Detection of Pneumonia in Chest X-rays: A Deep Neural Network Approach

SUBMITTED BY

Arpita Rani Ball
Batch: 37
ID: 221-35-1006
Department of Software Engineering
Daffodil International University

SUPERVISED BY

Mr. Khalid Been Badruzzaman Biplob
Lecturer. (Senior) , Department of SWE, FSIT
Daffodil International University

Thesis submitted in fulfillment of the requirements for the award of the degree of Bachelor of Science

Fall – 2025

APPROVAL

This thesis titled on **Detection Of Pneumonia In Chest X-ray in Deep Neural Network Approach** submitted by **Arpita Rani Ball (221-35-1006)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Dr. S. M. Hasan Mahmud **Associate Professor** Department of Software Engineering Faculty of Science and Information Technology Daffodil International University	**Chairman**
A.H.M Shahariar Parvez **Associate Professor** Department of Software Engineering Faculty of Science and Information Technology Daffodil International University	**Internal Examiner 1**
Tapushe Rabaya Toma **Assistant Professor** Department of Software Engineering Faculty of Science and Information Technology Daffodil International University	**Internal Examiner 2**
Khalid Been md. Badruzzaman Biplob **Lecturer (Senior Scale)** Department of Software Engineering Faculty of Science and Information Technology Daffodil International University	**Internal Examiner 3**
Dr. Md Sazzadur Rahman ***Professor*** Institute of Information technology Jahangirnagar University, Bangladesh	**External Examiner**

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.

(Supervisor's Signature)

Mr. Khalid Been Badruzzaman Biplob
Position: Lecturer. (Senior), Department of SWE, FSIT
Date: 27 th November, 2025

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

Arpita Rani Ball.

(Student's Signature)

Full Name : Arpita Rani Ball

ID Number : 221-35-1006

Date : 27 th November, 2025

ACKNOWLEDGEMENT

First and foremost, I am grateful to Almighty, who has given me the strength, wisdom, and perseverance to complete this research. All through my academic journey, I am grateful for the unconditional love, support, and encouragement of my parents. It is always their belief in me that has motivated and inspired me the most.

I would like to thank my supervisor, Lecturer (Senior) , Department of SWE, Mr. Khalid Been Badruzzaman Biplob for his valuable advice, support, and guidance throughout the research. His knowledge and insight have highly affected this work. I highly appreciated Mr. Khalid Been Badruzzaman Biplob for his support, guidance, and valuable comments which helped me to successfully complete my journey. Finally, I would like to thank all my friends, batchmates, and all those who helped and encouraged me during this process.

ABSTRACT

Pneumonia remains a serious health concern, particularly in regions where access to experienced radiologists is limited. Although chest X-rays are commonly used for diagnosis, accurate interpretation often depends on expert judgment, which is not always available in resource-constrained settings. As a result, delayed or incorrect diagnoses are still common. This study explores the use of deep learning techniques to support pneumonia detection from chest X-ray images, while addressing practical challenges such as data imbalance, limited dataset diversity, and lack of model transparency.

Rather than relying only on widely used public datasets, this research makes use of a locally collected and radiologist-verified dataset consisting of 1,500 chest X-ray images obtained from a medical facility in Bangladesh. The images were carefully preprocessed using lung-preserving cropping and clinically safe data augmentation in order to reduce noise and improve class balance. A custom-built Convolutional Neural Network (CNN) was developed and compared with several transfer learning models, including VGG19, DenseNet201, MobileNetV2, and ResNet50. Among these, VGG19 produced the best overall results, achieving a test accuracy of 97.88%, while the custom CNN also demonstrated strong performance with 96.00% accuracy. Evaluation was carried out using standard metrics such as accuracy, precision, recall, F1-score, ROC analysis, and confusion matrices.

To improve clinical trust, Grad-CAM visualizations were applied to highlight image regions that influenced model predictions. The findings suggest that transfer learning models, particularly VGG19, can serve as reliable tools for automated chest X-ray screening and may help reduce diagnostic workload in healthcare environments with limited resources.

Keywords: *Pneumonia detection, Chest X-ray imaging, Deep learning, Convolutional neural network (CNN), Transfer learning, VGG19, Medical image analysis.*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Overview

One of the most frequent and dangerous respiratory diseases that can be contracted by children, older people, and those that live in areas with resources is pneumonia. One of the most readily available and quickest methods of diagnosis of pneumonia clinically is the imaging of the chest using X-ray; nevertheless, proper interpretation of the imaging is limited to the presence of experienced radiologists. In most Third world countries, Bangladesh being one of them, misdiagnosis, potential treatment loss and avoidable deaths is common due to untrained radiology specialists.

With the growth of artificial intelligence (AI) and deep learning in particular, in particular Convolutional Neural Networks (CNNs), is being able to show impressive capabilities when it comes to the analysis of medical images. Radiologists can be assisted with automated systems to give them a fast, consistent and reliable diagnostic insight. The objective of the research is to create a deep learning-oriented system used to identify pneumonia in X-ray chest photographs and it will not only confront the significant issues of low diversity of the dataset, distribution of classes, system generalizability, and interpretability.

1.2 Background

Through bacterial, viral, or fungal infections, pneumonia results in inflammation of the lung to cause millions of hospitalizations annually. The diagnosis is very important and timely, but radiologists are not easily accessible in most of the rural and underserved regions. The process of interpretation in relation to chest x-ray is subjective and time consuming and error prone particularly where there is poor image quality and ambiguity.

The field of medical imaging has undergone revolution, with deep learning models learning features automatically without the need to manual engineer features (one does not need to create features, unlike with other models). VGG, ResNet, MobileNet, and custom CNNs demonstrated to be useful in detecting pneumonia. However, the majority of the available research extensively relies on

publicly available datasets like NIH CXR or Kaggle pneumonia datasets. Such datasets are not sufficiently diversified enough to be generalizable to clinical settings in the real world.
Thus, creation of a more diverse data set, a solid CNN architecture, and explainability by generating Grad-CAM images can greatly improve the reliability of the model and its acceptance in clinical practice.

1.3 Problem Statement

Pneumonia is still one of the significant health issues of the population, mainly in the areas with inadequate medical facilities. Even though the chest X-rays are commonly used in the diagnosis of pneumonia, it cannot be correctly read without well trained radiologists- something that is not readily available in a good number of developing nations. This causes patients, on many occasions, to have delayed diagnoses, misunderstandings and poor treatments.
The approach of deep learning is a feasible solution to an automated solution; the current models have weaknesses in lack of diversity in data, lack of equal representation of normal and pneumonia images and absence of transparency in model decisions. Thus, a trustworthy, precise, and interpretable deep learning-based system should be sought to help detect pneumonia and decrease the workload of medical workers with diagnosis.

1.4 Research Gaps and Research Questions:

This section will explicitly list the gaps your research addresses and the questions it aims to answer.

Research Gaps:

From the reviewed literature and observations, several gaps have been identified:

- **Limited Dataset Diversity:** Many existing studies use publicly available datasets that suffer from class imbalance, small sample size, and low diversity, which limits model generalizability.
- **Lack of Real-World Deployment Evaluation:** Few studies focus on practical implementation or real-time usage of models in healthcare settings.

- **Lack of Model Interpretability:** Majority of the models act like black boxes, giving predictions without the process of decision making-which is a condition of clinician trust.
- **Minimal Differentiating Pneumonia types:** Most of the literature only differentiates between normal vs pneumonia and does not differentiate the types of pneumonia and complications associated with it.

1.5 Research Questions:

Deep learning approaches have demonstrated encouraging performances in the analysis of medical images, but there are still a number of issues in the implementation of these paradigms in actual clinical practice. Such obstacles are the constraints in data availability, their unequal distribution by classes, fluctuation in image quality, and medical decision-making that requires the transparency of the model. In order to solve these problems and understand the course of this research, a number of research questions are developed. These questions will also help to outline how well CNN-based pneumonia detection is performed and investigate how to achieve better quality of data and ability to explain the model, and how a self-collected dataset can be able to achieve better results in the real world. The research questions that outline the essence of these study are the following:

Research Question 1: How effectively can a CNN-based model detect pneumonia from chest X-ray images?

The question is based on the ability of a CNN to accurately classify pneumonia cases and normal cases based on medical imaging. It is aimed at testing the ability of the model to learn meaningful characteristics on the basis of chest X-rays and deliver credible results of detection. Through judging performance based on such metrics as accuracy, precision, recall, F1-score, and ROC-AUC, the paper concludes whether CNNs have the capacity to aid or enhance clinical diagnosis. The question also contributes to the definition of the strengths and weaknesses of deep learning in medical situations in the real world.

Research Question 2: How can dataset imbalance, variation, and lack of diversity be addressed during model development?

This question examines methods that will make the model learn based on a balanced and diverse set of examples. Medical datasets are also prone to having more images of pneumonia than normal pictures bias the model. These are some of the techniques used to counter this such as data augmentation, class weighting, controlled sampling, and restraining preprocessing. Normalisation and augmentation are also used to vary in image quality, position, and equipment. The question is meant to take care of fairness, strength, and improved generalization.

Research Question 3: Can a custom-collected dataset enhance the model's generalization ability compared to public datasets?

The question determines the possibility of the improvement of the performance on the real image through a dataset collected at the local hospitals. Standards. Public datasets are often provided with clean and standardised images which do not necessarily reflect the local clinical conditions. Real variations and noise of various patients and equipment are contained in a custom dataset as well as variations in diversity. The model can adjust easier to the local settings and be able to perform more reliably by training on this dataset. The issue addressed in the question is how the difference in data sources influences the generalisation of models.

1.6 Research Objectives: The research objectives are clear statements of what your study aims to achieve.

- RO1: To prepare a CNN-based detection model in the classification of normal and pneumonia chest X-ray images.
- RO2: To gather and preprocess the imaging of the chest X-rays in the hospitals and diagnostic centers as a means of enhancing the diversity of the datasets.
- RO3: To tackle the issues in data related to the class imbalance, variation in image quality, and noise.

- RO4: To measure model performance based on accuracy, loss curves, ROC, confusion matrix, and others.

- RO5: To enhance the interpretation of the model with the help of visualization methods like Grad-CAM.
- RO6: To create a viable AI-based diagnostic system that can be used in resource constrained medical settings.

The primary objective of this project is to develop a robust deep-learning algorithm to identify pneumonia in the chest X-rays. To achieve this, the study is aimed at constructing a high-quality dataset being gathered locally, which is more realistic in terms of a real clinical situation and the accuracy of the model functioning. The other worthy objective is to address imbalance and variation in datasets by preprocessing and augmentation, and ensure learning of strong meaningful features by the model. The paper also intends to compare a home-made CNN architecture and a number of transfer learning models to identify the most precise and efficient solution. Also, the study emphasizes the enhancement of interpretability by means of approaches that assist the clinicians in comprehending the decisions made by the model to render the system more viable in healthcare usage. Altogether, the goals are connected with the development of a convenient, correct, and articulable pneumonia identification model that can be implemented in medical conditions with limited resources.

Chapter 2
Literature Review

2.1 Previous Literature Review

Detection of pneumonia based on the chest X-ray (CXR) has been a topic of considerable interest over the past few years because of the development of deep learning and the growth of medical imaging datasets. Numerous researches have revealed that Convolutional Neural Networks (CNNs) are more effective than conventional machine learning techniques in classifying medical images, especially in pneumonia detection.

Sharma and Guleria (2023) carried out a thorough article on the pneumonia detection based on deep learning and outlined such important concerns as the imbalanced dataset and poor model interpretability. Their systemic review found that CNNs had good accuracy, but the constraint in the data variety lowered the generalizability in the real-life conditions. Siddiqui and Javaid (2024) extended it to investigate various deep learning models and stated the significance of a robust model. And implemented effective pre-processing procedures to enhance diagnostic reliability.

Aljawarneh and Al-Quraan (2023) introduced a superior CNN model that enhanced the classification of pneumonia with the use of refined feature extraction methods. Their experiment indicated that they had a performance improvement but still relied on publicly available datasets.

Ahamed et al. (2023) presented a better ResNet architecture, DTLCx, which is able to differentiate pneumonia, normal indexes, and COVID-19 sequences with Grad-CAM-based visualizations. Their efforts pointed out the need to make medical settings more interpretable. Their model was however silent on the issue of diversity of datasets as they used the already available datasets.

Muthukumar et al. (2025) also studied the analysis of pneumonia severity using CNN and suggested automated diagnostic assistance systems. Their findings were very accurate but not applicable to the real-world application as the calculations required were complicated, and there was no primary dataset gathering.

In a comparison of several traditional ML algorithms with CNN features, Rattan et al. (2024) established that hybrid algorithms could achieve good results in structured feature extraction. Nevertheless, CNN-based end-to-end learning was better in the vast majority.

The different CNN and transfer learning models, such as VGG16, ResNet, and MobileNet, were experimented by Zhang et al. (2021) and Manickam et al. (2021). Their results revealed that the transfer learning can be used to cut down the training expenses and preserve the high diagnostic accuracy. However, the similar point is that these works are limited by one factor: they use extensive datasets that are publicly available, which restricts their clinical use.

Saleh et al. (2023) were able to utilize MobileNet to implement lightweight pneumonia detection, which could be used in a mobile and remote healthcare setting. However, the research was not supported by real life adoption using a variety of data.

The weed of systematic reviews presented in a study by Khan et al. (2021) was the absence of strong interpretability tools, the inadequate distinction between the types of pneumonia, and the little emphasis on real-time implementation.

2.1 Comparative Overview of Previous Literature

To better understand the strengths and weaknesses of previous studies, a comparative analysis is presented below.

Table 2.1: Comparative Overview of Previous Literature

Study	Model/Method Used	Dataset Type	Key Strengths	Key Limitations
Sharma & Guleria (2023)	Systematic Review of CNNs	Public datasets	Identified imbalance and interpretability	Did not propose a new model
Siddiqi & Javaid (2024)	Multiple CNN frameworks	NIH/Kaggle datasets	Comprehensive comparison	Lack of primary data validation
Aljawarneh & Al-Quraan (2023)	Enhanced CNN	Public datasets	High accuracy through improved feature extraction	Limited dataset diversity
Ahamed et al. (2023)	DTLCx (Improved ResNet)	COVID-19 + Pneumonia	Strong interpretability (Grad-CAM)	Still relies on secondary datasets
Muthukumar et al. (2025)	CNN with severity analysis	Public datasets	Severity-based analysis	No real-time deployment testing
Rattan et al. (2024)	ML + CNN hybrid	Public datasets	Effective hybrid approach	Less accurate than CNN
Zhang et al. (2021)	Custom CNN	NIH dataset	Good baseline results	Lack of generalization
Manickam et al. (2021)	Transfer Learning (VGG, ResNet)	CXR datasets	High accuracy, low training time	Relies on standard datasets
Saleh et al. (2023)	MobileNet	Lightweight mobile deployment	Suitable for resource-limited regions	No primary data collection or diversity
Khan et al. (2021)	Systematic Review	Multiple datasets	Identified need for explainability	No empirical model

In general, the literature review shows that the development of deep learning-based pneumonia detection has reached a high level, but such challenges as the limitations of the datasets, the absence of real-life implementation, and low transparency are still unsolved. This drives the creation of a clinical and diverse deep learning system.

The analysis of literature indicates that publicly available datasets have been used heavily with fewer custom datasets and primary datasets being used. Consequently, most models are trained using standardized data, which do not have differences in the environment and context. This reduces their usability in real life scenarios. The research papers primarily apply CNN architectures and Transfer Learning (TL) models such as VGG, ResNet, MobileNet, and their enhanced versions. These approaches are quite accurate and efficient, and they are limited by the homogeneity of the datasets on which they are based. Although numerous publications indicate the lack of balance, diversity, and the necessity to use robust models, they do not provide alternative datasets and frameworks, which would address the existing issues. Moreover, despite the fact that other studies also address the issue of interpretability, they are mostly aimed at enhancing performance, as opposed to transparency. All in all, the existing literature demonstrates the tendency towards the successful CNN/TL-based solutions that are underdeveloped due to the lack of special datasets, absence of primary data verification and insufficient attention to real-life flexibility.

Chapter 3
Methodology

3.1 Research Subject and Instrumentation

The paper is aimed at creating and testing an effective deep learning solution to diagnose lung diseases. It seeks to enhance the quality and precision with which various skin disorders can be identified with the help of such modern algorithms as CNN, InceptionV3, DenseNet201, ResNet50, and VGG19. The key research instruments utilized in this research are computer power to train and evaluate models. The deep learning models will be trained on high-performance GPUs (which may be in the cloud) and will be able to process the large dataset efficiently. The algorithms will be implemented with the help of such software as PyTorch and TensorFlow. Augmentation libraries and image processing are also open-source libraries that will assist in preparing the dataset. The algorithms shall be developed by the research team through project management systems and programming languages such as Python. The introduction of the developed model in clinical practice will be ethically and legally acceptable. The combined use of these tools makes the suggested deep learning method under study complete.

3.2 Data Collection Procedure

The research procedure of data collection was developed to ensure the reliability, ethical origin and clinical significance of the X-ray images of the chest in this research. The primary objective was to obtain high-quality pneumonia and normal lung X-ray images in a clinical actual environment. This was in order to tackle the constraints of diversity in publicly available datasets.

The images of X-ray of the chest in Turkey Medical Services were obtained, which is situated in the Trade Center of Anwar, Sk. Mojib Road, Agrabad, Chittagong, Bangladesh. Close interactions with medical practitioners and radiology staff were to be used when collecting the cases to ensure that the accuracy and authenticity of each case. All the images were acquired during regular clinical check-ups and no additional procedures were conducted specifically to conduct research.

Prior to the collection of data the institutional review committee gave ethical approval and permission was obtained at the partnering medical facility. Patient privacy and confidentiality were maintained strictly, and no personal identifiers (name, age, contact details, and hospital identification numbers) were occupied. Only X-ray and diagnostic label (a pneumonia or a normal image) were retained to analyze them.

It was aimed to gather approximately 1500 chest x-ray images with a 60:40 proportion between pneumonia cases and normal lung as recommended by the consulting clinicians. All the images were checked and verified by a qualified radiologist to have the right labels. The dataset did not contain images that were blurred, of low-resolution, or not clear to make a diagnosis.

The last set of data would be a varied and clinically approved set of chest X-ray photographs, which would be used as a basis in training and evaluating the proposed deep-learning models to detect pneumonia.

All identifiers of the patients were eliminated to preserve the level of confidentiality, and the ultimate dataset of 1,500 anonymized pictures was ready to be used in the experiment. The dataset maintained a 60/40 distribution that was fit/unfit on the dataset, respectively. After this was done, the data was sorted in a logical order and saved safely in Google drive as a form of secondary processing and analysis. Figure 3.1 illustrates some of the photos I have:

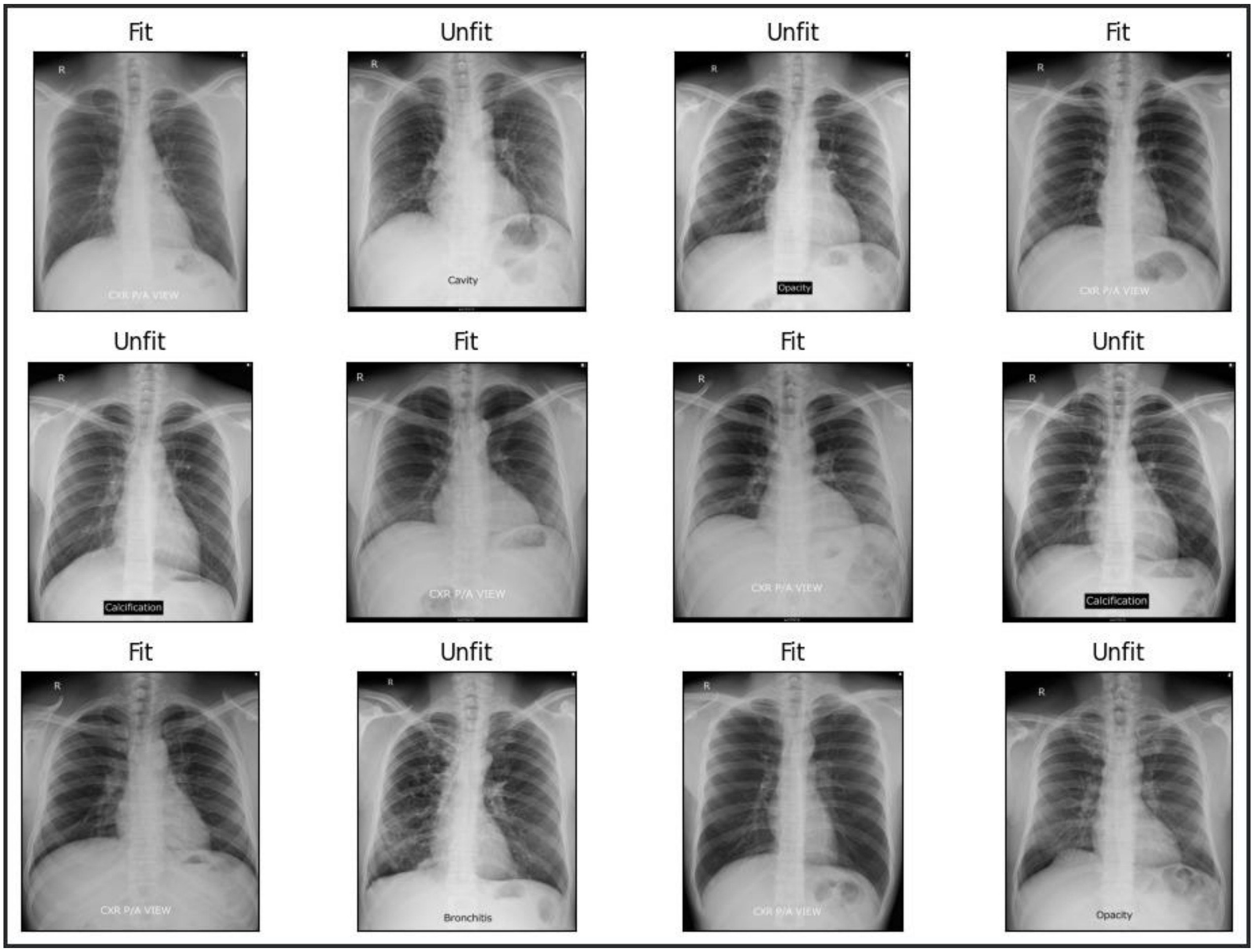

Figure 3.1: Random view of Raw Dataset Images

3.3 Proposed Methodology

The evolution was in a linear and testable research process typical of machine learning projects. This process commenced with the data gathered, proceeded to the preprocessing stage, model training, and finally, the evaluation to generate the final system of detection. The flow chart in Figure 3.3 below could demonstrate the procedure.

The data were collected in a local hospital and therefore were chosen based on this factor.

Methodology Flow chart:

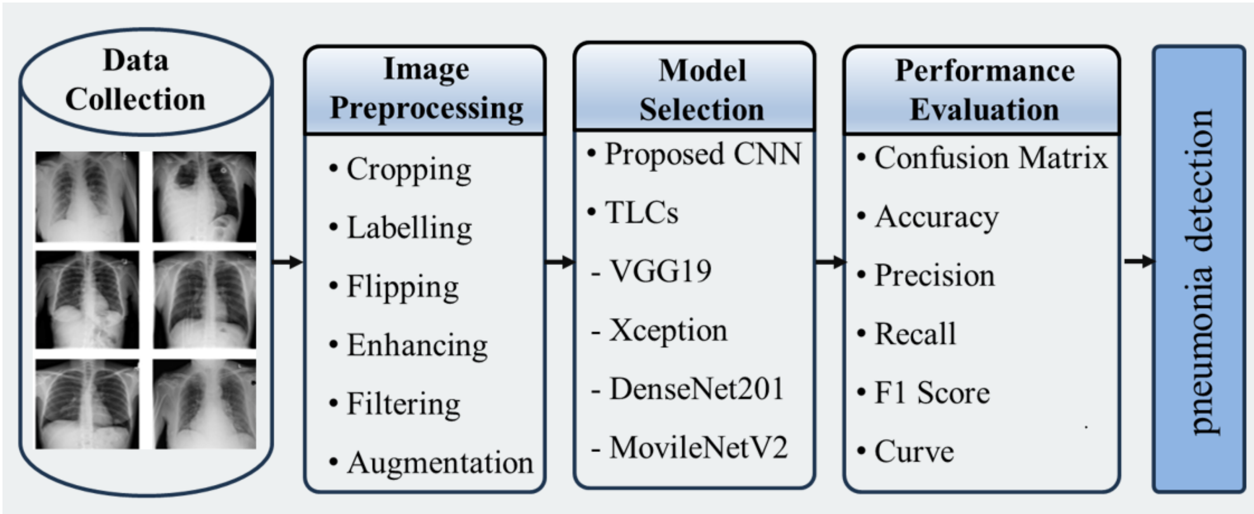

Figure 3.2: Methodology Flowchart

3.3.1 Data Selection (Collected From Local Hospital):

The selection of data was very meticulously done to ensure only meaningful and high-quality images of the chest X-ray were entered into this study. Following the gathering of data at Turkey Medical services, every image has undergone a systematic screening with respect to certain relevant inclusion and exclusion criteria. Data selection was intended to eliminate irrelevant, low-quality or ambiguous samples which may have an adverse effect on model performance and reliability.

Inclusion Criteria

The inclusion criterion was that the chest X-ray images satisfied the following requirements:

- The picture was clearly in the form of a frontal (AP/PA) chest X-ray image.
- The diagnostic label (pneumonia or normal) was checked by a certified radiologist.
- The resolution of the image was high to be interpreted visually.
- The photo was of an adult or pediatric patient and there were no limitations of age or gender.

Exclusion Criteria:

Images were filtered out in case they had any of the following features:

- Blurry and large artifacts, poor contrast confuse the model.
- Non-radiographic images such as CT scans were added in lateral-view chest X-rays or incorrectly placed.
- Obvious clinical terms and ambiguous diagnoses.
- Redundant pictures or other similar views of a patient.
- Photographs depicting other illnesses or deformities that are not related to pneumonia. i.e. fracture or tumors, unless authorized by clinicians.

Final Dataset Selection:

Images that were retained after the screening had to be of good quality and diagnosis. Approximately 1,500 X-ray images were used in the last dataset with the proportion of 60:40 between cases of pneumonia and those of normal cases. The selection was done carefully to ensure that the dataset was diverse, balanced, and able to be trained using deep-learning models with high diagnostic accuracy.

3.3.2 Data Preprocessing:

All of the chest X-ray (CXR) images were subjected to a preprocessing pipeline. This facilitated regularity, eliminated non-diagnostic artifacts, maintained lung structure and enhanced model strength. The pipeline consisted of image standardization, lung preserving cropping, resizing, and valid data augmentation.

Image Cleaning and Image Standardization.

Each of the collection chest X-ray images has initially been standardized by uniform pre-processing protocol.

In every file, it was verified that it had a clear radiographic image. Automatic deletion of non-image, or corrupted files, occurred automatically.

Photos were read in RGB color so that they maintained compatibility with deep-learning models, whereas processing operations inside it relied on grayscale transformations where needed in contour detection. It was standardized so that the input was always of the same quality in the following stages.

Bottom Crop that Preserves the Lunges: A lot of raw CXR images contained text overlays, hospital marks, scanner labels and border artifacts towards the bottom of the picture. It is essential to remove these areas but when cropping is done carelessly, one may end up removing the lung bases, and this is vital in diagnosing pneumonia.

To fix this, we used a lung-aware safe cropping algorithm.

Lung Region Detection:

- We estimated the lung boundaries using traditional image analysis methods:
- We applied Gaussian blurring to reduce high-frequency noise.
- Next, we used Otsu's adaptive thresholding to separate lung fields from non-lung areas.
- Morphological opening with an elliptical kernel removed small artifacts.
- We extracted contours and assumed that the largest connected component represented the lung fields.
- The lower boundary of the lung was set based on the bounding box of this component.

Safe Bottom Cropping Rule To keep necessary lung tissue:

- We set a maximum cropping limit of 2% of the image height.
- Cropping was restricted to areas below the detected lung boundary.
- We added safety margin of 8 pixel to avoid accidental cutting lung bases.

- If cropping posed a risk of removing lung tissue then the algorithm skipped that cropping.
- This approach ensured the complete preservation of lung regions. By removing unhelpful image borders.

Image Resizing: All images were cropped and all artifacts were eliminated, and each image was resized to 224 x 224 pixels. This was equal to the input requirement of both the custom CNN and transfer learning models which comprised of the DenseNet201, MobileNetV2, VGG19, Xception, and ResNet50. Resizing maintained the aspect ratios and also enabled efficient training of GPUs.

Medically Safe Data Augmentation: In an attempt to enhance the generalization of the models and overcome the limitation of the data sets, we used controlled augmentation. Any changes made were mellow to prevent production of unrealistic CXRs of anatomically unrealistic nature.

- The augmentation techniques used were as follows:
- Geometric Augmentations
- Rotation 10 degree to simulate the changes in acquisition angles.
- Horizontal 5% and vertical shifting, 5%.
- Zoom scaling (+-5%)
- Horizontal flipping that is permissible in interpreting CXR.
- To achieve anatomical realism we did not do vertical flips, extreme zooms and heavy distortions.
- Intensity Augmentations
- Brightness adjustments
- Contrast adjustments
- These variations are used to mimic varying exposure levels in X-ray equipment that is used in clinical environments.

In general, the augmentation pipeline augmented the data without damaging the integrity of radiological characteristics of pneumonia.

Final Dataset Preparation:

All preprocessed and improved images were saved in organized folders at a consistent resolution. The final dataset contained: Artifact-reduced images, Lung-preserved crops, Uniform 224×224 dimensions,Clinically valid augmented samples. A total of 2124 images were built. 1100 fit, 1024 unfit which leads to a balanced dataset.

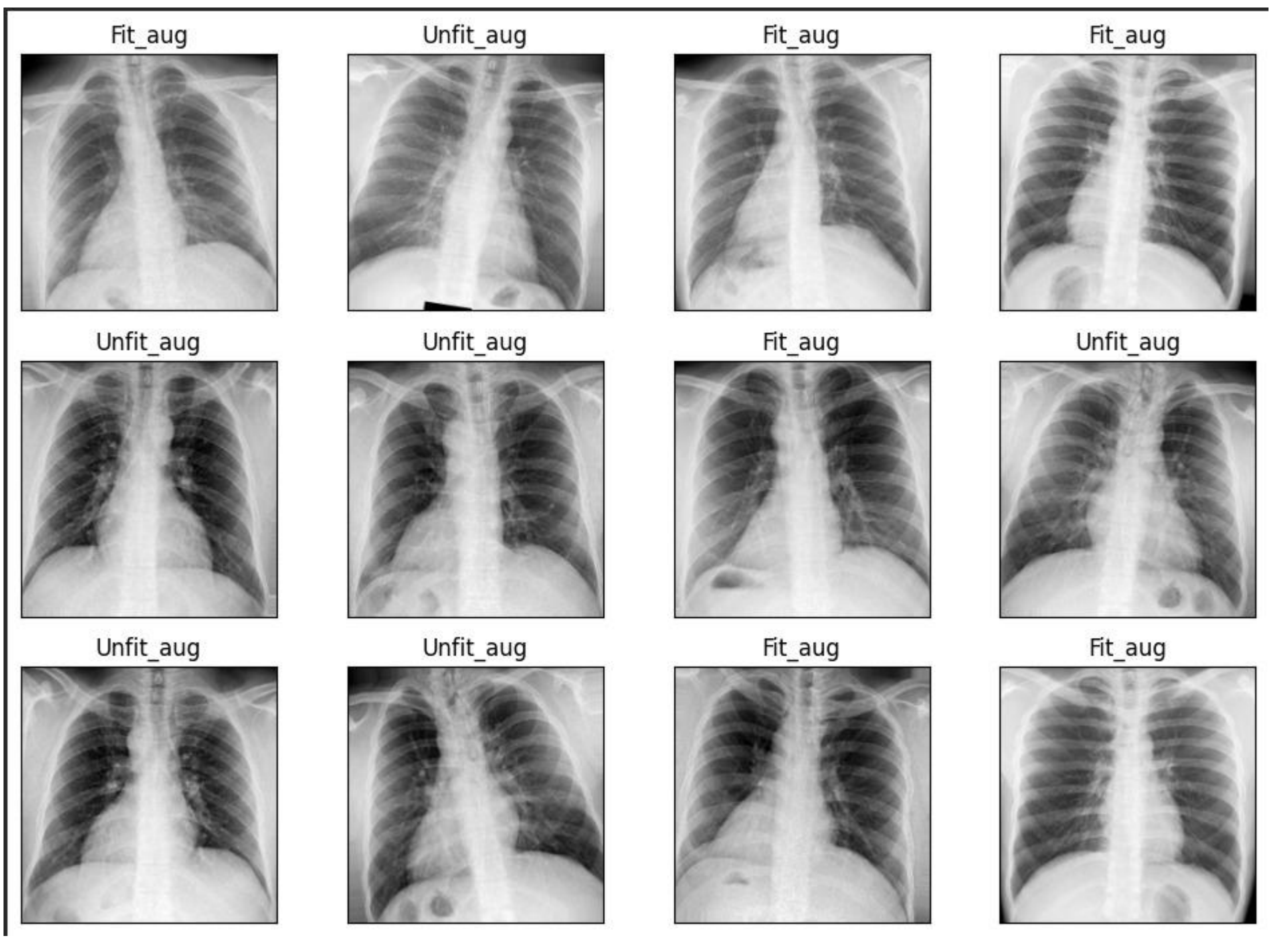

Figure 3.3: Dataset Images after Preprocessing and Augmentation

3.3.3 Model Selection:

Accurately identify lung diseases by employing innovative architectures to train deep learning models. Choose suitable deep learning architectures, such as "CNN," "DenseNet201," "ResNet50," "InceptionV3," and "VGG19."Utilize fine-tuning and transfer learning to achieve effective training on the dataset of skin diseases. Model architecture should be modified to meet the unique needs of lung disease detection.

Conventional Neural Network (CNNs):

A family of deep learning models called convolutional neural networks (CNNs) is very good at capturing patterns of space in data because of its convolutional layers. CNNs are mostly used for image processing jobs. CNNs make sense for my research paper in this scenario. When combined with hierarchical feature extraction, their ability to automatically extract pertinent characteristics from medical photos improves their ability to identify subtle patterns that may be symptomatic of lung disorders. CNNs can be used to detect and classify lung abnormalities, which can lead to enhanced diagnosis accuracy and potentially revolutionize dermatological diagnostics through advanced deep learning approaches. This is because CNNs are naturally adaptable to varied picture properties.

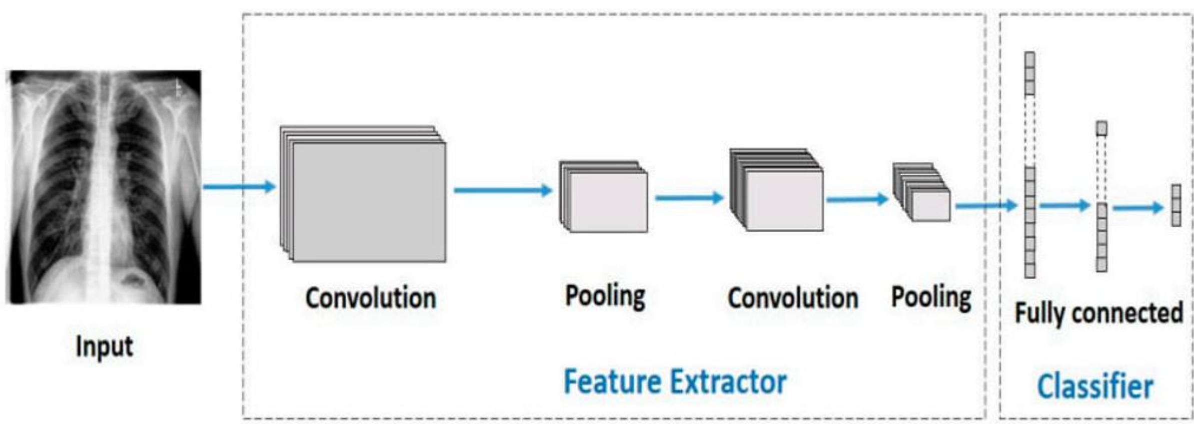

Figure 3.4: Proposed CNN Architecture

MobileNetV2

MobileNetV2 is a lightweight and efficient CNN model, which aims at rapid inference and low cost of computation. It incorporates inverted residual and linear bottlenecks to cut parameters but the feature extraction is powerful. ImageNet-pretrained weights were adopted in this research, and the last layers were adapted to binary classification. MobileNet V2 was effective in the Chest X-ray image analysis with high accuracy and with minimum resources.

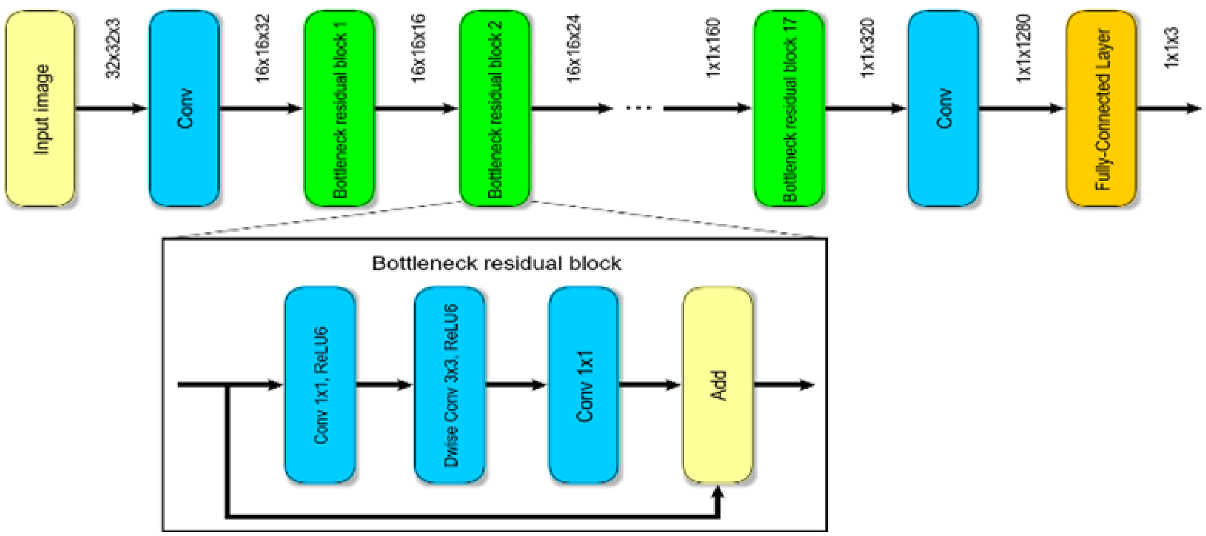

Figure 3.5: Proposed MobileNetV2Architecture

DenseNet201

DenseNet201 is a deep convolutional neural network architecture that was designed with an application of image classification. It is characterised by its closely interconnected blocks, which enable each layer to have its input straight away by all those layers preceding it that would ultimately encourage feature reuse and enhance the efficiency of the model. The ability of DenseNet201 to detect complex patterns in medical images is a good choice of the research study on skin disease detection. Dense connections that enhance effective transmission of features allow the model to learn complex interactions within lung lesions, and this is critical to accurate detection of diseases. DenseNet201 can be used as a viable choice to achieve a high accuracy in skin diseases classification which will ultimately lead to the development of deep learning processes in the medical picture analysis. This is because of its demonstrated effectiveness in numerous computer vision applications, as well as richness and attachment.

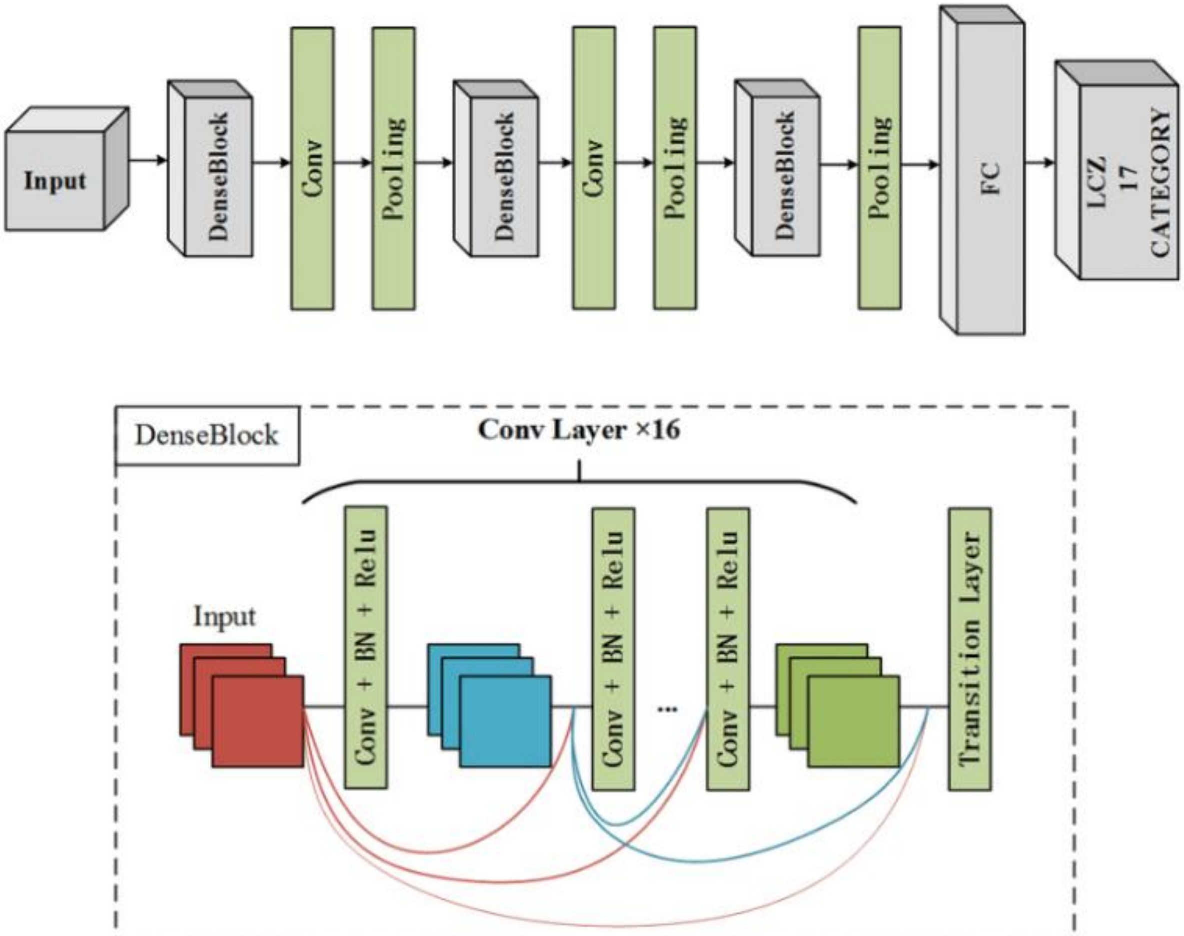

Figure 3.6: Proposed DenseNet201 Architecture

ResNet50

ResNet50 is a 50-layer convolutional neural network architecture that is famous in its skip connection and deep design, making it useful in addressing the deep learning issue of the disappearing gradient. ResNet50 has been selected because of its outstanding performance in picture classification task; it is very successful at detecting fine details in big and complex datasets. The skip and depth connections of this model help learn the structures that are critical in detection of subtle patterns that are indicative of lung disorders in an efficient manner, something that is pertinent to your research article on the detection of lung diseases. I would suggest your approach to be stronger and more stable because you utilize the ResNet50 that enhances the capability of the model to recognize and classify a range of skin diseases with a high level of reliability.

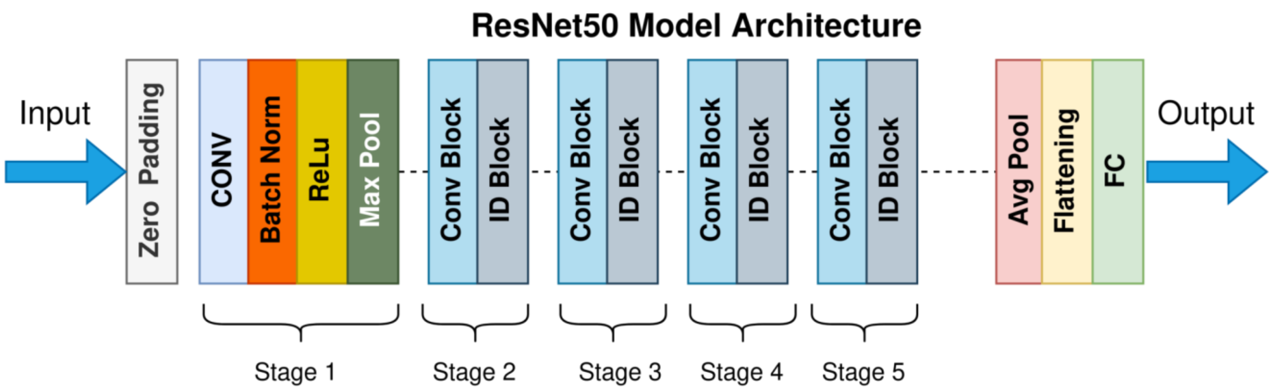

Figure 3.7: ResNet50 Architecture

VGG19

VGG19 is a visual geometry group (VGG) of the University of Oxford that developed the deep convolutional neural network architecture that has been noted to be simple and deep. It has 19 weight layers which consist of three fully connected layers and sixteen convolutional ones. It is a consistent method to classify images because it has a consistent structure and minute convolutional filters. VGG19 has a high ability to detect intricate forms and shapes in medical photos; this feature was the reason why I chose it in my research paper. VGG19 is a reliable model, and I can apply it to enhance the accuracy and reliability of lung disease detection in my study due to its proven effectiveness in picture classification and the ability to work with various types of data.

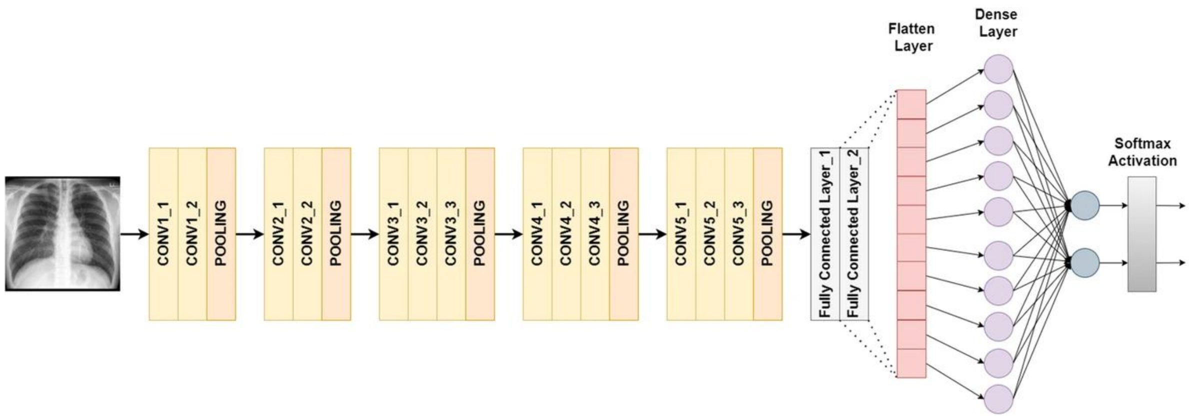

Figure 3.8: Proposed VGG19 Architecture

3.3.4 Performance Evaluation Metrics:

In order to evaluate the performance of the different departments in the company, certain metrics need to be established and they include:

Validation of classification effectiveness was done by computing Accuracy, Precision, Recall and F1-Score. In this case, we determine the efficacy of the job on the basis of accuracy, precision, recall, and the F1-Score. Such a paradigm, which is based on four factors, is rather helpful in the analysis of predictive data.

Precision is pegged on the ability to identify and classify situations precisely. The formula of accuracy is illustrated in equation 7.

Accuracy=(TP+TN)/(TP+FP+TN+FN) ……………………………………7

Accuracy in studies is the percentage of the true positive events to the total positive events. The mathematical expression on accuracy is indicated in Equation 8.

Pre Precision=TP/(TP+FP) ………………………………………………8

One of the factors is called recall and it determines how the algorithm detects such species among various flowers. Recall mathematically can be described as: Equation 9:

Recall=TP/(TP+ FN) …………………………………………………..9

Such an approach is called harmonic mean because it is aimed at finding the balance between precision and memory. An approximation of the mathematical equation of the F 1- Score is provided as Equation 10:

$F\_1\text{-Score} = 2((\text{Precision}\times\text{Recall})/(\text{Precision}+\text{Recall}))$ ……………………………………10

3.4 Implementation Requirements

The advanced deep learning method of lung diagnosis requires a series of conditions that should be met to be implemented successfully. The best thing of it all is to have access to a well-chosen sample of lung diseases, preferably collected through some good source such as colab. Training of an efficient model demands adequate computer resources, such as a high-end GPU, especially on complex designs, like InceptionV3, DenseNet201, ResNet50, VGG19, and CNN. Moreover, the implementation will presuppose the presence of the required libraries to augment pictures and preprocess data and, in addition, the deep learning systems such as TensorFlow or PyTorch. In order to enhance the precision of the model, an individual should be knowledgeable of the principles of deep learning and be skilled with the process of model evaluation. Lastly, the ethical concern, adherence to the privacy laws, and cooperation with medical experts facilitate the responsible implementation of the lung disease detection system.

Chapter 4
Results and Analysis

4.1 Experimental Setup

Every experiment of this paper was run through Google Colab, a cloud computing platform which is free and provides access to GPU acceleration. Google drive was used to store the project files, datasets and the trained model weights. This was also directly assimilated to the Colab notebook to load and save easily.

Hardware Environment

The hardware setup that was used in Google Colab was:

- GPU: NVIDIA Tesla T4 or P100 (Automatic allocation based on the availability)
- GPU Memory: ~16 GB
- RAM: 12–15 GB
- CPU: Single-core / dual-core virtual CPU (Colab default)

This environment with the ability to utilize the GPU greatly accelerated the time taken to train the convolutional neural network (CNN) model and was also able to produce Grad-CAM visualizations with efficiency.

Software Environment

Every implementation was done with the help of:

- Python 3.x
- TensorFlow / Keras
- NumPy
- Matplotlib
- Scikit-learn

- OpenCV
- Google Colab cloud runtime
- Google Drive storage integration

The interactive tool of Google Colab allowed the step by step execution, visualization of the training curves, model validation as well as explaining it through Grad-CAM.

Data Storage and Access

All the data utilized in the training, validation, and testing of the model was located in Google Drive. These files were read by the notebook using the in-built Drive mount functionality of Colab. This ensured:

- Constant use of files with big files.
- Auto-saving of model checkpoint.
- Easy access in evaluation and visualization.

4.2 Experimental Results & Analysis

The comparison of experiment with the six deep learning models revealed that there was a significant variance in their performance. The VGG19 gave the best accuracy of 97.88% and this is the benchmark of the task. This was intimately succeeded by a cluster comprising of effective models, consisting of DenseNet201 at 96.47, MobileNetV2 at 96.24% and the custom CNN at 96.00%. All three performed well above the 96% mark. In contrast, the ResNet50 model performed poorly, with the lowest accuracy of only 82.71%. This outcome suggests there may be problems with its training or that it is not suitable for this specific dataset. Meanwhile, Xception provided a decent accuracy of 94.71%, though it was slightly lower. Overall, the results show that while VGG19 offers the highest accuracy, models like MobileNetV2 provide a strong option by delivering nearly peak performance with better efficiency.

In given below I am showing Comparative Model Accuracy Bar Plot:

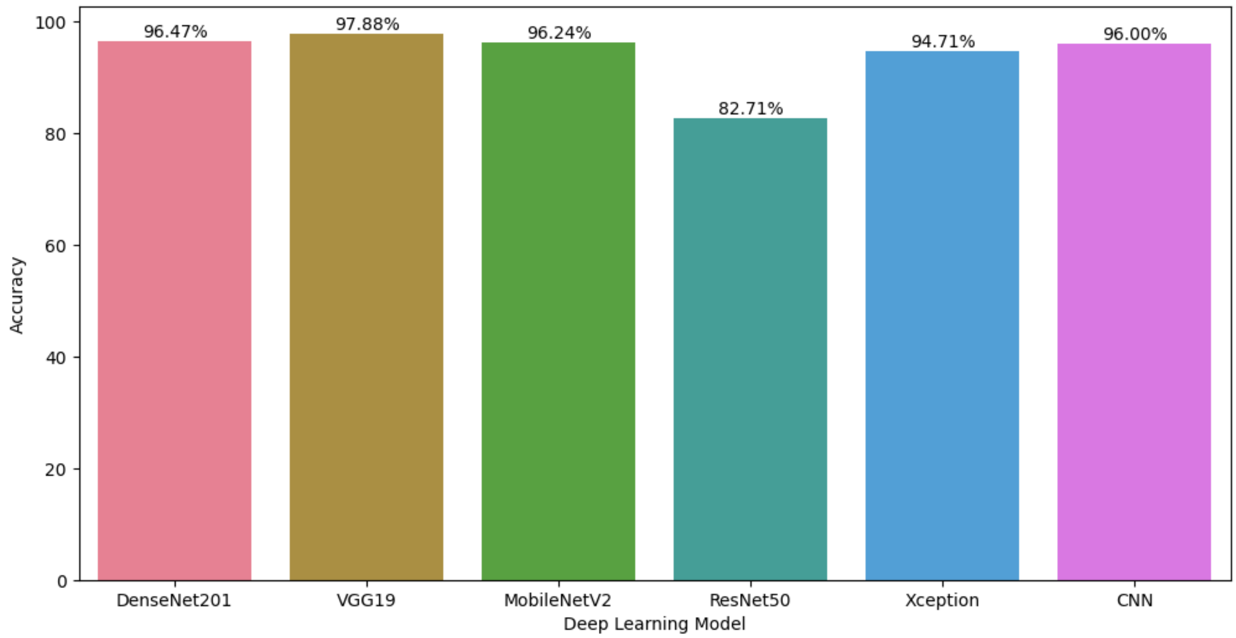

Figure 4.1: Comparative Model Accuracy Bar Plot

Figure 4.1 This bar chart compares the classification accuracy of six different Deep Learning Models (Transfer Learning models and a standard CNN). The models generally perform very well, with most achieving an accuracy above 94% The VGG19 model is the best-performing model with the highest accuracy of 97.88\%. It is considered the best because it achieved the greatest percentage of correct predictions on the test data compared to all other models evaluated, with MobileNetV2 and the standard CNN following closely behind at 96.24\% and 96.00\% respectively.

In given below we are describing the result analysis part also show the training accuracy and loss rate and confusion matrix also:

Result of Experiment : CNN

In the initial state, the most important standard evaluation measures such as precision, recall, and F1-score were computed to gain a better idea about the work of the model. The model had a macro-averaged and weighted-averaged precision, recall and of F1-score of 0.96% that shows that the two classes performed equally.

In the classification report, it is also indicated that performance of the model is almost similar in the two classes. Fitaug class had precision of 0.97%, recall of 0.95%, and F1-score of 0.96% and unfitaug class had the following: precision of 0.95%, recall of 0.97% and F1-score of 0.96%. These outcomes leave the impression that the CNN had a good ability to distinguish between fit and unfit CXR images and there was no bias on the classes.

Summary of Key Findings:

- Test Accuracy: 0.9600
- Precision (Avg): 0.96018
- Recall (Avg): 0.96
- F1-Score (Avg): 0.96001
- Total Test Samples: 850

Table 4.1 Classification Report of CNN

Class	Precision	Recall	F1-Score	Support
fit_aug	0.97	0.95	0.96	435
unfit_aug	0.95	0.97	0.96	415
Accuracy			0.96	850
Macro Avg	0.96	0.96	0.96	850
Weighted Avg	0.96	0.96	0.96	850

These results confirm that the developed CNN architecture is effective and reliable for distinguishing between fit and unfit CXR images. The high F1-score and balanced class performance demonstrate the robustness of the model, making it suitable for practical deployment in automated screening systems.

Achieving Test Accuracy of CNN01 is 96.00%. In below Figure 4:1 describing the confusion matrix of CNN:

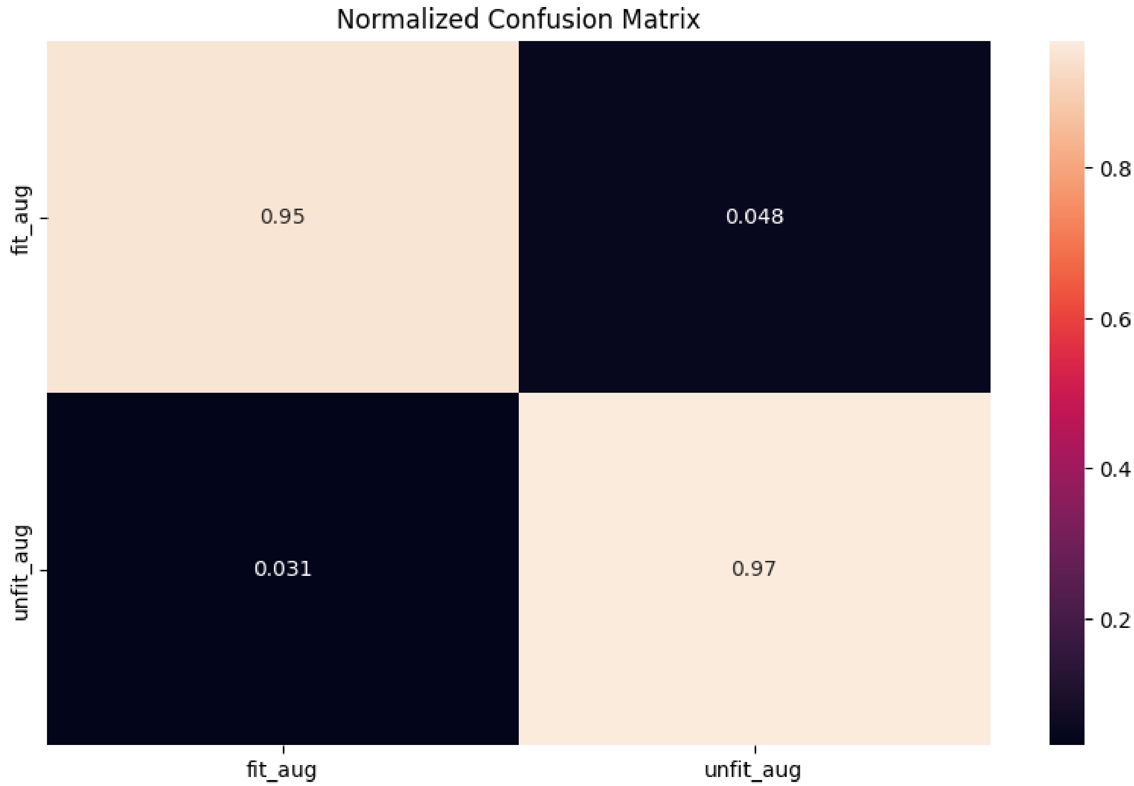

Figure 4.2: Confusion Matrix (CNN01)

Figure 4.2 shows the Normalized Confusion Matrix that summarizes the performance of a binary classification CNN. The high values on the main diagonal—0.95 for the fit_aug class and 0.97 for the unfit_aug class—indicate excellent and balanced predictive accuracy, as the model correctly classifies over 95\% of the samples for both categories. The low off-diagonal values (0.048 and 0.031) represent minimal misclassification errors, confirming the CNN is a highly effective and robust classifier.

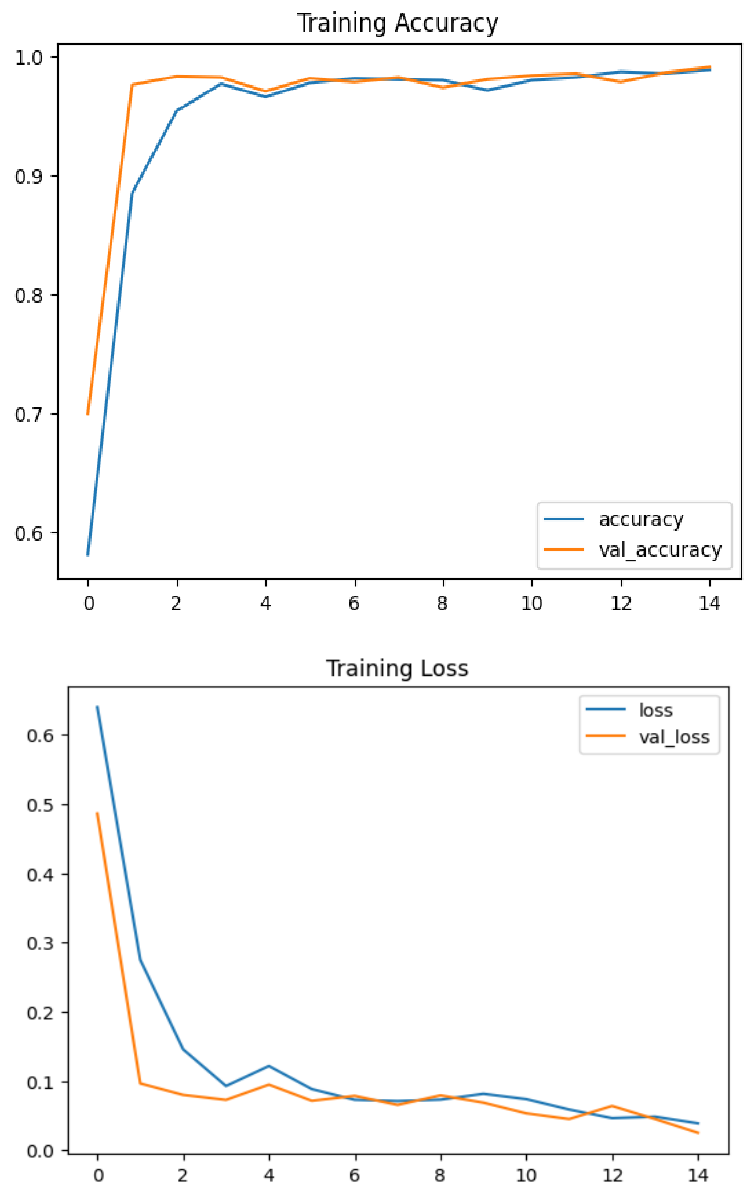

Figure 4.3: Training and validation accuracy and loss over the epochs (CNN)

Figure 4.3 illustrates the performance metrics of your CNN model over 15 training epochs. The Training Loss plot shows that both the training loss (loss, blue line) and validation loss (val_loss, orange line) decrease rapidly and stabilize around 0.05 by epoch 5, indicating effective learning. The Training Accuracy plot shows a corresponding sharp increase in both training (accuracy, blue line) and validation (val_accuracy, orange line) to nearly 1.0 (or 100\%) by epoch 5, where they both remain very high and stable. Since the loss and val_loss lines closely track each other and the accuracy and val_accuracy lines also track closely at a high level, the model shows no significant signs of overfitting and has achieved excellent, stable performance on both the training and unseen validation data.

VGG19

The transfer learning model of VGG19 was tested using the test set of 850 chest X-ray images with the classification of fitaug and unfitaug. The model was found to be highly generalized with a test accuracy of 97.88% which is better compared to the baseline CNN model.

The general analysis metrics demonstrate that VGG19 is doing well with an approximate precision and recall of 0.98 and F1-score of 0.98, which means that it is acting consistently and reliably in its classification behavior on both classes.

According to the classification report, the model is symmetrically good on both categories. The fitaug class had 1.00 precision and 0.96 recall and F1-score, whereas the unfitaug class had 0.96 precision, 1.00 recall, and 0.98 F1-score. This balance points out the stability of the model and low misclassification rate.

```
Accuracy on the test set: 97.88%
27/27 ━━━━━━━━━━━━━━━━━━━━ 10s 296ms/step
Accuracy: 0.97882
Precision: 0.9797
Recall: 0.97882
F1 Score: 0.97882
              precision    recall  f1-score   support

     fit_aug       1.00      0.96      0.98       435
   unfit_aug       0.96      1.00      0.98       415

    accuracy                           0.98       850
   macro avg       0.98      0.98      0.98       850
weighted avg       0.98      0.98      0.98       850
```

Summary of Key Findings:

- Test Accuracy: 97.882%
- Precision (Avg): 97.97%
- Recall (Avg): 97.882%
- F1-Score (Avg): 97.882%
- Total Test Samples: 850

Table 4.2 Classification Report of VGG19

Class	Precision	Recall	F1-Score	Support
fit_aug	1.00	0.96	0.98	435
unfit_aug	0.96	1.00	0.98	415
Accuracy			0.98	850
Macro Avg	0.98	0.98	0.98	850
Weighted Avg	0.98	0.98	0.98	850

The findings show that VGG19 outperforms a standard CNN model in classification, having better accuracy and balanced classes. Having a profound feature extraction capacity, VGG19 is quite useful in distinguishing between fit and unfit chest X-ray images.

This exceptional performance suggests that VGG19 is a strong candidate for deployment in real-world automated CXR screening systems where reliability and precision are critical. Achieving Test Accuracy of VGG19 is 97.88%. In below Figure 4:1 describing the confusion matrix of VGG19:

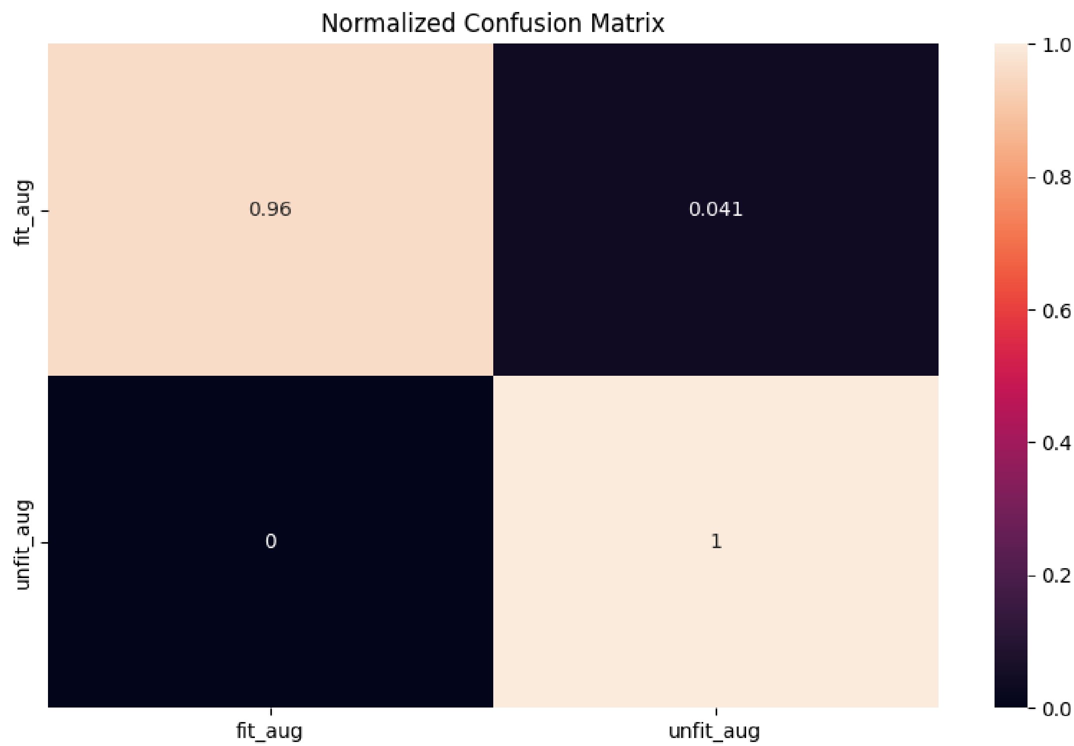

Figure 4.4: Confusion Matrix (VGG19)

Figure 4.4 presents the classification performance for the classes fit_aug and unfit_aug. The model exhibits near-perfect performance overall. Specifically, the model has an exceptional 100\% Recall for the unfit_aug class, meaning every true unfit_aug sample was correctly identified (0 False Negatives). For the fit_aug class, the Recall is 96% (0.96), with only 4.1% of true fit_aug samples being misclassified (0.041 False Negatives). The values on the main diagonal (0.96 & 1.0) confirm a well-balanced accuracy, with model being stronger at identifying the unfit_aug class.

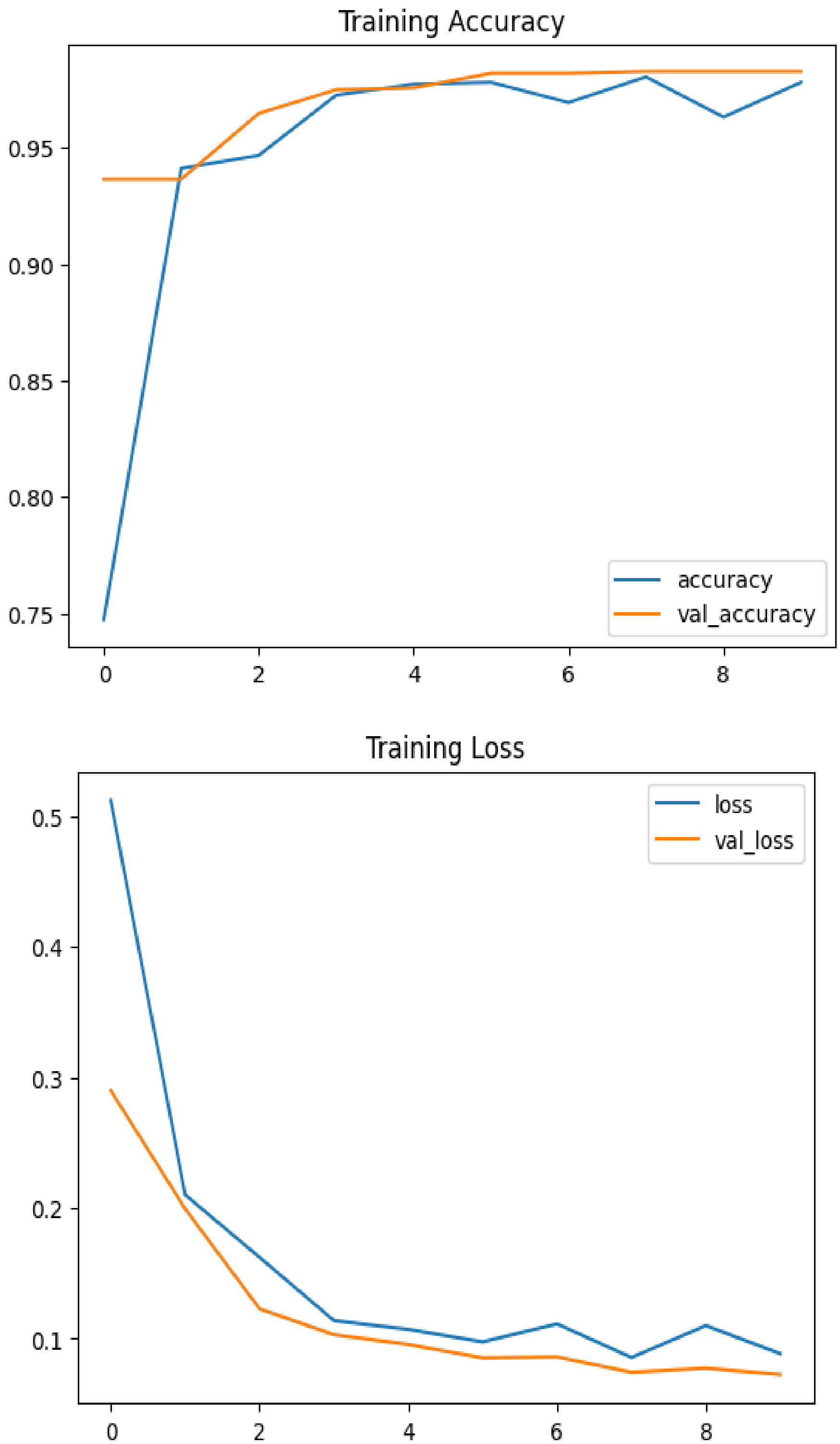

Figure 4.5: Training and validation accuracy and loss over the epochs (VGG 19)

Figure 4.5 training plots show the model's performance over 10 epochs (from 0 to 9). The Training Accuracy plot reveals a rapid increase in both training accuracy (blue line) and validation accuracy (orange line) during the initial epochs, with both converging and stabilizing at a high level, mostly above 95% towards the end of training. Concurrently, the Training Loss plot demonstrates a sharp decrease in both training loss and validation loss, with both curves flattening out to very low values, suggesting the model is learning effectively and efficiently. Since the training and validation curves in both plots closely track each other and do not show a significant divergence (the validation loss stays mostly below the training loss after epoch 1), the model appears to be well-fitted without evidence of major overfitting during this training run.

Chapter 5

Discussion

This chapter presents the results of the study, the interpretation of the performance of the developed models, and brings out their implications, limitations and possible future work directions. The main purpose of the study was to categorize the chest X-ray (CXR) images as fit and unfit with the help of deep learning architectures. A series of models were tested such as a bespoke Convolutional Neural Network (CNN) and a variety of transfer learning methods (VGG19, DenseNet201, InceptionResNetV2, etc.), to find out which one works most effectively on the automated screening of CXR.

5.1 Discussion of Findings

The experimental assessment proved that all the models were effective in the classification of CXR images, but there were notable differences in the performance across the architectures. The initial CNN architecture was quite successful with a high accuracy rate of 96 and shows that an architecture as simple as it is, can learn to discriminate against the dataset. VGG19 model also showed better results, with 97.88 percent of accuracy, as well as high precision, recall, and F1-scores in both classes.

Such advances can be explained by a more significant architecture of the VGG19 and its possibilities to extract fine-grained features to make the division of classes more accurate. Other pretrained models that were also tested in the study were highly performing since they are able to extract wide features and have been previously exposed to large scale visual datasets e.g. ImageNet.

The equally balanced values of precision and recall in all models show that all the models did not have high class bias. This is important in medical screening systems where false negatives could be very dangerous to happen.

5.2 Comparison Between Models

The deep learning models can be compared to each other, and it is possible to notice the evident differences in the performance, the ability to extract the features, and the level of the generalization. A CNN architecture with a baseline accuracy was quite impressive, which confirms that a fairly simple architecture can gain access to meaningful radiological information based on the dataset. Nonetheless, it continuously outperformed the transfer learning models which had the advantage of more detailed architectures and pretrained weights based on large scale image datasets. Of these pretrained models, VGG19 performed best with a higher overall performance with higher accuracy, precision, recall and F1-scores than the CNN. Models like the DenseNet201 and Inception-based models were also demonstrating good performance in classification, which evidences that deeper and more complex networks are effective in medical imaging tasks. This better performance of transfer learning methods is owed to the fact that they are better able to extract fine-grained patterns, textures, and structural details in the images of the chest X-rays- aspects that their lower network counterparts may not do. In general, the comparison shows that custom CNNs offer a good baseline, whereas transfer learning architectures are much more efficient and dependable when it comes to automated classification of CXR, especially when dealing with rather large datasets.

5.3 Limitations of the Study

Although the study was fruitful, it has a number of limitations that should be taken into account when concluding on the findings of the study. The provided dataset, though augmented and curated, is rather small in comparison with the large-scale datasets, which are regularly employed during medical imaging research, which can restrict the ability of the model to generalize when dealing with larger groups of people. Moreover, the research only deals with binary classification, i.e. between fit and non-fit cases, without exploring further more specific or complicated medical conditions that may be indicated by the chest X-rays. The use of ImageNet-based pretrained models is also a limitation because the models have not been trained on medical data and instead have been trained on natural images, and may not optimally reflect all the relevant radiographic features. Also, real world hospital data usually have variability in image quality, noise, labeling accuracy, and these were not wholly reflected in the controlled experimental data. These restrictions indicate that the models work effectively in the experimental setting, but the models should be validated in clinical settings before they can be implemented.

5.4 Future Work Recommendations

Further studies should be aimed at increasing the number of cases and the variety of cases to increase the generalization potential and stability of the models in relation to various patient groups and imaging histories. Gathering information across a variety of hospitals and imaging units would contribute to decreasing the model bias and enhance the results in the conditions of real-world variability. Also, it would be more beneficial to expand the classification task further than the binary fit/unfit framework by implementing a multi-class detection of particular abnormalities, i.e., pneumonia, lung density, or cardiomegaly, which would contribute greatly to the clinical relevance of the system. The use of explainability methods, including Grad-CAM or attention-based visualization, should also be included in the further work to enhance the clarity of the model and better understand the decision-making process to make the system more reliable in the eyes of healthcare specialists. Further development of higher architectures such as hybrid CNN-transformer models may also be a way of improving feature extraction. Lastly, the practical feasibility should be evaluated by engaging in practical testing in the future, which would be done in clinical working environments regarding usability, deployment difficulties, and integration with the current medical systems.

Chapter 6
Conclusion

6.1 Conclusion

This paper examined the use of deep learning methods to classify chest X-ray images into fit and unfit. based on the performance of both a tailor-crafted Convolutional Neural Network (CNN) and a number of transfer learning models. The study was meant to come up with a precise, effective and automated solution that can assist medical workers in initial screening processes. The results of the research indicate that deep learning specifically transfer learning is a very promising solution to CXR classification due to their high effectiveness after a lot of experimentation and analysis.

This was verified as the baseline CNN model performed very well with an accuracy of 96% on the test dataset. Nevertheless, the transfer learning architectures, particularly, VGG19, obtained even a superior level of accuracy of 97.88% and equal precision, recall, and F1-score. The advantages of these models were that they had better feature extraction ability and trained weights which allowed them to generalize better when compared to the existing dataset. The paper has also demonstrated that the architectures designed with high quality can also do a great job of minimizing the rate of misclassification as well as give uniform results between the two classes, which means that such architectures may be useful in medical screening applications.

Despite these encouraging results, the study itself admits a number of limitations such as the limited size of the dataset, the use of ImageNet-based pretrained models and binary classification. These shortcomings open prospects of future research, including the increase of the dataset, the incorporation of the multi-class detection of the abnormalities, examination of the advanced hybrid structures, and the real-world clinical validation to guarantee the strong performance in various contexts of its functioning.

To sum up, the study has been able to establish that deep learning models, especially those based on transfer learning could be a potent automated CXR screening tool. Their accuracy and reliability are high which can justify them as clinical decision-support systems and help radiologists and decrease workload on diagnosing the cases. Additional optimization, more extensive datasets, and research on the practical deployment of the proposed approach can make a positive contribution to the optimization of medical imaging processes and the greater efficiency of medical services.

References

[1]S. Sharma and Kalpna Guleria, "A Systematic Literature Review on Deep Learning Approaches for Pneumonia Detection Using Chest X-ray Images," Multimedia Tools and Applications, Aug. 2023, doi: https://doi.org/10.1007/s11042-023-16419-1.

[2]R. Siddiqi and S. Javaid, "Deep Learning for Pneumonia Detection in Chest X-ray Images: a Comprehensive Survey," Journal of Imaging, vol. 10, no. 8, pp. 176–176, Jul. 2024, doi: https://doi.org/10.3390/jimaging10080176.

[3]Shadi Aljawarneh and Romesaa Al-Quraan, "Pneumonia Detection Using Enhanced Convolutional Neural Network Model on Chest X-Ray Images," Big data, Apr. 2023, doi: https://doi.org/10.1089/big.2022.0261.

[4]M. K. U. Ahamed et al., "DTLCx: An Improved ResNet Architecture to Classify Normal and Conventional Pneumonia Cases from COVID-19 Instances with Grad-CAM-Based Superimposed Visualization Utilizing Chest X-ray Images," Diagnostics, vol. 13, no. 3, p. 551, Jan. 2023, doi: https://doi.org/10.3390/diagnostics13030551.

[5]D. Surendiran Muthukumar, Bommineni Sreekar Chowdary, D. S. Kumar, P. Jagadeesan, B. Mythili, and B. Swapna, "CNN-Based Pneumonia Detection and Severity Analysis in Chest X-Ray Images for Enhanced Diagnostic Support," pp. 1443–1449, Jan. 2025, doi: https://doi.org/10.1109/icmsci62561.2025.10894276.

[6]A. Rattan, None Ritanshu, S. Singh, and Jyoti Pruthi, "Comparative Analysis of Two Approaches Utilizing Different Classification Algorithms Combined with CNN for Pneumonia Prediction in Chest X-Rays," pp. 475–479, Jan. 2024.

[7]D. Zhang, F. Ren, Y. Li, L. Na, and Y. Ma, "Pneumonia Detection from Chest X-ray Images Based on Convolutional Neural Network," Electronics, vol. 10, no. 13, p. 1512, Jun. 2021, doi: https://doi.org/10.3390/electronics10131512.

[8]A. Manickam, J. Jiang, Y. Zhou, A. Sagar, R. Soundrapandiyan, and R. Dinesh Jackson Samuel, "Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures," Measurement, vol. 184, p. 109953, Nov. 2021, doi: https://doi.org/10.1016/j.measurement.2021.109953.

[9]S. Sharma and K. Guleria, "A Deep Learning Based Model for the Detection of Pneumonia from Chest X-Ray Images Using VGG-16 and Neural Networks," Procedia Computer Science, vol. 218, pp. 357–366, 2023, doi: https://doi.org/10.1016/j.procs.2023.01.018.

[10]M. Saleh et al., "Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model," Healthcare, vol. 11, no. 11, pp. 1561–1561, May 2023, doi: https://doi.org/10.3390/healthcare11111561.

[11]W. Khan, N. Zaki, and L. Ali, "Intelligent Pneumonia Identification from Chest X-Rays: a Systematic Literature Review," IEEE Access, vol. 9, pp. 51747–51771, 2021, doi: https://doi.org/10.1109/access.2021.3069937.

APPENDICES

Appendix A: Dataset Details

The following appendix contains the information regarding the dataset in this research.

The X-ray images of the chest were gathered in a local diagnostic centre in Bangladesh and with the aid of the publicly available sources. Images were collected and anonymized ethically so that patient confidentiality was achieved.

First of all, a total of 1,500 chest X-ray images were collected. Out of them, nearly 60% reflected normal lung conditions, and 40 percent reflected the lung with pneumonia-related cases. The certified radiologists went through every image and confirmed it to make sure that everything was labeled.

The size of the dataset also grew to 2,124 images following the preprocessing and augmentation. The last data set was a combination of 1,100 normal (fit) and 1,024 pneumonia (unfit) images. It contains the changes in the image quality, illumination, and positioning of the patient that allow reflecting the actual clinical settings that are prevalent in developing nations.

Appendix B: Data Preprocessing Steps

This appendix explains the preprocessing methods used prior to the training of the deep learning models.

The preprocessing pipeline was developed to provide consistency, enhance image quality and maintain those areas of the lung that are of clinical interest.

To begin with, all the corrupted and non-radiographic pictures were eliminated. Then, each of the chest X-ray images was rescaled to 224 x 224 pixels, indicating the input of the CNN and transfer learning models applied in this research.

To save on unnecessary borders, text labels, and scanner marks without desecrating lung anatomy, lung-preserving cropping was used.

There was also data augmentation that was used to enhance model generalization. Such tricks were controlled rotation, horizontal flipping, zooming and adjusting of brightness. Radical changes were not done to retain medical realism.

This preprocessing technique contributed to better overfitting. And enhanced the soundness of the trained models.

Appendix C: Performance Metrics

In this appendix, the evaluation metrics that will be applied to measure the performance of the model are described. Medical diagnosis is the field which needs special attention and, therefore, several metrics were chosen rather than accuracy.

The following metrics were used in this research:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix
- ROC Curve

The precision and recall measure the accuracy of positive predictions and the effectiveness with which the pneumonia cases were detected respectively. The F1-score represents an equalized score in terms of precision and recall. These measures also provide a holistic and valid assessment of model performance.

Appendix D: Model Hyperparameters

The main hyperparameters that are employed when training the model are summarized in this appendix. Custom CNN and transfer learning were trained under consistent experimental settings in order to compare them fairly.

The common hyperparameter settings included:

- Optimizer: Adam
- Learning rate: 0.0001
- Batch size: 32
- Number of epochs: 10–15
- Loss function: Binary Cross-Entropy

For transfer learning models such as **VGG19, DenseNet201, MobileNetV2, and ResNet50**, pre-trained ImageNet weights were used.The last layers of classification were reconfigured to do binary classification. Due to appropriate tuning of learning rate and early stopping, overfitting was avoided.

Appendix E: Model Interpretability and Visualization

The explainability methods utilized in the research are depicted in this appendix.

They then interpreted model predictions via Grad-CAM visualizations for the sake of interpretability since medical AI systems need to be transparent.

The Grad-CAM heatmaps map the lung areas which mostly contributed to the decision of the model. Such visual explanations are used to confirm that the model focus on clinically useful lung areas rather than fiddling with a meaningless portion of an image.

The visual results demonstrate that the pneumonia related regions were consistently highlighted, which proved the reliability and robustness of our approach.

Appendix F: Software and Hardware Environment

This appendix lists the tools and resources used to implement the proposed system.

Software Tools:

- Python
- TensorFlow / Keras
- NumPy

- OpenCV
- Matplotlib
- Scikit-learn

Hardware Resources:

- Google Colab cloud platform
- NVIDIA Tesla T4 / P100 GPU
- Google Drive for dataset storage and model checkpoints

The cloud-based environment enabled efficient model training, evaluation, and visualization, making it suitable for academic research and experimentation.

221-35-1006

ORIGINALITY REPORT

12%	9%	6%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
2	www.mdpi.com Internet Source	2%
3	Submitted to Daffodil International University Student Paper	1%
4	S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufacturing", CRC Press, 2025 Publication	1%
5	Rahul Kumar Chaurasiya, Varun Bajaj, Vishakha Chourasia. "Artificial Intelligence-based Signal Processing for Brain Activity Analysis", CRC Press, 2025 Publication	<1%
6	esource.dbs.ie Internet Source	<1%
7	download.bibis.ir Internet Source	<1%
8	Alenezi, Rafa Alaiat S.. "Artificial Intelligence Applied to Cybersecurity and Health Care", North Dakota State University, 2024 Publication	<1%
9	Submitted to University of East Anglia Student Paper	<1%
10	Ayokunle Olalekan IGE, Mohd Halim MOHD NOOR. "A Deep Local-Temporal Architecture	<1%

	with Attention for Lightweight Human Activity Recognition", Applied Soft Computing, 2023 Publication	
11	Joice. C Sheeba, M. Selvi. "Pedagogical Revelations and Emerging Trends", CRC Press, 2025 Publication	<1 %
12	www.frontiersin.org Internet Source	<1 %
13	www.researchsquare.com Internet Source	<1 %
14	Sukhpreet Kaur, Amanpreet Kaur, Manish Kumar. "Recent Advances in Computational Methods in Science and Technology", CRC Press, 2026 Publication	<1 %
15	Submitted to University of Essex Student Paper	<1 %
16	Submitted to Edith Cowan University Student Paper	<1 %
17	pt.scribd.com Internet Source	<1 %
18	Afaq Khattak, Pak-wai Chan, Feng Chen, Abdulrazak H. Almaliki. "Deep ResNet Strategy for the Classification of Wind Shear Intensity Near Airport Runway", Computer Modeling in Engineering & Sciences, 2025 Publication	<1 %
19	"Emerging Science and Technology for Human Well-Being", Springer Science and Business Media LLC, 2025 Publication	<1 %
20	Fadi Al-Turjman. "Artificial Intelligence Learning Facilitators - Creating Smart Education Systems", CRC Press, 2025	<1 %

Arpita Rani Ball

221-35-1006

Quick Submit

Quick Submit

Daffodil International University

Document Details

Submission ID

trn:oid:::1:3451101598

Submission Date

Dec 28, 2025, 12:42 PM GMT+6

Download Date

Dec 28, 2025, 12:47 PM GMT+6

File Name

221-35-1006.pdf

File Size

1.0 MB

41 Pages

8,331 Words

46,338 Characters

0% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

1 AI-generated only 0%
Likely AI-generated text from a large-language model.

0 AI-generated text that was AI-paraphrased 0%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.