DEEP LEARNING BASED CHEST X-RAY ANALYSIS FOR DETECTION OF LUNG DISEASES

BY

MD RAKIBUL ISLAM
ID: 201-15-3430

AND

SABBIR HOSSAIN RIFAT
ID: 201-15-3324

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mr. MD ASSADUZZAMAN
Lecturer (Sr.Scale)
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Mr. MOHAMMAD JAHANGIR ALAM
Lecturer (Sr.Scale)
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Deep Learning-Based Chest X-Ray Analysis for Detection of Lung Diseases", submitted by Md Rakibul Islam and Sabbir Hossain Rifat to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 15-07-2024.

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori — **Board Chairman**
Professor & Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Amit Chakraborty Chhoton — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Mohammad Jahangir Alam — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Arshad Ali — **External Examiner**
Professor
Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Md Assaduzzaman, Lecturer (Sr.Scale), Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

14.7.24

Mr. Md Assaduzzaman
Lecturer (Sr.Scale)
Department of CSE
Daffodil International University

Co-Supervised by:

14.7.24

Mr. Mohammad Jahangir Alam
Lecturer (Sr.Scale)
Department of CSE
Daffodil International University

Submitted by:

Rakib 14.07.24

Md Rakibul Islam
ID: - 201-15-3430
Department of CSE
Daffodil International University

Rifat 14.07.24

Sabbir Hossain Rifat
ID: - 201-15-3324
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Mr. Md Assaduzzaman, Lecturer (Sr.Scale)**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Machine learning, Deep learning, NLP" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Dr. Sheak Rashed Haider Noori, Professor And Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Early and accurate lung disease prediction and detection is very important for our patient. chestX-ray diagnosis methods are timeconsuming and sesitive process This research explores deep learning techniques to automate lung disease classification from chest X-ray images, enhancing diagnostic efficiency and accuracy. We focus on five conditions: Edema, Pneumonia, Tuberculosis, COVID-19, and the normal state.Due to privacy concerns in obtaining X-ray images directly from medical facilities, we utilized a Kaggle, nis.gov and v7labs.com datasets of 14,631 chest X-ray images, verified by medical professionals. The dataset was separated into 80% learning., 10% for testing,10% for assurance. To ensure a robust model, data augmentation techniques such as gamma correction, Image resizing, data augmentation, histogram equalization, noise reduction were applied, enhancing the dataset and improving model performance.We evaluated several CNN model, including (CNN), ResNet50, VGG16, and DenseNet. Each model was assessed based on its training and validation accuracies. DenseNet is became a very good model, gaining a training accuracy of 99.01%, testing accuracy of 89%, and validation accuracy of 88%, outperforming the other models. VGG16 and CNN also demonstrated high performance, with accuracies around 87%, while ResNet50 gained an accuracy of 80%.Our work underscores the potential of advanced deep learning models in classifying and identifying lung conditions based on chestX-ray pictures, highlighting a significant improvements in diagnostic efficiency and accuracy that these technologies can offer.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Overview

Medical inquiry improved much in the last few years. They play a significant role in every field but at the same time, new diseases are emerging. In the medical field where lung is a very serious organ in the human body.[1] [2] Since the release of the novel Covid-19, numerous research projects have been launched worldwide to forecast it accurately. Covid-19 and the previous lung illness pneumonia are intimately associated since excessive chest congestion or pneumonic condition, was a major cause of death for many.[3] Deep learning techniques are being used and are advised by numerous studies in this field. our agenda of the study is apply deep learning architecture that leverage CT and X-ray scan datasets where we pick five classes to objectively evaluate the most recent methods for detecting lung diseases. So we built a model that would not only recognize old lung diseases but also discover new diseases with increased accuracy.

1.2 Background and Present State

In the ever-evolving landscape of medicine, the historical context remains vital as novel diseases continue to emerge with alarming frequency. The medical community is acutely aware of this reality, perpetually striving to stay ahead of the curve. Among the plethora of health challenges faced globally, respiratory conditions stand out prominently.[4] Lung cancer, asthma, tuberculosis (TB), Infections of the lower lungs with chronic pulmonary obstructive condition are designated by the World Health Organization (WHO) as significant contributors to global mortality rates. Astonishingly, statistics reveal staggering numbes: 334 million individuals grapple with asthma, 65 million are burdened by COPD, and 10 million suffer from TB. The WHO grimly predicts an annual toll of 1.6 million deaths from lung cancer, 3 million from COPD, and 1.4 million from tuberculosis.

In light of these alarming figures, the imperative to develop robust diagnostic tools is unmistakable. Hence, our overarching objective engineer sophisticated deep learning design ready for not only accurately classifying respiratory symptoms but also detecting and identifying the onset of these debilitating conditions.

1.3 Problem Statement

The The rationale of the study originates from the urgent necessity to handle escalating burden of respiratory diseases amidst the backdrop of continuous medical advancements and the emergence of novel health challenges. With respiratory conditions such as lungcancer, asthma,tuberculosis, Infections of the lower lungs with chronic pulmonary obstructive condition looming large as significant contributors to global mortality rates, there exists an urgent call to action. The staggering statistics provided from(WHO) underscore severity situation, with millions of individuals worldwide grappling with these conditions and millions more succumbing to their adverse effects annually. Against this backdrop, the study seeks to harness the power of CNN, specifically to develop robust diagnostic tools capable of accurately catched respiratory symptoms and detecting the onset of these debilitating diseases. By leveraging vast datasets comprising CT and X-ray scans, the study endeavors to contributeto the burgeoning of medical imaging, thereby paving the path for more potent attacks in disease management and patient care.

Our project aims to find the gap in the lung disease classification detect lung disease and also find the condition. We apply Deep learning techniques. The resulting model and dataset can be an importantsource of information for studies and development. Our main aim is to make sure our techniques perform well to find the lung condition and disease. If we talk about the agenda of our project then we try to improve metrics by fine-tuning the parameters of Deep learning techniques, which will boost accuracy. We will achieve our project's goal if we complete it on schedule.

1.4 Objectives

Even in the present, the medical community is aware of its history. since new diseases progressively emerge every Lung cancer, asthma, tuberculosis (TB), Infections of the lower lungs with chronic pulmonary obstructive condition are designated by the World Health Organization (WHO) classifies having a big major impact in global mortality. According to data, 334 million people have asthma, 65m infected COPD, and 10m obtain TB. WorldHealth Organization analyze 1.6m people will die for lung infection, 3 million from COPD, and 1.4 million by tuberculosis annually. Therefore, our agenda is to make a deep learning model which is good for us to classify, identify lung symptoms in addition to detecting.

1.5 Scope and Limitations

In the recent day or past many Authors have worked in this field and also used as many algorithms as possible for their achievement but after doing our observation we got many decreases in their work firstly they use a maximum of 4 classes of diseases for classification. we know that day by day our evolution is upgrading but some Author can not clarify their future work properly so we decided that we merge five classes of different diseases which are chest cancer, pneumonia, COVID-19, tuberculosis, and pulmonary asthma, and normal. after that, we classified that actual dataset, and then we applied multiple deep learning algorithms. At last, we also clarify our future work and show some particular algorithms that performed better than others algorithms.

1.6 Report Organization

In this report Organization we fullfill all requirement by introduce our work and also distinguish our related work.Then we also compare our related work and we clarify our approaches and also we make our methodology information with proper result identification and at last we add our last 3 important thing of our work.

Chapter1: Introduction

The first part sets the stage by explaining why detecting lung diseases is important and how deep convolution-al neural networks can help. It also highlights the need to compare various detection and data enhancement methods. The chapter outlines the main research questions, goals, and the structure of the study.

Chapter2: Literature Review

For this portion we make the difference between our precious related work and find their main speech and we research of their working methodology and then we make a difference table for our quick view option and we also clear our agenda and contribution after awar about their work.

Chapter3: Methodologys

In this portion we describes the unique things used in our work. It explains the work design, data collection steps , model building, and the reasons for selecting specific way. In this phase we also addresses ethical problem and any limitations that that is very harmful for our work.

Chapter4: Implementation

This stage we describe about our four algorithm respectively CNN, DenseNet, Resnet50, VGG16 architecture with CNN based model activity information. Accuracy information, precision info, recallresult, F1__scoreresult, and confusions matrixs formula which we used in our work.

Chapter5: Result and Analysis

In this portion we discuse about our applying four model and our highest performed model result and analysis and also compier their work and which model performed good or bad.At last we attached a model analysis picture end of this portion.

Chapter 6: Impact on Society,Environment and Sustainability

In this portion we describe societal, Environment and Sustainability. In this part examines some of the investigation's larger implications for medical testing and wellness. It talks about how better lung disease detection road can benefit treatment results, caregivers, and society in general. Future technological developments and their effects on society are taken into account.

Chapter 7: Conclusion and Future Work

In this finale slot we discuses about our whole work information in a short way and then we make the difference our work from others.At last paragraph we also attached our future planning or work of our work.

1.7 Summery

Important advancements are being achieved in medical investication lately. with a particular focus on the critical role of the lungs in the human body. The emergence of new diseases, especially in the context of the novel COVID-19, has spurred global research initiatives aimed at accurately predicting and understanding the virus. The link between COVID-19 and previous lung illnesses, such as pneumonia, is emphasized, given the historical impact of chest congestion and pneumonic conditions on mortality rates. The use of deep learning techniques, employing CT and X-ray scan datasets, to objectively assess the latest methods for detecting lung diseases. Main thing of these studies is to make deep learning algorithms capable of not only identifying existing lung diseases but also discovering new ones with enhanced accuracy.

CHAPTER 2
Literature Review

2.1 Overview

The combined summary presents us with many types of strategies mentioned by many Authors for the identification and categorization of lung illnesses, with a primary emphasis on pneumonia, COVID-19, and lung cancer.These methods make use of cutting-edge technology, such as deep learning, machine learning, and image processing methods, on datasets like COVID-19, ChestX-ray8, and LIDC-IDRI. For precise predictions, the approaches include elements like feature extraction, adaptive ROI estimation, and soft computing techniques. Evaluation indicators include specificity, F1-score, accuracy, recall, and precision; high accuracy rates, ranging have been observed.Common issues include scarce datasets and doubts about the accuracy and effectiveness of the models. Addressing these drawbacks, investigating novel architectures, and highlighting the significance of more thorough research is the lack of more research they work for the highest 4 types of disease. Some authors work for only one disease detection.

2.2 Related Works

Goyal, S. [1] introduces an unique framework for detecting lung infection, such as pneumoniaandCOVID-19, using patient chestX-ray pictures. The process includes feature extraction, adaptive ROI estimation, dataset adjust, image ecellence , illness prediction. For precise detection, it uses a large collection of characteristics, improved region growth methodologies, and publicly accessible chest X-ray datasets. Classification is done using soft computing techniques .

Analysis of their proposed big methodology using precisione, recalle, correctness, F,1-score, and detection accuracy.comparison of the accuracy and computation time of the current approaches. And got 95.05% accuracy by applying model.

S. Taha Ahmed [2] uses CT and X-ray scan datasets to conduct a rigorous analysis of deep learning algorithms used in lung disease detection processes. It examines current research from 2015 to 2021 to determine the most suitable approaches, constraints, and precision levels. The key drawbacks are the lack of standard datasets, big training sets, high data dimensionality, and feature independence. For picture datasets continue to be state-of-the-art. Lung disease diagnostic technologies can be spread to remote locations and are essential for medical professionals to make better decisions. It is best to introduce experts, doctors, radiologists, and patients to offer guidance. Deep learning model is learn for better datasets and have improved activation maps to increase their recognition rate.

Ibrahim [3] applied multi-classification model that has been developed in detect to the covid-19pandemic to diagnose covid-19, pneumonia, and lung infection using C-T and chestX-rays pictures. The goal of this model is to improve classification accuracy and dataset size. They considered four architectures: ResNet + (GRU), ResNet152V2 +(Bi-GRU), and VG G19-CNN. With 98.75% accuracy, 97.05% recall, 92.43% precision, 98.5% specificity, and 99.3% negative predictive value, the VG G19+CN N architecture play better than others. This method avoids patient death and is essential for early discovery. Increasing images from whole data, lengthening the learning epochs, and applying deep methods like for augmentation help to boost the result of the suggested algorithm.

L. Pham [4] presents a deep learning algorithm for investigation, agenda to detect problem in respiratory cycles and catch diseases from sound recordings ICBHI benchmark dataset. The architecture uses front-endfeature collection aim and a back-end network to classify spectrogram features. Experiments show theframework outperforms current methods and offers a Teacher-Student scheme for real-time applications. they divided multiple categories for analysis and got 91% for teachers and 88% for students. they gained a Teacher andStudent learning view define to reduce algorithmcomplexity and achieve. They should explore model compression and quantization.

Al-Antari [5] prepared deep learning computeraided diagnosis system based on the yolo predictor which detects and diagnoses covid-19 based on the chestX-ray image. For this study, they used covid-19 and ChestX-ray8 datasets. For classification, they used a CAD-based YOLO predictor. They applied DarkCovidNet, Covid-Net, VGG-19, and CoroNet algorithms for predicting X-ray images and got 87.02%,92.40%,93.48%, and 89.60% respectively. In this study, they classified and predicted 2 diseases Covid-19 and Pneumonia but they did not talk about other classes They can't talk about the limitation or future work.

Chaturvedi et al.[6] proposed Computer-aided diagnosis techniques using machine training based on machine training ways to detect lung Cancer. In this study Image processing techniques and segmentation methods are used for prediction. Logisticregression, decisiontree classifier, gradient boosting model are implemented for photo preprocessing law are apply to improve system performance and accuracy. They used various types of datasets like LUN A16, SuperBowl Dataset and LIDC-IDRI. CT-Scan, X-Ray, and MRI Scan images they used in their study.By using GaussianBlur, Otsu Thrashold, Mobilenet, InceptionV3, and VGG8 their Best gaining among via efficiency, recollection, sensitivity, an F1-s and detection accuracy.comparison of the accuracy and computation time of the current approaches.In this study, they talk or work on how to boost the accuracy rate and show the comparison. However, they can't work for new lung diseases or disease classes for better prediction and identification.

Kasinathan et al.[7] focused on constructive cloudbased lung growth detection and phase correctness. The system allowed virtual monitoring and e-diagnosis. The applying method achieved good performance, recall, precision for only lung tumor classification. For detecting lung tumour detection and phase correctness they used the LIDCIDRI dataset of many low info.CloudLTDSC hybridframework with lung tumour detector, segmentations, and stageclassifier modules.Active model for lung tumour segmentations . Multilayer convolutional neural networks in this stage classification.Utilization of lungs CT DICOM images. Average lung tumour stage classification accuracy of 97.02% to 99.1% using multiple C-NN algorithm. They got 97.01% accuracy for using M-CNN on the LIDC-IDRI/PET dataset where false positive rate 2.1.they should apply this method in a large dataset and analysis system capability. they predicted the early stage of the lung tumor but did not work for other stages of the tumor.

Souid et al.[8] worked on the correctness and indentifying of lung diseases for chestX-ray images. They made Transferlearning with metadata argument and resampling the dataset for improved performance. They also made a big difference between performance other stateof-the-art methods using AUC statistics with an Average AUC of 0.811. They Resampled the dataset and improved model performance. They used the NIHChestXray-14 dataset also for their work and applied Mobile-Net V2 with the CNN layers algorithm and got 90% accuracy. The dataset resampling could improve model performance. The model is designed for low computing power devices. The work is only focused on Lung tumors.

Nageswaran [9] tries to build and discover an accurate correctness and detecting and prediction of lung infection by actual technology. At first, they gathered photos, and for processing preprocessing they used a geometric mean filter. For segmentation, they used the k-means algorithm. They used KNN and ANN in 80+ CT scans from 60+ of patients utilized as the datasets. That research aims to improve the accuracy of lung infection detection systems. But they did not show the accurate accuracy percentage of KNN which they mention that after the applying KNN they got better accuracy. They did not provide specific details about the limitations of the machine training and image utilizing techniques used for lung infection correctness and detection.

Ismael [10] proposed deep learning strategy, such as deep featureextraction, finetuning of pretrained convolutionali neuralnetworks (CNN), and endto end training of their made in CNN model, for the classification of covid-19 and normal chestX-ray picture. they experiment with pretrained deep CNN models (ResNet-18, Reset50, Reset101, VG G16, and VG G19) for deep feature extraction and classification using SupportVector Machines with various kernel functions. they used 180 COVID-19 and 200 normal datasets for their work. The Resnet50 model with the SVM classifier achieved the highest accuracy .On the other hand, only the ResNet50 model achieved a good and unbreakable accuracy. In same dataset, the CNN model got 91.6% accuracy. they use a quite small dataset for their work. they work for only one disease. They did not discuss other dangerous diseases.

Gupta [11] applying a Net-19 model for classifying covid-19 in chest X-ray images. they used Transfer Learning (Feature Extraction) IntegratedStacking Net-19 Classification Model for classification The proposed model InstaCov-Net Achieved 99.08% accuracy on three class classification and 99.53% on two of these class classification. Achieved 99% recall, F1 score, and precision on ternary classification.Gaining a good precisionand a good recall on binary correctness .InstaNet-19 can detect covid-19 without human attraction at an economicalcost. They did not mention the dataset that they used in their research work.

Ibrahim [12] applaying approaches to deep learning with classifying covid-19, viralpneumonia, bacterialpneumonia, and normal CXR scansX-ray . where they apply the CXR image dataset for their work and applied an AlexNet deep learning-based algorithm.

A proposed approach using a pretrained AlexNet model for correctness. they measure the actual accuracy for one - four-way classification. At last for fourway classification, the model achieved an good accuracy alsso sensitivity is also good, and specificity is good. Molecular testing for distinct pathogen strains are not up to par. The RT-PCR sequencing approach is very extra expensive and can leada false results. Underconstructed countries and remote places lack learning kitsand equipment. The use of X-rays can be tedious for qualified radiologists. they use a small dataset of pneumonia.

Hussain [13] prepared hybrid strategy connected to fuzzy thinking integration that is built on modular neural network technology. The use of an integrated method facilitated the diagnosis of lungnodules andpneumonia. the applying method called CoroDet uses a 22-layer CNN model Evaluation based on correctness, pracision, racall, F_1 score, specificiting, sansitivity, and confution matrix. they used the largest COVID-19 datasets which were divided by Xray image and CT picture. Lastly, also Achieved classification good accuracy for two class correctness, 94.2% for 3 class classifications, and good for 4 class correctness. They did not mention the dataset used for evaluation like size, diversity, or representativeness. They used only two classes for implementation and detection. they should work for more than four diseases.

Varela-Santos [14] apply a neural network work for lung disease for lung disease classification. they used a Classification using soft reaction integration based on various quantitative improvement of features. They used the largest ChestX-ray picture for classification. They applied some hybrid models for classification such as Fuzzy response integration, Multiple objective feature optimization, Image preprocessing, segmentation, Feature extraction, rulebased system, mechine learning architecture, neuralnetworks, Deep convolution neuralnetworks (DCNNs) for feature achiving and classification. If we talk about the result, they got 96.84% for hybrid model-1 and 79.34% for hybrid model-2 cross-dataset results. They achieved the highest accuracy in the JSRT database was 96.76%. They were not able to use multiple datasets in their work and also they should work for more classes or features of lung disease.

Nafisah [15] make an digital system for early identifying of TB applying deep learning and The transferlearning mechanism with a deep CNNs for TB detection. They compared the classification performance of different CNN models using CXR datasets. The actual proposed system achieved a high accuracy of 99.1% in TB detection by using EffitiantNet B3. They work for only TB not to mention other Diseases. They should work for 2 or more Diseases. The paper does not discuss any potential limitations or biases associated with the use of publicly available CXR datasets for correcting the classification result of deep rain models.

Sarkar [16] built a multiscale CNN framework with lung disease classification using CXR images. They used for their work a Dataset of 650CXR picture categorized in 7 disease groups. In this research they used a deep learning architecture called (MS-CNN) Then they trained the model on binari, threeclass, fourclass, fiveclass, sixclass, sevenclass datasets. At last, they Make the comparison between those models through comparative analysis using a performance matrix. At last, they gain 96.05% accuracy for seven class classifications. The study focused on lung-related diseases and did not include data on various type of lung infections, limiting the scope of the research. The MS-CNN model used in the study did not build a lighthweight structure, requiring more hardware resources for running the model.

Singh [17] apply a architecture called deepconvolutional neaural network (CNN) with Resnet50 configuration for detecting COVID-19. They achieve 97% accuracy in detecting COVID-19. In this study, they use ChestX-ray datasets and C T datasets. In paper acknowledges that the ResNet50 and ResNet101 models used in the study have limitations in terms of efficiency and accuracy, with ResNet50 delivering only 90% efficiency and ResNet101 delivering 80% efficiency. The study also mentions that there is a lack of research in the area of COVID-19 identification using machine learning algorithms due to limited awareness and database availability, indicating a potential limitation in the existing literature.

2.3 Comparison between existing works

The current landscape of lung disease detection research reveals a predominant focus on utilizing deep learning models, particularly CNNs, with varying methodologies and emphasis on diseases such as covid-19, pneumonias, and lung infection. Notable works, such as those referenced as [3], [5], and [10], employ CNNs but differ in their approaches and achieved accuracies. employs a VGG19+CNN multi-classification model achieving impressive accuracy (98.05%) for detecting covid-19, pneumonias, and lung infection. [11] Conversely, presents a deep-learning CAD system utilizing diverse algorithms, yielding accuracies ranging between 87.02% and 93.48% for covid-19, pneumonias. In contrast, [12] explores pretrained CNN models for covid-19 and normal chestX-ray classification, with Resnet50 gain a good accuracy (94.7%) using SVM. Despite these high accuracies, a critical research gap remains in the absence of comprehensive comparative analyses, limiting the determination of the most effective model for lung disease detection in this domain.

These studies often overlook crucial aspects such as interpretability, dataset biases, and real-world clinical applicability, which are vital for enhancing the reliability and practical utility of AI-driven diagnostic tools. Interpretability concerns the capacity to comprehend and believe in the judgments made by AI models, which is crucial for gaining acceptance and adoption by medical professionals. Additionally, biases within datasets, such as those found in ChestX-ray8, can skew model performance and generalizability, warranting careful consideration and mitigation strategies. Moreover, The models' suitability for practical clinical settings is essential for their integration into healthcare workflows, necessitating further exploration and validation beyond controlled research environments.

In summary, while recent studies in lung disease detection showcase promising advancements in deep learning methodologies, there exists a significant gap in comprehensive comparative analyses and discussions on interpretability , dataset biases , and real-world clinical applicability. Addressing these gaps is crucial for advancing the reliability and practical utility of AI-driven diagnostic tools in the field of lung disease detection.

TABLE 2.3.1: Table of comparative evaluation

Studies	**Imagetype**	**CNNModel**	**Gaining Accuracy**
Goyal, S. [1]	ChestX-Ray	KNN,SVM,ANN, F-RNN-LSTM	F-RNN-LSTM = 95.05%
S. Taha Ahmed.[2]	digitalx-ray picture	CNN	-
Ibrahim. et al. [3]	Cheest X-ray	ResNet152V2 + GRU , ResNet152V2 + Bi-GRU,VGG-19 + CNN	VGG-19 + CNN = 98.-5
L. Pham. [4]	lung images	CNN-MoE	For Student = 98% For Teacher = 88%
Al-Antari .[5]	chest X-ray	VG G-19	VG G-19 = 93.48%
Chaturvedi [6]	chestx-ray images	VGG-8	VGG-8 = 97.85%
Kasinathan [7]	chestX-ray	M-CN N	M-CNN = 97.01%

Solid. [8]	ChestX-ray picture	Mobile-Net V2	Mobile-Net V2 = 90%
Nageswaran. [9]	ChestX-ray (CXR) Picture	ANN	ANN = 92.09%
Ismael. [10]	ChestX-ray (CXR) images	Resnet50 + SV M	ResNet50 + SVM = 92.6%
Gupta. [11]	chest radiographs (CXR)	InstaCov-Net	InstaCov-Net = 99.08%
Ibrahim. [12]	ChestX-ray (CXR)	AlexNet	AlexNet = 93.42%
Hussain. [13]	ChestX-ray (CXR)	22-layer CNN	22-layer CNN for 3 class classification = 94.02%
Varela-Santos. [14]	ChestX-ray (CXR)	DCNNs hybride model-1	DCNNs = 96.84%
Nafisah. [15]	ChestX-ray (CXR)	EffitiantNet B3	EffitiantNet B3 = 99.1%
Sarkar. [16]	ChestX-ray (CXR)	MS-CN-N	MS-CNN = 96.05%
Singh. [17]	ChestX-ray (CXR)	ResNet50	ResNet50 = 90%

2.4 Open Issues

In the Recent studies have focused on leveraging deep learning techniques, particularly CNNs, to classify lung diseases, notably targeting COVID-19 detection. These studies share a common objective of achieving accurate disease classification and often utilize datasets like ChestX-ray8. Despite achieving accuracies ranging from 87.02% to 98.05%, variations exist in their methodologies and dataset handling. For instance, one study employs a multi-classification model, achieving high accuracy across COVID-19, pneumonia, and lung cancer. Another study utilizes a deep learning CAD system with various algorithms, achieving commendable accuracies for COVID-19 and pneumonia. Additionally, another study focuses on classifying covid-19 and normal chestX-ray pictures, experimenting with pretrained models to achieve high accuracy. These differences underline the necessity for a thorough comparative analysis to catch good effective methodology in lung infection identification, highlighting nuances in approach, algorithm choice, and model training.

2.5 Summary

The corpus of research on lung disease detection demonstrates spread area of working side and developments in area of artificial intelligence as they relate to medical diagnosis. Researchers have used novel frameworks and deep learning mechine learningtechniques to find the problems presented illnesses like covid-19 ,pneumoniea,lung cancar and so more diseasese. Frequently, these ways use advanced models such as CNNs, hybrid systems, and ensemble classifiers, which result in remarkable accuracy rates between 90% and 99.1%. The significance of standardized datasets and their relevance in training robust models are repeatedly emphasized by the studies. Accurate prediction is mostly dependent on feature extraction methods, image preprocessing, and classification algorithms. Despite the acknowledged drawbacks—such as dataset size and model efficiency. Although there are certain drawbacks, such as dataset size and model efficiency, these studies together add to the increasing amount of data demonstrating artificial intelligence's potential to completely transform medical diagnostics, especially when it comes to lung disorders. The focus on important diseases, along with the investigation of new methods and continuous debates about their shortcomings, represents a changing field of research intending to enhance healthcare using advanced technology.

CHAPTER 3
Methodology

3.1 Overview:

Our work agenda is to predict lung diseases through image classification and deep learning, employing a meticulous methodology that includes requirements analysis, data preprocessing, and implementation for key deeplearning model like CN N, ResNet50, VGG16, DenseNet. The model utilizes a diverse chest X-ray dataset acquired primarily from Kaggle, nis.gov and v7labs.com augmented for robustness through various techniques. Managed using an agile approach, the project undergoes iterative development cycles overseen by a dedicated project manager. Financial analysis considers budget allocation for infrastructure, personnel, and computational resources, ensuring cost-effectiveness throughout. The expected outcome is an effective Chest X-ray Analysis system capable of initial lung condition identification and assessment, contributing significantly to healthcare advancements.

3.2 Proposed Methodology

In This study proposes a good methodology for an digital classification of chestX-ray pictures into multiple disease features, including covid-19, pneumonea, normal, edema, tuberculosis.

Our applying methodology attached of three core stages:

Data Preprocessing: In data preprocessing phase we apply gamma correction, Image resizing, data augmentation, histogram equalization, noise reduction concept. that's why after applying these concepts we get more accurate and clear data for training our models.

Deep Learning Algorithms: In the second stage, the approved picture are fed into several deep taining model. We utilize Convolutionail Naural Networks (CNN) along with advanced architectures like ResNet50, DenseNet, and VGG16. These models are specifically chosen for their proven efficacy in image recognition tasks. They train and gaining more good features by the Xray pictures to distinguish between different disease conditions.

Training and Classification: The final stage involves training the deep training models with preprocessed pictures. In learned features are then passed through fullyconnected layers to play the correctness task. The output is a set of predictions categorizing the images into one of the specified disease conditions.

This methodology aims to leverage the stamina of deeplearning to give a reliable and digital diagnostictool, carefuly attaching healthcare expert in the nearly and properly detection chest-related diseases.

Such famous neural network technique that has been extensively utilized for training and processing picture is the convoulutional naural system (CNN). The Convoulution layer's matrix arrangement is designed with image filtering in mind. The input layer, convolutionallayer, fullyconnected layer, poolinglayer, dropoutlayer for CNN building, and a final linked dataset classification layer are the layers used in the data learning process of the Convolution Neural Network. By mapping a set of calculations to the input test set, each layer has a specific function.

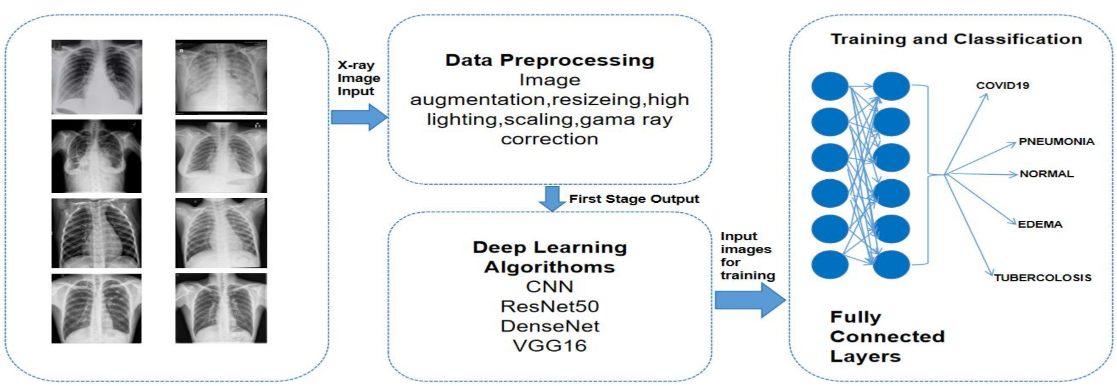
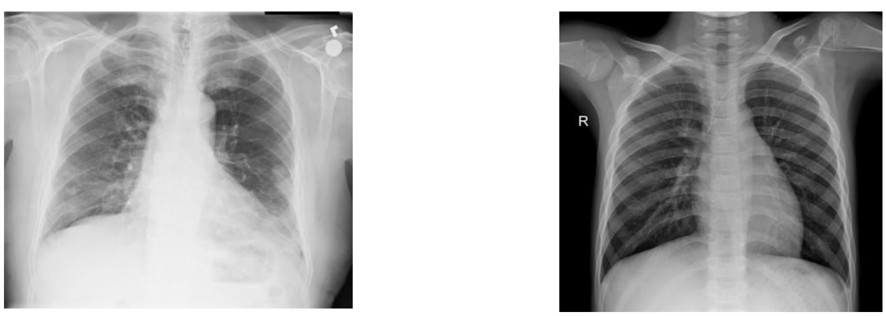


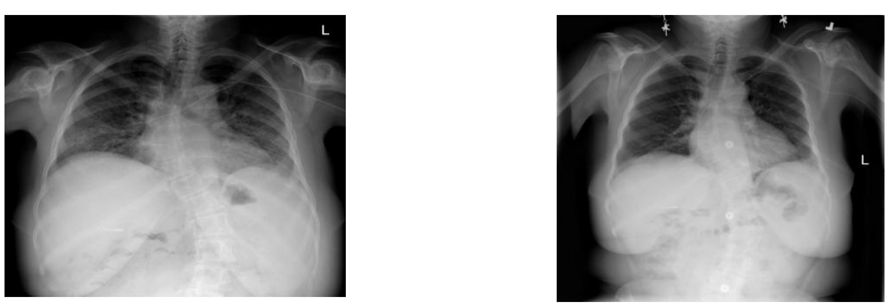
Figure 3.2.1: Architecture of workingprocess.

A.Datasets

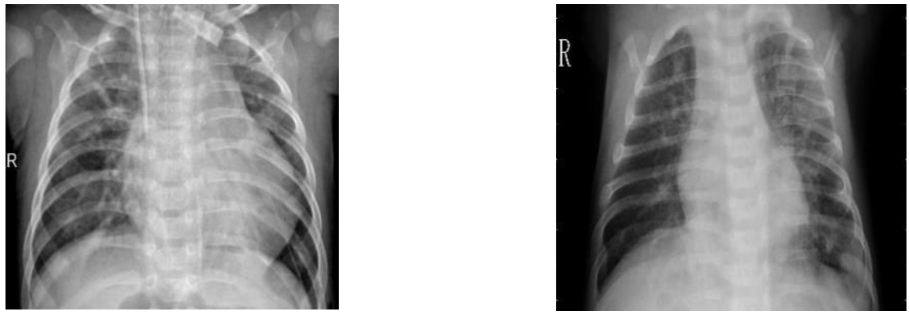
In our research, initial efforts to gather chest X-ray images directly from medical facilities were impeded by privacy concerns. We turned to kaggle, nis.gov and v7labs.com where we obtained diverse datasets covering lung conditions such as chest pulmonary edema, pneumonia, normal, tuberculosis, and COVID-19. With the assistance of a doctor, we ensured a thorough understanding of the diseases. The dataset consisted around 14,631 images, distributed across learning, experiment, and correnctess folders at 80 %, 10 %, and 10 %, respectively. The pictures included varying numbers for each disease class: 3087 for pulmonary edema, 4262 for COVID-19, 4421 for pneumonia, 1174 for tuberculosis, and 1687 normal chest X- rays. All images were exclusively sourced from kaggle, nis.gov, v7labs.com are reputable platform for medical imaging data.



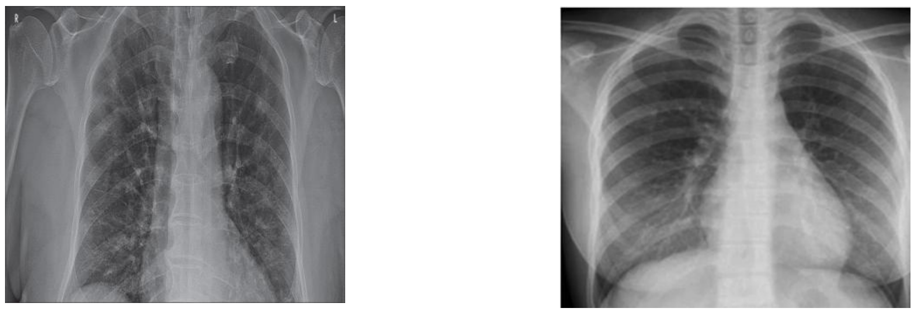
A. Normal



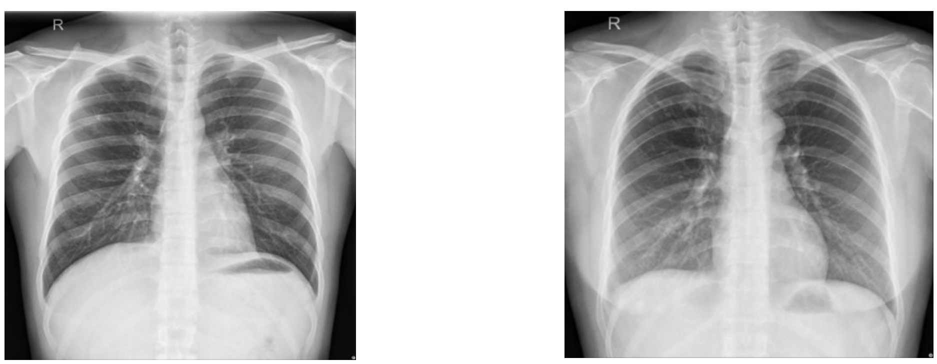
B. Pulmonary edema



C. Pneumonia



D. COVID-19



E. Tuberculosis

Figure 3.2.2: Some actual image of Datasets.

We have an uncle who is a doctor who helped us understand the diseases and their consequences in the dataset. He helped us in many ways so that we could bring up a good result.Moreover, the invaluable guidance of a doctor, who happens to be the researcher's uncle, significantly enriched our understanding of the diseases and their clinical implications within the dataset.His expertise brings a vital things in ensuring our work of thoroughness and accuracy of our research findings.Throughout the study, all images were sourced exclusively from Kaggle, nis.gov and v7labs.com widely recognized platform for data acquisition and analysis in the field of medical imaging.

TABLE 3.2.1: Images used in the learning, test and verification sets.

	Pulmonary edema	Covid_19	Pneumonia	Tuberculosis	Normal
Train	2176	4076	3875	1030	1341
Test	653	106	390	86	234
Validation	258	80	156	58	112

A. Experiment Process

We collect our information from the Xray picture is the first step in the suggested architecture. The issue of not good enough data is occasionally solved by using dataaugmentation technology, since deep naural networks require additional data for learning and good performance. The process planned to go as follows:

Picture prepearing:

x-ray Data for the model test was provided by Image Archive. Tailored visuals for the website were employed in this phase. A detailed analysis of the collected photographs was conducted to ensure white backgrounds. Images in colour are set against white back.

Picture Augmentation process:

At this moment, the enhancement to images is applied. Adding new data to an existing dataset while maintaining the description data is known as photo enrichment. Expanding the data set enhances the generalization of the model, making it more accommodating to

Our data training needs were met with the use of data enrichment. However, techniques including brightness, contrast, saturation, scaling, cropping, flipping, and revolution were applied to improve color and position. Arbitrary spins from 15 to fifteen degrees of freedom above unplanned 90-degree rotations, distortion, bending, vertical and horizontal reversals, skating, and luminous intensity conversion were among the data enhancement strategies used. Every initial image was enhanced ten times. A mixed picture is improved by certain types of modifications.

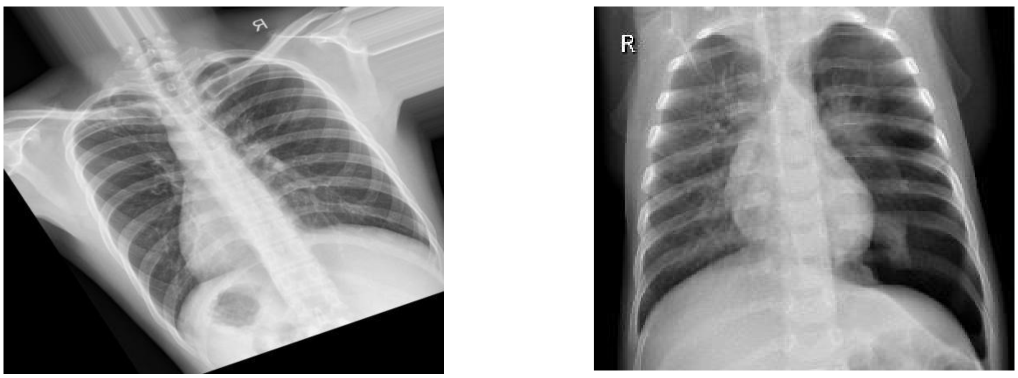


Figure 3.2.3: Augmented Images

Classification:

Here,we apply and build a classification diagram to understanding our work and also the stage. Figure 3.2.4 spread our experiment process.

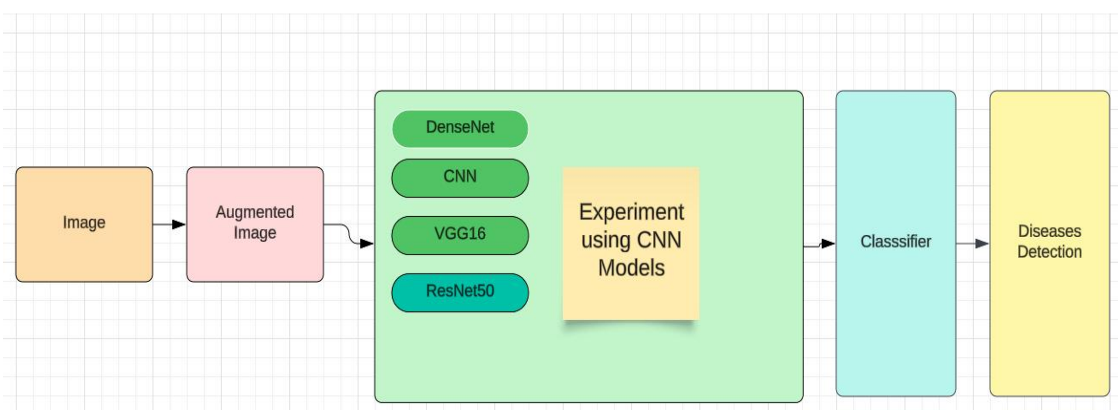


Figure 3.2.4: Experiment architecture model process.

3.3 Hardware/ Software Requirement

The implementation of the research on "ChestX-ray Analysis for Predicting DeepLearning with Image Classification and DeepLearning" comprises particular specifications to guarantee an effective completion. The execution needs cover concerns for data gathering and assessment of models in addition to software and physical pieces.

A. HardwareRequirements:

- **Supplies for powerful computing:** possession of an IT infrastructure capable of handling the intricate duties required in image interpretation and deep learning, with enough storage and computational capacity.
- **GraphicsProcessing Unit (GPU):** In order to dramatically cut down on the amount of time needed for model construction, deepnaural networks (CNNs) must be learned quickly, which requires a GPU.

B. Software Requirements:

- **Deeplearning Architecture:** Utilization of popular deep learning frameworks such as TensorFlow for the development, training, and evaluation ofthe CNN models.
- **Python programing languages:** Python offers a flexible and popular environment for data analysis and deep learning algorithm implementation.
- **picture processinglibraries:** The initial processing and health picture augmentation through the integration of image processing goods

3.4 Project Management and Financial Analysis

A. Project Management

Learning with Image Classification and Deep Learning comprises particular specifications to guarantee its effective completion. Our needs will be implemented with hardware as well as software factors, along with data collecting and appraisal of models concerns.

The project is managed using an agile methodology, with iterative development cycles overseen by a dedicated project manager. Key stages include data gathering, algorithm development, system integration, user interface design, testing, and deployment. The project objectives are to conduct a comprehensive literature review to identify current trends in deeplearning modals for healthcare picture process, like chestX-ray classification; evaluate existing datasets to determine the most suitable for learning and checking the deep learning modal and investigate also select appropriate deep learning frameworks and tools. The project timeline involves defining research scope and objectives (2-3 weeks), conducting a literature review (5-6 weeks), evaluating datasets of (4-5 weeks), researching and choosing deep learning frameworks (6-7 weeks), and at last finalizing the research plan and documenting methodology (4-5 weeks).

Resource planning includes accessing academic databases for the literature review, using high-performance computers for dataset evaluation, and utilizing development servers for framework testing. Training and skill development ensure the research team has a strong foundation in relevant fields, with training sessions on specialized tools. Risk management involves conducting a risk analysis to anticipate potential challenges and developing contingency plans. The communication plan includes weekly research team meetings to discuss progress and challenges, and bi-weekly research planning meetings to review the plan and adjust timelines as necessary.

Project Time-line:

A. Phase 1: Define Research Scope and Objectives (2-3 week)

B. Phase 2: Conduct Literature Review and Compile Findings (5-6 weeks)

C. Phase 3: Evaluate Datasets and Select Suitable Data Sources (4-5 weeks)

D. Phase 4: Research and Choose Deep Learning Frameworks (6-7 weeks)

E. Phase 5: Finalize Research Plan and Document Methodology (4-5 week)

B. Financial Analysis

The budget is allocated to pay for infrastructure, people's wages, data collection instruments, and computational resources. Impact on healthcare, market acceptance, and research contributions are examples of potential rewards. Throughout the project, cost-effectiveness is ensured by regular financial evaluations.The estimated cost for the "ChestX-ray processing for Predicting DeepLearning with Image Classification, DeepLearning" research ranges from 56,100 to 75,900 BDT. This includes 45,000 to 60,000 BDT for hardware, 1,000 to 1,500 BDT for stakeholder visits, 1,000 to 2,000 BDT for software and tools, 3,000 to 4,000 BDT for data collection and processing, and 1,000 to 1,500 BDT for documentation. A contingency fund of 5,100 to 6,900 BDT is also included to cover unexpected costs.

Table 3.4.1: Estimated Cost for our deep learning-based project

SN	Using Components	All Estimated Costs (BDT)
01.	Hardware/Infrastructure	45000-60000
02.	Visiting Stakeholders	1000-1500
03.	Software and Tools	1000-2000
04.	Data collecting and analytical	3,000-4000
06.	Document and actual Report make	1,000-1500
08.	Protect plan(10 % of our total amount)	5100-6900
Total Estimated Cost		**56100-75900**

3.5 Summary

In summary, our project agenda to make a modal in a way to predicting lung diseases using picture classification and deeplearning rules. In our applying methodology we will use data preprocessing, model development, training, testing, and evaluation. For doing our project methodology we faced various types of problems like class identification issues, dataset collection, and methodology mapping. but at last, we were able to build our methodology and whole process also including project management and financial management of our project. The expected outcome of this project will be early identification and diagnosis of lungdiseases based on ChestX-ray images which can help medical professionals.

CHAPTER 4
Implementation

4.1 Overview

Our work agenda is the foundational principles also advancements in ConvolutionalNeural Networks (CNNs), emphasizing their application in image classification, particularly in healthcare diagnostics. We explore ResNet50, DenseNet, and VGG16 as exemplars of CNN architectures, highlighting their respective contributions to overcoming challenges in deep learning. The evaluationmetrics discussed—accuraacy, precisione, recal, and F1score—demonstrate effectiveness and suitability of CNNs for tasks ranging from disease detection to image recognition. By examining these models through rigorous testing and analysis, our research contributes to advancing the field of deep learning, with implications for improving diagnostic accuracy and efficiency in clinical settings.

4.2 Train Model/ Prototype Design

Convolutional Neural Networks (CNNs)

The ConvolutionalNeural Network (CNN) algorithm has transformed computer vision by emulating the hierarchical pattern recognition of the human visual system. CNNs excel in automatically learning data patterns through their distinct convoloutional, polling, and fullyconnected layers. [23] Convoloutional layers serve as feature extractors, identifying basic elements like form and Designs, when deeperlayers detect complex features like shapes and objects. Pooling layers reduce dimensionality of featuremaps by downsampling, highlighting the most significant features. Ultimately, fuly connected layer integrate these features to make predictions, enabling tasks like object recognition in images or sentiment analysis in text.

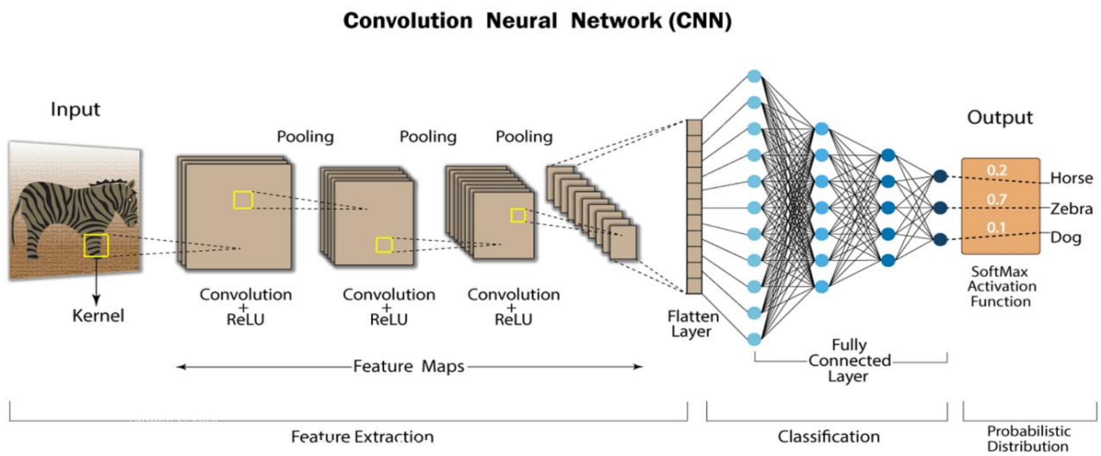


Figure 4.1.1: ConvolutionalNeuralNetwork (CNN)

key advantages for CNN is cop up to train spatial hierarchies of sensetive things directly from actual data, eliminate for handcrafted featureengineering. end-to-end learning process makes CNNs is very important for picture classification, objectdetection, and sagmentation. CNNs exhibit parametersharing, is same set of weights is applied across different phase of the input, making them computationally efficient and scalable to large datasets. This characteristic allows CNNs to process high-dimensional data efficiently.

Resnet50: Resnet50, build for ResidualNetwork which have 50 layers, is a deep convolutionalneural network model designed to solve vanish gradient problem depend on deep networks. Developed by Kaiming and colleagues in 2015, ResNet50 introduces residual learning, where each layer focuses on learning the residual mapping relative to the input rather than the desired mapping directly. [24] This method allows ResNet50 to train deeper networks without performance degradation since additional layers can learn the identity mapping if needed. ResNet50 uses skip connections or shortcuts that bypass one or more layers, facilitating , thereby enhancing training efficiency and effectiveness.

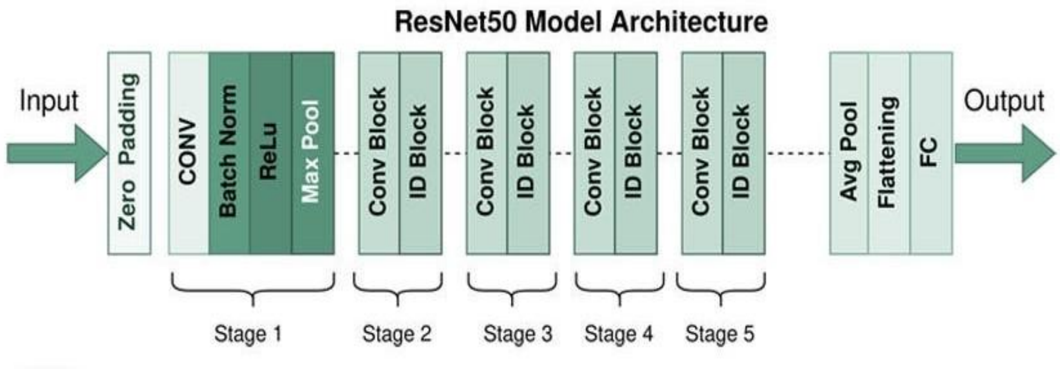


Figure 4.1.2: ResNet50 Model Architecture.

With its innovative design, ResNet50 has become a cornerstone in the phase of deeplearning, generely for computer vision like photo correctness, object detection, and semantic segmentation. Its architecture, which effectively addresses the challenges of training very deep networks, has been widely adopted and serves as the foundation for many stateoftheart models across various domains. The success of ResNet50 in maintaining high performance with increasing network depth has significantly influenced the development of subsequent deep learning models and techniques.

DenseNet: Densely Connected Convolutional Networks (DenseNets) are a deep learning architecture that enhances flow of informations and gradien among the network by connecting every layer to any other layer in a feedforward manner.

Alleviates the vanishinggradient problem, enabling the training of very deep networks. [25] leading to efficient parameter usage and improved learning efficiency. DenseNets typically require fewer parameters and achieve better performance than traditional CNNs, making them more compact and effective.

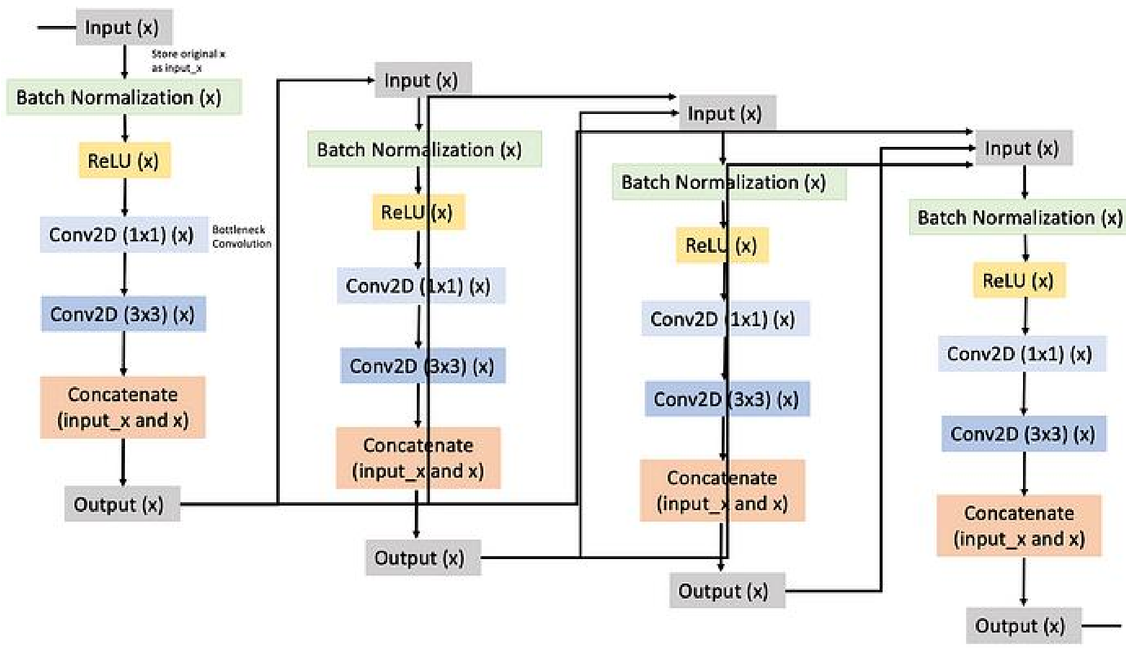


Figure 4.1.3: DenseNet Model Architecture.

Inception Despite the increased computational complexity and memory requirements, DenseNets are easier to train and often yield superior accuracy in tasks like photo clasification and sagmentation, thanks to dense connections and improved information flow.

VGG16 : It gained prominence with performance in the Imagenet Large Scale of Visual RecognitionChallenge (ILSVRC) 2014, where it achieved top results. VGG16's architecture add of 16 layers, including 13 convolutionallayers and 3fullyconnected layers, along with max-pooling layers interspersed between convoulutional layers to remove the spatial demensions. This structured design allows VGG16 to achieve remarkable accuracy in recognizing and classifying images.

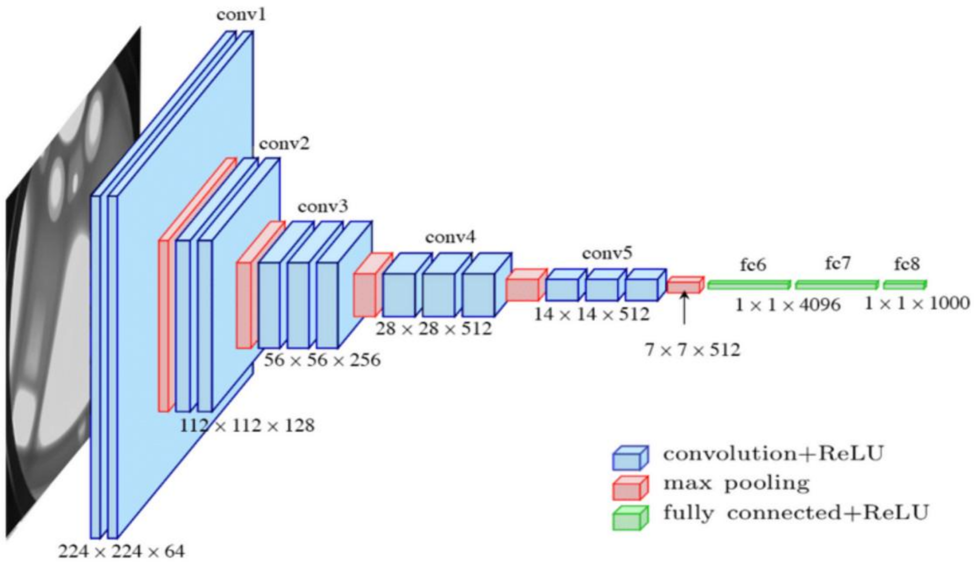


Figure 4.1.4: VGG16 Model Architecture.

The convolutional layers of VGG16 use small receptive fields a stride of 1 and padding to preserve the vital resolution of the input image. This design choice helps in capturing fine details in the images.[26] Additionally, VGG16 employs a uniform architecture, where each convolutional layer is followed by a Rectified Linear Unit (ReLU) activation function. This promotes non-linearity, enhancing the model's ability to learn complex patterns. The combination of small kernels and ReLU activations makes VGG16 particularly effective in extracting and learning intricate features from images.

4.4 Summary

In our exploration of Convolutional Neural Networks (CNNs), we've delved into their pivotal role in computer vision, particularly in automating feature extraction through layers dedicated to convolution, pooling, and fully connected operations. ResNet50 and DenseNet stand out among the architectures studied, each offering unique advancements in deep learning. ResNet50's introduction of residual learning addresses challenges in training deep networks, while DenseNet's dense connections enhance information flow and parameter efficiency. VGG16, known for its simplicity and effectiveness, underscores the significance of architectural design in achieving high accuracy in image classification tasks.

CHAPTER 5
Result & Analysis

5.1 Overview

In order to facilitate thorough investigations, our research on "Chest X-ray Analysis for Predicting Deep Learning with Image Classification and Deep Learning" uses an experimental setup that is painstakingly designed with hardware, software, and data resources. High-performance gear, such as a powerful central processing unit (CPU), 16 GB of RAM, and a graphics processing unit (GPU) called GTX 1660 Super 6 GB OC, support the computational backbone. The models are implemented using a Python programming environment and deep learning frameworks, mainly TensorFlow and Keras. Different chest X-ray images representing five classes are included in the experimental dataset, which was obtained from Kaggle, nis.gov and v7labs.com. To guarantee this dataset is subjected to extensive preprocessing, standardization, and normalization.A key component of the model construction is the utilization of CNN architectures that have already been trained, together with fine-tuning tailored to each lung illness. To uphold ethical standards, data usage agreements and patient data protection are just two examples of the ethical considerations that are incorporated into the setup. To support reprodocibility, and future collaoboration in the field in lung disease diagnastics, extensive documentation is kept throughout the experimental process, including code details, model architecture, hyper parameters, and findings.

5.2 Experimental/ Simulation Result

Result of Experiment 1:

TABLE 5.2.1: Acuracy for Clessification of Individual CNN Networks in Detecting Lung Diseases.

Architecture	Training Accuracy	Testing Accuracy	Model Accuracy
VGG16	**97.04%**	**91%**	**87%**
ResNet50	**86.01%**	**84%**	**80%**
DenseNet	**99.01%**	**89%**	**88%**
CNN	**97.49%**	**89%**	**87%**

A comparison of different pretrained deep Convoulutional Neural Network (CNN) architectures, DenseNet, VG G16, ResNet50, a bespoke CNN model, shows significant variations in their lung disease prediction and chest X-ray analysis accuracies in table 5.2.1. With a model accuracy of 88%, DenseNet was the best-performing model, as the table illustrates, proving its better ability to recognize and categorize lung illnesses from chest X-ray pictures. This suggests that the design of DenseNet is very good at extracting complex features required for precise diagnosis. VGG16 and customized CN N model both attained an 87% model accuracy, which is marginally less than DenseNet's performance but still has strong diagnostic abilities. With a training accuracy of 99.01%, DenseNet highlights the high generalization it can achieve with the right training.Despite this, when the model was used with unknown data, it performed poorly. In actual use, the bespoke CNN model performed similarly to VGG16 and demonstrated strong training accuracy (97.04%).showed the least accuracy of any model (80%), indicating that it was less useful in this particular application. Its training accuracy of 86.01% suggests that there may be challenges in extracting the critical characteristics from this dataset that are required for precise lung disease classification. The crucial nature of model selection in medical image analysis is highlighted by this performance gap, since architectural subtleties can have a considerable impact on diagnosis accuracy.

TABLE 5.2.2: Pricision, Recal, F1_Score, and Support (n) for CNN networks (depending on thenumber of pictures, n=numbers).

VGG16				
	Precision	**Recall**	**F1-Score**	**Support**
Normal	0.86	0.88	0.87	210
Pneumonia	0.96	0.97	0.97	437
edema	0.79	0.54	0.64	50
Covid19	0.63	0.83	0.72	64
tuberculosis	0.94	0.64	0.77	46
Accuracy			0.89	807
Macro Avg	0.84	0.77	0.79	807
Weighted Avg	0.90	0.89	0.89	807
ResNet50				
	Precision	**Recall**	**F1-Score**	**Support**
Normal	0.79	0.80	0.80	210
Pneumonia	0.94	0.96	0.95	437
edema	0.56	0.62	0.59	50
Covid19	0.66	0.58	0.62	64
tuberculosis	0.74	0.54	0.62	46
Accuracy			0.84	807
Macro Avg	0.74	0.70	0.72	807
Weighted Avg	0.84	0.84	0.84	807
DenseNet				

	Precision	Recall	F1-Score	Support
Normal	0.81	0.90	0.85	210
Pneumonia	0.99	0.96	0.97	437
edema	0.62	0.76	0.68	50
Covid19	0.84	0.73	0.78	64
tuberculosis	0.99	0.70	0.82	46
Accuracy			0.90	807
Macro Avg	0.85	0.81	0.82	807
Weighted Avg	0.91	0.90	0.90	807
CNN				
	Precision	**Recall**	**F1-Score**	**Support**
Normal	0.79	0.92	0.85	210
Pneumonia	0.99	0.98	0.98	437
edema	0.73	0.64	0.68	50
Covid19	0.90	0.72	0.80	64
tuberculosis	0.94	0.65	0.77	46
Accuracy			0.90	807
Macro Avg	0.87	0.78	0.82	807
Weighted Avg	0.91	0.90	0.90	807

Based on the comparative analysis detailed in Table 5.2.2, DenseNet and a custom CNN model have emerged as the top performers for lung diseases detection, achieving an impressive accuracy of 90%. DenseNet exhibited particularly high precision (0.99) and recall (0.96) for 'Pneumonia' and showed strong performance across other disease categories like 'COVID-19' and 'Tuberculosis'.

Similarly, the custom CNN model demonstrated outstanding results for 'lung diseases detection' (precision 0.99, recall 0.98) and solid metrics for 'Normal' cases (precision 0.79, recall 0.92). VGG16 achieved an accuracy of 89%, excelling notably in 'Pneumonia' detection with precision of 0.96 and recall of 0.97. ResNet50, with an accuracy of 84%, also performed well for 'Pneumonia' (precision 0.94, recall 0.96), highlighting the critical role of model selection in achieving optimal performance. These findings underscore the superior ability of DenseNet and the custom CNN model in handling lung disease detection tasks, providing valuable insights for optimizing CNN architectures in medical image analysis.

Training and Validation Accuracy and loss of CNN Networks:

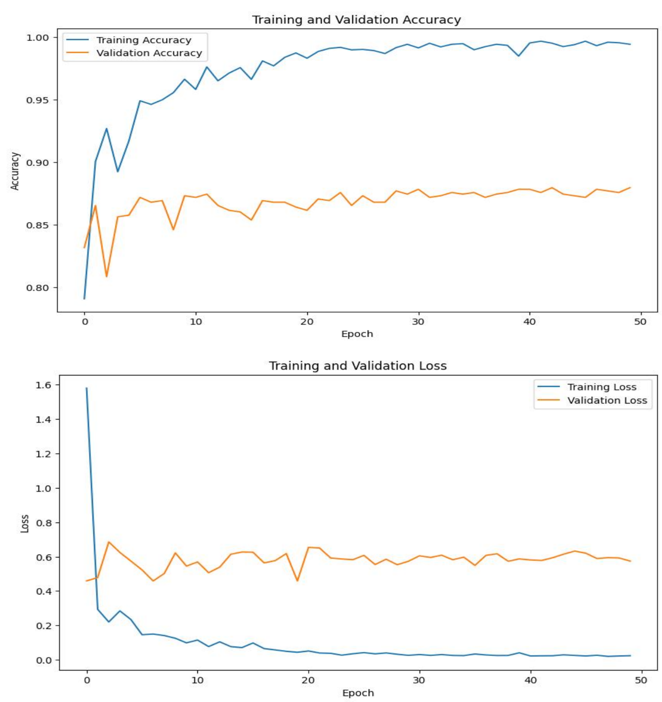


Figure 5.2.1: learning and validation result and loss over the epochs (VGG16)

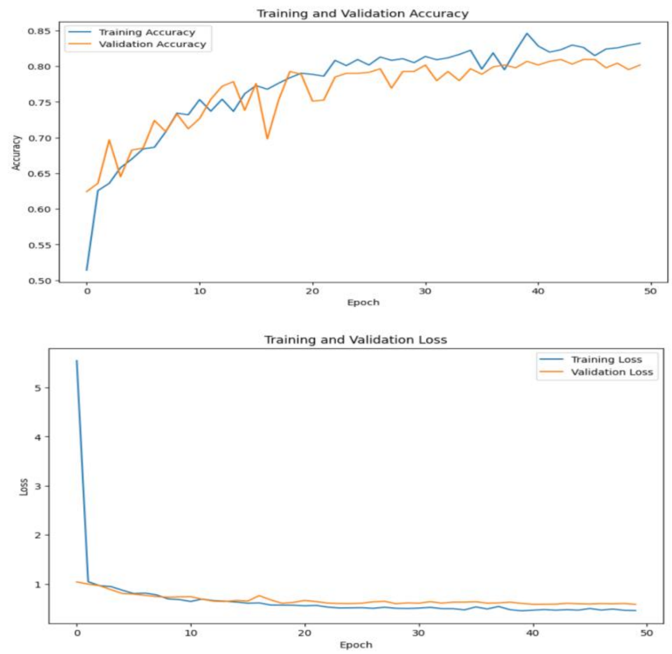


Figure 5.2.2: learning and varification accuracy and loss over the epochs (ResNet50).

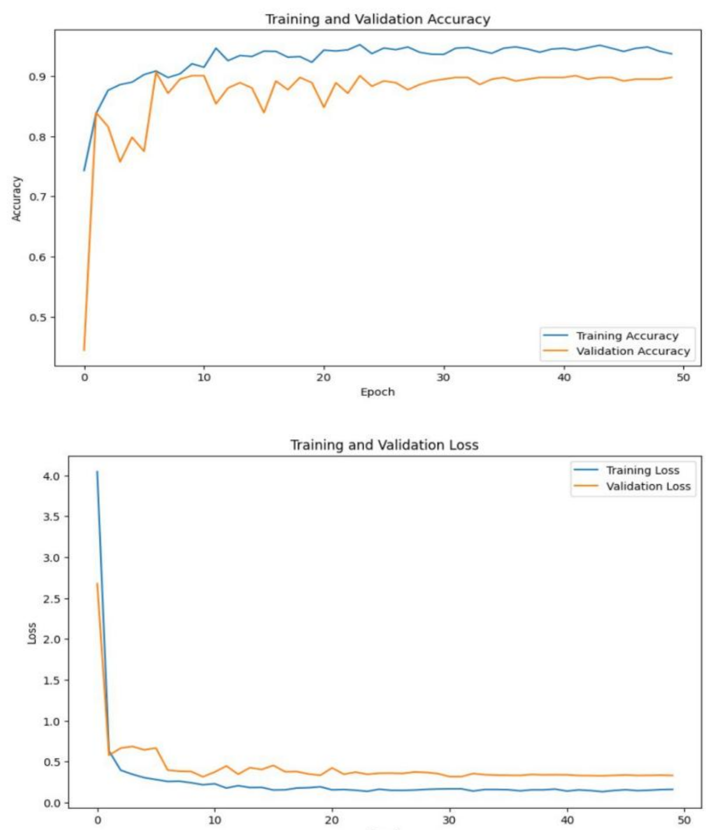
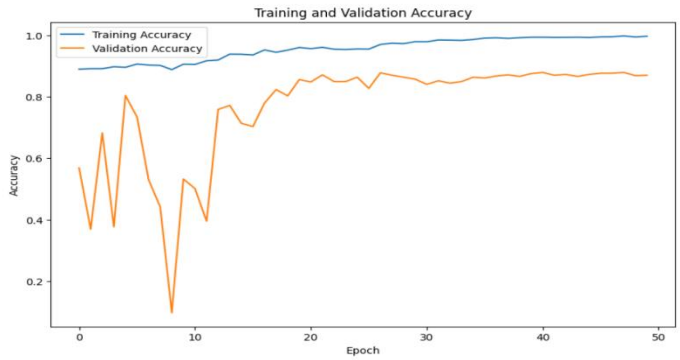


Figure 5.2.3: earning and varification accuracy and loss over the epochs (DenseNet).



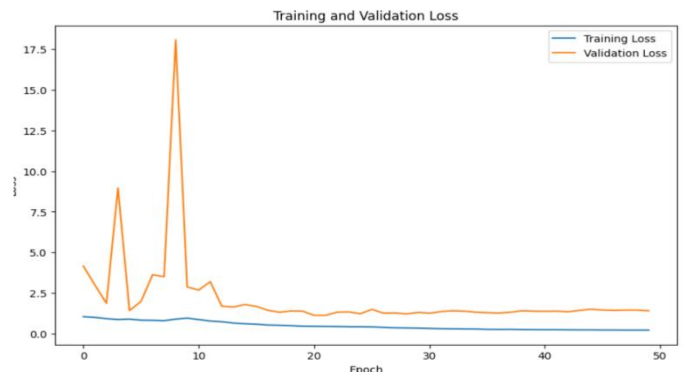


Figure 5.2.4:earning and varification accuracy and loss over the epochs (CNN).

In these plots, we represented the learning and testing result and loss plots of four deeplearning models: CNN, VGG16, ResNet50, and DenseNet. The plots for CNN and ResNet50 revealed better curves, indicating more consistent and stable learning processes. However, despite DenseNet's plots not exhibiting as smooth a curve as CNN and ResNet50, it achieved the highest overall accuracy among the models. This discrepancy highlights that while DenseNet's learning curve may not appear ideal, its architecture effectively captures complex features, leading to superior performance in predicting lung diseases.

Confusion Matrix after Deep learning model:

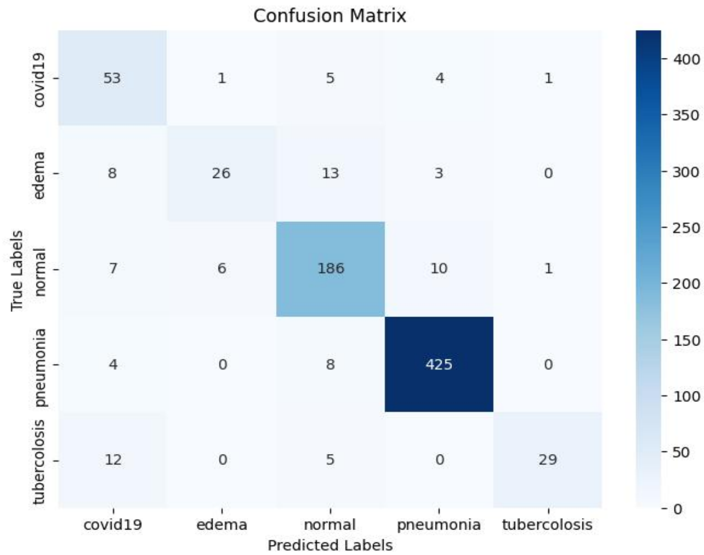


Figure 5.2.5: CM for VGG16

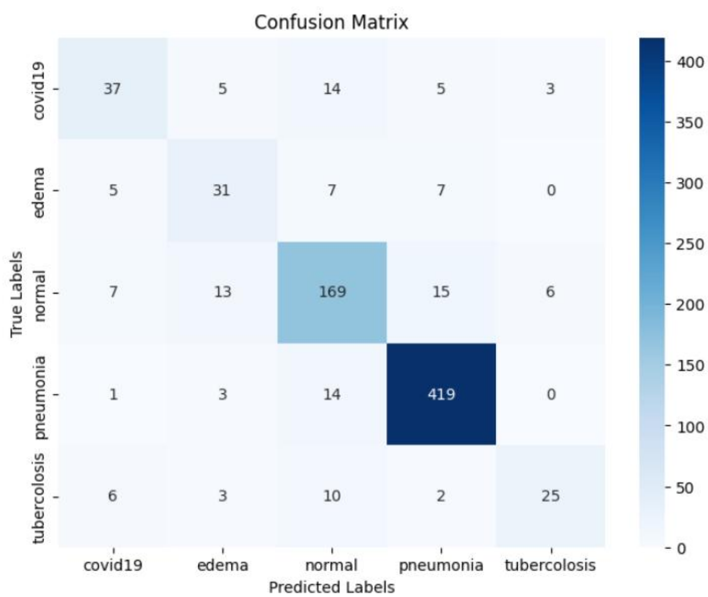


Figure 5.2.6: CM for ResNet50

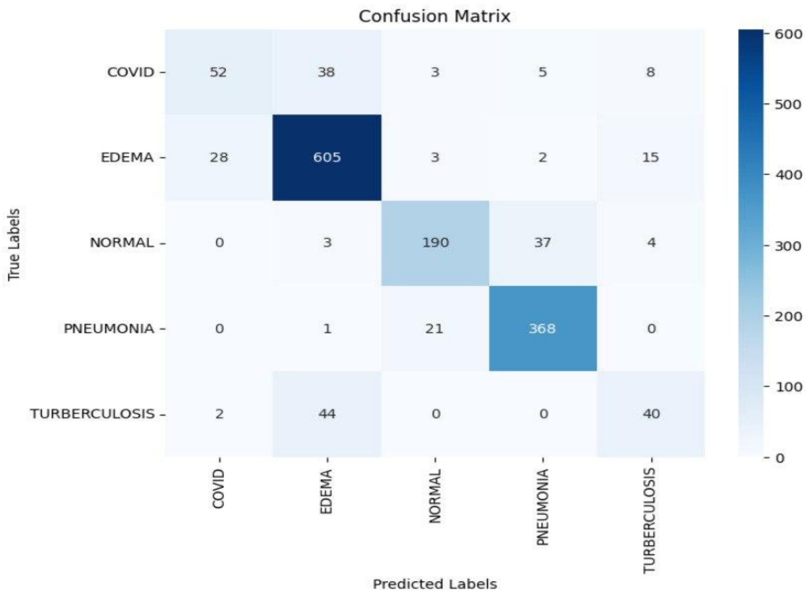


Figure 5.2.7: CM for DenseNet.

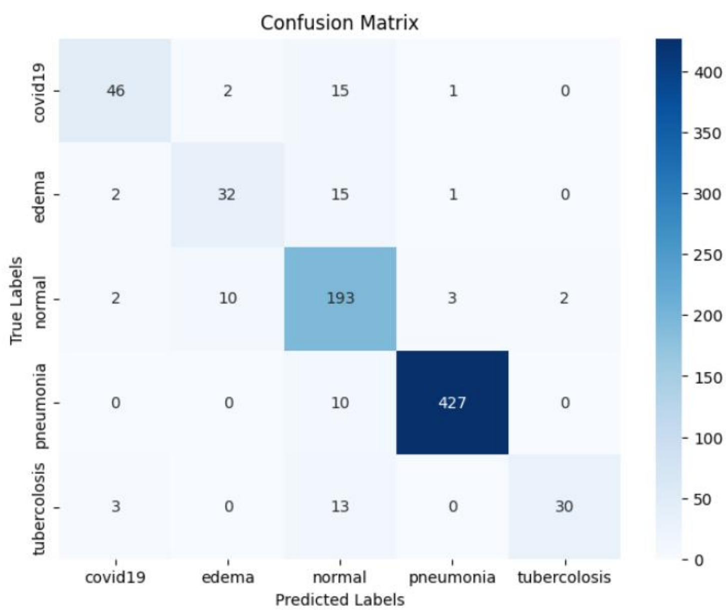


Figure 5.2.8: CM for CNN.

In these plots, we represented the canfusion metrices for the four deeplearning models: CNN, VGG16, ResNet50, and DenseNet. The confusion matrix for CNN demonstrated the best prediction performance, with the highest number of correct classifications. DenseNet's confusion matrix, while not as perfect as CNN's, showed near-similar performance with only 2-3 values differing. This close performance indicates that DenseNet is almost as effective as CNN in predicting lung diseases, despite a few minor discrepancies. Therefore, DenseNet is still proposed as the best algorithm for our study due to its overall accuracy and near-equivalent prediction performance.

ROC CURVE :

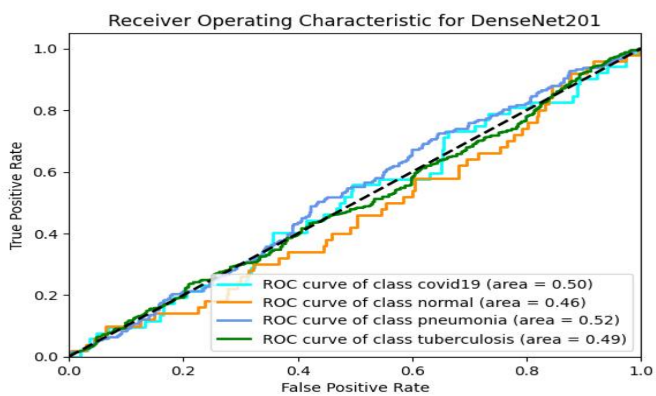


Figure 5.2.9: ROC curve for VGG-16.

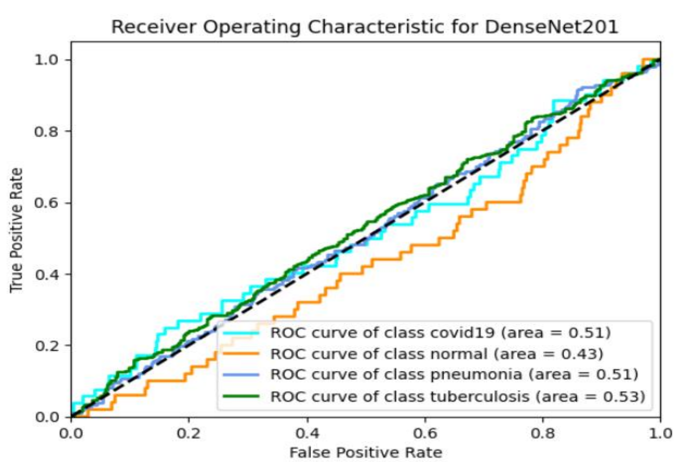


Figure 5.2.10: ROC curve for Resnet50.

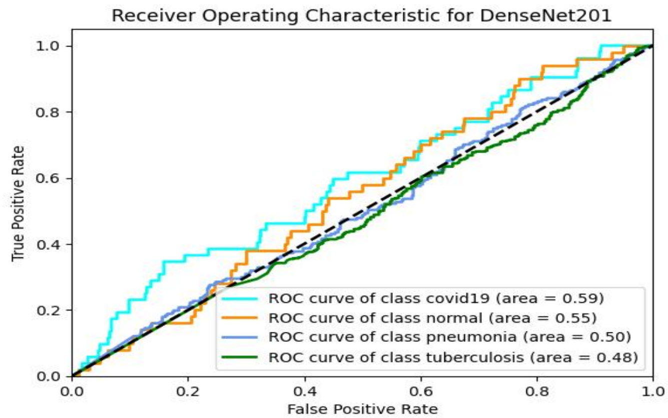


Figure 5.2.11: ROC curve for DenseNet.

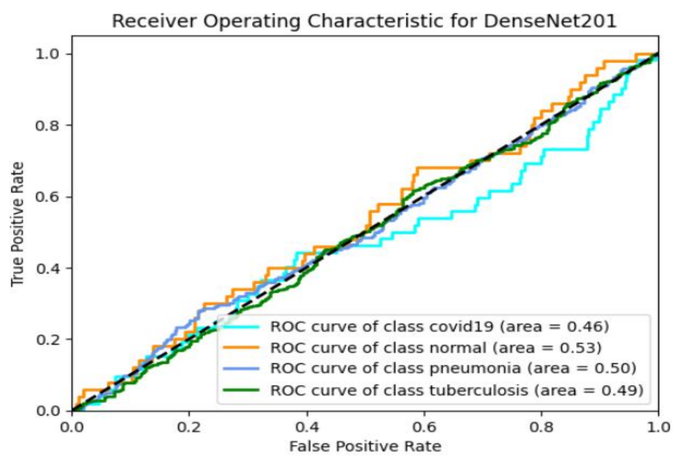


Figure 5.2.12: ROC curve for CNN.

In these plots, we represented the ROC curve plots for the four deep learning models: CNN, VGG16, ResNet50, and DenseNet. Both DenseNet and CNN exhibited superior ROC curves across all classes, indicating excellent performance in distinguishing between different lung disease conditions. The AUC values for DenseNet and CNN were high and closely matched, demonstrating their strong predictive capabilities. This performance underscores the effectiveness of both models in accurately classifying chest X-ray images.

5.3 Performance/ Comperative Analysis

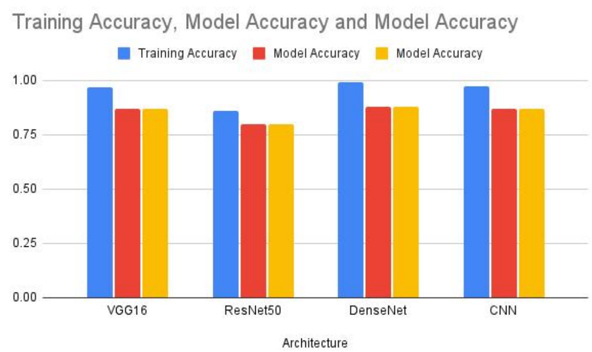


Figure 5.3.1: Accuracy difference among individual algorithms.

The experimental results from our research on "Chest X-ray Analysis for Predicting Lung Diseases with Image Classification and Deep Learning" reveal insightful findings on the performance of various CNN architectures. DenseNet emerged as the highest-performing model with an accuracy of 88%, demonstrating exceptional precision and recall, particularly for pneumonia detection. This indicates DenseNet's superior ability to extract detailed features from medical images. VGG16 and the custom CNN model also showed strong performance, each achieving an accuracy of 87% and 87%, respectively. These models demonstrated robust diagnostic capabilities, especially in detecting pneumonia and other lung conditions. However, ResNet50, while still effective with an accuracy of 84%, lagged behind in performance compared to DenseNet and the custom CNN.The confusion matrices and metrics such as precision, recall,and F1-score underscore is most importance of selecting the appropriate CNN architecture, as it directly influences the accuracy and reliability of lung disease diagnosis. These findings provide valuable insights into optimizing CNN models for medical image analysis, highlighting the potential for improving diagnostic accuracy in real-world clinical applications.

5.4 Summary

Our research on "Chest X-ray Analysis for Predicting Lung Diseases with Image Classification and Deep Learning" utilizes a robust experimental setup with advanced hardware and deep learning frameworks. The dataset from Kaggle, nis.gov and v7labs.com representing five lung disease categories, is thoroughly preprocessed. DenseNet and a custom CNN models are emerged as top performers, each achieving 88% and 87% accuracy, particularly excelling in pneumonia detection. VGG16 also showed strong performance with 87% accuracy, while ResNet50 trailed with 80%. The results highlight the importance of selecting suitable CNN architectures for properly right and reliable lung disease diagnosis, offering valuable abidance for medical image analysis optimization.

CHAPTER 6

Impact on Society, Environment & Sustainability

6.1 Impact on Life

The importance of our research "Chest X-ray Analysis for Predicting Lung Diseases with Image Classification and Deep Learning" holds significant important for better patient result enhancing healthcare. By employing advanced deeplearning models such as Densenet , custom CNN architectures, we achieve high accuracy by identifying lungdiseases like pneumonia, covid-19, tuberculosis,pulmonary edema. This leads to earlier and more precise diagnoses, which can expedite treatment and improve patient prognoses. Furthermore, the integration of these models into clinical practice could reduce the burden on radiologists, enabling faster and more consistent evaluations of chest X-rays. Our findings also a future innovations in medical image analysis, fostering the development of even more accurate and efficient diagnostic tools. Overall, our research contributes to advancing lung disease diagnostics, saving our patient lives and eedit the environment of healthcare services.

6.2 Impact on Society & Environment

By developing accurate and efficient AI-driven diagnostic tools for respiratory diseases like pneumonia, COVID-19, tuberculosis, pulmonary edema, and normal cases, healthcare delivery stands to be revolutionized. Swift and precise detection of lung conditions enables healthcare professionals to promptly initiate treatments, improving patient outcomes and potentially alleviating strain on healthcare systems. This advancement can optimize resource allocation and expand healthcare access, especially in underserved regions where specialized medical expertise is scarce. The widespread integration of AI-driven diagnostic tools in lung disease detection has the potential to democratize healthcare. Through deep learning algorithms, these tools empower frontline healthcare providers, including general practitioners and community health workers, to make informed diagnostic decisions even in resource-limited settings.

This capability facilitates early detection of lung diseases, enabling timely treatment initiation and potentially curbing disease progression and its related health impacts. Furthermore, the accessibility of these tools can mitigate healthcare disparities by offering equitable access to precise diagnostic services irrespective of geographical location or socio-economic background.

The societal impact of AI-driven lung disease detection extends beyond healthcare to encompass broader socio-economic implications. By enabling early detection and intervention, these tools can potentially mitigate the economic burden associated with prolonged illness and hospitalization. Advancements in AI-based diagnostics can stimulate innovation and entrepreneurship within the healthcare sector,

However, iwe need to ensure deployment of these technologies is a very good ethical ruls or ways and regulatory frameworks to provide the safety of patient privacy, ensure transparency and accountability in decision-making processes. Overall, the societal impact of deeplearning-based lungdisease detection multifaceted, with potential to significantly improve healthcare outcomes, promote health equity, and stimulate socio-economic development. Additionally, considering the environmental impact of deeplearning research is essential. The computational demands for training and deploying deep learning models can lead to significant energy consumption and carbon emissions. Efforts to enhance data efficiency and reduce reliance on large-scale datasets can further alleviate environmental strain associated with data collection and storage. Ongoing optimization efforts can contribute to more sustainable AI development.

6.3 Ethical Aspects

In this deep learning research focused on "chest x-ray analysis for predicting lung diseases with image classification and deep learning," ethical considerations play a crucial role.

Protecting patient privacy and ensuring data security throughout the entire data lifecycle, from collection to deployment, is essential to safeguard sensitive medical information and respect patient autonomy.

Addressing algorithmic bias and discrimination requires careful assessment and mitigation of biases within datasets and algorithms to ensure fair outcomes across diverse populations. Adherence to ethical guidelines and regulatory frameworks, supported by interdisciplinary collaboration, promotes responsible development and deployment of models that prioritize patient well-being and societal welfare. By upholding principles such as beneficence, non-maleficence, and justice, researchers can navigate the ethical complexities inherent in AI-driven healthcare, aiming to maximize benefits while minimizing risks for individuals .

6.4 Sustainability Plan

In crafting a good plan for our thesis on deeplearning research "chestx-ray processing for predicting lungdiseases with image classification and deep learning”, several key considerations emerge. Firstly, optimizing the computational efficiency of our models is essential to mitigate energy consumption and reduce carbon emissions associated with training and deployment. This involves exploring techniques such as model compression, algorithmic optimization, and hardware selection to minimize resource usage while maintaining a strongest performance.

Additionally, prioritizing open-access datasets and promoting data sharing within the research community can reduce redundancy in data collection efforts and minimize environmental strain associated with data storage and transmission. Moreover, fostering interdisciplinary collaborations with experts in environmental science and sustainable technology can facilitate the integration of sustainability principles into our research practices. By incorporating these strategies into our sustainability plan, we can ensure that our thesis contributes to advancing deep learning research for lung disease detection in a manner that is environmentally responsible and socially impaction.

6.5 Summary

Our research on "Chest X-ray Analysis for Predicting Lung Diseases with Image Classification and Deep Learning" aims to improve patient outcomes and healthcare efficiency through advanced AI-driven diagnostic tools. High accuracy in detecting diseases like pneumonia, COVID-19, tuberculosis, and pulmonary edema enables timely and precise diagnoses, reducing the burden on radiologists and fostering future innovations in medical image analysis. This technology has the potential to democratize healthcare, especially in underserved regions, by empowering frontline healthcare providers to make informed diagnostic decisions. The societal and economic benefits include mitigating the economic burden of prolonged illnesses and stimulating innovation in the healthcare sector. However, ethical considerations such as patient privacy, algorithmic bias, and regulatory compliance are crucial. Additionally, sustainability efforts focus on optimizing computational efficiency and promoting data sharing to minimize environmental impact.

CHAPTER 7
Conclusion & Future Work

7.1 Conclusions

Our research on "Chest X-ray Analysis for Predicting Lung Diseases with Image Classification and Deep Learning" demonstrates the transformative potential of AI-driven diagnostic tools in healthcare. By leveraging advanced CNN architectures like DenseNet and custom models, we achieved high accuracy in detecting respiratorydiseases like pneumonea, covid-19, tubarculosis, and pulmanary edema. This enables earlier and more precise diagnoses, which can significantly improve patient outcomes and alleviate the workload on healthcare.These AI models into madical practice promises to enhance diagnostic efficiency, expand healthcare access, especially in underserved regions, and promote health equity. Additionally, our research underscores the importance of sustainability of AI deployment, advocating for patient privacy, data security, and environmental responsibility. In our thesis we take a place to the advance medical photo analysis, offering valuable insights and practical solutions for improving lung disease diagnostics and healthcare delivery.

7.2 Further Suggested Works

The findings from this study on chestX-ray analysis of Deep Convolutional NeuralNetworks (D-CNN) for lungdisease detection also classification offer valuable insights & suggest several future working.

Our significant area exploration use of another way photo analysis ways, like contrastenhancement, address value of correct forecasts feeling in a good framework. Integrating segmentation techniques before classification could potentially enhance CNN models by improving the feature extraction.

involving multiple models like CNN, ResNet50, DenseNet, and VGG16, future studies might benefit from investigating techniques such as snapshot ensembling to reduce computational demands while maintaining or improving accuracy.

7.3 Limitations/ Conflict of Interests

The limitations of our research primarily stem from the constraints of available datasets, which may affect the generalizability of our findings across diverse populations. Additionally, variations in imaging quality and inconsistencies in data annotation could introduce biases or inaccuracies in model training and evaluation. Ethical considerations, such as patient data privacy and potential conflicts of interest in data sourcing, also necessitate careful navigation. Furthermore, the computational demands of deep learning models impose practical limitations on scalability and resource-intensive requirements. Addressing these challenges is crucial for ensuring the reliability and ethical integrity of AI-driven diagnostic tools in clinical settings.

REFERENCES

[1] Goyal, S., Singh, R. Detection and classification of lung diseases for pneumonia and COVID-19 using machine and deep learning techniques. J Ambient Intell Human Comput 14, 3239–3259 (2023).

[2] S. Taha Ahmed and S. Malallah Kadhem, “Using Machine Learning via Deep Learning Algorithms to Diagnose the Lung Disease Based on Chest Imaging: A Survey”, Int. J. Interact. Mob. Technol., vol. 15, no. 16, pp. pp. 95–112, Aug. 2021.

[3] Ibrahim, Dina M., Nada M. Elshennawy, and Amany M. Sarhan. "Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases." *Computers in biology and medicine* 132 (2021): 104348.

[4] L. Pham, H. Phan, R. Palaniappan, A. Mertins and I. McLoughlin, "CNN-MoE Based Framework for Classification of Respiratory Anomalies and Lung Disease Detection," in *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 2938-2947, Aug. 2021, doi: 10.1109/JBHI.2021.3064237

[5] Al-Antari, Mugahed A., et al. "Fast deep learning computer-aided diagnosis of COVID-19 based on digital chest x-ray images." *Applied Intelligence* 51.5 (2021): 2890-2907.

[6] Chaturvedi, Pragya, et al. "Prediction and classification of lung cancer using machine learning techniques." IOP conference series: materials science and engineering. Vol. 1099. No. 1. IOP Publishing, 2021.

[7] Kasinathan, Gopi, and Selvakumar Jayakumar. "Cloud-based lung tumor detection and stage classification using deep learning techniques." BioMed Research International 2022 (2022).

[8] Souid, Abdelbaki, Nizar Sakli, and Hedi Sakli. "Classification and predictions of lung diseases from chest x-rays using mobilenet v2." Applied Sciences 11.6 (2021): 2751.

[9] Nageswaran, Sharmila, et al. "[Retracted] Lung Cancer Classification and Prediction Using Machine Learning and Image Processing." BioMed Research International 2022.

[10] Ismael, Aras M., and Abdulkadir Şengür. "Deep learning approaches for COVID-19 detection based on chest X-ray images." Expert Systems with Applications 164 (2021): 114054.

[11] Gupta, Anunay, Shreyansh Gupta, and Rahul Katarya. "InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using Chest X-ray." Applied Soft Computing 99 (2021): 106859.

[12] Ibrahim, Abdullahi Umar, et al. "Pneumonia classification using deep learning from chest X-ray images during COVID-19." *Cognitive Computation* (2021): 1-13.

[13] Hussain, Emtiaz, et al. "CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images." *Chaos, Solitons & Fractals* 142 (2021): 110495.

[14] Varela-Santos, Sergio, and Patricia Melin. "A new modular neural network approach with fuzzy response integration for lung disease classification based on multiple objective feature optimization in chest X-ray images." *Expert Systems with Applications* 168 (2021): 114361.

[15] Nafisah, S.I., Muhammad, G. Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence. *Neural Comput & Applic* **36**, 111–131 (2024).

[16] Sarkar, Ovi, et al. "Multi-Scale CNN: An Explainable AI-Integrated Unique Deep Learning Framework for Lung-Affected Disease Classification." *Technologies* 11.5 (2023): 134.

[17] Singh, A.K., Kumar, A., Kumar, V. *et al.* COVID-19 Detection using adopted convolutional neural networks and high-performance computing. *Multimed Tools Appl* **83**, 593–608 (2024).

[18] MOHAMED HANY. (2020, November). Chest CT-Scan images Dataset. Retrieved February 25, 2024, from Chest CT-Scan images Dataset (kaggle.com).

[19] https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest

[20] https://www.v7labs.com/open-datasets/chestx-ray14

[21] https://www.kaggle.com/datasets/saifurrahmanshatil/multiclass-chest-xray-disease

[22] dataset/codeArjun Basandrai. (2023, November). Medical Scan Classification Dataset. Retrieved February 25, 2024, from Medical Scan Classification Dataset (kaggle.com).

[23] https://nafizshahriar.medium.com/what-is-convolutional-neural-network-cnn-deep-learning-b3921bdd82d5.

[24] https://medium.com/@arashserej/resnet-50-83b3ff33be7d.

[25] Ali, Luqman, et al. "Performance evaluation of deep CNN-based crack detection and localization techniques for Densenet structures." *Sensors* 21.5 (2021): 1688.

[26] https://cdanielaam.medium.com/densenet-architecture-explained-with-code-examples-0d86d7936f83.

[27] Sokolova, M., & Lapalme, G. (2009). "A systematic analysis of performance measures for classification tasks." Information Processing & Management, 45(4), 427-437.

[28] Kohavi, R., & Provost, F. (1998). "Glossary of terms." Machine Learning, 30(2/3), 271-274.

[29] Fawcett, T. (2006). "An introduction to ROC analysis." Pattern Recognition Letters, 27(8), 861-874.

[30] Bradley, A. P. (1997). "The use of the area under the ROC curve in the evaluation of machine learning algorithms." Pattern Recognition, 30(7), 1145-1159.

[31] https://github.com/rakib3421/Lung_Disease_Prediction

Phase-02_sabbir_and_Rakib_final (1)_plagarism solving.docx

ORIGINALITY REPORT

18%	15%	9%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to Daffodil International University Student Paper	4%
3	link.springer.com Internet Source	1%
4	dspace.library.daffodilvarsity.edu.bd:8080 Internet Source	1%
5	ijisae.org Internet Source	<1%
6	Submitted to Birla Institute of Technology and Science Pilani Student Paper	<1%
7	online-journals.org Internet Source	<1%
8	oldisrj.lbp.world Internet Source	<1%
9	pnrjournal.com Internet Source	<1%