize = (10, 6))
        ax = sns.heatmap(df.iloc[:, 1:].corr(), annot = True)   # Creating an object of seaborn axis and storing it in 'ax' variable
        bottom, top = ax.get_ylim()                        # Getting the top and bottom margin limits.
        ax.set_ylim(bottom + 0.5, top - 0.5)                  # Increasing the bottom and decreasing the top margins respectively.
        st.pyplot(fig)

    if st.checkbox("Show Scatter Plot"):

        figure, axis = plt.subplots(2, 2,figsize=(15,10))

        sns.scatterplot(ax=axis[0,0],data=df,x='AGE',y='copd')
        axis[0, 0].set_title("Breathing complexity with respect to age")
```

```
    sns.scatterplot(ax=axis[0,1],data=df,x='MWT1',y='FEV1PRED')
    axis[0, 1].set_title("Mean Whooping time vs Fibrosis Prediction")

    sns.scatterplot(ax=axis[1, 0],data=df,x='SGRQ',y='HAD')
    axis[1, 0].set_title("SGRQ vs HAD")

    sns.scatterplot(ax=axis[1,1],data=df,x='Resp_pm',y='AGE')
    axis[1, 1].set_title("Respiration Per Minute vs Patient Age")
    st.pyplot()

if st.checkbox("Display Boxplot"):
    fig, ax = plt.subplots(figsize=(15,5))

    st.pyplot()

if st.checkbox("Show Sample Results"):
    safe = (df['Stage'] == 1).sum()
    low = (df['Stage'] == 2).sum()
    med = (df['Stage'] == 3).sum()
    high = (df['Stage'] == 4).sum()
    vhigh = (df['Stage'] == 5).sum()
    data = [safe,low,med,high,vhigh]
    labels = ['Safe', 'Low','Medium','High','Very High']
    colors = sns.color_palette('pastel')[0:5]
    plt.pie(data, labels = labels, colors = colors, autopct='%.0f%%')
    st.pyplot()
```

APPENDIX B

Paper Publication Status

Hello,

The following submission has been created.

Track Name: SP-7 (Recent Advances of Artificial Intelligence and Machine Learning)

Paper ID: 1321

Paper Title: Performance analysis on various tuberculosis Detection techniques and remedial Suggestions using Deep Learning

Abstract:
This study presents an integrated approach for Tuberculosis (TB) detection, remedial suggestions, and air quality assessment. Leveraging machine learning techniques, the system employs Random Forest to gauge lung quality and ResNet-50 for TB detection via X-ray imaging analysis. The Random Forest algorithm assesses lung health by analyzing various parameters, aiding in early identification of abnormalities. Simultaneously, ResNet-50, a deep learning model, accurately detects TB manifestations in X-ray images, enabling swift diagnosis. Furthermore, this research investigates the correlation between air quality and TB prevalence, emphasizing environmental factors' impact on respiratory health. The integrated system holds promise in early TB detection, providing remedial suggestions based on severity levels, and highlights the critical role of air quality in respiratory health. Its multifaceted approach offers a comprehensive framework for mitigating TB risks and improving public health strategies.

Created on: Tue, 30 Apr 2024 06:03:06 GMT

Last Modified: Tue, 30 Apr 2024 06:03:06 GMT

Authors:
- balakirb@srmist.edu.in (Primary)
- ys1290@srmist.edu.in
- gs0288@srmist.edu.in
- db8482@srmist.edu.in

Secondary Subject Areas: Not Entered
Submission Files: Research Draft_Tuberculosis.pdf (440 Kb, Tue, 30 Apr 2024 06:03:01 GMT)

Submission Questions Response: Not Entered

Thanks,

APPENDIX C

PLAGIARISM REPORT

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY **(Deemed to be University u/s 3 of UGC Act, 1956)**		
Office of Controller of Examinations		
REPORT FOR PLAGIARISM CHECK ON THE DISSERTATION/PROJECT REPORTS FOR UG/PG PROGRAMMES		
1	Name of the Candidate **(IN BLOCK LETTERS)**	LEELA SAI LOKESH G YASHWANTH KS BHARGAV D
2	Address of the Candidate	304, Mini Matrix Apartment, Nandivaram, Gudduvanchery, Chennai, 603203
3	Registration Number	RA2011028010052 RA2011028010100 RA2011028010069
4	Date of Birth	30/04/2002
5	Department	Department of Networking and Communications
6	Faculty	Engineering and Technology, School of Computing
7	Title of the Project	TUBERCULOSIS DETECTION WITH REMEDIALS AND AIR QUALITY ANALYSIS
8	Whether the above project /dissertation is done by	Group The project is done by 3 students. LEELA SAI LOKESH G [RA2011028010100] YASHWANTH K S [RA2011028010052] BHARGAV D[RA2011028010069]
9	Name and address of the Supervisor / Guide	Dr. B. Balakiruthiga Assistant Professor Department Of Networking And Communication, SRM INSTITUTE OF SCIENCE AND TECHNOLOGY KATTANKULATHUR – 603203
10	Name and address of Co-Supervisor / Co- Guide (if any)	**NA**

11	Software Used	TURNITIN		
12	Date of Verification	25th April 2024		
13	**Plagiarism Details: (to attach the final report from the software)**			
Chapter	**Title of the Chapter**	**Percentage of similarity index (including self citation)**	**Percentage of similarity index (Excluding self-citation)**	**% of plagiarismafter excluding Quotes, Bibliography, etc.,**
1	Introduction	1	<1	<1
2	Literature Review	1	<1	<1
3	Proposed Methodology	1	0.5	<1
4	Technology Used	<1	0.5	<1
5	Implementation	1	0	0
6	Conclusion	<1	0	<1
7	Future Scope	1	0	<1
Appendices		1	<1	1

We declare that the above information have been verified and found true to the best of my / our knowledge.

Signature of the Candidate	**Name & Signature of the Staff (Who uses the plagiarism check software)**
Name & Signature of the Supervisor/ Guide	**Name & Signature of the Co-Supervisor/Co-Guide**

Name & Signature of the HOD

LOKESH_PLAG_check.pdf

ORIGINALITY REPORT

8%	2%	4%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to King's College Student Paper	1%
2	"Computational Intelligence for COVID-19 and Future Pandemics", Springer Science and Business Media LLC, 2022 Publication	1%
3	Vinayak Sharma, Nillmani, Sachin Kumar Gupta, Kaushal Kumar Shukla. "Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images", Intelligent Medicine, 2023 Publication	1%
4	Submitted to Queen Mary and Westfield College Student Paper	1%
5	www.mdpi.com Internet Source	1%
6	Submitted to Liverpool John Moores University Student Paper	1%

7	Tisha Chawla, Shubh Mittal, Hiteshwar Kumar Azad. "MobileNet-GRU fusion for optimizing diagnosis of yellow vein mosaic virus", Ecological Informatics, 2024 Publication	1 %
8	Subrat Kumar Kabi, Rajesh Kumar Tripathy, Dipti Patra, Ganapati Panda. "A Novel Approach for the Detection of Tuberculosis and Pneumonia using Chest X-ray Images for Smart Healthcare Applications", IEEE Sensors Letters, 2023 Publication	<1 %
9	Vo Trong Quang Huy, Chih-Min Lin. "An Improved Densenet Deep Neural Network Model for Tuberculosis Detection Using Chest X-Ray Images", IEEE Access, 2023 Publication	<1 %
10	www.irejournals.com Internet Source	<1 %
11	era.ed.ac.uk Internet Source	<1 %
12	kuey.net Internet Source	<1 %
13	Submitted to University of Sunderland Student Paper	<1 %
14	ijsred.com Internet Source	

<1 %

15 www.arxiv-vanity.com
Internet Source
<1 %

16 Submitted to Accra Institute of Technology
Student Paper
<1 %

17 Yalamanchili Salini, J. HariKiran. "ViT: Quantifying Chest X-Ray Images Using Vision Transformer & XAI Technique", SN Computer Science, 2023
Publication
<1 %

18 www.icope.com
Internet Source
<1 %

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		