Comparative Analysis of Different Deep Learning Models to Classify Infectious Diseases from Chest Radiological Images

By

Md Hasibur Rahman

211-15-14616

and

Kashfea Kafi

211-15-14578

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

Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Md. Abdullah-Al-Kafi
Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

Dhaka, Bangladesh

January 12, 2025

APPROVAL

This Project titled **Comparative Analysis of Different Deep Learning Models to Classify Infectious Diseases from Chest Radiological Images**, submitted by Md Hasibur Rahman, ID No: 211-15-14616 and Kashfea Kafi, ID No: 211-15-14578 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 12 January, 2025.

BOARD OF EXAMINERS

Md Taimur Ahad 12/1/2025

Dr. Md. Taimur Ahad — **Chairman**
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mushfiqur Rahman — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Rahmatul Kabir Rasel Sarker — **Internal Examiner**
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Sadat Hasan — **External Examiner**
Data Scientist (Senior Principal Officer)
Risk Management Division
Brac Bank

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md Sadekur Rahman, Assistant Professor**, 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:

Sadik

Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Sadik

Md. Abdullah-Al-Kafi
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Md Hasibur Rahman
Student ID: 211-15-14616
Department of Computer Science and Engineering
Daffodil International University

Kashfea

Kashfea Kafi
Student ID: 211-15-14578
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project(FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Md. Sadekur Rahman**, **Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Computer Vision** 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 themat all stages have made it possible to complete this project.

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

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

Recent rapid advancement in deep learning has shown promise in manifold applications, with medical image analysis as one of its prime focus areas. A study was undertaken with the aim of comparing the effectiveness of traditional CNNs-VGG19, ResNet50, and a Custom-designed CNN architecture-with the newly evolving ViTs for categorizing chest radiological images of infectious respiratory conditions. Through an extensive EDA, we found that there are inherent challenges and complexities with medical imaging datasets. Addressing these challenges, tailored image pre-processing methodologies were used, emphasizing the importance of zoom and noise reduction in enhancing model efficacy. Our study findings have demonstrated the robustness and adaptability of CNNs, with VGG19, ResNet50, and Custom CNN outperforming the Vision Transformer on various performance metrics. However, besides accuracy, the importance of model interpretability was underlined. By applying gradient-based visualization and attention map methodologies, we tried to shed light on the "black box" nature of deep learning models and possibly open up new perspectives for enhancing cooperation between AI systems and healthcare professionals. This research underlines both the potential and challenges of AI in medical imaging and forms a foundation for further studies that conjoin technological innovation with clinical expertise.

Table Of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter provides a comprehensive introduction to the research, setting the foundation for the study. It discusses the motivation behind the research, its objectives, the adopted methodology, expected outcomes, and the structure of the report.

1.1 Introduction

Respiratory illnesses including COVID-19, tuberculosis, and pneumonia pose significant threats to global health, characterized by alarming rates of morbidity and mortality worldwide (WHO, 2020). Chest radiographic images have consistently been essential non-invasive diagnostic instruments that facilitate the prompt identification and continuous assessment of these diseases. Furthermore, these images are crucial in shaping patient management approaches and, in turn, affecting health outcomes. However, the traditional method of manually analyzing such images is labor-intensive and also subject to variability, often leading to inconsistent or even incorrect results [24].

1.2 Motivation

During the last decade, the field of deep learning has shown outstanding progress across many application domains, including in the field of medical imaging. CNNs have emerged as a promising solution for improving the accuracy of disease detection from chest X-rays, as supported by multiple recent investigations [2][3]. Advancing this development, transformer models, hitherto significant mainly within the realms of natural language processing, are establishing their presence in the field of medical imaging. Initial studies have suggested that these models can compete on an equal footing with CNNs regarding performance in particular tasks [4]. In the backdrop of such progress, there intrinsically remains an impetus to explore more in-depth capabilities of transformer models, especially for the classification of various chest radiological images.

1.3 Objectives

Overall Aim of the Research

The core objective of this research effort is to explore and evaluate Convolutional Vision Transformers for the task of classifying infectious diseases using chest radiological images. Special attention is given to satisfying two major objectives: improving model interpretability without compromising its performance levels.

Detailed Objectives

- Design and implement a range of deep learning architectures, including VGG19, ResNet50, and a Custom CNN to classify chest radiological images across five key categories: COVID-19, Lung Opacity, Normal, Viral Pneumonia, and Tuberculosis.
- Optimize the performance of these models to achieve superior classification accuracy in diagnosing infectious respiratory diseases.
- Focus on improving accuracy and efficiency specifically with the Custom CNN model.
- Identify areas for further improvement to advance the field of chest radiological image classification.

1.4 Methodology

The methodology for this research is organized into several key steps. First, data preparation involves quality checks, image preprocessing, and data augmentation to enhance the dataset's diversity and reliability. Following this, models such as VGG19, ResNet50 are developed and fine-tuned. Custom Convolutional Neural Networks (CNNs) are also created to improve classification accuracy. Models are evaluated using metrics like accuracy, precision, recall, F-score, and ROC-AUC. Finally, model introspection is conducted through convolution and attention map visualizations to analyze the decision-making process and validate model performance comprehensively.

1.5 Project Outcome

Accurate classification of chest radiological images: Successful implementation of VGG19, ResNet50, and a Custom CNN will enable the accurate classification of chest radiological images across five categories: COVID-19, Lung Opacity, Normal, Viral Pneumonia, and Tuberculosis.

Improved diagnostic capabilities: Optimized deep learning models will demonstrate superior classification accuracy in diagnosing infectious respiratory diseases, leading to better diagnostic performance compared to traditional methods.

Enhanced Custom CNN performance: The Custom CNN model will show improved accuracy and efficiency in classifying chest radiological images, highlighting its potential as a more effective tool in medical imaging tasks.

Benchmarking and identification of gaps: Through comparison with traditional deep learning architectures, the study will establish benchmarks for model performance and identify areas for future research, contributing to advancements in chest radiological image classification.

1.6 Organization of the Report

The report is organized into six detailed chapters. Chapter 1 introduces the project, providing a comprehensive overview of its motivation, objectives, methodology, and anticipated outcomes. Chapter 2 delves into the background of the research, conducting a literature review and gap analysis to establish the context and necessity of the study. Chapter 3 focuses on the research methodology, detailing the proposed approach, design specifications, project plan, and task allocation. Chapter 4 discusses the implementation and results of the project, explaining the environment setup, testing, evaluation, and comparative analysis. Detailed results and their interpretations are also provided. Chapter 5 addresses engineering standards and design challenges, ensuring compliance with software, hardware, and communication standards. Additionally, it discusses the societal impact, sustainability, ethical considerations, project management, and complex engineering problem-solving aspects of the project. Finally, Chapter 6 concludes the report by summarizing the findings, identifying limitations, and providing recommendations for future work. References to relevant studies and standards are included throughout the report to support the analysis and enhance its credibility.

Chapter 2

Background

Convolutional Neural Networks (CNNs) have shown promise in classifying infectious diseases from chest radiological images. Advanced architectures like ResNet, DenseNet, and hybrid models offer accurate and efficient disease diagnosis, including pneumonia, tuberculosis, and COVID-19. However, challenges remain, such as limited generalization, lack of interpretability, and insufficient exploration of alternative architectures.

2.1 Introduction

Chest radiography, especially chest X-rays, remains one of the most integral parts of medical diagnostics. It has been a key to the diagnosis and evaluation of a variety of respiratory conditions since its inception, from the erstwhile diseases of tuberculosis and asthma to the more recent pandemic-related viruses such as SARS-CoV-2, which causes COVID-19. Technological advancement in medicine has not diminished the importance of X-rays. Manual conventional analyses—developed through extensive expertise gathered over time—carry some advantages but are increasingly being outweighed by the concerns of inefficiency in terms of time, variation in interpretation, and the potential for human error, especially in high-pressure situations such as outbreaks and pandemics.

2.2 Literature Review

Table 2.1: Summary of Literature Reviewed.

Authors	Year	Title	Methodology	Key Findings
BK Gowru et al.	2023	Hierarchical Bayesian optimization based convolutional neural network for chest X-ray disease classification	Proposed a CNN model with Bayesian optimization for enhanced disease classification using chest X-rays	Improved disease classification with high accuracy in diverse datasets.
S KB, A CV	2024	Leveraging Compact Convolutional Transformers for	Integrated CNN with transformers	Improved interpretability and accuracy in

		Enhanced COVID-19 Detection in Chest X-Rays	and Grad-CAM visualization to classify COVID-19 from chest X-rays	COVID-19 diagnosis.
A Velayudham et al.	2023	IoT enabled smart healthcare system for COVID-19 classification using optimized robust spatiotemporal graph convolutional networks	Introduced IoT-based CNN models for COVID-19 detection with smart healthcare solutions	High precision in detecting COVID-19 from chest X-rays using IoT-driven methodologies.
R Bagri, A Shah	2024	Predictive Diagnosis of Lung Diseases Using Artificial Intelligence	Developed a CNN-based predictive model for diagnosing lung abnormalities	Accurate detection of lung nodules and diseases in radiographic images.
G Anita, S Singarapu	2023	Automated Detection and Classification of Pneumonia using Deep Learning and CNN	Automated pneumonia classification using CNN with augmented datasets	Achieved substantial accuracy improvements and reduced misclassifications.
V Asha et al.	2024	Enhanced ResNet-101 Model for Effective Classification of COVID-19 Disease from CXR Images	Leveraged ResNet-101 architecture with fine-tuning for COVID-19 classification	Enhanced classification performance compared to baseline models.
M Patankar et al.	2024	A novel dense-net deep neural network with enhanced feature selection for TB classification	Proposed a DenseNet model with feature selection for tuberculosis stage classification	Improved classification accuracy across various stages of TB.
NO Adiwijaya et al.	2024	CNN-Based Classification of Infectious Lung Diseases using Thorax X-Ray Analysis	Utilized CNN for infectious lung disease classification using Kaggle datasets	Demonstrated high precision in identifying multiple lung diseases.

N Asadoorian et al.	2024	Pre-trained Quantum CNN for COVID-19 Disease Classification Using CT Images	Combined pre-trained quantum CNN with transfer learning for CT-based COVID-19 classification	Enhanced computational efficiency and accuracy in disease detection.
O. Olabode et al.	2024	Deep Learning Models for Classification of COVID-19 Severity Levels	Implemented transfer learning-based CNN for classifying COVID-19 severity using chest X-rays	Reliable severity classification with potential for clinical integration.
Yogesh H.	2023	Deep Convolutional Neural Network-Based Covid-19 Classification From Radiology X-Ray Images For IoT-Enabled Devices.	Deep convolutional neural network classifies X-ray images into COVID-19, pneumonia, and normal categories with high accuracy for IoT devices.	Achieved accuracy of 90.95% in classifying COVID-19, Pneumonia, and Normal patients.
Aya Elagili	2022	Deep Learning to Improve COVID-19 Detection: Using CNN-Based Transfer Learning from Chest X-Ray Images.	The model demonstrates the application of transfer learning, achieving an impressive 91% accuracy with the VGG19 architecture.	Transfer learning from chest X-rays improves COVID-19 detection accuracy.

1. A CNN model optimized with hierarchical Bayesian techniques outperforms traditional methods in classifying chest X-ray images for diseases like pneumonia and tuberculosis.

While demonstrating superior accuracy and scalability, the model's robustness across diverse datasets remains a limitation [11].

2. Combines CNN and transformers with Grad-CAM for better interpretability in detecting COVID-19. Improves accuracy but has complexity and generalization issues [13].

3. Employs CNNs for lung disease classification, achieving accuracy in identifying lung nodules but lacks disease specificity and interpretability [15].

4. Same as above, highlighting CNN efficacy but with concerns about specificity and lack of interpretability [17].

5. Implements CNN with data augmentation for robust pneumonia classification but lacks evaluation for other infections and real-world variation [18].

6. Fine-tunes ResNet-101 for COVID-19 detection, offering high accuracy but with higher computational costs and limited generalization [19].

7. Uses DenseNet with enhanced feature selection for tuberculosis classification, excelling in accuracy but risking overfitting and limited disease scope [20].

8. Employs CNN for infectious lung diseases using Kaggle data. Effective but constrained by dataset limitations and absence of explainability [22].

9. Quantum CNNs with transfer learning offer efficient COVID-19 classification but are resource-intensive and focused on CT images [25].

10. Applies CNN-based transfer learning to classify COVID-19 severity. Reliable but lacks dataset diversity and exploration of other diseases [27].

11. Uses DenseNet121 and VGG16 for COVID-19 detection with 91% accuracy but lacks explainability and comparative model performance analysis [24].

Utilization of Deep Learning Techniques in Medical Imaging:

Machine learning, through its subfield of deep learning, has heralded a new era for medical imaging. Computational techniques, grounded in the ability of machines to learn patterns, are transformative. They overcome the challenges thrown up by manual methods, offering solutions that are not only faster but also consistently more accurate. In the field of chest radiography, several algorithms have been developed that can very quickly analyze images, highlighting possible areas of interest with a precision that is equal to or even surpasses that of human experts [2][3]. This speed and precision are especially crucial where early diagnosis is critical.

An Overview of Convolutional Neural Networks (CNNs)

Leading the successes of deep learning in image analysis are Convolutional Neural Networks (CNNs). These networks are designed to specially process data that has a grid-like structure, as is the case for images. In other words, CNNs can extract and learn intricate patterns from raw image data through several layers of convolution, pooling, and fully connected layers [7]. Architectures like VGG19, ResNet50, and Xception have set benchmarks in many image-related tasks; still, their application to the field of medical imaging, specifically to chest X-rays, has been revolutionary [10][6][5]. However, one glaring criticism remains: these powerful networks, despite their accuracy, often act as "black-boxes," making their decision-making process opaque, which can be a significant barrier in a field where understanding 'why' is often as crucial as knowing the 'what' [9].

Evolution of Transformer Models in Computer Vision

Although they were originally developed for sequence management in natural language processing tasks, transformer models have broadened their application. Because of their unique architecture in highlighting the interrelationships between different parts of an input, they have had a great influence on the computer vision area. One such development is the Convolutional Vision Transformer (ViT), which leverages the strength of convolutional neural networks (CNNs) with the holistic contextual understanding provided by transformers [4]. Transformers are defined by the inclusion of self-attention mechanisms, which can be very informative about the parts of an image that the model finds most informative, thus bridging the gap between high accuracy and the need for model interpretability [12].

Chest Radiograph Image Processing Methods

Before the images are subjected to the scrutiny of advanced algorithms, their quality has to be ensured first. Pre-processing techniques such as CLAHE become a very important tool in enhancing the finer details in an X-ray and ensuring that subtle but important information is not lost [1]. Further, the sharpening algorithms emphasize the boundaries and regions of interest for a more defined image ready for analysis. With zooming, it becomes possible to enlarge those areas of a picture that may hold an issue, this guarantees algorithms a much clearer view. The application of those techniques lays the foundation such that deep learning models, be it CNNs or transformers, are operating on optimized data for maximal diagnostics.

2.3 Gap Analysis

CNN-based models for classifying infectious diseases from chest radiological images often face challenges in generalization. Most models are trained and validated on specific datasets that lack diversity in terms of patient demographics, radiological equipment, and clinical conditions. Consequently, these models may underperform when applied to new populations or different imaging setups. This limitation poses a critical barrier to their widespread adoption in global healthcare systems. Addressing this gap requires using diverse and multi-institutional datasets to improve the robustness and adaptability of CNN models, ensuring they are suitable for various clinical environments and demographic groups.

CNNs have been successful in medical image classification but are often criticized for their lack of transparency. The "black box" nature of these models limits clinicians' trust, as they cannot easily understand the features and patterns driving predictions. This gap in interpretability poses challenges in clinical decision-making, where explainability is critical for accountability and validation. Techniques such as Grad-CAM and SHAP have been proposed to address this issue, but their integration into diagnostic workflows remains limited. Future research should focus on developing intuitive and robust explainability frameworks to enhance the trustworthiness of CNN-based diagnostic tools.

Although CNNs have demonstrated remarkable performance in medical image classification, the exploration of alternative architectures such as ResNet, DenseNet, and Inception remains underdeveloped. Most studies focus on standard architectures without evaluating the comparative advantages of newer or hybrid approaches. This gap restricts the discovery of more efficient and accurate models tailored for specific clinical tasks. By systematically assessing and benchmarking various architectures across diverse datasets, researchers could uncover models with superior performance and adaptability. Such efforts would optimize CNN applications in medical imaging and address current challenges in accuracy, efficiency, and generalizability.

2.4 Summary

This discussion focused on applying Convolutional Neural Networks (CNNs) to classify infectious diseases from chest radiological images, aligning with your thesis topic. After identifying a suitable dataset from Harvard's Dataverse, a literature review of 12 studies was conducted, summarizing their methodologies and key findings, including various CNN implementations like ResNet and DenseNet for diagnosing diseases such as pneumonia, tuberculosis, and COVID-19. Detailed insights were provided for each study, highlighting advancements and limitations. Additionally, three critical gaps were analyzed: limited generalization across diverse datasets, lack of interpretability in CNN predictions, and insufficient exploration of alternative architectures. These insights establish a robust foundation for addressing gaps and advancing your thesis research.

Chapter 3

Research Methodology

This section provides a detailed account of our data handling practices, model development strategies, evaluation metrics, and introspective analysis techniques.

3.1 Methodology

The methodology for this study is designed to systematically utilize convolutional neural networks (CNNs) to classify infectious diseases from chest radiological images. It begins with data preparation, which ensures high-quality inputs through steps like quality checks, image preprocessing for normalization and enhancement, and data augmentation to increase dataset diversity and robustness. In the model development and training phase, pre-trained models such as VGG19, ResNet50 are fine-tuned, and a Custom CNN is developed to optimize performance. Next, model evaluation and comparison are conducted using metrics like accuracy, precision, recall, F-score, and ROC-AUC to determine the best-performing model. Finally, model introspection and analysis employs techniques like convolution visualization and attention map visualization to analyze the model's decision-making process, ensuring transparency and reliability. This methodology ensures a structured and effective approach to improving classification accuracy while addressing data quality, model performance, and interpretability challenges.

3.1.1 Overview

This study presents a structured methodology to classify infectious diseases in chest radiological images using convolutional neural networks (CNNs). The process begins with data preparation, focusing on ensuring high-quality inputs through quality checks, image preprocessing, and data augmentation to enhance the dataset's diversity and robustness for training. Following this, model development and training involves leveraging pre-trained models such as VGG19, ResNet50 as well as developing and fine-tuning a custom CNN to optimize performance. The models are then subjected to evaluation and comparison using key performance metrics, including accuracy, precision, recall, F-score, and ROC-AUC, to identify the most effective model. To ensure transparency and interpretability, the final step, model introspection and analysis, uses visualization techniques like convolution and attention map visualization to analyze how the models make predictions. This methodology provides a comprehensive framework for achieving reliable and accurate disease classification.

3.1.2 Proposed Methodology

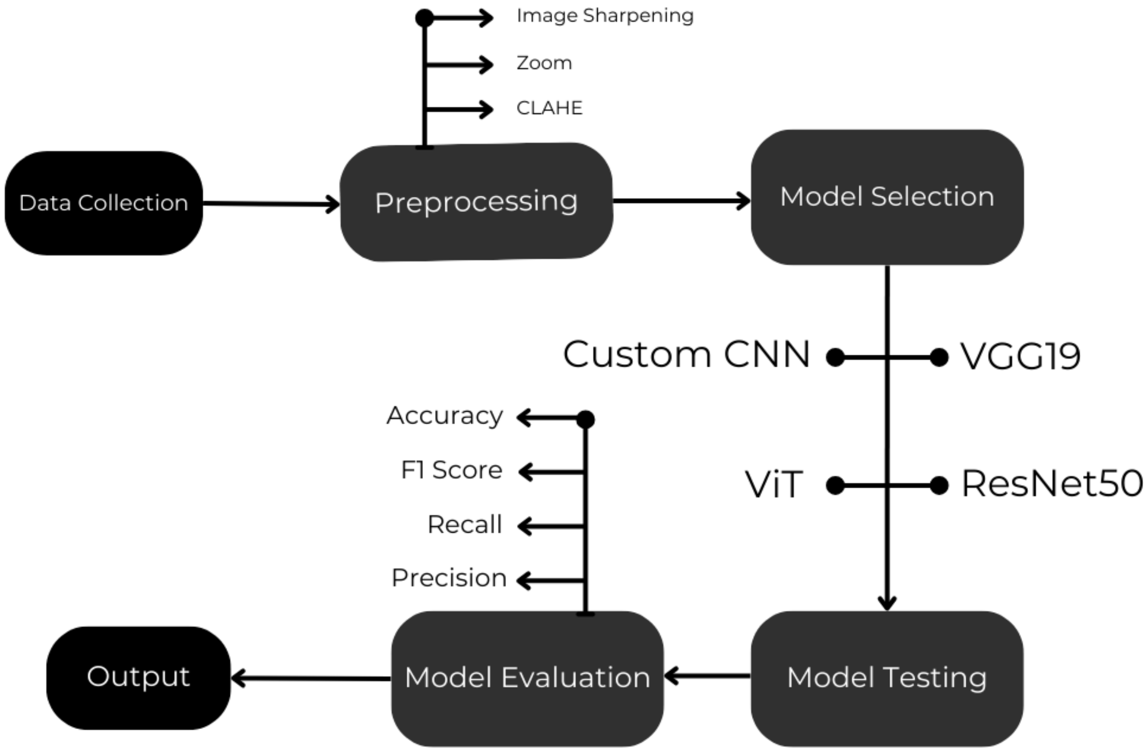

Figure 3.1: Proposed methodology diagram.

3.2 Detailed Methodology and Design

Data Preparation:

This project uses the open-source, free-to-access chest radiological image dataset hosted online [28]. Classified into COVID-19, Lung Opacity, Normal, Viral Pneumonia, and Tuberculosis, the database presents a wide array for analysis. The quality verification ensures that all images used here are frontal views, since this is diagnostically relevant and most consistent. All images go through a label accuracy check followed by preprocessing for standardization of size and format. The other steps under Data Preparation involve EDA, Image Pre-processing, Batching, Shuffling, which are discussed next. Following are a few example images from the training set and the test set:

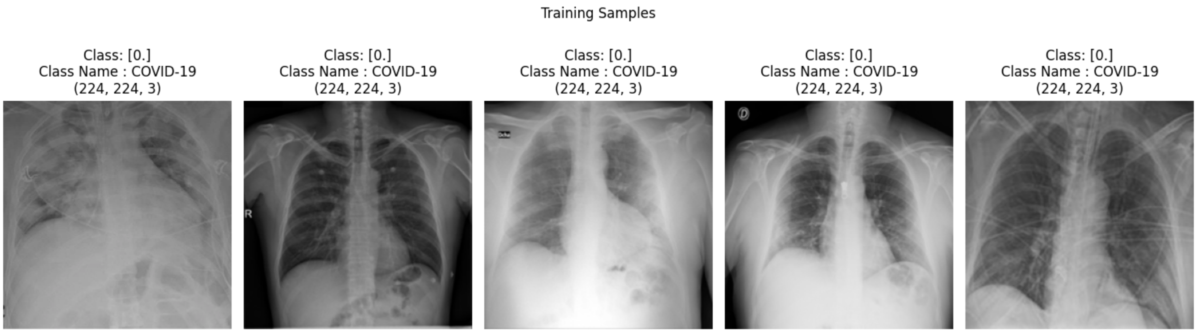

Figure 3.2 Sample images from the training dataset.

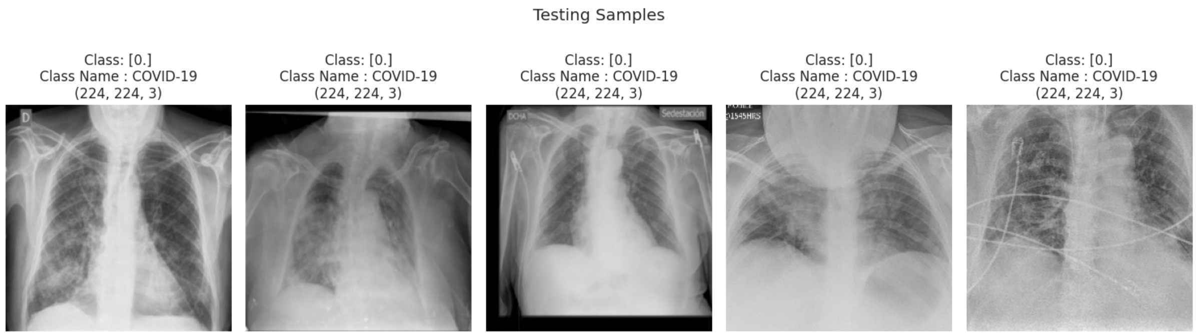

Figure 3.3: Sample images from the testing dataset.

Exploratory Data Analysis:

Distribution of classes across various datasets.

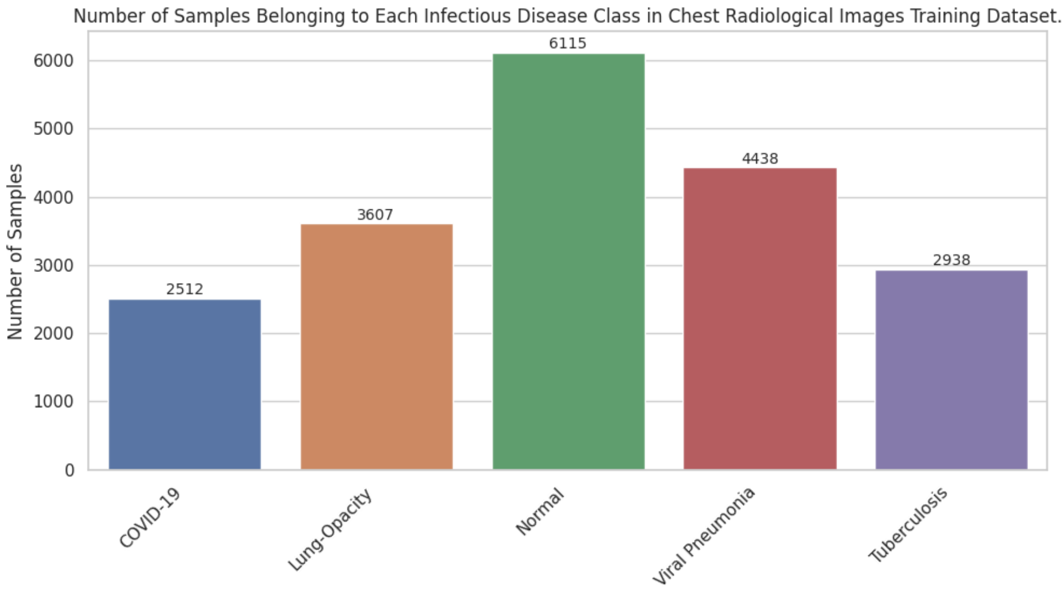

Figure 3.4 Class distribution in the training dataset.

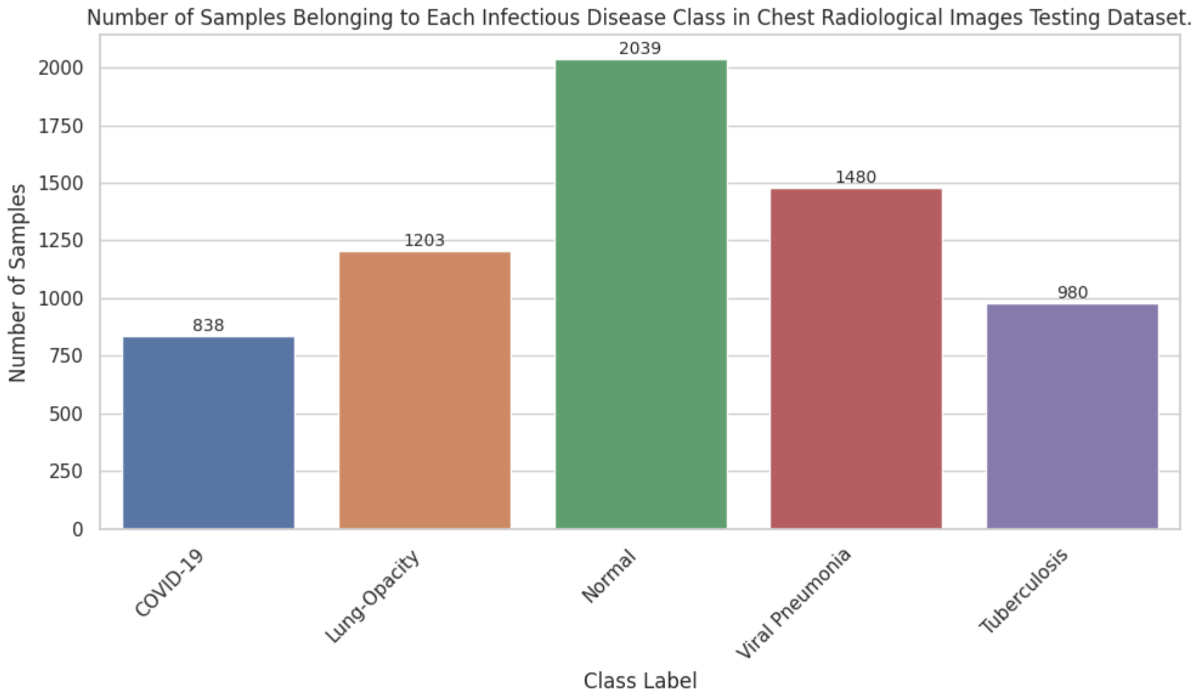

Figure 3.5 Class distribution in the testing dataset.

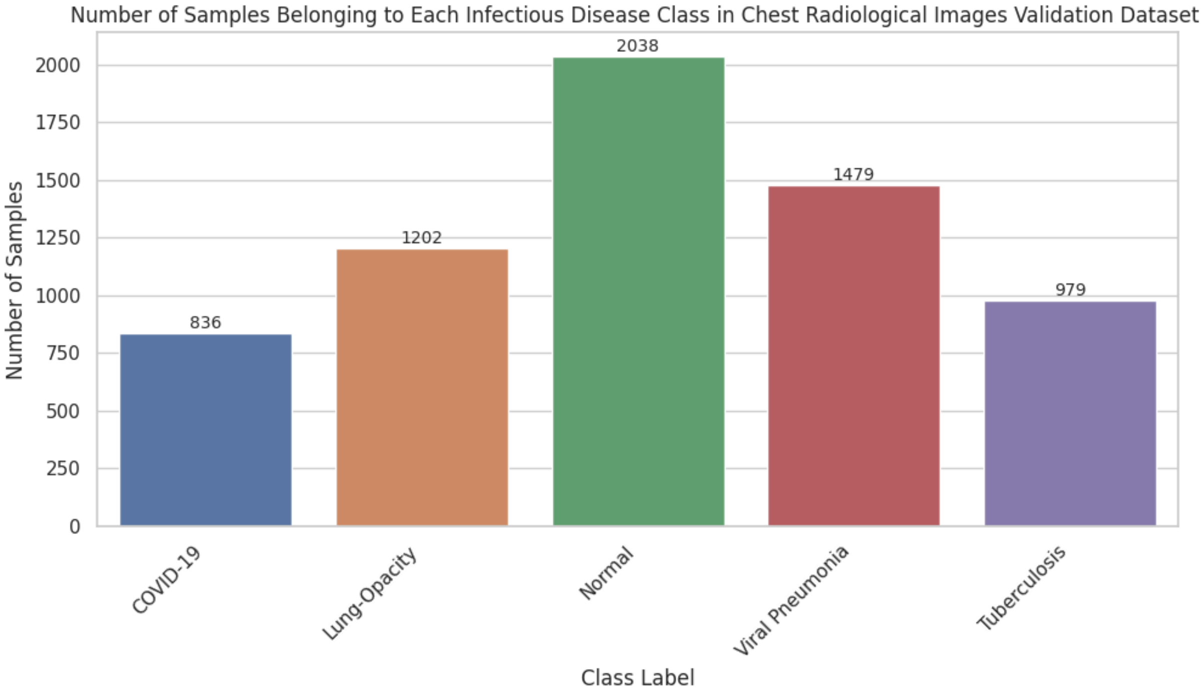

Figure 3.6 Class distribution in the validation dataset.

- The distribution of the classes in the training, testing, and validation datasets has been analyzed; as one may see from figures 3.4, 3.5, and 3.6, the different classes are highly skewed. The 'normal' class is very dominant especially in the training dataset as opposed to the 'covid19' and 'tuberculosis' classes that manifest a relatively lower number of samples.
- The training, testing, and validation datasets have the same distribution of classes, but the absolute count differences may create substantial differences in model training and performance assessment.

	Class Weights:
0	1.5613057324840764
1	1.0873301912947047
2	0.6413736713000817
3	0.8837314105452907
4	1.334921715452689

Figure 3.7 Class based weights to use during training.

Results of Class Imbalance:

- This might just be a class imbalance problem, where models can tend to train in a way that they naturally would rather predict the majority class (in this case, 'normal') more often, since it gets many more "rewards" when it predicts the majority class correctly.

- Models developed utilizing imbalanced datasets may exhibit deceptively elevated accuracy rates; however, they often underperform in practical applications, particularly when tasked with predicting minority classes. This limitation is especially problematic as such a model may struggle to reliably identify 'covid19', which raises significant concerns regarding patient care and public health outcomes.

Justification for using class weights:

- Class weight is the common mechanism to deal with class imbalance by assigning higher weights to the under-represented classes and lower weights to the over-represented classes. This way, it gives a balanced "penalty" during model training such that errors in predicting minority classes are more "penalized" compared to the majority classes.
- Weighing the class weights that are introduced, the class 'covid19' has the highest weight so that the model will face a heavier penalty for misclassifying the examples of 'covid19' compared to misclassify examples that fall under 'normal'.
- The application of these weights can make the model more sensitive to minority classes, thus increasing its generalization ability and performance regarding under-represented categories.

Average pixel intensity distribution across different datasets.

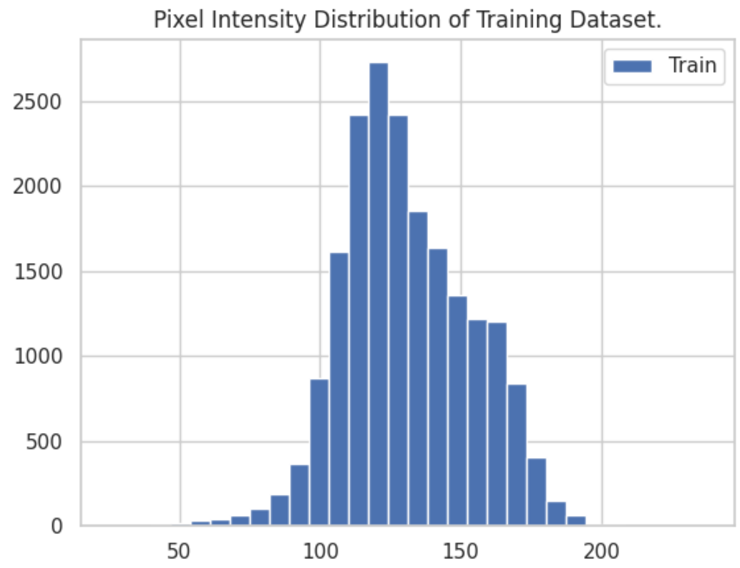

Figure 3.8 Training images pixel intensity distribution.

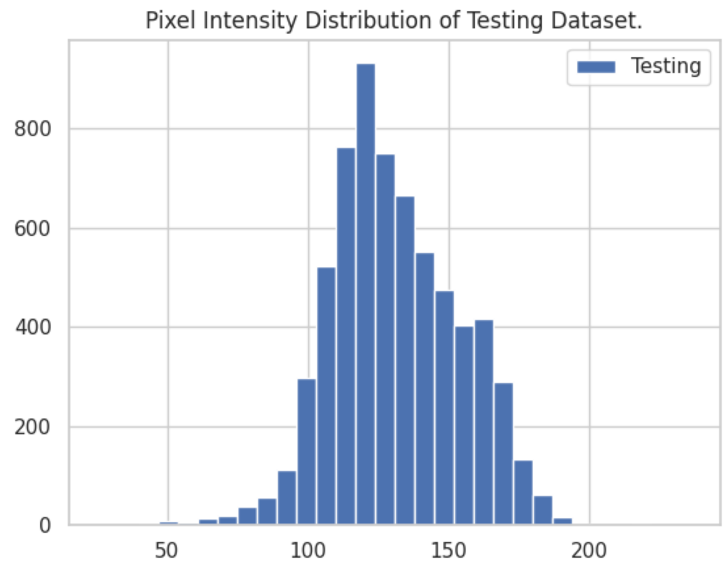

Figure 3.9 Testing Images pixel intensity distribution.

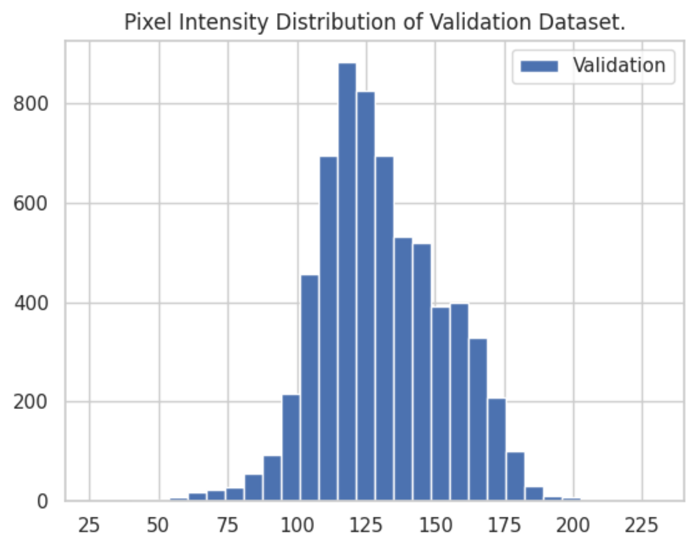

Figure 3.10 Validation Images pixel intensity distribution.

Consistency Among Sets:

A thing that is important is the homogeneity seen in the pixel intensity distribution of training, testing, and validation datasets, as shown by Figures 3.8, 3.9, and 3.10, respectively. This will show evidence that the radiological image acquisition and preparation protocols established are being regularly adhered to for all datasets. This would thus assure generalizability and reliability of the results by training, validating, and testing the models over comparable distributions of image intensity.

Distribution of pixel intensities across classes and datasets.

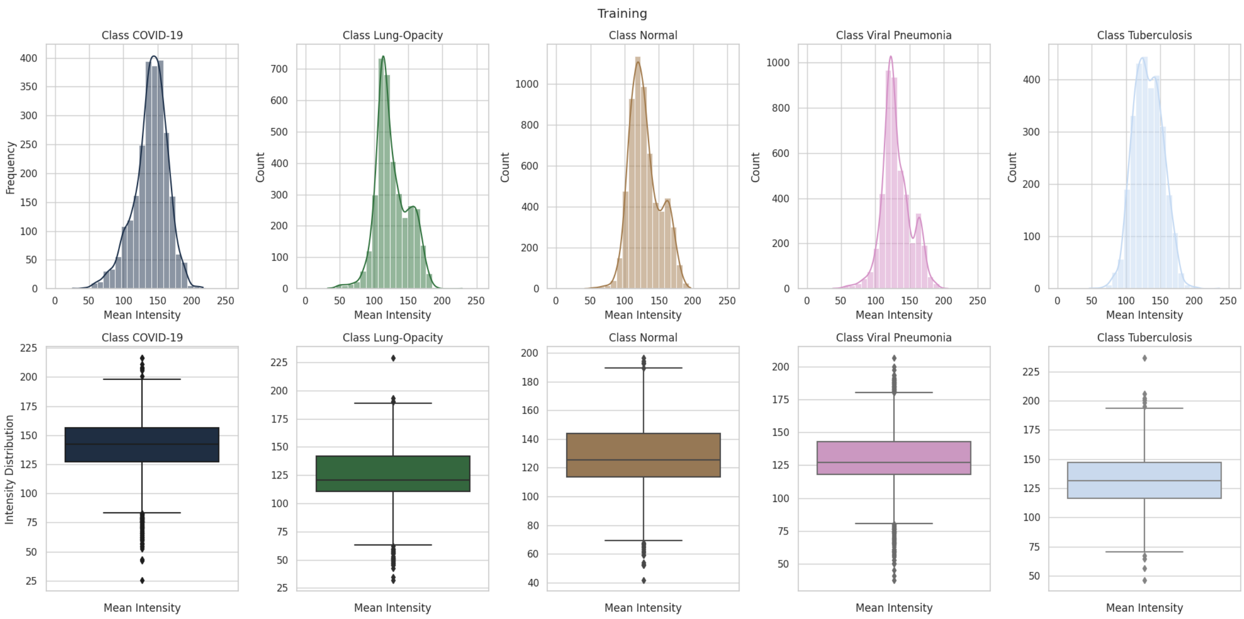

Figure 3.11 Training Pixel Intensity Distribution by Class.

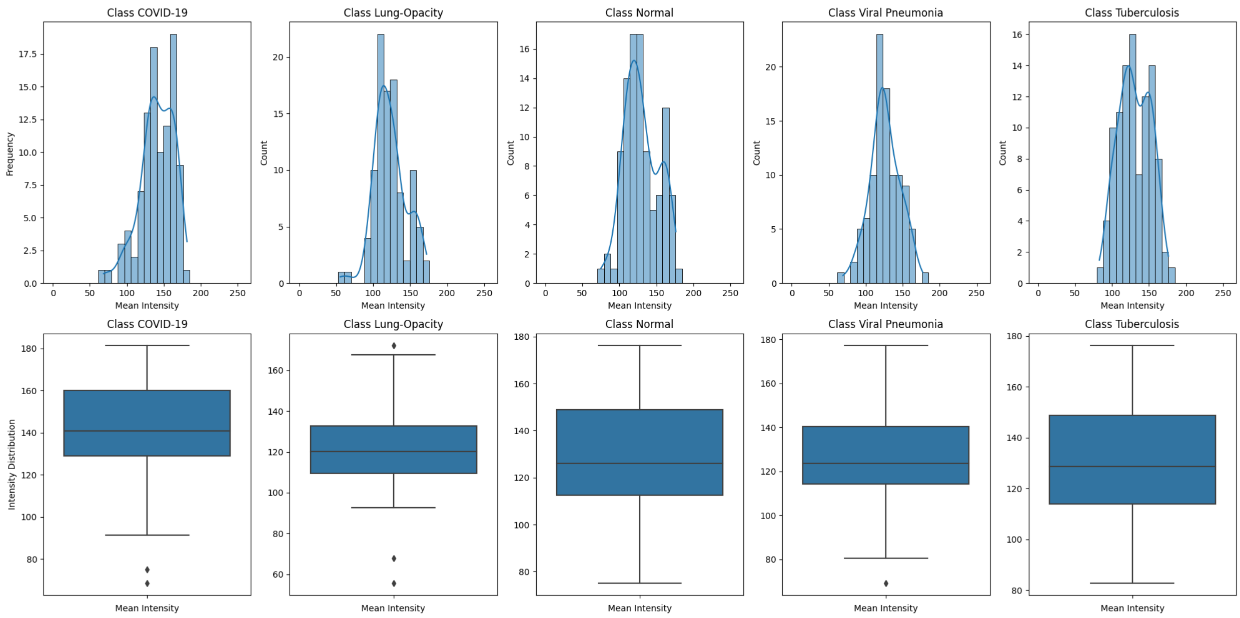

Figure 3.12 Testing Pixel Intensity Distribution by Class.

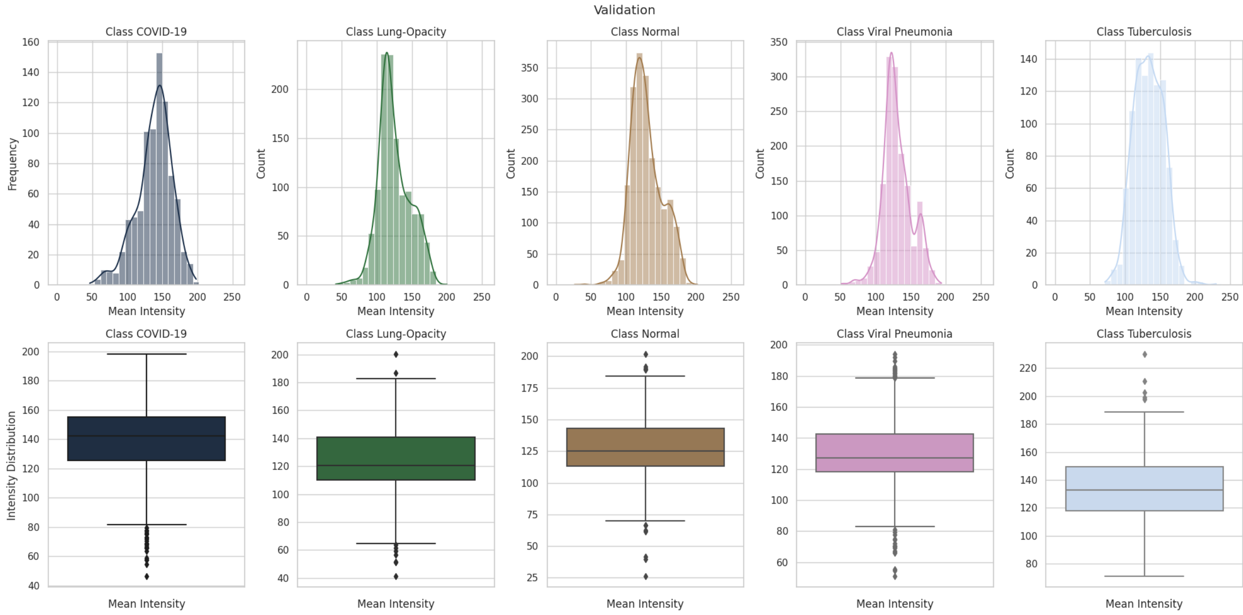

Figure 3.13 Validation Pixel Intensity Distribution by Class.

Overview of Intensity Distributions:

The variation in pixel intensity distributions across the different classes, as shown by figures 3.11, 3.12, and 3.13, gives us important insights into how distinctive each disease's radiological manifestation might be. Probably, these variations are associated with specific pathophysiological changes related to each condition.

Problems with 'Lung Opacity' and 'Normal' Classes:

The nearly identical distributions of pixel intensity of the 'Lung Opacity' and 'Normal' classes are a challenge in themselves. This may be an indication of the existence of inherent similarities in radiological characteristics between these classes, which could make the model hardly distinguishable from one another. The finding is consistent with the expectation of possibly reduced precision or recall for 'Lung Opacity,' suggesting the need for improved feature extraction or specialized modeling techniques in order to better distinguish them.

Class-specific Observations:

- **COVID-19:** A near-normal distribution suggests this class is well represented, likely because of even imaging or manifestation of disease.
- **Lung Opacity:** A less frequent, secondary peak after the descent from the mean in the intensity distribution could represent a subcategory within the wider class and may suggest that there are multiple radiological patterns related to this pathology.
- **Normal:** Its similarity with 'Lung Opacity' and the secondary peak, though pronounced, suggests the possibility of overlapping with pathological conditions.
- **Viral Pneumonia:** The steep ascent and descent with a subsequent spike post-mean could be representative of a mixture of typical and atypical manifestations of the disease within the dataset.

- **Tuberculosis:** The twin peaks centered around the mean indicate the presence of two distinct radiographic patterns for this condition, both of which are well-represented in the dataset.

1. Image Pre-Processing:

Pre-processing and standardization techniques encompass a range of methods aimed at transforming raw data into a format suitable for analysis and interpretation. These techniques include:

Contrast Limited Adaptive Histogram Equalization:

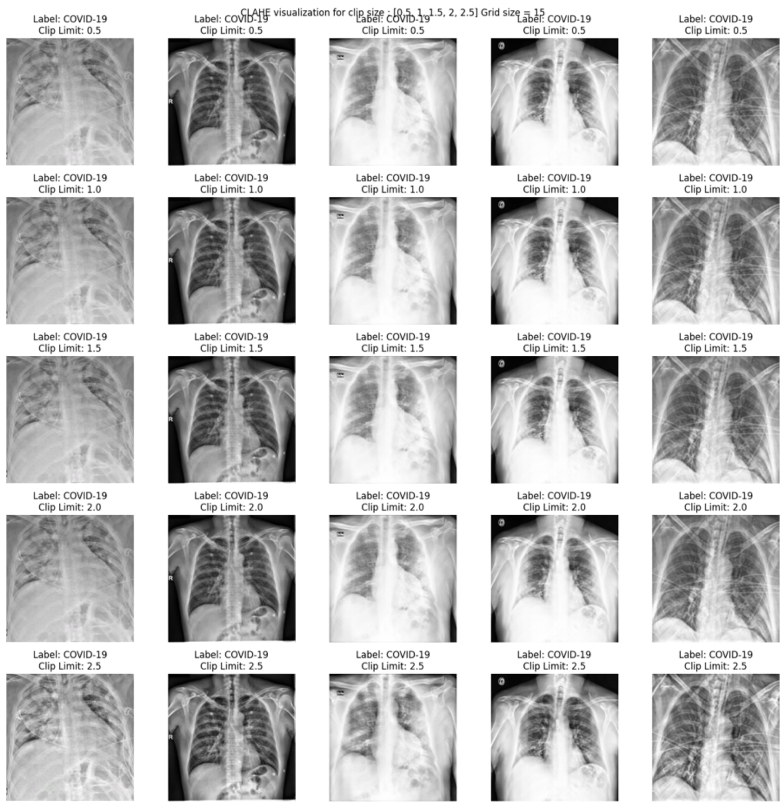

Figure 3.14 Experimenting with clip sizes when applying CLAHE.

CLAHE enhances contrast in localized regions of images, preventing over-amplification of noise and limiting the intensification of contrast.

Zoom

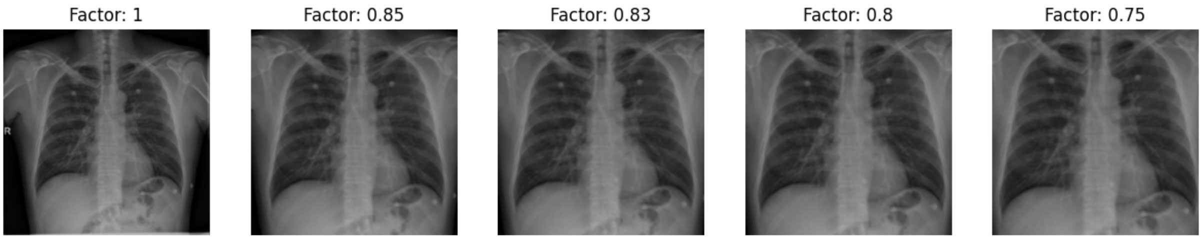

Figure 3.15 Experimenting with zoom factors.

Zooming, a pre-processing technique for chest X-ray images, offers two benefits:

1. **Artifact and Noise Elimination:** Healthcare facilities and equipment manufacturers often place labels, brandings, or diagnostic details on radiological images, which are "noise" for deep learning models. These markings, concentrated around the image's periphery, introduce unwanted variances and biases into the dataset. By zooming into the chest X-ray's core region, we can remove or reduce these extraneous details, focusing the model's attention on the medically relevant parts.

2. **Avoiding Misleading Model Training:** A model's success should be based on recognizing and differentiating pathology from normalcy based on medically relevant features, not irrelevant artifacts. Deep learning models, especially in medicine, may sometimes use these artifacts or noise for predictions, leading to inaccurate predictions and reduced generalizability and reliability. By eliminating these potential pitfalls, we ensure our model learns from underlying pathology, not visual cues.

Image Sharpening

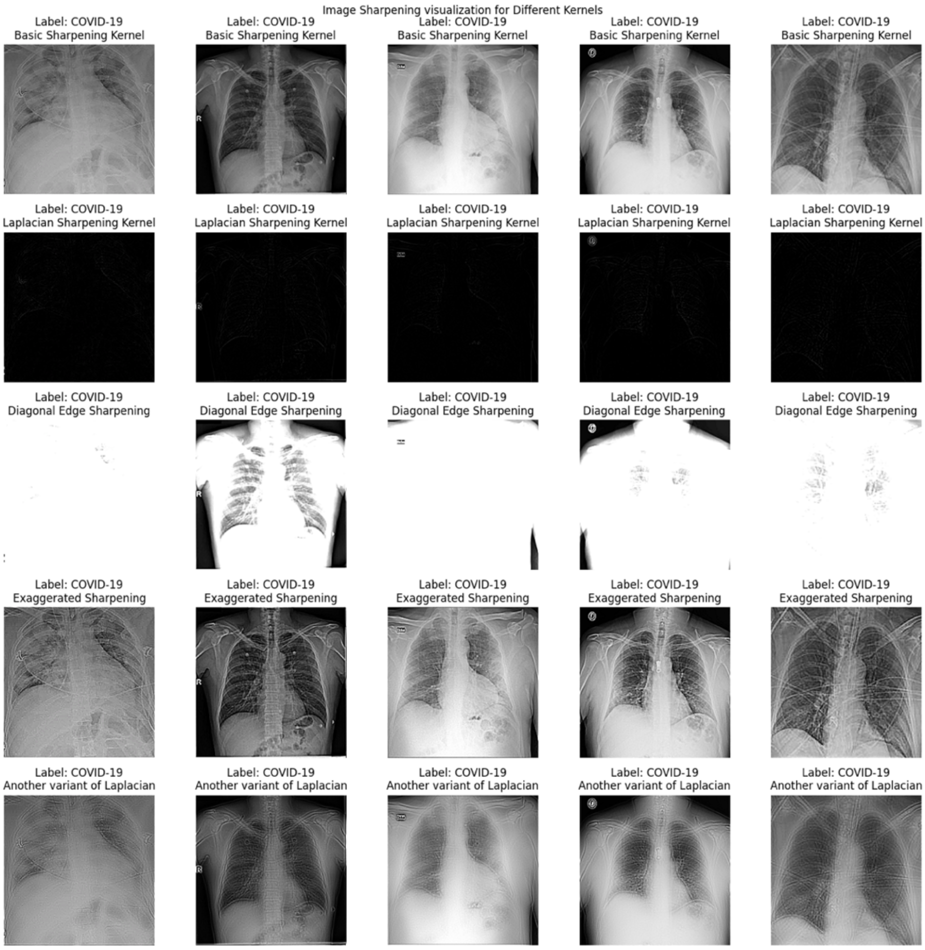

Figure 3.16 Experimenting with image sharpening kernels.

Ensuring clarity and precision in medical imaging is crucial. Image sharpening enhances boundaries and delineates intricate structures, aiding accurate diagnosis and analysis. We evaluated multiple sharpening kernels on a given dataset, as shown in Figure 3.16.

Pre-Processed Images

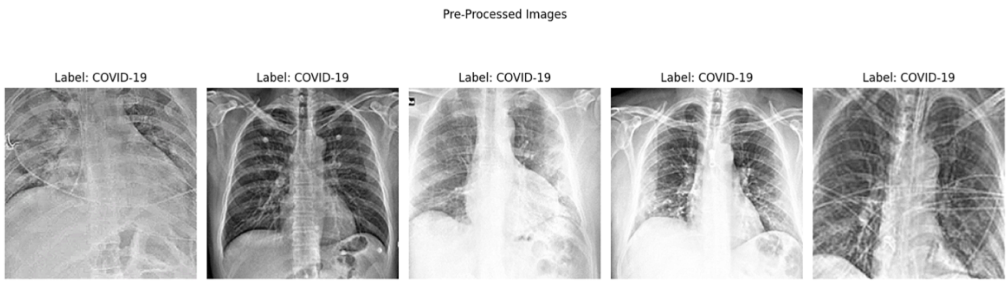

Figure 3.17 Zoom, CLAHE and Image Sharpening.

Three fundamental techniques were combined on selected images. Figure 3.17 shows the compounded effects and potential synergy.

Shuffling and Batching:

The initial dataset was provided in .npz format. While .npz format is versatile in storing arrays, it has the disadvantage of being memory-intensive when dealing with large datasets such as radiological images. This, therefore, makes data manipulation and processing difficult, especially on systems with constrained RAM.

To lighten these computational constraints, the dataset was serialized with a pixel intensity data type (dtype) conversion to float32. The choice of float32 was strategic; it strikes a balance between precision and memory consumption, ensuring that the images retain enough detail for accurate analysis while consuming less memory.

The serialized data was then saved in hdf5 after that. The hdf5 format has been developed to store and organize high-volume data efficiently. Data can be accessed efficiently and does not require loading data into memory, hence quite computationally friendly for this application and much better than the use of the .npz format.

Having procured the data in hdf5 format, further enhancements were brought to the handling of the data using the tf.data.Dataset API in TensorFlow. The data pipeline operations were made even smoother using the Map Dataset and Prefetch Dataset variants from tf.data.Dataset.

Batch size 128. The data is batched and, very importantly, shuffled: much of the machine learning literature assumes this step of randomness, especially when augmented by some features of the tf.data.Dataset API of TensorFlow in Prefetch Dataset.

Importance of Shuffling in Machine Learning:

Shuffling datasets prevents overfitting and ensures uniform distribution of data, mitigating bias from data order.

2. Model Development and Training:

Deep learning models, including VGG19 and ResNet50, are fine-tuned on a curated dataset with adjusted hyperparameters for improved performance.

- Proprietary CNN uses depth-wise separable, dilated, residual, attention, and batch normalization.
- Specialized Convolutional Vision Transformer model developed for image classification.
- Training Protocol: There are dedicated training sets the pre-trained models go through. A validation set allows parameter tuning. The Adam optimizer runs based on the categorical cross-entropy loss function. The idea here is to invoke transfer learning by using the architecture from pre-trained architectures: VGG19 and ResNet50. In this initial stage, the top parts of the models behave as feature extractors. Next, once the gradients start adjusting, full-network training takes place. Such a two-phase protocol matters where preliminary model adaptation should happen followed by extensive training of the network.

3. Model Evaluation and Comparison:

Models evaluated on separate test set using metrics like accuracy, precision, recall, F-1 score, and ROC-AUC. Comparative analysis determines most effective model.

Accuracy:

- **Definition**: The ratio of correctly predicted instances to the total instances is known as the accuracy of a prediction.

$$\text{Accuracy} = \frac{\text{Number of Correct Prediction}}{\text{Total Number of Prediction}}$$

Precision:

- **Definition**: Ratio of correctly predicted positives to total predicted positives.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall:

- **Definition**: Ratio of correctly predicted positives to all actual positives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

F-score (F1 Score):

- **Definition:** F1 Score balances precision and recall, considering false positives and negatives.

$$\text{Recall} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

ROC-AUC (Receiver Operating Characteristic - Area Under the Curve):

- **Definition:** ROC curve plots true positive rate against false positive rate. AUC quantifies model's ability to discriminate between positive and negative classes.

4. Model Introspection and Analysis:

Deep learning model interpretability, especially in medicine, requires visualization tools.

1. Grad-CAM (Gradient-weighted Class Activation Mapping):

- Grad-CAM produces localization maps highlighting important image regions for concept prediction.
- Technique highlights significant regions in medical images, aiding diagnosis and model trust.

2. Guided Grad-CAM:

- Method combines guided backpropagation and Grad-CAM for high-resolution visual explanations [9].
- Guided Grad-CAM enhances radiological image resolution, revealing intricate details and granular understanding of influential features.

3. Spatial Attention Layers Visualization:

- CNNs use spatial attention to focus on specific image regions [21].
- Visualizing spatial attention in medical imaging models can reveal regions of significance, potentially correlating with disease manifestation.

3.3 Project Plan

This project aims to explore the potential of Convolutional Neural Network for classifying infectious diseases such as COVID-19, pneumonia, tuberculosis, and lung opacity using chest radiological images. The study will benchmark traditional Convolutional Neural Networks (CNNs) and custom deep learning architectures. Additionally, it emphasizes improving model interpretability through visualization techniques, which is crucial in medical imaging.

Objectives

- **Dataset Preparation:** Acquire, preprocess, and balance datasets for effective training.
- **Model Development:** Implement and fine-tune traditional CNNs (VGG19, ResNet50), a Custom CNN, and a ViT model for classification tasks.
- **Evaluation:** Compare model performance using metrics such as accuracy, precision, recall, F1-score, and AUC.
- **Interpretability:** Enhance model transparency using Grad-CAM, Guided Grad-CAM, and attention visualizations.
- **Insights and Recommendations:** Draw clinical and computational insights to improve disease detection accuracy.

Timeline:

Table 3.1: Project timeline table

Phase	Tasks	Duration
1. Planning and Setup	Define research questions, set objectives, acquire resources, and prepare the environment.	2 weeks
2. Dataset Preparation	Collect datasets, perform exploratory data analysis (EDA), and preprocess images.	3 weeks
3. Model Implementation	Implement VGG19, ResNet50, Custom CNN, and Vision Transformer architectures.	4 weeks
4. Training and Fine-Tuning	Train models using transfer learning and fine-tuning strategies.	4 weeks
5. Model Evaluation	Evaluate models on test data using classification metrics.	2 weeks
6. Visualization & Analysis	Generate visualizations and interpret model predictions with attention mechanisms.	2 weeks

3.4 Task Allocation

In this research, tasks are distributed between Team Member 1 and Team Member 2 to align with their respective strengths and contributions. Team Member 1 takes primary responsibility for data handling, including collection, preprocessing (e.g., CLAHE, zooming, sharpening), and addressing class imbalances through techniques like data augmentation and applying class weights. Also trains the majority of the models, evaluates their performance metrics, and generates Grad-CAM visualizations for model interpretability. Additionally, Team Member 1 contributes to the report by documenting the data preparation and methodology sections. Team Member 2 trains one model and supports Team Member 1 with the literature review by collecting relevant research papers and summarizing their findings. These summaries provide overviews that are incorporated into the study. Team Member 2 leads the report writing, documenting the literature review, evaluation results, and future recommendations. Both team members collaborate on finalizing the report to ensure consistency and accuracy. Regular discussions between Team Member 1 and Team Member 2 maintain alignment throughout the research. Together, they prepare the final presentation and defense of the study.

3.5 Summary

This study explores the use of deep learning, particularly Convolutional Neural Networks (CNNs), for classifying infectious diseases such as COVID-19, pneumonia, tuberculosis, and lung opacity from chest radiological images. The methodology includes data preparation with rigorous preprocessing to address class imbalance, using techniques like normalization and augmentation. Models such as VGG19, ResNet50, a Custom CNN, and a Vision Transformer are trained and evaluated using metrics like accuracy, precision, recall, F1-score, and ROC-AUC, with a focus on mitigating challenges posed by radiologically similar classes. To enhance transparency, interpretability tools like Grad-CAM and Guided Grad-CAM are employed to visualize decision-making processes, aligning the models with clinical diagnostic needs. The research aims to combine robust performance with interpretable AI solutions for improved reliability in medical imaging diagnostics.

Chapter 4

Implementation and Results

This chapter provides a detailed overview of the experimental environment and analyzes the outcomes of the research. It includes the environment setup, testing and evaluation metrics, comparative analysis of different models, and a discussion of the results.

4.1 Environment Setup

To implement and replicate the research, the environment setup requires a MacBook Air M1 or similar hardware with a minimum of 16GB RAM (32GB recommended), 512GB SSD, and an operating system like Ubuntu 20.04 LTS, Windows 10/11, or macOS. The software setup includes Python 3.8 or higher and IDEs such as Google Colab, Visual Studio Code, PyCharm, or Jupyter Notebook. CUDA and cuDNN are essential for NVIDIA GPUs. Required Python libraries include numpy, pandas, TensorFlow, Keras, PyTorch, OpenCV, Pillow, matplotlib, seaborn, and plotly for computations, deep learning, image processing, and visualizations. Additional tools like tensorflow-datasets, h5py, scikit-learn, albumentations, and joblib are recommended for dataset handling, data augmentation, and model evaluation. These resources enable efficient data processing, model training, and analysis for research purposes.

4.2 Comparative Analysis

A comparison of the performance between the Convolutional Vision Transformers (ViT) and traditional Convolutional Neural Networks (CNNs), which include VGG19 and ResNet50 and custom CNN architecture in chest radiological image classification, provides a number of lessons:

Table 4.1: Comparison performance metrics of VGG19, ResNet50, Custom CNN and ViT model

Sl	Model Name	Loss	Test Accuracy	Precision	Recall	AUC	F1 Score
0	VGG19	0.15	0.94	0.94	0.94	0.99	0.94
1	CustomCNN	0.20	0.97	0.92	0.92	0.99	0.92
2	ResNet50	0.22	0.93	0.93	0.93	0.99	0.93
3	ViT	0.66	0.79	0.81	0.76	0.95	0.78

Performance Metrics:

- **ViT** has an accuracy of 0.78, which is relatively lower than other models, its loss value was way higher at 0.66, indicating possible struggles during the training phase.

- **VGG19** was the best performer with the highest accuracy of 0.94 and lowest loss value of 0.15.

- **CustomCNN** was very close with an accuracy 0.97, but that could be a typo in view of its loss is higher than VGG19. ResNet50 achieved a competitive accuracy of 0.93, which was between VGG19 and ViT.

- **ResNet50** ResNet50 achieved a competitive accuracy of 0.93, which was between VGG19 and ViT.

In a nutshell, the Vision Transformers demonstrated great strides in computer vision, yet their application on the classification tasks of chest radiological images, under the current dataset and computational constraints, seems less optimal in comparison with the conventional CNNs like VGG19, ResNet50, and the Custom CNN model.

Deep learning has revolutionized medical imaging, but existing research faces limitations like narrow disease focus and lack of interpretability. This research addresses these gaps by employing a comprehensive approach that integrates multiple architectures and focuses on a wide range of infectious diseases. Below we will see why our models works better than the existing works:

Deep CNN for chest X-ray classification but lacked comparison and explainability. Our research benchmarks multiple architectures, addresses more diseases, and ensures transparency for clinical trust [11].

Aya Elagili's study uses transfer learning for COVID-19 detection but lacks model diversity and clinical interpretability. Our research addresses these limitations by incorporating various architectures and explainability techniques [24].

Velayudham's IoT-enabled COVID-19 classification framework faces scalability and security challenges. Our research addresses multiple diseases, incorporates transparency through Grad-CAM, and utilizes Vision Transformers, offering a more robust and adaptable approach [15].

4.3 Results and Discussion

The main objective of this work is to evaluate several deep learning models on a large scale, in terms of their capability of classifying different conditions from chest radiology images. Models are evaluated using a wide range of metrics, from loss to the detailed AUC curves. To complement numerical metrics, visualization methods also give insights into models' focus when making a prediction.

Before getting into individual model evaluations, it's useful to know the general framework underpinning their training:

1. Models Evaluation:

- VGG19
- ResNet50 (both transfer learning and fine-tuning approaches)
- A CustomCNN leveraging depth wise separable convolutions and spatial attention skip connections.
- Vision Transformer

2. Training Dynamics:

- Batch size: 128
- Initial Learning Rate: 0.0001
- Optimizer: Adam
- Loss Function: Cross-entropy
- Activation: Softmax (Post one-hot encoding)

VGG19 Assessment

VGG19: The convolutional neural network architecture developed by the Visual Geometry Group, University of Oxford, is probably the most popular for simplicity in image classification tasks owing to its performance. This network was inspired by an approach toward deeper layers, starting with a very small kernel of 3×3 convolutions and held great influence over the models used in deep learning research or practice. Below illustrates a breakdown of how each is done with the applied results.

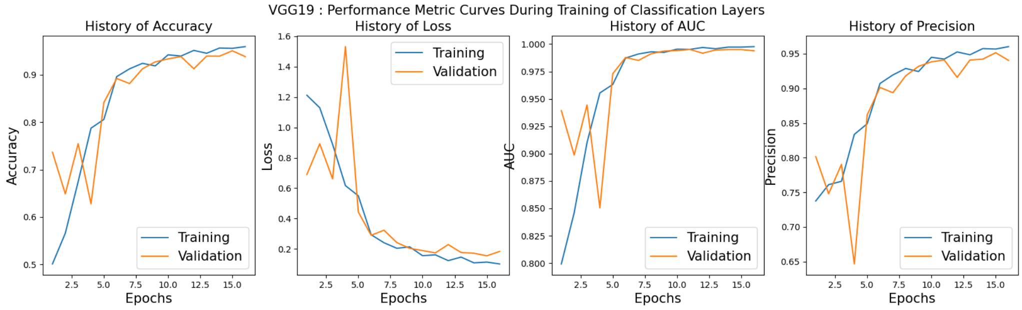

Figure 4.1 VGG19 loss accuracy during transfer learning.

Transfer Learning Phase: During this phase, as shown in figure 4.1, there was a decrease in loss for both the training set and the validation set. The metrics like accuracy, precision, and recall increased; hence, the model got better at classifying images correctly.

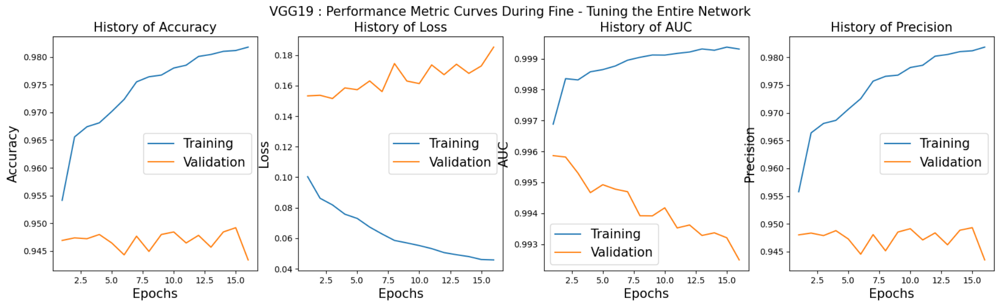

Figure 4.2 VGG19 loss accuracy during fine-tuning.

Fine-Tuning Phase: During fine-tuning as shown in figure 4.2, the model overfit. Training set metrics improved, but validation performance declined, especially in loss, accuracy, precision, and recall.

Classification Report

	precision	recall	f1-score	support
0-Covid 19	0.94	0.95	0.95	838
1-Lung Opacity	0.89	0.92	0.90	1203
2-Normal	0.94	0.93	0.94	2039
3-Viral Pneumonia	0.98	0.96	0.97	1480
4-Tuberculosis	0.97	0.98	0.98	980
accuracy			0.95	6540
macro avg	0.95	0.95	0.95	6540
weighted avg	0.95	0.95	0.95	6540

Figure 4.3 VGG19 Classification Report.

Classification Report Analysis:

- Viral Pneumonia and Tuberculosis classes showcased impressive precision and recall values, suggesting the model's strong capability to detect and correctly classify these conditions.
- Lung opacity and normal images, however, pose a unique challenge. As indicated in our exploratory data analysis, these classes showed quite similar distributions of pixel intensities and were therefore harder to classify. This fact was reflected in their precision and recall values. While the values are still very good, they are a little lower compared to the other classes, which underlines the inherent challenge in distinguishing between these two classes.
- Moreover, the Covid 19 classes also showed impressive precision and recall but not to the level of Viral Pneumonia or Tuberculosis. This can be attributed largely to the similarities of chest X-ray or chest radiology images of covid 19 to that of lung opacity or normal control. This will come out clearly below under discussion on confusion matrix heatmap.

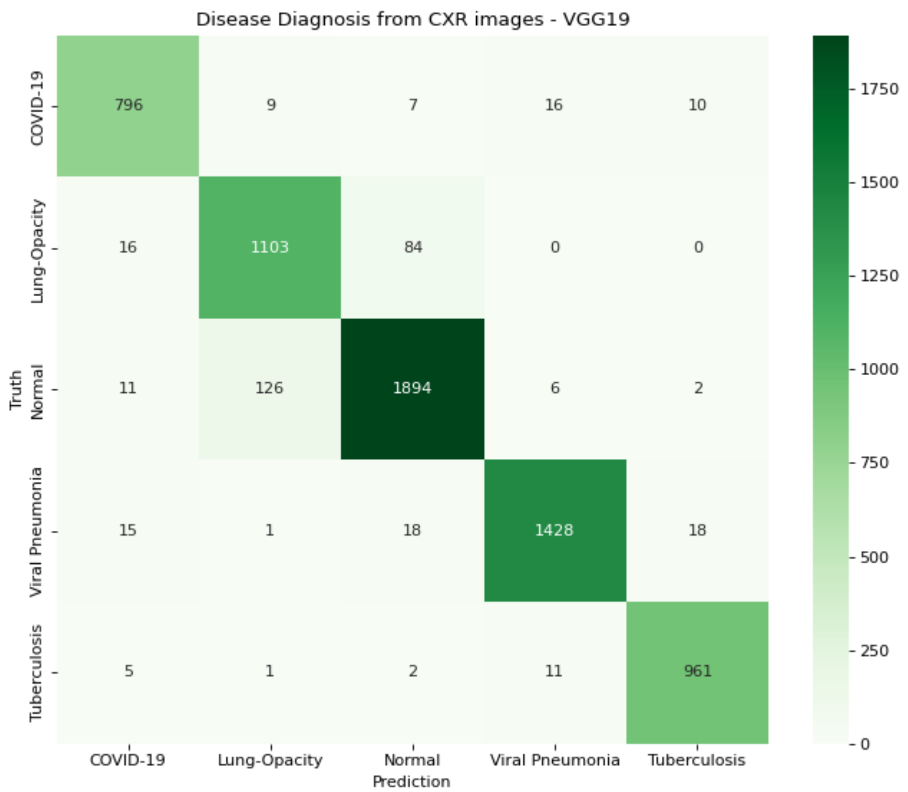

Figure 4.4 VGG19 Confusion Matrix

Confusion Matrix: Figure 4.4: The confusion matrix supports our initial hypothesis from the EDA: the overlapping pixel intensity distributions between lung opacity and normal images might reduce a model's discriminative power. That is also supported by the off-diagonal elements in the confusion matrix, particularly between these two classes. While the model performed exceptionally well, largely or overall, as shown by the diagonal elements of the confusion matrix of figure 4.4, the largest portion of misclassifications was between lung opacity and normal control. 126 Normal controls were miss-classified as lung opacity. While 84 were miss-classified as normal, making it the largest number from any other classes. Covid 19 also had some proportionately high numbers of misclassifications to other classes, such as Viral Pneumonia or Tuberculosis.

Model Activation Visualizations: To implement both grad-cam and guided grad-cam heatmaps, the pre-requisite is the activation flows running through the last convolutional layer in the model. Also to note, the preprocessing of the VGG19 dataset did not include the zooming functionality. The lack of the zooming step basically renders the visualizations overwhelmed with the peripheral noises or artefacts usually around chest X-rays as typically indicated by figure 4.5 shown below.

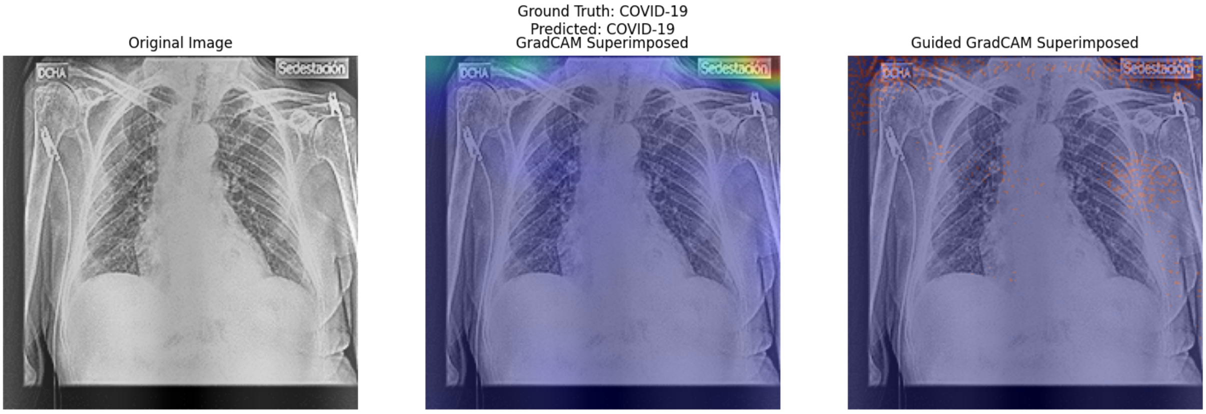

Figure 4.5 VGG19 , comparing GradCam with guided GradCam

Grad CAM: This visualization technique highlights the areas in the image that highly activate the neurons in the deeper layers of the model. In our case, however, Grad CAM strongly highlighted peripheral artifacts of the images rather than clinically informative areas. This could be due to the lack of zooming during pre-processing, which would have otherwise reduced such distractions.

Guided Grad CAM: This approach used backpropagation and Grad CAM together to give higher resolution maps. It would be interesting to find, however, that in paying attention to the clinically relevant features, it doesn't totally ignore the artifacts—it can look past them. Because guided Grad CAM does not visualize only the output from a final convolutional layer, but also the importance along this feature through the entire network. Hence, after finding out that late layers started focusing on noise—it resulted in edges—finding how much earlier layers catch is important for clinically related pattern capture.

ResNet50 Assessment.

ResNet50 architecture uses deep residual learning for image recognition.

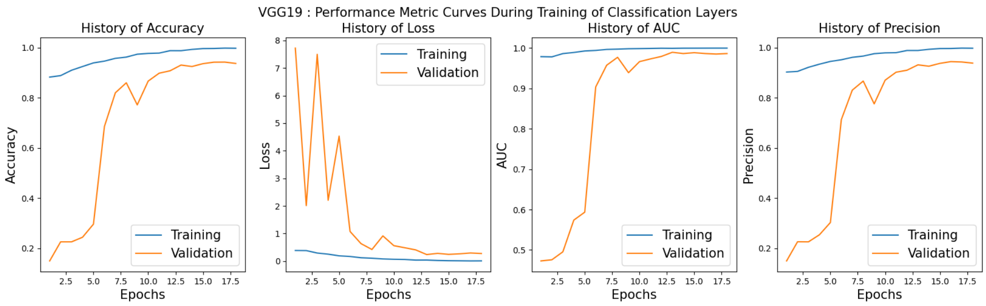

Figure 4.6 ResNet50, Loss accuracy curve during transfer learning.

Transfer Learning Phase: ResNet50, used in the transfer learning setup, showed very good adaptability, similar to that of the VGG19 model, as illustrated by figures 4.1. All the performance metrics continued to improve, especially accuracy, precision, and recall, which showed how well this model had learned from previous knowledge and was able to adapt that to our dataset.

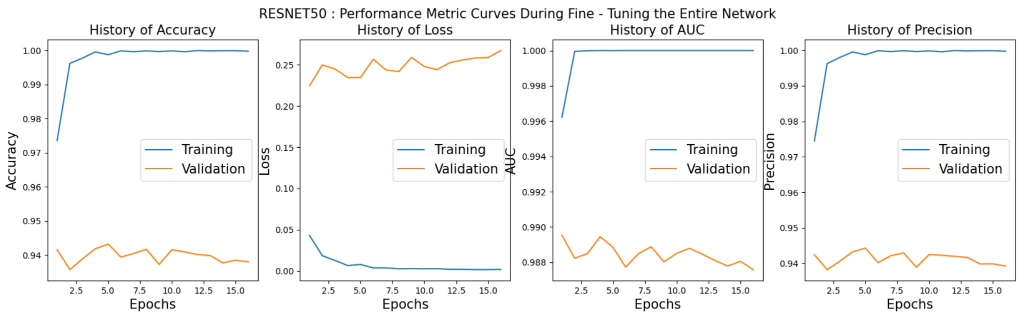

Figure 4.7 ResNet50, Loss accuracy curve during fine tuning.

Fine-Tuning Phase: But it was noticed during fine-tuning that ResNet50 is prone to overfitting, which also became a trend from the VGG19 evaluation through figure 4.2. The training metrics just kept increasing and increasing a normal trend for models learning and becoming experts on the training set. But this expertise did not change much on the validation set. A slight increase in validation loss, along with a mild decrease in metrics such as accuracy, precision, and recall, indicated that the model might be over-specializing on the training data and losing its generalization capability on unseen data.

Classification Report

	precision	recall	f1-score	support
0	0.97	0.91	0.94	838
1	0.94	0.87	0.90	1203
2	0.91	0.97	0.94	2039
3	0.93	0.99	0.95	1480
4	0.99	0.89	0.94	980
accuracy			0.94	6540
macro avg	0.95	0.93	0.94	6540
weighted avg	0.94	0.94	0.94	6540

Figure 4.8 ResNet50, classification report.

Classification Report Analysis:

- **Covid19 (Class 0):** The model had an excellent precision of 0.97 but with a recall of only 0.91, it means it was highly precise in most Covid19 predictions made, yet it did miss predicting some actual cases of Covid19. Given the real-world scenario right now, it is very critical to maximize the recall for this class, as missing the actual positives will have dangerous consequences.
- **Lung Opacity (Class 1):** A class which was more interesting from the EDA perspective: considering the visual similarity between 'Lung Opacity' and 'Normal' seen in EDA, a precision of 0.94 and recall of 0.87 are highly commendable for this model. But since the recall is a bit lower than precision, the model sometimes classifies 'Lung Opacity' as some other class; most likely 'Normal', due to their similar pixel intensity distribution seen in EDA.
- **Normal (Class 2):** The model had a higher recall of 0.97 than precision, indicating that it has a strong proclinity toward classifying images as 'Normal'. This might explain the reduced recall for 'Lung Opacity', suggesting possible false negatives for 'Lung Opacity' classified as 'Normal'.
- **Viral Pneumonia (Class 3):** With a recall of 0.99, the model almost always identifies 'Viral Pneumonia' when it's present, demonstrating its prowess in the recognition of this particular pathology.
- **Tuberculosis (Class 4):** High precision of 0.99 means that whenever the model predicts an image as Tuberculosis, it almost always is; though with a recall of 0.89, it is not catching all the cases of Tuberculosis present, which hints at possible misclassifications.

Confusion Matrix:

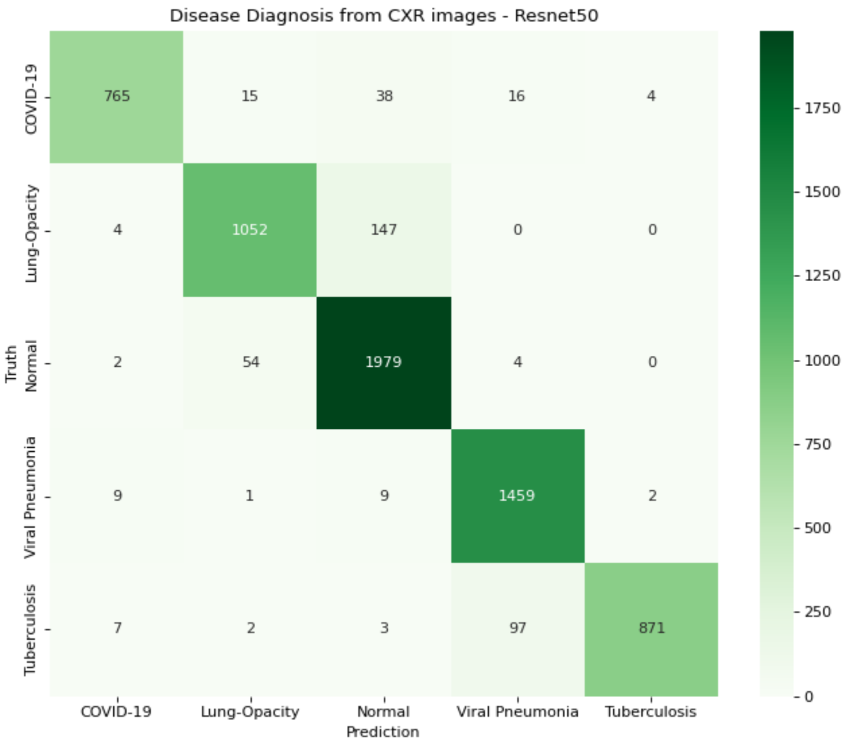

Figure 4.9 ResNet50, Confusion Matrix Heatmap.

Lung Opacity vs. Normal: The largest number of misclassifications for the model were between 'Lung Opacity' and 'Normal', with 147 and 54 instances, respectively. This agrees with our previous EDA findings where these two classes presented analogous pixel intensity distributions. More likely than not, such statistical similarity confuses the model, leading to mistakes that could prove clinically meaningful.

Covid19 Misclassifications: Covid-19 presented a relatively high rate of misclassification, being mainly classified as 'Lung Opacity', 'Normal', and 'Viral Pneumonia'. It is important to appreciate the clinical implications of these misclassifications. Covid-19 can manifest as pneumonia and have radiological features similar to both 'Lung Opacity' and 'Viral Pneumonia,' which may be why the model gets confused between these classes. The overlap in radiological features makes it very difficult to differentiate Covid-19 from quite a few other pneumonias, even for seasoned radiologists.

Overall Model Performance: While there are some areas in the results that need improvement, it must be emphasized that, based on the on-diagonal components of the confusion matrix, the general efficacy of the ResNet50 model is amazing. Its high level of accuracy shows a good aptitude for generalizing much and not being too susceptible to overfitting due to its core underlying architecture. The off-diagonal components hint at those particular areas where the model's performance can be improved.

Model Activation Visualizations: Figures 4.10 and 4.11 contain a visual experiment that covers the generation of Grad-CAM and Guided Grad-CAM superimposed images using the ResNet50 model, respectively. To look into the effect of different blending strengths on the clarity, interpretability, and visual aesthetics of the superimposed heatmap, the alpha parameter can be varied incrementally between 0.1 and 0.6 with a step size of 0.1.

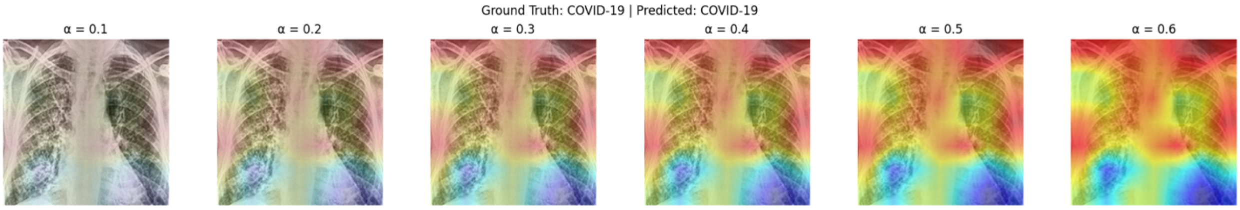

Figure 4.10 ResNet50, experimenting with different values of alpha for superimposed grad-cam image generation.

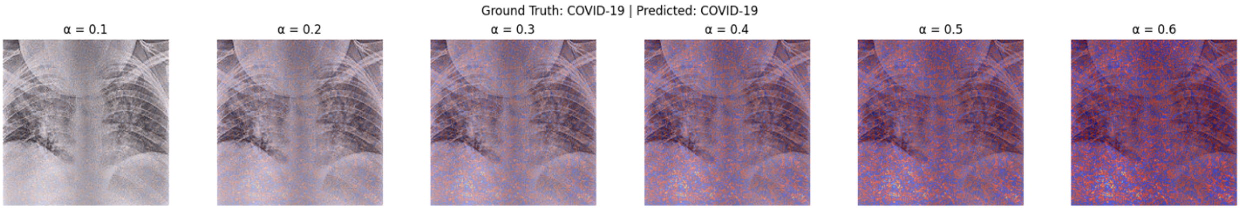

Figure 4.11 ResNet50, Experimenting with different values of alpha for guided Grad Cam visualization.

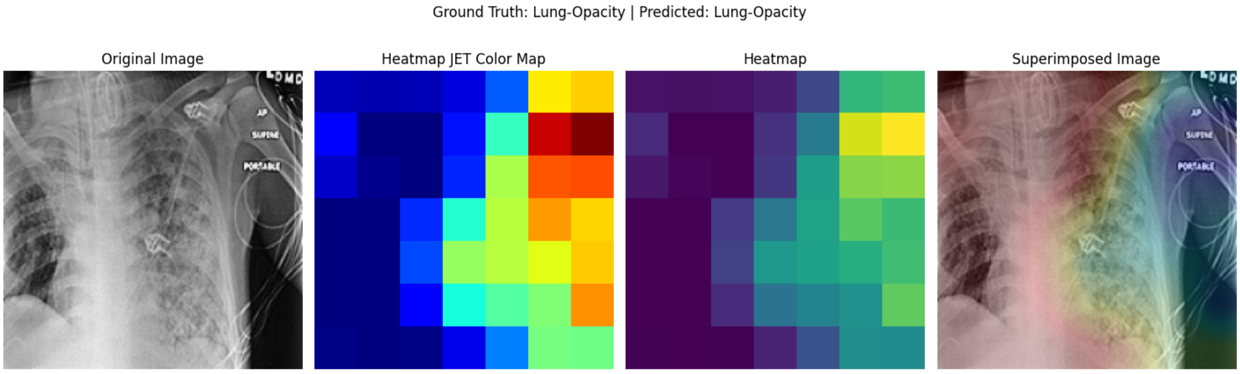

Figure 4.12 ResNet50, Visualizing the original image, the gradcam image in different color maps and the superimposed image.

Figure 4.12 sequentially displays the original medical image, its corresponding Grad-CAM visualizations across various color maps, and the final superimposed image. This juxtaposition highlights the transformation from raw diagnostic information to areas of model attention, culminating in a consolidated image. The diverse color maps in Grad-CAM offer flexibility in visual interpretation, allowing viewers to select the most clinically intuitive palette. The superimposed image synthesizes raw and attention-driven data, providing a holistic view for practitioners to correlate model predictions with discernible features in the diagnostic image. This progression enhances model transparency and reinforces the potential of visual aids in clinical decision-making.

Custom CNN

Custom CNN architecture combines spatial attention, depth wise convolutions, and skip connections for efficient training and feature extraction.

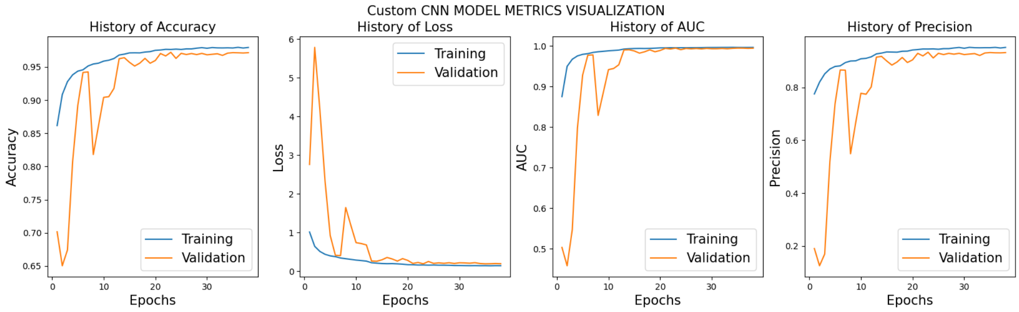

Figure 4.13 Custom CNN loss accuracy curve during training.

Training from Scratch

Loss Convergence: In contrast to the VGG19 and ResNet50 models that exhibited overfitting during fine-tuning, the Custom CNN demonstrated favorable convergence behaviors for both training and validation sets. The training loss exhibited a consistent downward trend, aligning with the anticipated optimization patterns as depicted in Figure 4.13.

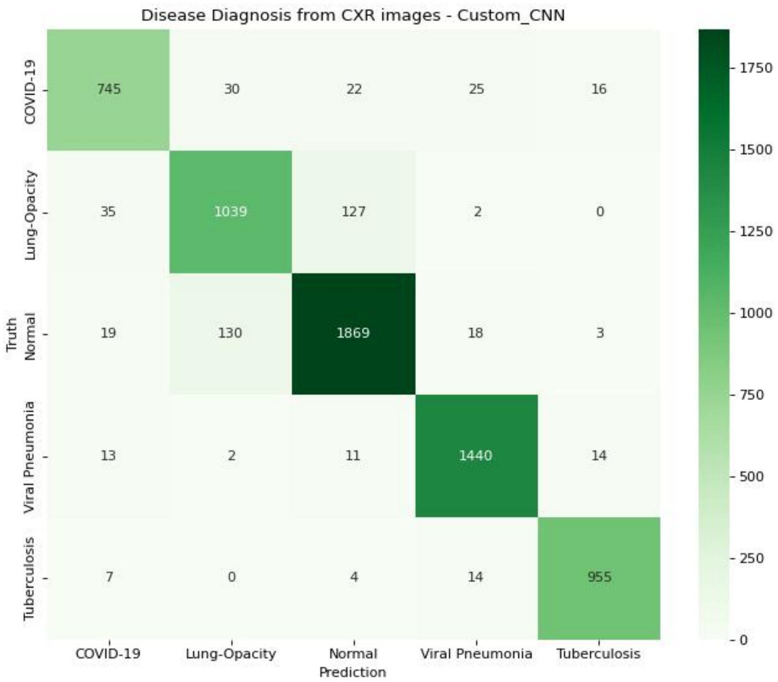

Figure 4.14 Custom CNN confusion matrix heatmap.

Confusion Matrix Heatmap: The Custom CNN model demonstrated classification patterns similar to those observed in the VGG19 and ResNet50 models, particularly in the task of distinguishing between the ‘Lung Opacity’ and ‘Normal’ classes.

1. Lung Opacity versus Normal: The distinction between ‘Lung Opacity’ and ‘Normal’ samples remains the most significant classification challenge. The substantial number of misclassifications, as evidenced by the VGG19 and ResNet50 models, underscores the intricate differentiation between these two classes. Clinically, the boundary between what is considered ‘normal’ and what is identified as ‘lung opacity’ can be subtle. The comparable average pixel intensity distributions, as observed during the exploratory data analysis (EDA), provide quantitative confirmation of this challenge.

2. COVID-19 Misclassifications: The misclassifications associated with the ‘COVID-19’ class further emphasize the intricate nature of the disease. COVID-19, being a novel pathogen, can manifest in X-ray images in ways that closely resemble other conditions. For instance, it may exhibit features common to ‘Lung Opacity’, ‘Normal’, or ‘Viral Pneumonia’. These shared radiographic patterns across different conditions underscore the significance of augmenting deep learning models with clinical expertise and supplementary diagnostic tools.

Classification Report

	Precision	Recall	F1-Score	Support
0	0.91	0.89	0.90	838
1	0.87	0.86	0.86	1203
2	0.92	0.92	0.92	2039
3	0.96	0.97	0.97	1480
4	0.97	0.97	0.97	980
----------	----------	--------	---------	---------
Accuracy			0.92	6540
Macro Avg	0.92	0.92	0.92	6540
Weighted Avg	0.92	0.92	0.92	6540

Figure 4.15 Custom CNN classification report

Classification Report Analysis: Custom CNN model’s classification metrics analyzed for chest X-ray classes.

1. General Performance: An overall accuracy of 0.92 indicates that the Custom CNN model performs exceptionally well in all aspects. The macro and weighted averages for precision, recall, and F1-score consistently range from 0.92 to 0.93, demonstrating a balanced performance across all classes without any discernible bias.

2. Class-specific Insights:

- **Class 0 (COVID-19):** The precision and recall for this class are relatively close, with 0.91 and 0.89, respectively. This indicates that the model can identify COVID-19 cases with a high degree of confidence. However, there is a slight tendency to miss a few actual cases, as indicated by the recall of 0.89.
- **Class 1 (Lung Opacity):** This class has the lowest F1-score of 0.86. Given the challenges previously highlighted about distinguishing between "Lung Opacity" and "Normal" samples, this result is not entirely unexpected. The similarity in pixel intensity distributions for these two classes may contribute to this slightly reduced performance.
- **Class 2 (Normal):** The model achieves a balanced precision and recall for normal samples, with both metrics attaining a value of 0.92. This implies that the model exhibits an equal likelihood of generating false positives and false negatives for this particular class.
- **Class 3 (Viral Pneumonia) & Class 4 (Tuberculosis):** Both of these classes demonstrate the highest F1-scores at 0.97, indicating a robust performance by the model in identifying these conditions.

3. Comparison with Other Models: When compared to the VGG19 and ResNet50 models, the Custom CNN demonstrates comparable performance. Nevertheless, the persistent difficulty across all models in distinguishing 'Lung Opacity' from 'Normal' samples highlights the inherent complexity of the task.

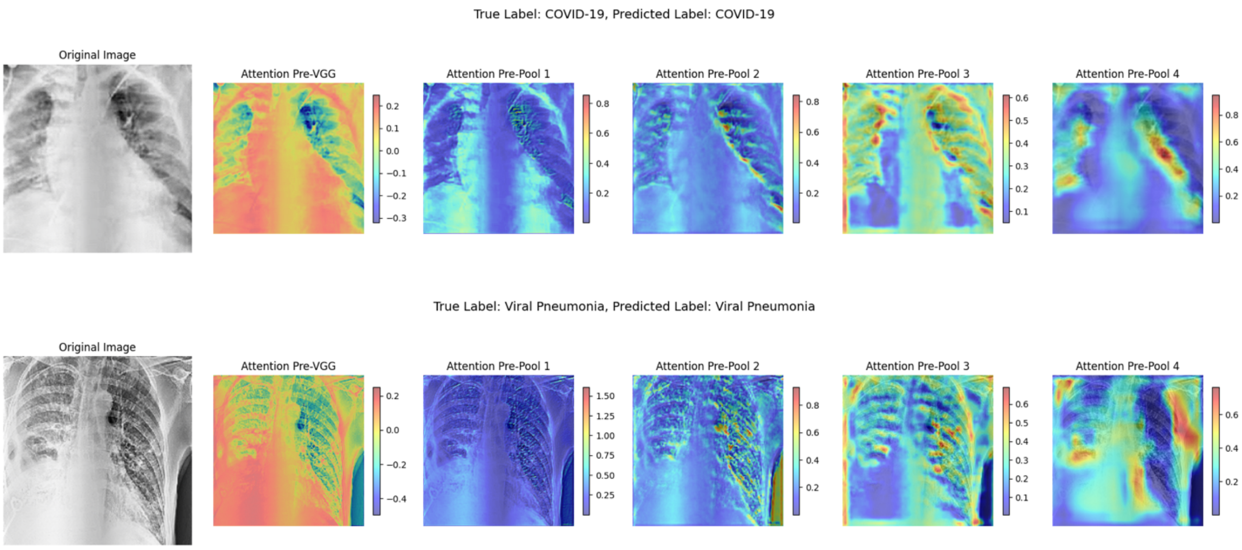

Figure 4.16 Custom CNN, Visualizing the spatial attention maps in different layers.

Attention Map Visualization: Figure 4.16 above elucidates the functioning and significance of spatial attention mechanisms within the Custom CNN architecture, providing a deeper comprehension of how attention dynamically evolves across various layers to enhance interpretability and performance.

Vision Transformer Model

Vision Transformers (ViT) have emerged as a revolutionary approach in computer vision, departing from the conventional convolution-based architectures. The fundamental concept is to treat images as sequences of patches, analogous to words in a sentence. Subsequently, transformer mechanisms, which have demonstrated remarkable success in natural language processing (NLP) tasks, are employed for image classification.

Architecture: Adapting the ViT b16 variant from Dosovitski et al.'s seminal work entitled "An image is worth 16x16 words," the model divides images into predetermined-sized patches, linearly embeds them, and subsequently processes them through self-attention mechanisms. This paradigm shift enables the model to capture long-range dependencies within the image, a capability that traditional CNNs may encounter difficulties in achieving without employing deeper architectures or specialized modules [4].

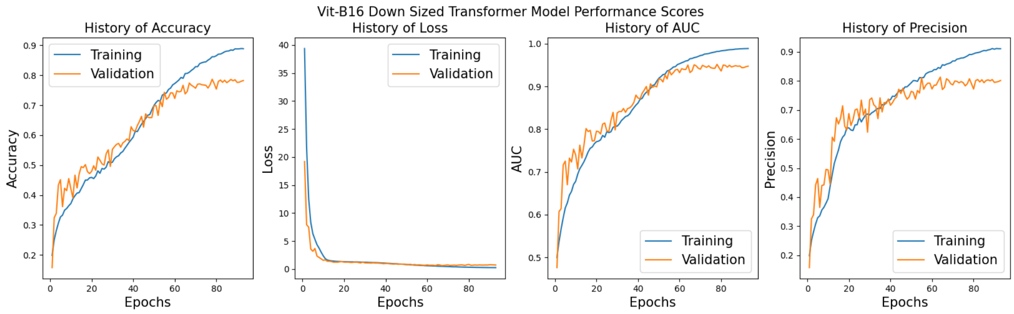

Figure 4.17 Vision transformer loss accuracy curve.

Training Dynamics: Figure 4.17 presents a visualization of the training dynamics of the ViT model on the chest X-ray dataset. As observed, the model initially exhibits an improvement in metrics such as loss, accuracy, precision, and recall. However, after a certain point, the model reaches a plateau, indicating the challenges associated with training ViT on datasets that are not sufficiently extensive.

Vision Transformers, by their design, are data-intensive entities. Their success stories, as reported in literature, often depend on substantial datasets. When confronted with datasets that are comparatively smaller, even if they may be considered large by general standards, the capacity of the ViT frequently exceeds the available data. This mismatch can result in suboptimal convergence or even overfitting. Additionally, ViTs, with their attention mechanisms, are computationally demanding. The training process can be protracted and necessitates robust hardware infrastructure.

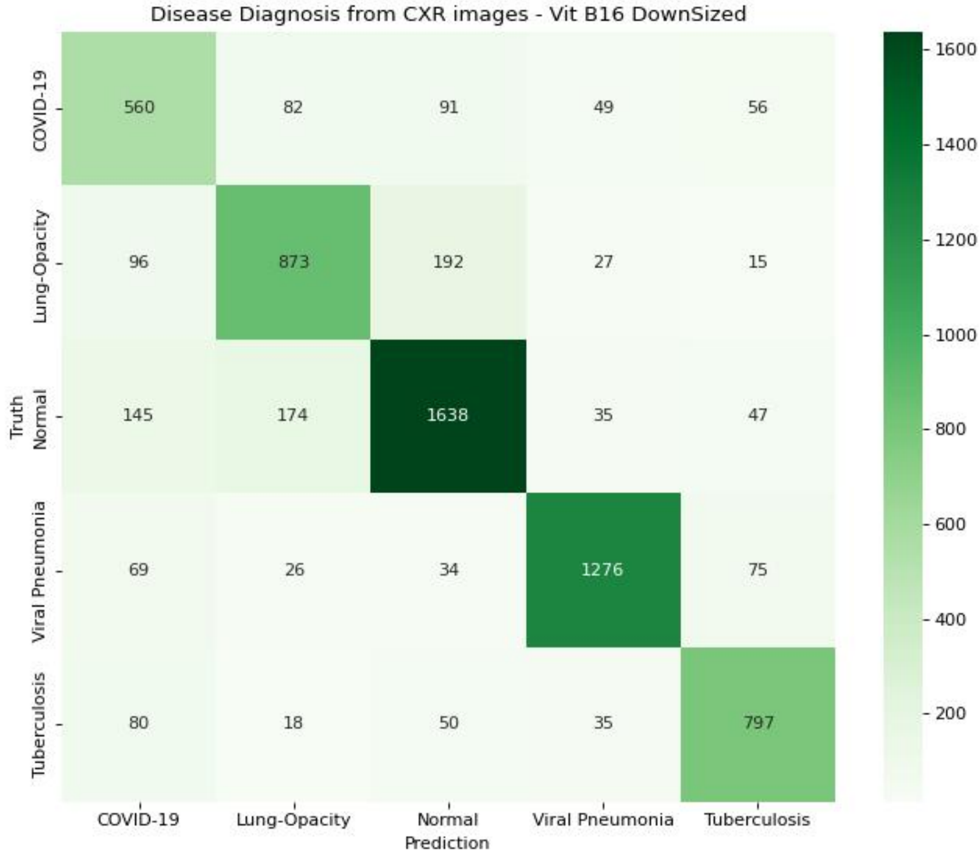

Figure 4.18 Vision transformer confusion matrix heatmap.

Confusion Matrix Heatmap: Figure 4.18 presents the confusion matrix for the Vision Transformer model's performance in chest X-ray image classifications. A thorough analysis of this matrix and the accompanying observations unveils several key insights:

Comparative Performance: The Vision Transformer (ViT) exhibits inferior performance compared to conventional convolutional models such as VGG19, ResNet50, and a Custom CNN. As previously mentioned, ViTs require substantial computational resources and, consequently, often necessitate extensive datasets to attain their purported efficacy. The limited availability of chest X-ray datasets, particularly those of sufficient size and diversity, clearly impedes the ViT's potential in this domain.

Classification Report

	Precision	Recall	F1-Score	Support
0	0.59	0.67	0.63	838
1	0.74	0.73	0.73	1203
2	0.82	0.80	0.81	2039
3	0.90	0.86	0.88	1480
4	0.81	0.81	0.81	980
----------	--------	-----	-----	-----
accuracy			0.79	6540
macro avg	0.77	0.77	0.77	6540
weighted avg	0.79	0.79	0.79	6540

Figure 4.19 Vision Transformer classification report.

Classification Report Analysis: Figure 4.19 presents the performance metrics of the Vision Transformer (ViT) model in the classification of chest X-ray images. A comprehensive analysis of the classification report reveals several salient findings.

Overall Performance: Upon a superficial examination, an accuracy of 0.79 may seem reasonably satisfactory. However, when compared against the performance of previous models such as VGG19, ResNet50, and the custom CNN, the relative ineffectiveness of the ViT becomes apparent.

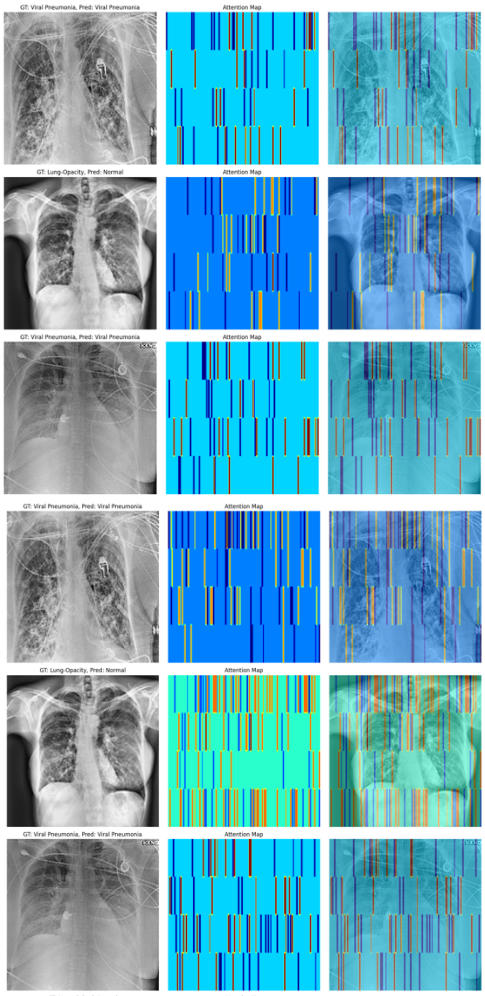

Figure 4.20 Vision transformer, block 1 attention map , block 5 attention map.

Horizontal Lines in Attention Map: The visualization of horizontal lines in the attention maps indicates an anomaly. Ideally, attention maps should highlight regions in the input image that the model perceives as most informative. In the context of chest X-ray images, we anticipate areas of disease manifestation, edges of lungs, or other clinically relevant regions to be illuminated. However, horizontal lines do not align with any natural or clinically relevant patterns in such images.

4.4 Summary

In this chapter we discussed the implementation and results of our study. The implementation section covers the environment setup, including hardware and software components, Python libraries, and image processing techniques utilized in our work. Following this, we explained the testing and evaluation process, emphasizing our model's performance and comparing it with other related works.

We trained four different models and performed a comparative analysis to determine which model performed the best. In the results and discussion section, we provided a detailed examination of each model's accuracy, predictions, and confusion matrices, highlighting the strengths and weaknesses of each. This chapter not only showcased the results but also critically analyzed the outcomes, providing a clear understanding of the model's effectiveness in achieving the study's objectives.

Chapter 5

Engineering Standards and Design Challenges

This chapter explores the engineering standards and challenges encountered during the project, focusing on compliance with industry norms, societal and environmental impact, and complex problem-solving. It also delves into project management aspects and provides a summary of the findings.

5.1 Compliance with the Standards

5.1.1 Communication Standards

Selected Standard: HL7 (Health Level Seven) Standards

Alternatives:

DICOM (Digital Imaging and Communications in Medicine):

- Pros: Industry standard for storing and transmitting medical imaging data; supports interoperability.
- Cons: Primarily designed for image storage and transfer; does not address broader data communication needs.

FHIR (Fast Healthcare Interoperability Resources):

- Pros: Modern standard for exchanging healthcare data; supports web-based communications.
- Cons: May require additional adaptations for integration with imaging systems.

5.2 Impact on Society, Environment and Sustainability

The research significantly impacts society by enhancing diagnostic accuracy for infectious diseases, leading to improved patient outcomes and reduced healthcare burdens. The adoption of automated medical imaging systems fosters equitable access to diagnostic tools, particularly in resource-constrained regions. Environmentally, the reliance on digital processing reduces the need for physical resources, such as film-based radiology. The integration of sustainable practices, including energy-efficient computational resources and cloud-based solutions, minimizes environmental footprints. Moreover, the project emphasizes long-term sustainability through robust and interpretable AI models, aligning technological advancements with ethical and environmental considerations for widespread, sustainable healthcare solutions.

5.2.1 Impact on Life

This research contributes significantly to improving the quality of life by enhancing the accuracy and speed of diagnosing infectious diseases like COVID-19, pneumonia, and tuberculosis. Automated medical imaging systems reduce the reliance on manual diagnostics, minimizing errors and ensuring consistent results. Early and accurate detection leads to timely medical interventions, which can lower morbidity and mortality rates, particularly for respiratory diseases that require prompt attention. Additionally, the integration of AI tools into clinical workflows reduces the workload on healthcare professionals, allowing them to focus on patient care. This project also benefits patients in remote or underserved regions by providing diagnostic capabilities in areas where specialized medical expertise may not be available. Overall, the research enhances healthcare delivery, ensuring better outcomes, reduced diagnostic disparities, and improved access to quality care.

5.2.2 Impact on Society & Environment

The implementation of AI-based diagnostic systems addresses healthcare disparities by providing affordable and scalable solutions, especially in low-resource settings. By automating disease detection, the project promotes equity in healthcare access, empowering communities with limited medical infrastructure. Environmentally, the shift from traditional film-based radiology to digital diagnostics reduces waste and resource consumption, making healthcare practices more sustainable. Furthermore, energy-efficient computation techniques and cloud-based processing minimize the environmental impact of deploying AI in healthcare. The project's emphasis on inclusive and environmentally friendly practices ensures a positive societal impact, supporting both equitable healthcare delivery and sustainable development goals.

5.2.3 Ethical Aspects

Ethics play a central role in this research, ensuring the models developed are transparent, fair, and accountable. The use of explainability techniques, such as Grad-CAM and Guided Grad-CAM, ensures that AI predictions can be interpreted and trusted by medical professionals. The research also tackles biases in training datasets to ensure equitable diagnostic performance across diverse populations. Privacy and data security are prioritized, aligning with healthcare regulations like GDPR and HIPAA. Furthermore, ethical AI practices are integrated to prevent over-reliance on automation by ensuring AI tools complement, rather than replace, medical expertise. This approach builds trust among stakeholders and ensures ethical deployment in clinical settings.

5.2.4 Sustainability Plan

The sustainability plan focuses on the long-term viability and adaptability of the proposed AI models. By incorporating energy-efficient computational frameworks and cloud-based solutions, the research minimizes its environmental footprint. Regular updates to training datasets and algorithms ensure the models remain relevant to evolving clinical needs and diverse demographics. Additionally, the project emphasizes collaborative efforts to deploy these solutions in low-resource settings, ensuring that the benefits are globally accessible. By balancing technological innovation with environmental and societal considerations, the sustainability plan ensures the research supports ongoing improvements in healthcare without compromising future resources.

5.3 Project Management and Financial Analysis

Project Management: The project followed a structured and iterative approach to manage tasks efficiently. Key stages included data collection and preprocessing, model development and training, testing and evaluation, and final documentation. Agile methodologies were employed, allowing for flexibility and adaptability at each stage. Weekly team meetings facilitated effective communication, progress tracking, and timely issue resolution. Task allocation was based on individual expertise, with Team Member 1 focusing on data handling and model training, while Team Member 2 contributed to literature review, report writing, and training one model.

Financial Analysis: The financial aspect of this project was carefully managed to balance quality with cost-efficiency. Key financial considerations included expenses for computational resources, data acquisition, and tools required for model development and evaluation.

Analysis of Costs: Development tools, storage solutions, and computational resources are the main components of the project. A breakdown of the projected budget is shown below:

The primary budget:

Table 5.1: Financial analysis

Category	Description	Cost
Computational Resources	Google Colab Pro Subscription ($10/month for 6 months)	$60
Storage Options	Google Drive (1 TB storage plan) for global server models and client data	$10
Cost	Total cost of computational resources and storage	$60
Total Primary Budget	Total budget for the project	$130

5.4 Complex Engineering Problem

The project faced challenges with imbalanced datasets and generalization across diverse conditions. Advanced data preprocessing techniques and the development of multiple architectures were employed to address these challenges.

5.4.1 Complex Problem Solving

Table 5.2: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP 3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
√	√	√				

EP1: Depth of Knowledge

Details: The project demonstrates a strong foundation in advanced deep learning techniques, specifically Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and custom architectures. It applies domain-specific preprocessing (e.g., CLAHE, augmentation) and interpretability methods like Grad-CAM. This knowledge extends to handling imbalanced datasets, enhancing model generalization, and ensuring interpretability for medical imaging applications.
Justification: Our thesis reflects deep expertise in AI for medical imaging, integrating advanced methods for disease classification, data processing, and model transparency, contributing to state-of-the-art solutions.

EP2: Range of Conflicting Requirements

Details: The project addresses several conflicting requirements, such as:
Achieving high model accuracy while maintaining computational efficiency.
Balancing interpretability (e.g., Grad-CAM) with model complexity.
Ensuring generalizability across diverse datasets while managing class imbalances.
Justification: These conflicting demands were resolved through strategic techniques like class weights, lightweight architectures (e.g., VGG19), and using ViTs for global feature extraction, demonstrating careful consideration of trade-offs.

EP3: Depth of Analysis

Details: Comprehensive analysis was conducted across multiple models (e.g., CNNs, ViTs), evaluating performance metrics such as accuracy, precision, recall, and ROC-AUC.

Visualization methods like Grad-CAM provided deeper insights into the models' decision-making processes.

Justification: This in-depth evaluation ensured reliable conclusions about model strengths and weaknesses, while addressing challenges like dataset imbalance and class similarity (e.g., 'Normal' vs. 'Lung Opacity').

Mapping with Knowledge Profile for EP1

Table 5.3: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
√				√

K3 Engineering Fundamentals

Definition: This refers to a comprehensive understanding of the core principles and theories underlying engineering, such as mathematics, physics, and computing.

Justification:

Our thesis integrates engineering fundamentals in machine learning, including deep learning architectures (CNNs, Vision Transformers). These models leverage mathematics (linear algebra, calculus) for convolution operations, backpropagation, and optimization techniques (Adam optimizer).

Image preprocessing techniques like CLAHE, sharpening, and zero-centering employ computational techniques rooted in engineering principles to enhance the data.

Exploratory Data Analysis (EDA) incorporates statistical methods to analyze pixel intensity distributions and class imbalances, which are core engineering fundamentals.

K8 Research Literature

Definition: The ability to critically review, analyze, and apply existing literature and research to inform and guide engineering practice.

Justification:

Our thesis is heavily informed by existing research, as evident in the cited works [4][8] discussing CNNs, Vision Transformers, and pre-processing techniques like CLAHE.

Our compared our findings with prior studies, such as the application of ResNet50 to medical imaging, grounding your research in well-documented literature.

The inclusion of visualization techniques like Grad-CAM and Guided Grad-CAM reflects our analysis and application of advanced concepts discussed in recent research papers.

5.4.2 Engineering Activities

Table 5.4: Mapping with knowledge Profile.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
√	√		√	

EA1 Range of resource justification attainment: The study utilized diverse resources, including a comprehensive chest X-ray dataset with over 32,000 samples across five classes, processed using techniques like zooming and CLAHE to ensure data quality. Advanced machine learning frameworks, such as TensorFlow, and visualization tools like Grad-CAM were employed, alongside specialized methods like class weighting and learning rate scheduling to optimize model performance. Pre-trained models (VGG19, ResNet50) were leveraged for transfer learning, minimizing computational demands and training time while improving accuracy. Additionally, the development of a Custom CNN integrated innovative features like depth wise separable convolutions and spatial attention mechanisms, showcasing advanced engineering techniques.

EA2 Level of Interaction justification attainment: The study demonstrates strong interdisciplinary collaboration by integrating deep learning principles with medical imaging to address both technological innovation and healthcare challenges. The dataset was curated with input from medical professionals, ensuring clinical relevance and accuracy. Additionally, pre-trained models like VGG19 and ResNet50 were fine-tuned, effectively leveraging established architectures to enhance feature extraction capabilities and improve performance.

EA4 Consequences for Society and Environment: The study addresses significant healthcare challenges by enabling accurate and efficient diagnosis of infectious respiratory diseases, potentially saving lives and easing the burden on healthcare systems. By utilizing open-source datasets and pre-trained models, it fosters the development of affordable AI solutions, particularly benefiting under-resourced regions. Additionally, the use of transfer learning reduces computational demands and energy consumption, promoting environmental sustainability.

5.5 Summary

This thesis explores the application of deep learning models, including CNNs and Vision Transformers, for classifying infectious respiratory diseases from chest X-rays. The study tackles challenges like noisy data, class imbalance, and overlapping pixel intensities using advanced preprocessing techniques such as CLAHE, zooming, and sharpening. A Custom CNN was developed with innovative features like depthwise separable convolutions and spatial attention, alongside the fine-tuning of pre-trained models like VGG19 and ResNet50. The research demonstrates interdisciplinary collaboration by integrating engineering and medical imaging, supported by clinical inputs to ensure dataset relevance. It emphasizes societal and environmental impacts by promoting affordable AI solutions and reducing energy consumption via transfer learning. Evaluation metrics like precision, recall, and interpretability tools (Grad-CAM) ensure robust validation. By combining cutting-edge AI with medical diagnostics, the study advances model performance and transparency, addressing critical healthcare challenges while promoting innovation and sustainability in engineering solution.

Chapter 6

Conclusion

This study compared traditional CNNs (VGG19, ResNet50), Vision Transformers (ViTs), and a Custom CNN architecture for classifying infectious respiratory diseases in chest radiological images. The results highlighted the superior performance of traditional CNNs and the custom architecture over ViTs, while emphasizing the importance of interpretability through visualization techniques to bridge AI and clinical practice.

6.1 Summary

The application of deep learning models, especially Convolutional Neural Networks (CNNs) and the newly born Vision Transformers (ViTs), in medical imaging has brought about challenges and opportunities. This paper, therefore, attempts to compare the performances of the traditional CNN architectures in the likes of VGG19 and ResNet50 with the Convolutional Vision Transformers in the classification of chest radiological images for infectious respiratory diseases. This study further explored a customized convolutional network architecture that made use of spatial attention mechanisms, depth-wise convolutions, as well as employing blocks similar to VGG and ResNet50 in its architecture.

From the results, it became clear that, whereas Vision Transformers have made substantial strides in computer vision, their application to specific medical imaging tasks, given the dataset size and computational resource constraints, was suboptimal. Traditional CNN architectures, on the other hand, showed superior performance because of their spatial hierarchies and design tailored for image data. Additionally, custom CNN architecture with spatial attention layers showed benefits of model specialization for a particular task.

Through an example, the importance of model interpretability was established using gradient-based visualization methods and attention maps. Attention maps in visualization provide much more insight into the model's decision mechanisms, paving the way for better collaborations between machine learning systems and practitioners in healthcare.

6.2 Limitation

The study has several limitations that must be acknowledged. First, the dataset used, while comprehensive, suffers from class imbalance, with certain categories like 'COVID-19' and 'Tuberculosis' being under-represented compared to 'Normal.' This imbalance may affect the generalizability of the models, despite attempts to address it through class weights and augmentation. Second, the reliance on chest radiological images alone may overlook complementary diagnostic modalities, such as CT scans, which could provide additional information for improved classification accuracy.

Third, Vision Transformers (ViTs) showed suboptimal performance, likely due to the limited size of the dataset. These models typically require vast amounts of data and computational resources to achieve their full potential, which posed a constraint in this research. Fourth, the similar pixel intensity distributions of some classes, such as 'Lung Opacity' and 'Normal,' presented challenges for differentiation, reducing precision and recall for these categories.

Additionally, while interpretability techniques like Grad-CAM and Guided Grad-CAM provided insights into model decision-making, their effectiveness in clinical settings requires further validation. Lastly, the computational complexity of the models, particularly ViTs and the Custom CNN, necessitates substantial hardware resources, which may limit the scalability of the approach in resource-constrained environments. Future work should address these limitations to enhance the study's impact and applicability.

6.3 Future Work

Dataset Expansion: Given that Vision Transformers generally benefit from larger datasets, investigating their performance with an augmented and more comprehensive dataset would provide further insights into their applicability to medical imaging.
Further Refinement of Image Pre-processing Techniques: The findings underscored the substantial influence of pre-processing techniques, particularly zoom on the model's performance. Future research endeavors should investigate the combination of sophisticated pre-processing techniques to mitigate noise and artifacts.
Clinical Integration and Validation: Collaborate closely with radiologists and healthcare practitioners to establish an iterative feedback loop. This ensures that the models closely align with clinical realities and can be relied upon for real-world applications.
Improving Model Interpretability: In addition to gradient and attention visualizations, investigating other state-of-the-art interpretability techniques would enhance our

comprehension of model decision-making processes. This would foster greater trust among end-users.

Real-time Application: Future research can investigate the feasibility of integrating these models into real-time diagnostic tools for radiologists through optimization of model design and computational efficiency.

References

[1] S. M. Pizer, et al., "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[2] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, et al., "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLoS Medicine*, vol. 15, no. 11, p. e1002686, 2017.

[3] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2097–2106.

[4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[5] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, et al., "Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[9] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020.

[10] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[11] B. K. Gowru, A. R. Giduturi, and M. J. Babu, "Hierarchical Bayesian optimization based convolutional neural network for chest X-ray disease classification," *International Journal of ...*, 2025. [Online]. Available: https://search.ebscohost.com

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., “Attention is all you need,” in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[13] S. K. B. and A. C. V., "Leveraging Compact Convolutional Transformers for Enhanced COVID-19 Detection in Chest X-Rays: A Grad-CAM Visualization Approach," *Frontiers in Big Data*, 2024. [Online]. Available: https://www.frontiersin.org

[14] N. Maulana, M. F. R. Kholiq, and M. D. H. A., "Klasifikasi Citra Medis Penyakit Pneumonia dengan Convolutional Neural Network," *OKTAL: Jurnal Ilmu ...*, 2024. [Online]. Available: https://journal.mediapublikasi.id

[15] A. Velayudham, R. Karthick, and A. Sivabalan, "IoT enabled smart healthcare system for COVID-19 classification using optimized robust spatiotemporal graph convolutional networks," *Signal Processing and Internet of Things*, 2025. [Online]. Available: https://www.sciencedirect.com

[16] R. Bagri and A. Shah, "Predictive Diagnosis of Lung Diseases Using Artificial Intelligence," *EC Pulmonology and Respiratory Medicine*, 2024. [Online]. Available: https://ecronicon.net

[17] R. Bagri and A. Shah, "Predictive Diagnosis of Lung Diseases Using Artificial Intelligence," *EC Pulmonology and Respiratory Medicine*, 2024. [Online]. Available: https://ecronicon.net

[18] G. Anita and S. Singarapu, "Automated Detection and Classification of Pneumonia using Deep Learning and CNN," *Journal of Intelligent Systems and Internet Computing*, 2025. [Online]. Available: https://americaspg.com

[19] V. Asha, J. C. Mathew, T. T. Rao, and S. Mondal, "Enhanced ResNet-101 Model for Effective Classification of COVID-19 Disease from CXR Images," *Proceedings of IEEE International Conference on Self-Adaptive and Self-Organizing Systems*, 2024. [Online]. Available: https://ieeexplore.ieee.org

[20] M. Patankar, V. Chaurasia, and M. Shandilya, "A novel dense-net deep neural network with enhanced feature selection for TB classification," *Multimedia Tools and Applications*, 2024. [Online]. Available: https://link.springer.com

[21] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 3–19.

[22] N. O. Adiwijaya, B. S. Miranda, and D. I. Swasono, "CNN-Based Classification of Infectious Lung Diseases using Thorax X-Ray Analysis," *ResearchGate Publications*, 2024. [Online]. Available: https://www.researchgate.net

[23] Y. H. Bhosale, S. Zanwar, Z. Ahmed, M. Nakrani, D. Bhuyar, and U. Shinde, “Deep Convolutional Neural Network Based Covid-19 Classification From Radiology X-Ray Images For IoT Enabled Devices,” *IEEE Xplore*, Mar. 01, 2022. https://ieeexplore.ieee.org/document/9785113 (accessed Aug. 08, 2023).

[24] A. Elagili Elagili, I. Ighneiwa Ighneiwa, and Z. Rajab Rajab, “Deep Learning to Improve COVID-19 Detection: Using CNN-Based Transfer Learning from Chest X-Ray Images,” *The 7th International Conference on Engineering & MIS 2021*, Oct. 2021, doi: https://doi.org/10.1145/3492547.3492616.

[25] N. Asadoorian, S. Yaraghi, and A. Tahmasian, "Pre-trained Quantum CNN for COVID-19 Disease Classification Using CT Images," *PeerJ Computer Science*, 2024. [Online]. Available: https://peerj.com

[26] Ç. Bakır and M. Babalık, "Tüberküloz Hastalığının Tespiti için Derin Öğrenme Yöntemlerinin Karşılaştırılması," *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Dergisi*, 2024. [Online]. Available: https://dergipark.org.tr

[27] O. Olabode, K. G. Akintola, and E. Adeyemi, "Deep Learning Models for Classification of COVID-19 Severity Levels," *Journal of Information Technology Innovations*, 2024. [Online]. Available: https://gexinonline.com

[28] A. Basu, S. Das, S. Ghosh, S. Mullick, A. Gupta, and S. Das, “Chest X-Ray Dataset for Respiratory Disease Classification,” *Harvard Dataverse*, Jan. 2021, doi: https://doi.org/10.7910/dvn/wnq3gi

[29] World Health Organization, “The top 10 causes of death,” *World Health Organization*, Aug. 07, 2024. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

Comperative Analysis of Different Deep Learning Models to Classify Infectious Diseases from Chest Radiological Images

ORIGINALITY REPORT

24%	18%	16%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
2	O. Olabode, K. G. Akintola, and E. Adeyemi. "Deep Learning Models for Classification of COVID-19 Severity Levels", Journal of Information Technology Innovations, 2024 Publication	3%
3	Submitted to Daffodil International University Student Paper	2%
4	Submitted to United International University Student Paper	1%
5	www.mdpi.com Internet Source	1%
6	Abdelaziz Testas. "Distributed Machine Learning with PySpark", Springer Science and Business Media LLC, 2023 Publication	1%