PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David
SWANSEA - ABERTAWE

School of
Applied Computing

IMPROVING MODEL GENERALIZATION OF PNEUMONIA DETECTION FROM CHEST XRAY IMAGES USING DEEP LEARNING AND TRANSFER LEARNING

AKHIL KOCHAMPARAMPIL ANIL

Supervisor Dr. Seena Joseph

September 2025

STUDENT NO: 2409841

COURSE: MSc Software Engineering and Artificial Intelligence

Declaration of Originality

I, Akhil Kochamparampil Anil, hereby certify that I am the sole author of this project. All sources of information have been properly acknowledged, and the work presented is entirely my own. Furthermore, I confirm that the completed dissertation adheres to the prescribed word limit.

This work has not been submitted previously in fulfilment of the requirements for any other degree or qualification.

Signed:...

10-09-2025

Date: ....

CONSENT FORM

I, Akhil Kochamparampil Anil, authorize that my MSc Project in Software Engineering and Artificial Intelligence may be made available for inter-library loan and photocopying (in compliance with copyright regulations), and that the title and abstract may be shared with external organizations.

Signed:..

10-09-2025

Date: ....

Acknowledgement of Copyright

I confirm that the copyright of this project report, as well as any outputs produced during the course of the project, is vested in the University of Wales Trinity Saint David, Swansea.

ABSTRACT

Pneumonia continues to be a major global health threat and one of the leading causes of mortality, with chest X-rays serving as the primary diagnostic tool. Despite their promise, deep learning models for pneumonia detection often face limitations in generalization, performing strongly on familiar datasets but losing accuracy when applied to unseen data from different clinical environments. This study addresses that challenge by applying transfer learning with ResNet152 and DenseNet201 to improve generalization in pneumonia detection using chest X-ray images. Publicly available datasets, including the COVID-19 Radiography Database and the Chest X-ray (COVID-19 & Pneumonia) dataset, were merged, pre-processed, and balanced prior to training. Model optimization involved freezing and unfreezing layers, adjusting learning rates, and applying various data augmentation strategies. Performance was evaluated with metrics such as accuracy, precision, recall, F1-score, error rate, and AUC-ROC, and further validated using an independent external dataset. Experimental results showed that DenseNet201 consistently outperformed ResNet152, reaching 93.2% accuracy with an AUC of 0.9916 on the main dataset and 88.3% accuracy with an AUC of 0.9742 on the external dataset. By comparison, ResNet152 achieved 86.4% accuracy (AUC 0.9665) on the main dataset and 84.6% accuracy (AUC 0.9573) on the external dataset. Confusion matrix analysis revealed that most errors occurred between Normal and Pneumonia classes, mirroring real-world diagnostic challenges. In conclusion, DenseNet201 demonstrated superior robustness and generalization compared to ResNet152, highlighting its suitability for clinical application and underscoring the necessity of external validation in medical AI research.

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I, Akhil Kochamparampil Anil, would like to sincerely thank my supervisor, Dr. Seena Joseph of the University of Wales Trinity Saint David, for her outstanding guidance, encouragement, and support throughout the course of this project. Her expertise in deep learning, academic writing, and critical evaluation played a central role in shaping this research. The constructive feedback and continuous motivation she provided were instrumental in refining my ideas, strengthening the methodology, and bringing this work to completion.

My heartfelt gratitude also goes to my family for their unwavering support, patience, and encouragement, which have been a constant source of strength during this journey. Their belief in me and continued reassurance enabled me to stay committed and overcome the challenges of this demanding study.

Finally, I am grateful to the University of Wales Trinity Saint David for providing access to its digital library and online resources. The scholarly articles available through these platforms offered a strong theoretical foundation that informed my exploration of pneumonia detection and model generalization.

1. CHAPTER 1 – INTRODUCTION

1.1 RESEARCH BACKGROUND

Pneumonia is a prevalent and often life-threatening respiratory illness that contributes to high rates of concerning disease and fatality, the condition disproportionately affects children, older adults, and individuals with weakened immune systems, while chest radiography remains a primary diagnostic tool one of the most widely used diagnostic techniques due to its accessibility and effectiveness in revealing lung abnormalities. Nonetheless, interpreting these X-rays manually can be time-consuming, subject to variability between radiologists, and dependent on specialized expertise that is not always available in resource-limited settings. These limitations have spurred increasing interest in applying artificial intelligence (AI), and deep learning in particular, to automate the classification of medical images and support clinicians in achieving earlier and more accurate diagnoses.[1, 2]. Chest X-rays and other medical images play a vital role in diagnosing respiratory illnesses, as they offer detailed insights into lung anatomy. Their widespread application in clinical practice makes them a suitable focus for deep learning models aimed at enhancing both diagnostic precision and efficiency.

Deep learning techniques, especially Convolutional Neural Networks (CNNs), have shown outstanding performance in image classification applications, such as identifying pneumonia from chest X-ray image [3]. Convolutional Neural Networks (CNNs) can automatically learn and extract features from images, making them highly effective at identifying complex patterns that might be missed by the human eye. Despite these advances, a major challenge in medical AI is the lack of generalization. This means a model trained on one dataset may not perform well on data from a different source. This is a significant problem in clinical practice, where image data can vary widely between hospitals, different types of equipment, and diverse patient populations.

One approach to overcome this challenge is transfer learning [4]. This technique utilizes models that have been pre-trained on large, general datasets and then adapts them to domain-specific tasks, such as chest X-ray analysis. Architectures like ResNet and DenseNet are among the most commonly applied for transfer learning in medical imaging[4, 5]. However, even with transfer learning, many studies [6-8][1-3, 6-8] evaluate their models only on the training dataset, without testing generalization across external datasets [5].

1.1.1 Deep Learning in Medical Imaging

Medical imaging has been transformed by deep learning, enabling precise and automated interpretation of complex visual information. Convolutional Neural Networks (CNNs) form the foundation of many diagnostic applications, particularly within radiology. By emulating the human visual system, these models are capable of learning layered features—from simple edges and textures to more complex structures such as organs and pathological lesions. In clinical practice, deep learning has been effectively applied to identify conditions like pneumonia, tuberculosis, lung cancer, and COVID-19 across different imaging techniques, including X-rays, CT scans, and MRI [7].

A key strength of deep learning in medical imaging is its capacity to analyse large volumes of data rapidly and consistently, while avoiding fatigue or the subjective bias that can affect human interpretation [5]. CNNs can assist radiologists by highlighting abnormal regions and supporting early diagnosis, especially in overburdened healthcare systems. Additionally, deep learning systems are scalable and can be deployed in rural or underserved areas where specialist expertise is limited. Generalization is essential to guarantee that deep neural networks maintain reliable performance on previously unseen data. Transfer learning supports this goal by utilizing pre-trained knowledge and adapting it to specialized medical imaging tasks, particularly when only limited data is available.

To improve generalization, researchers [1-3] have adopted transfer learning involves adapting a model that has been leveraging prior training on large benchmark datasets such as ImageNet, by fine-tuning it for a targeted medical imaging application. This helps leverage learned features and improves training efficiency, particularly in domains with limited labelled data. ResNet and DenseNet are widely adopted architectures in transfer learning due to their depth, strong performance, and capacity to reduce overfitting. They have become central to developing reliable medical diagnostic systems, with transfer learning playing a key role in enabling adaptability across different datasets. In medical imaging, this approach allows models trained on large-scale datasets to be fine-tuned for specialized diagnostic tasks, enhancing both accuracy and efficiency, even when only limited medical data is available [8].

1.1.2 Transfer Learning

Transfer learning is an approach used to boost model performance, particularly in situations where labelled medical data is scarce. It involves using a model that has been pre-trained on a huge dataset, such as ImageNet, and fine-tuning it for a new application, like identifying pneumonia in chest X-ray images. By fine-tuning the model, better results can be produced even with smaller datasets[1, 3, 4, 9, 10].

Models like ResNet and DenseNet are often used in transfer learning because they're great at extracting useful features from images [4, 5, 10].

Even with transfer learning, many studies [4, 5, 10] still train and test their models on the same dataset. This can create illusion of high performance, but it often doesn't hold up in real-world settings where data can vary significantly in factors like image quality and patient demographics. For example, a model trained on two datasets, such as Covid19 radiography dataset, and chest Xray covid19 pneumonia might not perform well when applied to another, like the Chest Xray pa dataset. [11, 12]. This research aims to tackle that problem by improving how well models generalize across different datasets.

1.1.3 Role of Transfer Learning in Medical Imaging

Transfer learning has emerged as a highly effective approach in medical imaging, particularly in cases where annotated datasets are scarce or costly to obtain. The approach relies on models previously trained using large-scale datasets, such as ImageNet and tailoring them to specialized medical imaging tasks, including disease detection from chest X-ray scan[4, 5]. This approach allows models to transfer previously acquired fundamental features, including contours and textures, while fine-tuning higher-level features that are specific to medical applications. In areas like pneumonia, COVID-19, or tumour detection, transfer learning provides substantial benefits, enhancing accuracy, training efficiency, and convergence speed compared to building models from the ground up. Popular architectures such as ResNet, DenseNet, and VGG are frequently applied and fine-tuned for tasks including classification, segmentation, and anomaly detection across imaging modalities like chest X-rays, CT scans, and MRIs[6, 12]. The success of transfer learning in these applications demonstrates its potential to build accurate and generalizable diagnostic tools, especially in environments where medical data is limited.

1.2 RESEARCH PROBLEM STATEMENT

While deep learning has significantly improved pneumonia detection using chest X-rays, many existing studies [5, 10–14] focus solely on evaluating model performance on the same dataset used for training. Even studies that apply transfer learning [4, 5, 14] often evaluate their models on only one dataset. This leads to high accuracy in controlled environments but poor performance when the models are applied to real-world scenarios, where image data can vary significantly due to differences in patient demographics, imaging equipment, and clinical settings. This lack of generalization limits the practical and clinical applicability of AI-based diagnostic systems.

1.3 RESEARCH AIM AND OBJECTIVES

The central goal of this investigation is to enhance generalization in pneumonia detection by building a transfer learning–based deep learning model capable of accurately identifying pneumonia from chest X-ray images.

To achieve this aim following objectives are identified

1. To review and analyse existing literature to identify deep learning and transfer learning models used for medical image disease detection.

2. To Select and train two deep learning models using multiple datasets and apply transfer learning techniques for diagnosing pneumonia through chest X-ray analysis.

3. To assess the effectiveness of the trained models using benchmarking dataset and well-known evaluation metric.

.

1.4 CONTRIBUTION OF THE STUDY

This study contributes to the field of AI in medical imaging by evaluating and comparing the generalization performance of widely used Convolutional Neural Network (CNN) architectures, namely ResNet and DenseNet, across multiple chest X-ray datasets. It explores how deep learning combined with transfer learning can help overcome the frequent issue of limited generalization when models are applied to data originating from diverse clinical environments.

Additionally, the study provides a practical framework for evaluating cross-dataset generalization in pneumonia detection. This framework can be used and expanded upon in future research to develop more robust and clinically reliable diagnostic tools.

1.5 STRUCTURE OF THE STUDY

This research is structured into six main sections. The first part focuses on understanding the theoretical foundations of deep learning, transfer learning, and their applications in medical imaging, with an emphasis on pneumonia detection from chest X-rays. It also involves reviewing existing studies on convolutional neural networks (CNNs), challenges in achieving model generalization, and evaluation across diverse datasets.

Chapter Two delivers a detailed review of the literature on deep learning and transfer learning in medical imaging. It examines prior approaches for pneumonia detection using chest X-rays, highlights gaps in model generalization research, and establishes the rationale for selecting ResNet152 and DenseNet201.

Chapter Three describes the research methodology adopted in this study. It outlines the datasets used, preprocessing steps, model architectures, training procedures, and evaluation metrics. This methodological framework is designed to ensure reproducibility and support the study's objective of improving generalization.

Chapter Four presents the experimental setup and the results obtained from training and evaluating the models. It reports key performance metrics, confusion matrices, and ROC analyses for both the main and external datasets, followed by a discussion comparing the strengths of DenseNet201 and ResNet152.

Chapter Five summarizes the conclusions within this work, addressing its limitations and suggesting directions for future research. It emphasizes the superior generalization performance of DenseNet201 over ResNet152 while also identifying areas for further improvement.

Finally, Chapter Six provides a reflective account of the research process. It discusses challenges encountered, skills acquired, and lessons learned, while also highlighting the importance of effective project planning, supervision, and self-assessment in successfully completing the project.

2. CHAPTER 2 – REVIEW OF LITERATURE

This chapter reviews prior research and recent developments in applying employing deep learning techniques to identify pneumonia in chest X-ray scans. Existing approaches can be broadly categorized into traditional machine learning methods and modern deep learning approaches. Convolutional Neural Networks (CNNs) are widely regarded as the dominant approach due to their high accuracy and strong capability in extracting complex patterns from medical images. The discussion also examines significant studies and widely used architectures, such as ResNet and DenseNet, which are frequently applied with transfer learning for medical image classification. In addition, the chapter addresses the challenge of achieving model generalization in clinical practice and surveys strategies aimed at improving performance across diverse datasets. Particular emphasis is placed on the role of transfer learning in enhancing cross-dataset generalization and its significance in building reliable diagnostic systems for real-world use.

2.1 PNEUMONIA AND CHEST X-RAY IMAGING

Pneumonia is a respiratory illness that causes inflammation of the alveoli within one or both lungs. It poses serious health risks and can become life-threatening, particularly for vulnerable groups such as paediatric groups, aging populations, and immunocompromised individuals. As a significant global health challenge, timely and accurate diagnosis is essential for effective treatment. Chest X-ray imaging remains the most widely used method for diagnosing pneumonia, as it is affordable, widely available, and effective in revealing lung abnormalities[13]. Radiologists analyse chest X-rays by looking for indicators such as lung consolidation, opacities, or fluid buildup to diagnose pneumonia. However, this interpretation process can be slow, demands specialized knowledge, and is susceptible to human error and differences in clinical judgment[1]. These limitations have led to increasing interest in automated methods, particularly those powered by artificial intelligence, to assist or augment pneumonia diagnosis from chest X-rays.

2.2 DEEP LEARNING-BASED DISEASE DETECTION

Deep learning, especially through Convolutional Neural Networks (CNNs), has emerged as a key approach for automating disease detection in medical imaging. These networks are capable of learning intricate patterns directly from raw data, enabling them to identify abnormalities that may not always be consistently detected by human experts. CNNs excel at capturing spatial features such as edges, textures, and complex structures, which are critical for identifying disease-related visual markers. They have shown strong success in diagnosing conditions including pneumonia, lung cancer, tuberculosis, and COVID-19 using imaging modalities like chest X-rays and CT scans[14].

One main advantage of deep learning models in disease detection is their ability to process large volumes of medical data with speed and precision[14]. In clinical practice, these models can support radiologists by highlighting abnormal cases for closer examination, helping to speed up diagnosis and improve patient care. Additionally, the scalability of deep learning enables its use in resource-constrained environments where access to skilled medical professionals is limited. Pre-trained models, especially when refined through transfer learning, further boost efficiency by minimizing the time and computational resources required to build accurate diagnostic systems from the ground up[14-16].

Despite their advantages, applying deep learning models in real-world healthcare still presents several challenges. A major concern is limited generalization, as models trained on one dataset often perform poorly when tested on data from other hospitals or imaging equipment. This issue arises from differences in image quality, patient populations, and imaging protocols.

2.3 DEEP LEARNING BASED PNEUMONIA DETECTION

In recent years, deep learning has become a prominent approach for pneumonia detection, especially through the analysis of chest X-ray images. These methods have been widely applied and studied [17] applied to automate the detection process, delivering better results than conventional machine learning methods. For example, Manoj Kumar et al. carried out a study on pneumonia detection and classification from chest X-ray images using a DenseNet121-based deep learning model[17]. The main goal of the study was to develop an automated system for the diagnosis of pneumonia using chest X-ray scans. The researchers utilized a publicly available dataset containing 5,863 X-ray scans. Their model achieved 90% accuracy when assessed on the test dataset, the model achieved 89% sensitivity demonstrating a specificity of 90%. The method successfully combined transfer learning, data augmentation, and fine-tuning strategies to improve diagnostic accuracy and support fast, automated preliminary screening of pneumonia.

Sanskruti Gaurkhede et al.[18] a study was reviewed that focused on classifying respiratory diseases—namely pneumonia, COVID-19, and tuberculosis—using deep learning approaches. The research investigated the effectiveness of advanced neural network deep learning architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks

(RNNs), in capturing complex patterns from medical images, particularly CT scans. Several deep learning models were tested across different datasets, with ResNet-18 identified as the best-performing architecture for accurately classifying CT images of the three conditions. The study also addressed challenges such as class imbalance, limited data availability, and model interpretability, while highlighting the benefits of incorporating multimodal data to enhance diagnostic accuracy. Overall, the findings indicated that well-trained deep learning models can support radiologists in achieving earlier and more reliable diagnoses, ultimately aiding more effective clinical decision-making.

Vivek et al.[19] A study introduced a novel deep learning method pointed at enhancing pneumonia detection from chest X-ray images. They developed a custom Convolutional Neural Network (CNN) architecture, refined through extensive hyperparameter optimization and data augmentation to improve accuracy and robustness. The model was trained and validated on a large labelled chest X-ray dataset, achieving better performance than traditional diagnostic approaches. Evaluation metrics—including accuracy, sensitivity, and specificity—showed significant gains. The work also examined how transfer learning and ensemble strategies could further boost model effectiveness. Overall, the findings emphasize the potential of deep learning to automate pneumonia detection, providing a dependable and efficient tool for clinical applications.

Anitha et al.[20] a study was carried out on pneumonia detection using chest radiographs, where transfer learning was applied for feature extraction and machine learning algorithms were used for classification. Pre-trained deep learning models helped capture relevant patterns from the X-ray images, which were then classified using different machine learning techniques. The approach was tested on a benchmark dataset and achieved high classification accuracy. The findings demonstrated that integrating transfer learning with conventional machine learning methods is an effective strategy for pneumonia detection, enhancing diagnostic accuracy and aiding better healthcare decision-making.

Wang et al.[21]introduced the ChestX-ray14 dataset and demonstrated the potential of CNNs in multi-label classification of thoracic diseases, including pneumonia. Similarly, Rajpurkar et al.[22] developed the CheXNet model, a 121-layer DenseNet trained on ChestX-ray14, and reported performance comparable to radiologists for pneumonia detection[22] these studies demonstrate the ability of deep learning models to automatically extract complex features from raw image data, eliminating the need for manual feature engineering.

Despite these advances, a key challenge in current deep learning-based approaches lies in their limited generalizability. Many models are evaluated only on the dataset they were trained on, which may not reflect real-world scenarios where data can vary significantly between healthcare settings. As a result, their accuracy often decreases when exposed to new datasets. To address this, researchers have begun exploring transfer learning techniques, where models pretrained on large datasets such as ImageNet are fine-tuned for pneumonia detection using domain-specific datasets like RSNA and CheXpert [16, 22, 23]. While these approaches have improved performance in small or imbalanced datasets, the problem of cross-dataset generalization remains underexplored. Therefore, more research is needed to develop robust models that perform reliably across varied clinical environments.

2.4 TRANSFER LEARNING IN DISEASE DETECTION

Transfer learning has emerged as an essential method in medical imaging, especially when labeled data is scarce. It works by adapting models that were pre-trained on large, general-purpose datasets like ImageNet and fine-tuning them for specialized tasks, such as detecting diseases in medical images. This strategy minimizes the demand for extensive medical datasets and heavy computational resources while enhancing performance. Research has shown that transfer learning helps models achieve faster convergence and stronger generalization in cases where annotated medical data is limited or unevenly distributed [24, 25].

Kakkar et al. carried out a study focused on improving the classification of retinal diseases using Optical Coherence Tomography (OCT) images. The research targeted the detection of conditions such as Age-related Macular Degeneration (AMD), Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), as well as normal cases. To address the shortcomings of traditional non-deep learning methods, which were often affected by noise and inefficiency, the authors employed a transfer learning approach with pre-trained deep learning models. Their method achieved notable improvements in classification accuracy and robustness, demonstrating that transfer learning offers a more dependable and effective solution for OCT-based retinal disease diagnosis [26].

Eshika Jain et al. [26] conducted a study was reviewed that focused on the urgent need for early, non-invasive detection of colon cancer, which remains one of the primary contributors to cancer-associated deaths globally. The researchers proposed an advanced deep learning framework built on the EfficientNetB3 architecture to classify histopathological images of colon cancer with high accuracy. By applying transfer learning alongside advanced preprocessing methods, the model demonstrated an outstanding accuracy of 98.86%, with strong precision and recall, thereby reducing both false positives and false negatives. To further improve robustness and generalization, techniques such as data augmentation, batch normalization, and the Adamax optimizer were employed. The findings demonstrated the framework's effectiveness in distinguishing adenocarcinoma from normal tissue. This work underscores the potential of EfficientNetB3 as a scalable, lightweight, and dependable AI-driven diagnostic tool, supporting early detection and enhancing clinical outcomes in colon cancer diagnosis.

Gopi et al. [27] Conducted a study was carried out on the detection of rice leaf diseases using deep learning and transfer learning methods to support early diagnosis and reduce the risk of crop loss. The research focused on identifying common bacterial, viral, and fungal infections that severely impact rice yield. The performance of well-known transfer learning models was tested using two strategies: frozen layer feature extraction and fine-tuning. Results showed that DenseNet169 with frozen layers achieved a testing accuracy of 99.66%, while Xception, when fine-tuned, surpassed all other models with a testing accuracy of 99.99%. These findings confirmed that transfer learning, particularly fine-tuning, is highly effective for classifying rice leaf diseases, even when faced with challenges such as varying backgrounds and inconsistent image capture conditions.

Despite its advantages, transfer learning also faces challenges when applied to disease detection. A major issue is the domain shift between natural image datasets (e.g., ImageNet) and medical datasets. Features learned from non-medical images may not fully capture the complexity of pathological patterns in radiological images. Therefore, while transfer learning improves performance, it may not be optimal without careful tuning and adaptation. Researchers have proposed various solutions such as domain-specific pretraining, additional fine-tuning layers, and using medical-specific pretrained models to mitigate this gap [28, 29].

Furthermore, the use of transfer learning has opened the door to exploring model generalization across different datasets. For instance, models trained on one dataset (e.g., ChestX-ray14) can be evaluated on others like RSNA or CheXpert to assess their robustness and cross-domain performance. This is crucial in real-world clinical applications, where data can vary significantly between hospitals or imaging devices. Improving generalization using transfer learning aligns with the growing need for reliable and scalable diagnostic tools in medical imaging, especially for diseases like pneumonia, tuberculosis, and COVID-19 [14, 30].

2.4.1 Transfer Learning models

Transfer learning plays a crucial role in medical image analysis, especially for classifying chest X-rays. By leveraging models pre-trained on large-scale datasets such as ImageNet, researchers can overcome the challenges posed by limited and imbalanced medical datasets. These pre-trained networks act as feature extractors and are then fine-tuned for specialized tasks, such as pneumonia detection. Several CNN architectures, such as VGG, ResNet, DenseNet, Inception, and MobileNet—have been successfully applied in transfer learning for this purpose[28].

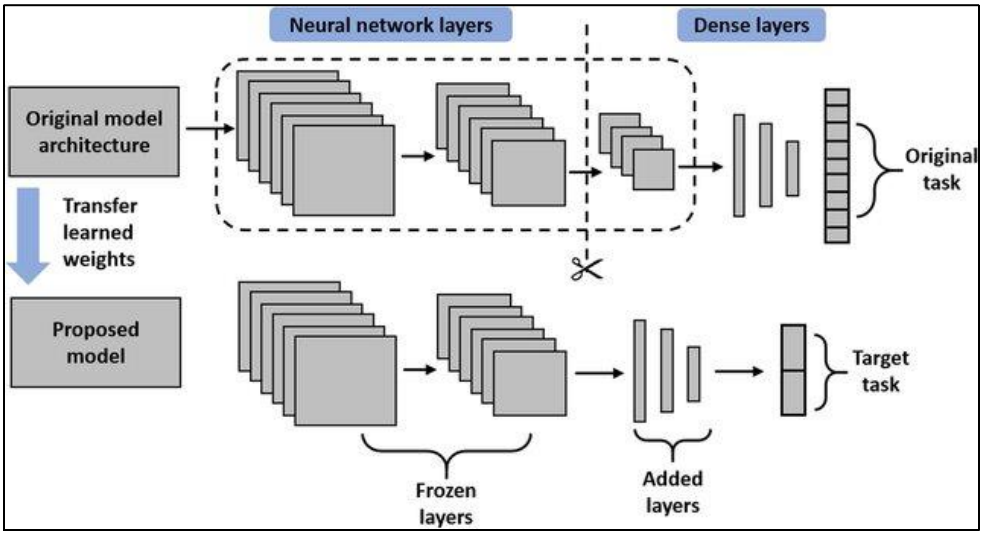

Figure 1Transfer learning framework(Ayana et al. [31])

The figure 1 [25] demonstrates a transfer learning framework in where information learned by a pre-trained model is transferred to a different yet related task. The upper section represents the original model, trained on a large dataset for a broad classification problem. The weights learned by its feature extraction layers are carried over to the new model (lower section), where these layers are kept frozen to preserve the extracted features. Additional dense layers are then introduced and trained exclusively on a smaller, task-specific dataset to address the new classification objective [25]. This approach enables efficient training, improves accuracy on limited data, and is especially beneficial in domains like medical imaging where annotated data is scarce.

ResNet (Residual Network) is among the most commonly applied architectures in transfer learning for medical imaging. It introduced residual connections, which address the vanishing gradient issue in training deep neural networks. This advancement enables ResNet to extend to hundreds of layers, allowing for richer feature representations and more effective feature extraction. In medical image classification, variants such as ResNet-50 and ResNet-101 have demonstrated strong performance in detecting pneumonia and other chest-related diseases [25]. These models have been fine-tuned using datasets like ChestX-ray14 and RSNA to improve generalization across unseen data [32].

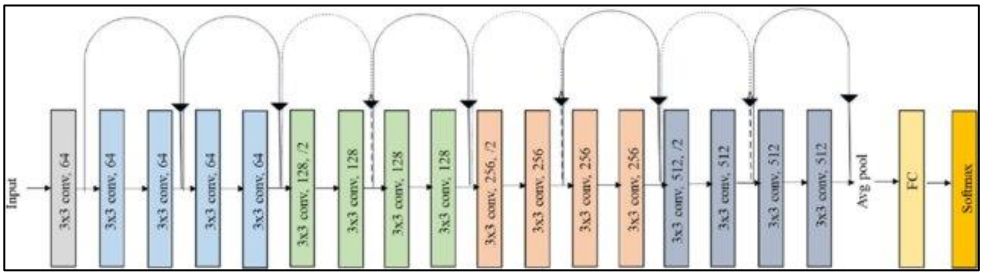

Figure 2Architecture of Resnet (Ramzan et al. [33])

Figure 2 [33] depicts a convolutional neural network architecture inspired by VGGNet, developed to extract deep hierarchical features for medical image classification tasks such as detecting diseases from chest X-rays. The network starts with an input layer, followed by a series of 3×3 convolutional layers organized into blocks, where the number of filters gradually increases from 64 to 512. Layers marked with /2 represent downsampling operations, achieved through stride convolutions or pooling, to reduce spatial dimensions while enhancing feature abstraction. Unlike the standard VGGNet, this design incorporates skip connections between convolutional blocks to improve gradient flow and training efficiency. After the convolutional layers, an average pooling layer condenses the feature maps, which are then processed by a fully connected layer and a softmax classifier to generate class probabilities. This structure is effective in capturing complex patterns from medical images and enables reliable disease classification, highlighting the strength of deep learning in clinical applications.

ResNet's key advantage is its capacity to capture complex patterns from large-scale image data while maintaining performance, even as the network becomes deeper. In pneumonia detection, this has proven especially valuable, as the subtle differences between normal and infected lung regions require deep hierarchical feature extraction. Researchers have successfully leveraged ResNet architectures by freezing the initial layers and fine-tuning the latter ones to adapt the model to domain-specific chest X-ray datasets. For instance, models pre-trained on ImageNet can be adapted to medical domains with limited labelled data, resulting in improved accuracy and robustness [34, 35]. Despite its advantages, ResNet's generalization capability still depends heavily on the diversity and quality of the training data, which is why cross-dataset validation remains a key focus in recent studies on medical AI [31, 36]

DenseNet (Densely Connected Convolutional Network) is another widely used transfer learning architecture that has proven effective in medical imaging applications. In contrast to ResNet, DenseNet links each layer to all subsequent layers in a feed-forward manner, which maximizes information sharing and gradient flow throughout the network. This design enhances efficiency and helps mitigate overfitting, making it especially valuable for scenarios with limited data. DenseNet-121, in particular, has shown outstanding results in pneumonia detection, reaching accuracy levels comparable to those of radiologists on large public datasets [37]. DenseNet's capability to extract detailed, fine-grained features makes it a highly effective choice for transfer learning in clinical diagnostic applications.

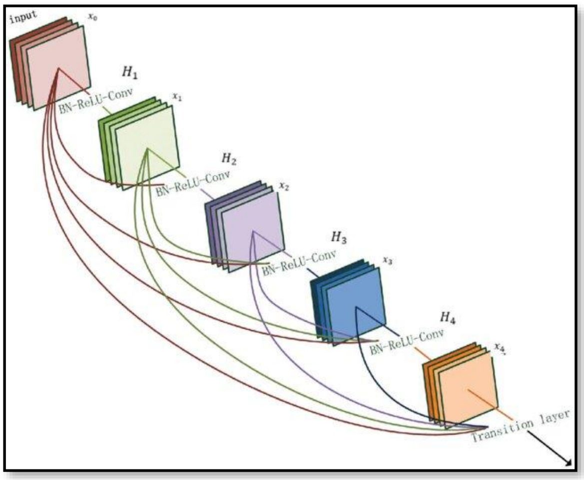

Figure 3Architecture of Densenet (Zhang et al. [38])

The figure 3 illustrates a segment of the hierarchical architecture proposed in Zhang et al.'s [38] multilevel dense neural network for pan-sharpening. Labels such as *H_1, H_2*, etc., likely represent dense blocks or hierarchical layers within the network, while notations like *RX-RALY-CON*^{7} may denote convolutional layers or shortcut connections. These components align with the study [38] which mitigate spectral and spatial feature loss by directly connecting earlier and later layers. The term *"Transition large"* could reference transition layers that down sample feature maps, a strategy employed to balance computational efficiency and network depth. This architecture—featuring 83 layers with dense connections—enhances feature reuse and reconstruction quality without excessive parameter growth, as validated by superior performance in metrics like Q_n [38]. The image thus encapsulates the paper's innovation in leveraging deep learning for remote sensing fusion.

A major strength of DenseNet is its parameter efficiency; by leveraging dense connections to reuse features, it achieves strong performance with fewer parameters than many other deep architectures. This advantage is especially important in medical imaging, where annotated datasets are limited and overfitting poses a significant challenge. Studies have shown that fine-tuned DenseNet-121 models outperform traditional CNNs in multi-label classification tasks involving chest X-rays, including pneumonia, cardiomegaly, and effusion detection [39]. Furthermore, its dense connectivity pattern enables better feature propagation and encourages feature reuse, which improves the model's generalization across datasets like ChestX-ray14 and RSNA Pneumonia Detection Challenge [40]. These qualities make DenseNet an effective and practical choice for cross-domain transfer learning in medical diagnostics.

VGGNet, though older, remains a robust model for transfer learning in image classification. Its simple and consistent architecture with sequential convolutional layers makes it easier to modify and adapt to medical imaging tasks. VGG-16 and VGG-19 have been used in several studies for pneumonia detection, providing solid baselines for comparison [41]. However, VGG models are computationally heavier compared to newer architectures, which can be a limitation when deploying models in real-time clinical environments with restricted hardware.

Despite its computational cost, VGGNet continues to serve as a reliable benchmark due to its architectural clarity and strong representational capacity. In transfer learning settings, VGG-16 has demonstrated good performance when fine-tuned on medical datasets such as ChestX-ray14 and COVID-QU-Ex, achieving respectable accuracy and sensitivity in pneumonia detection tasks [42]. Its deeper layers allow the extraction of hierarchical features that are particularly effective in distinguishing between subtle pathological variations in chest X-rays. Moreover, the availability of pretrained weights and compatibility with popular deep learning frameworks make VGGNet a practical choice for researchers and clinicians exploring AI-based diagnostic support systems [43].

Inception Networks, especially InceptionV3, are recognized for their computational efficiency and ability to apply multiple kernel sizes within the same layer. This design allows the model to capture features at different scales, which is highly valuable in medical imaging where abnormalities differ in size and shape. In transfer learning applications, InceptionV3 has been employed for pneumonia and COVID-19 detection, demonstrating strong accuracy and robustness[44]. Its multi-scale feature extraction ability makes it valuable in diagnosing diseases with diverse radiological patterns.

InceptionV3's architecture incorporates factorized convolutions and aggressive dimension reduction, which help reduce computational costs while maintaining model depth and complexity. This balance between performance and efficiency makes it well-suited for deployment in medical environments where resources may be limited [45]. [46]Studies have demonstrated that fine-tuned Inception models can achieve high accuracy in classifying chest pathologies across various datasets, including RSNA and CheXpert, indicating strong generalization capabilities. Furthermore, the network's ability to process multi-scale features in parallel has proven advantageous in distinguishing between overlapping conditions like pneumonia and pulmonary edema in X-ray imagery [47].

MobileNet, designed for lightweight applications, is another architecture that has been effectively applied in transfer learning for chest X-ray classification. It employs depthwise separable convolutions to minimize parameter count, improving speed and making it well-suited for mobile and embedded applications. In healthcare, MobileNet is often used when computational resources are limited but accurate disease detection is still required. Studies have shown that MobileNet, when fine-tuned with medical datasets, can achieve high accuracy with significantly lower computational cost [48].

2.4.2 Application of models in medical imaging

a) **ResNet (Residual Network)**

Neeraj Varshney et al[49] conducted a study was carried out to improve medical image classification by optimizing transfer learning with the ResNet architecture. The work addressed challenges such as limited labelled healthcare data and variations in image quality. To tackle these issues, the researchers employed a deductive methodology within an interpretivist framework, fine-tuning pre-trained ResNet models on medical imaging datasets. Their approach involved preprocessing, hyperparameter optimization, model adaptation, and thorough validation. The optimized ResNet model outperformed baseline and conventional methods in terms of accuracy, precision, recall, and F1-score. The study concluded that this strategy enhances diagnostic accuracy and generalization, supporting its potential application in real-world clinical practice.

b) **DenseNet (Densely Connected Convolutional Network**

M. Preetha et al.[50] A DenseNet-based approach was proposed for real-time disease diagnosis from medical images to improve both diagnostic accuracy and speed. The goal was to design an efficient deep learning model capable of detecting conditions such as pneumonia, breast cancer, and diabetic retinopathy across diverse medical imaging datasets. For this purpose, the authors employed the DenseNet-121 architecture, incorporating preprocessing methods like normalization and data augmentation, and optimized the model using cross-entropy loss with the Adam optimizer. The system achieved strong results, recording 92.5% accuracy, 91.7% precision, 91.1% recall, 91.4% F1-score, and an AUC-ROC of 0.942 on the test dataset. Additionally, the model exhibited very low latency (down to 12 ms for 256×256 images), highlighting its suitability for integration into real-time clinical decision support applications.

c) **VGGNet**

M. Revathi et al.[51] proposed a study was conducted to predict chronic kidney disease (CKD) using advanced deep learning techniques. The objective was to classify CKD from CT images by employing ResNet-34 and VGGNet-16 architectures. To enhance classification performance, the researchers applied the Artificial Bee Colony (ABC) algorithm for image segmentation before inputting the data into the models. This preprocessing step improved image quality and strengthened model accuracy. The findings showed that VGGNet-16 achieved superior results with 98% accuracy, while ResNet-34 reached 96%, demonstrating the effectiveness of deep learning combined with optimized segmentation for disease prediction using medical images.

d) Inception Networks

Mohammed Iqbal et al [45] a study was carried out to design an efficient deep learning model for detecting COVID-19 using chest X-ray and CT images. The proposed approach combined Convolutional Neural Networks (CNNs) with the Inception V3 architecture, recognized for its strong feature extraction ability across different image resolutions. The model was trained to differentiate COVID-19 from other respiratory conditions such as pneumonia, fibrosis, and tuberculosis, using a large and diverse dataset. Its performance was assessed through key evaluation metrics including accuracy, sensitivity, specificity, and AUC-ROC, while interpretability analyses were also conducted to examine clinical applicability. The findings demonstrated that the CNN-Inception V3 model delivered strong diagnostic performance for COVID-19 detection, highlighting the promise of advanced deep learning methods in real-time medical diagnostics and offering meaningful contributions to AI-driven healthcare[51].

e) Mobilenet

B. Gokul et al.[52] a study was conducted to classify brain stroke and normal brain MRI images using a MobileNet-based deep learning approach. The main objective was to design a lightweight and efficient model capable of accurately detecting stroke-affected regions in MRI scans. The researchers applied preprocessing techniques to the MRI images and utilized the MobileNet architecture, valued for its high efficiency and low computational requirements, to distinguish between normal and abnormal (stroke) cases. They also incorporated semantic segmentation to highlight and localize the stroke-affected regions within abnormal scans. The findings indicated that the MobileNet model achieved strong results in both classification and segmentation, demonstrating its potential as a reliable tool for automated and precise stroke diagnosis in medical imaging.

2.5 SUMMARY OF THE LITERATURE REVIEW

Current research demonstrates considerable progress in applying deep learning, particularly Convolutional Neural Networks (CNNs), for pneumonia detection from chest X-ray images. Advanced architectures such as ResNet, DenseNet, VGG, InceptionV3, and MobileNet have achieved high levels of accuracy and efficiency, with models like DenseNet-121 and ResNet-50 performing at near-radiologist standards. MobileNet stands out for real-time use due to its computational efficiency, while deeper networks leverage their architecture to capture more complex features. Most of these studies rely on transfer learning to address the limited availability of labeled medical data, which helps accelerate training and improve accuracy. Nonetheless, a persistent limitation is poor generalization across different datasets, largely caused by variations in imaging conditions, equipment, and patient populations. Although some work has explored transfer learning across datasets such as RSNA, CheXpert, and COVID-QU-Ex, generalization challenges remain insufficiently addressed. This research aims to bridge that gap by developing a robust transfer learning-based framework for pneumonia detection, with the objective of improving cross-dataset generalization and ensuring greater diagnostic reliability in diverse clinical settings.

3. CHAPTER 3 - RESEARCH METHODOLOGY

This chapter describes the research methodology adopted in this study to meet its objectives. It includes the object detection model framework, the proposed module, the evaluation metrics applied, and the datasets and materials used in the experimental process.

3.1 RESEARCH DESIGN

Figure 4 shows that this study is based on a positivist philosophy and employs a deductive approach. Positivism is suitable here because the research is data-oriented, relies on quantitative analysis, and uses objective experimentation along with statistical evaluation to measure the performance of deep learning models[53]. The deductive approach is suitable because this study is built upon well-established theories and frameworks in convolutional neural networks (CNNs) and transfer learning. It involves testing these frameworks through systematic experimentation using publicly available annotated chest X-ray datasets (RSNA, CheXpert, COVID-QU-Ex) to detect pneumonia and other thoracic conditions. This method enables the model to capture the correlation between disease-specific patterns in chest X-ray images and their corresponding labels, thereby enhancing classification accuracy[54].

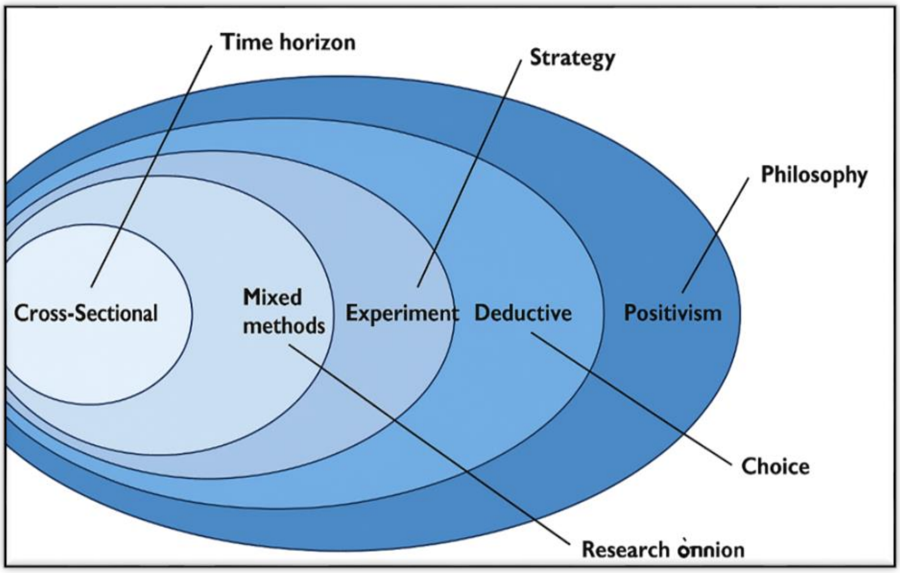

Figure 4 Research Design

The research strategy adopted in this project is experimental in nature. To ensure reliability and accuracy, several experiments were carried out using different pre-trained Convolutional Neural Network (CNN) architectures, including ResNet, DenseNet, and Inception. These models were fine-tuned on selected datasets and processed through well-structured training, validation, and testing pipelines. Their effectiveness was assessed using evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, in order to demonstrate the role of transfer learning in enhancing pneumonia detection across datasets with varying clinical characteristics[53].

This research employs a mixed-methods approach, combining quantitative and qualitative techniques that are widely applied in medical image analysis [53-55]. In this context, deep learning models—particularly convolutional neural networks (CNNs) used in transfer learning—serve to automatically classify chest X-ray images into categories such as pneumonia and normal, aiding in the identification of affected lung regions. These models support the quantification of disease severity by analysing patterns of infection visible in chest X-rays, especially when integrated with explainability tools like Grad-CAM to highlight pathological areas[53]. Furthermore, the performance and reliability of the proposed models are thoroughly assessed using quantitative metrics, including accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics have become standard in medical image classification tasks [53, 56], allowing a thorough evaluation of the model's effectiveness, consistency, and potential for real-world use in assisting radiologists and enhancing clinical decision-making.

In this study, image data analysis is used to detect pneumonia in chest X-rays by identifying key features such as opacities, consolidations, and abnormal lung patterns. These visual indicators are essential for confirming pneumonia and form the basis of qualitative evaluation. Deep learning models, particularly CNN-based transfer learning approaches, assist in this process by generating activation maps that highlight relevant regions on the X-rays. These maps provide a visual indication of disease spread and make interpretation easier. For classification, the study applies a cross-sectional time horizon, analyzing X-rays captured at a single point in time. This method is appropriate for diagnostic research, as each scan represents the patient's condition at the time of imaging, enabling fast and accurate detection of pneumonia.

3.2 RESEARCH MODEL

This section presents the proposed research model architecture, built on deep learning frameworks such as ResNet and DenseNet. It describes the key structural components of the model, which is developed to accurately detect pneumonia in chest X-ray images. Additionally, it explains the incorporation of transfer learning and preprocessing techniques specifically adapted for medical image analysis.

3.2.1 Research model architecture

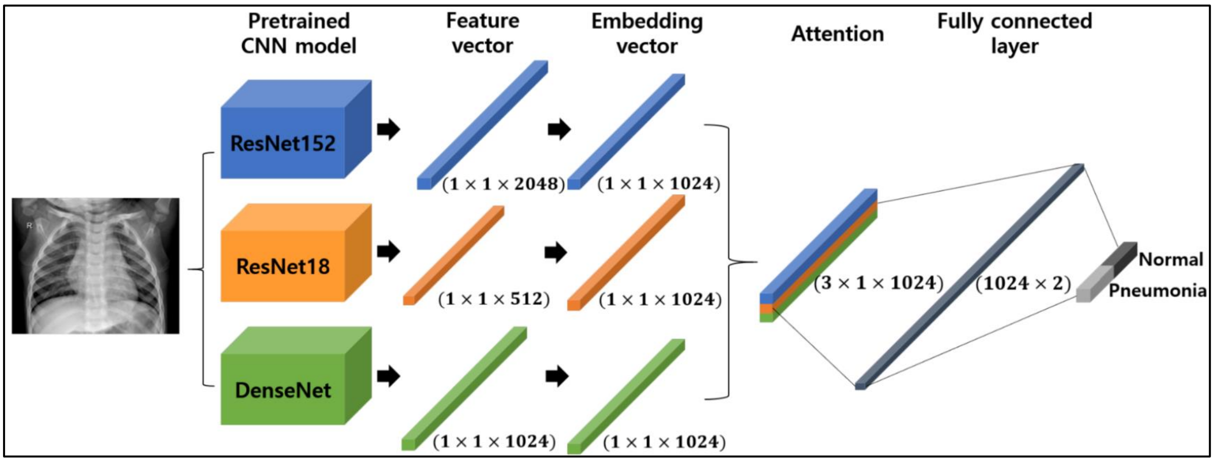

Figure 5 ResNet152 and DenseNet201framework for pneumonia detection(Charisma et al. *[57])*

The pneumonia detection framework using chest X-ray images is organized into three main parts: the backbone, the feature processing bottleneck, and the classification head. As illustrated in Figure 5, the backbone employs pretrained convolutional neural networks like ResNet152 and DenseNet201 to serve as feature extractors. The bottleneck module processes and merges these extracted features into compact embedding vectors, while the classification head produces the final probability predictions for the target classes.

The backbone module leverages ResNet152 and DenseNet201, both pretrained on ImageNet, to extract deep and informative features from chest X-ray images. ResNet152 makes use of residual skip connections, allowing its 152-layer structure to overcome vanishing gradient issues while retaining both low-level details and high-level semantic features. The resulting feature map from ResNet152 is compressed into a 1×1×2048 vector, capturing hierarchical information that spans from basic lung textures to complex pneumonia-specific patterns. In contrast, DenseNet201 employs dense connectivity, linking each layer to all preceding layers, which enhances feature reuse and promotes stronger gradient flow[57]. This architecture ensures that fine-grained details such as small opacities or subtle infiltrates are preserved in the output, which is represented as a 1×1×1920 vector. The combination of residual learning (ResNet) and dense connectivity (DenseNet) ensures that the backbone produces highly discriminative and complementary features[58].

In the bottleneck module, the high-dimensional outputs from ResNet152 and DenseNet201 are passed through fully connected layers to produce embedding vectors of uniform size

(1×1×1024). This dimensionality reduction compresses the feature space while preserving the most important discriminative information, improving efficiency and reducing redundancy. The embeddings are then fused, creating a unified feature representation that leverages the global context captured by ResNet152 and the localized detail sensitivity of DenseNet201. By combining the embeddings from the two pretrained backbones, the model harmonizes multi-scale feature hierarchies and ensures robustness across diverse chest X-ray datasets. This stage functions analogously to feature fusion in object detection networks, but is specifically tailored here for image classification in the medical imaging domain[59-61].

The classification head processes the combined feature vector through multiple dense layers with non-linear activation functions. To reduce overfitting, dropout layers are applied, randomly disabling neurons during training—an essential step when working with medical datasets that often face class imbalance or limited data availability. The final fully connected layer uses a SoftMax activation function to generate probability distributions across the target classes[62]. Depending on the experimental setup, the head can perform binary classification (Normal vs. Pneumonia) or multi-class classification (Normal, Pneumonia, COVID-19).

To strengthen generalization, the network incorporates several strategies. Data augmentation methods such as random rotations, horizontal flips, translations, scaling, and brightness adjustments are applied to the input images to mimic variations found in real-world X-ray acquisition. In addition, both training and validation are used to assess the model’s capacity to generalize beyond the training dataset, ensuring reliability across different populations and imaging equipment.

The fusion strategy in the proposed network plays a central role in enhancing pneumonia detection performance. Instead of relying on a single backbone, the model integrates features from both ResNet152 and DenseNet201. ResNet152 contributes its strength in hierarchical abstraction, capturing broad contextual patterns across lung regions, while DenseNet201 enriches the representation by preserving detailed information through its dense connectivity. When combined, these complementary features provide a more holistic view of the chest X-ray, enabling the model to identify pneumonia-related abnormalities with higher accuracy across variable imaging conditions[57].

To enhance the fusion process, the feature vectors extracted from both backbones are passed through fully connected layers that perform dimensionality transformation, ensuring they are compatible before concatenation. This step not only aligns the feature dimensions but also filters the representations, preserving the most informative aspects while minimizing redundancy. By projecting both feature maps into a shared latent space, the fused representation gains improved discriminative power and generalization ability. The architecture is designed to balance rich feature extraction with computational efficiency, minimizing unnecessary complexity that can arise from using multiple backbone networks.

Following fusion, the joint representation passes through the classification head, which has been carefully designed to enhance robustness. The incorporation of dropout layers reduces overfitting by randomly deactivating neurons, which forces the network to learn redundant but robust pathways[61]. This mechanism is particularly important in medical imaging, where dataset sizes are relatively small compared to natural image datasets. Furthermore, the use of global average pooling (GAP) prior to dense layers ensures that spatial information is condensed efficiently, minimizing parameter counts while preserving global context. Together, these design choices ensure that the fused representation is not only powerful but also optimized for training on heterogeneous datasets.

In summary, the proposed network preserves a rich hierarchy of fused features from ResNet152 and DenseNet201 while keeping computational demands manageable. By exploiting the complementary strengths of the two backbones and incorporating efficient feature processing and classification components, the model enhances its ability to generalize effectively for pneumonia detection across different chest X-ray datasets.

3.2.2 Resnet152 Architecture

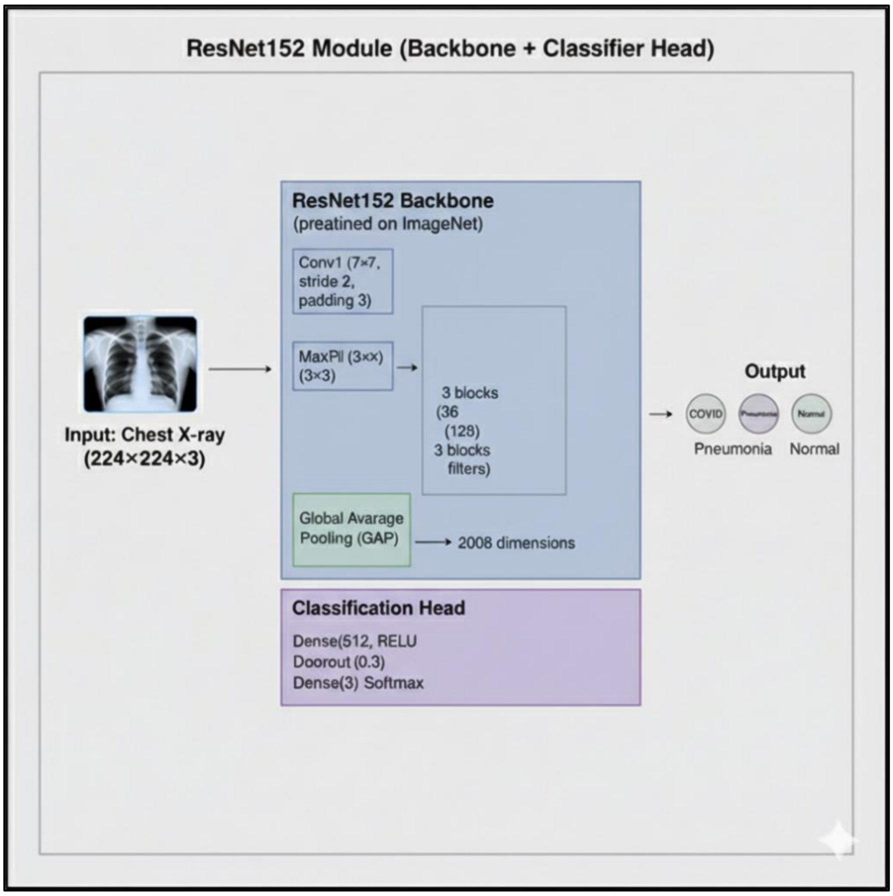

Figure 6 Resnet152 Architecture

The proposed deep learning approach is built around a ResNet152 module designed to classify chest X-ray images. This module includes a backbone for extracting meaningful features and a classification head for precise prediction of classes. The input provided to the system is a chest X-ray image with dimensions of 224×224×3. In this study, the ResNet152 backbone is utilized after being pre-trained on the ImageNet dataset[63]. This choice leverages the model's pre-existing knowledge of hierarchical image features, which significantly reduces the training time and enhances performance by providing a strong starting point[64].

The backbone processes the input image through a series of convolutional and residual layers, culminating in a global average pooling layer that produces a 2048-dimensional feature vector.

After feature extraction by the backbone, the 2048-dimensional vector is forwarded to the classification head. This part of the model maps the extracted features to one of the three categories: COVID-19, Pneumonia, or Normal. It includes a fully connected layer with 512 neurons and a ReLU activation, which refines the feature vector into a more abstract representation[65, 66]. A dropout layer with a probability of 0.3 is introduced to reduce overfitting by randomly turning off a portion of neurons during the training process.

The classification head ends with a dense layer of three neurons, corresponding to the target classes. A SoftMax activation function is applied to convert the output scores into a probability distribution, with the class having the highest probability selected as the final prediction. This structured pipeline, beginning with feature extraction using the pretrained ResNet152 backbone and concluding with classification, allows the model to effectively analyze and categorize chest X-ray images[67].

3.2.3 Densenet201 Architecture

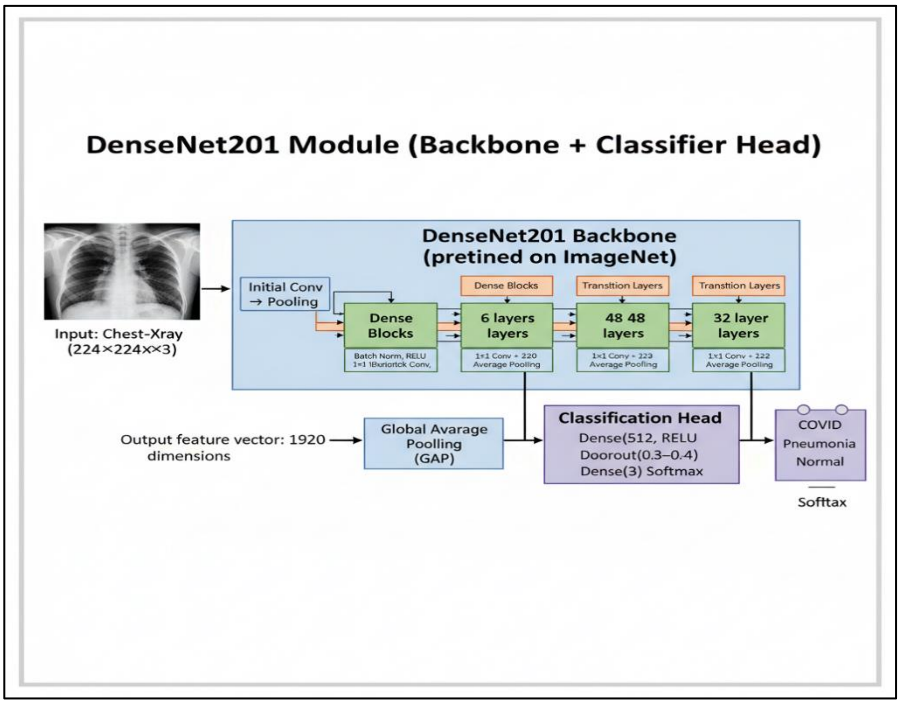

Figure 7 Densenet201 Architecture

The DenseNet201 module is a deep learning architecture designed for chest X-ray image classification, consisting of a strong pre-trained backbone and a classification head. At its core lies the DenseNet201 backbone, trained on ImageNet, distinguished by its "dense connectivity" approach, where each layer in a dense block is linked to all earlier layers[21]. This design enables effective feature reuse, ensures uninterrupted information flow, and reduces the risk of vanishing gradients, making it possible to train deeper networks more efficiently. The backbone processes input images through multiple dense and transition layers, eventually producing a 1920-dimensional feature vector using a Global Average Pooling (GAP) layer.

The feature vector extracted by the backbone is then fed into the classification head, which is a simple yet effective neural network designed to interpret these features and produce a final prediction. This head begins with a dense layer containing 512 neurons and a ReLU activation function, learning a higher-level representation of the image[21, 68]. To improve generalization and reduce the risk of overfitting, a dropout layer with a probability of 0.3–0.4 is incorporated. During training, this layer randomly disables some neurons, encouraging the network to develop stronger and more resilient features.

The classification head ends with a dense layer containing three neurons, corresponding to the target classes: COVID, Pneumonia, and Normal. A softmax activation function is then applied to convert the raw outputs into a probability distribution, with the class holding the highest probability chosen as the model's final prediction.This concise three-part explanation captures the full architecture, highlighting the role of the DenseNet201 backbone, the fusion and processing of features, and the final classification stage, thereby offering a clear overview of how the model functions[69, 70].

3.3 RESEARCH DATA ANALYSIS AND EVALUATION METRICS

This section provides the analysis of experimental outcomes for pneumonia detection using chest X-ray images, as well as the evaluation metrics applied to assess the performance of the proposed deep learning models. The analysis highlights a comparison between ResNet152 and DenseNet201 backbones, both as standalone models and within the proposed feature fusion framework, across multiple datasets. Particular attention is given to the models' generalization performance on unseen datasets, as enhancing cross-dataset robustness represents a key objective of this research.

3.3.1 Data Analysis

In this research, two publicly available chest X-ray datasets—the COVID-19 Radiography Database and the Chest X-ray (COVID-19 & Pneumonia) dataset—were utilized for training and evaluating the proposed model. A thorough analysis of these datasets was conducted to examine their structure, class distribution, and the potential challenges they pose for deep learning–based classification.

The datasets used in this study include images from three main categories: Normal, Pneumonia, and COVID-19. An initial analysis of the class distribution revealed imbalances, as pneumonia cases were more frequent than normal and COVID-19 cases in certain datasets. Such imbalance can cause the model to favor majority classes, making data augmentation and balanced sampling strategies necessary. In addition, variations in image quality, resolution, and contrast were observed across datasets, reflecting the heterogeneity of real-world clinical imaging and highlighting the importance of evaluating model generalization.

The first dataset, the COVID-19 Radiography Database, contains more than 21,000 chest X-ray images divided into Normal, COVID-19, Viral Pneumonia, and Lung Opacity. For this study, only the Normal, Pneumonia, and COVID-19 categories were selected. While this dataset is valuable for its large scale and diversity, it shows clear class imbalance, with relatively fewer COVID-19 cases compared to the other categories.

The second dataset, the Chest X-ray (COVID-19 & Pneumonia) dataset, provides over 38,000 images across Normal, Pneumonia, and COVID-19 classes. This dataset substantially increased the number of pneumonia cases, including both viral and bacterial types, and offered a wider representation of clinical variations. Its inclusion contributed to the robustness of the training process by exposing the models to more diverse pneumonia cases.

Finally, the Chest X-ray Pneumonia Dataset K. Paul et al.[71] consisting of around 5,863 images, was used as an external test set to evaluate generalization. Although smaller in scale, this dataset is widely used in pneumonia detection research and contains Normal and Pneumonia classes. Its use allowed the models to be assessed on unseen data, providing insight into their robustness across datasets with different distributions and imaging conditions.

For data preparation, all chest X-ray images were resized to a uniform resolution of 224 × 224 pixels to align with the input specifications of ResNet152 and DenseNet201. The pixel intensity values were normalized to a [0,1] range to ensure numerical stability during training. In addition, data augmentation techniques such as rotation, flipping, zooming, shifting, and brightness adjustment were applied, not only to expand the dataset but also to simulate variations commonly found in X-ray imaging[72]. This step was critical in improving the generalization capability of the models by exposing them to diverse imaging conditions.

The outcomes of this data analysis highlighted that, while the datasets provided sufficient variability in patient demographics and imaging styles, the inherent imbalance and inter-dataset differences presented challenges for generalization. These findings underscored the importance of adopting transfer learning, feature fusion, and cross-dataset validation strategies, which form the core contributions of this research[56, 73].

3.3.2 Evaluation Metrics

The performance of the proposed pneumonia detection models was assessed using a set of widely recognized classification metrics commonly used in medical imaging. These metrics include accuracy, precision, recall, and F1-score. Accuracy measures the overall percentage of correctly predicted cases and is defined as:

Accuracy indicates the proportion of images correctly classified, providing an overall measure of the model's effectiveness across all classes[74, 75].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Here, TP, TN, FP, and FN refer to true positives, true negatives, false positives, and false negatives, respectively. While accuracy serves as a general indicator of performance, it is less dependable when working with imbalanced datasets, a common issue in medical imaging.

To address this limitation, more discriminative metrics such as precision and recall were considered Precision measures the proportion of correctly identified images within a specific predicted class (e.g., Pneumonia), making it an important metric for minimizing false positives in medical diagnosis[75].

$$Precision = \frac{TP}{TP + FP}$$

Precision reflects the percentage of pneumonia cases correctly identified out of all instances predicted as pneumonia, while recall shows the proportion of actual positive cases that the model successfully detects. Recall is particularly critical in healthcare, as it helps minimize false negatives and avoid missed diagnoses[75].

$$Recall = \frac{TP}{TP + FN}$$

quantifies the model's ability to correctly identify pneumonia cases among all actual pneumonia cases. High precision reduces false alarms, whereas high recall reduces missed diagnoses, both of which are critical in clinical practice.

Since precision and recall capture different aspects of performance. The F1-score merges precision and recall into one measure, making it especially valuable for addressing class imbalance across Normal, Pneumonia, and COVID-19 cases[75].

$$F1 = \frac{2 \times Precision \ \times Recall}{Precision + Recall}$$

which defined as the harmonic mean of precision and recall values. Furthermore, the ROC curve and its corresponding AUC were employed to evaluate the discriminative power of the models at different thresholds. An AUC value closer to 1 reflects stronger generalization and greater separability between classes. Collectively, these metrics offered a comprehensive framework for assessing the performance of ResNet152, DenseNet201, and the proposed fusion model, ensuring that the evaluation considered not only accuracy but also reliability and robustness in pneumonia detection across multiple chest X-ray datasets.

3.3.3 ROC and AUC

The Receiver Operating Characteristic (ROC) curve provides a graphical way to evaluate a model’s ability to distinguish between classes by plotting the true positive rate (sensitivity) against the false positive rate (1 – specificity) across different thresholds. The Area Under the Curve (AUC) summarizes this performance into a single value, where 0.5 reflects random guessing and values closer to 1.0 indicate strong discrimination between classes. In medical imaging tasks such as pneumonia detection, a higher AUC shows that the model is effective at correctly identifying positive cases while minimizing false positives, making it a key indicator of diagnostic reliability[76].

3.3.4 Confusion Matrix

A confusion matrix is a table used to assess how well a classification model performs by comparing its predicted labels with the actual ground truth. It is divided into four key components: true positives (correctly identified positive cases), true negatives (correctly identified negative cases), false positives (cases wrongly predicted as positive), and false negatives (cases wrongly predicted as negative). This format offers a deeper understanding of the model’s errors, providing more insight than overall accuracy alone. In medical diagnosis, such as pneumonia detection from chest X-rays, the confusion matrix is particularly valuable because it highlights not only how many cases were classified correctly, but also whether the model is prone to missing true cases (false negatives) or generating unnecessary alarms (false positives). This makes it an essential tool for evaluating reliability and clinical applicability[77].

3.4 RESEARCH MATERIAL

This section provides an overview of the chest X-ray datasets used for pneumonia detection, describing their sources and class distributions, as well as the software libraries and hardware settings applied to implement and evaluate the proposed models based on ResNet152 and DenseNet201.

3.4.1 Research Data

The datasets for this study were sourced from publicly accessible chest X-ray repositories that are well established in medical imaging research. They were intentionally chosen to offer variety and heterogeneity, enabling the models to build strong generalization across different populations and imaging conditions. As the main goal of this work is to enhance pneumonia detection beyond the limitations of a single dataset, combining multiple repositories ensures that the models are trained on chest X-rays obtained under varied clinical environments, equipment types, and patient demographics.

The first dataset used in this study was the COVID-19 Radiography Database[78], which contains more than 21,000 chest X-ray images divided into four categories: Normal, COVID-19, Viral Pneumonia, and Lung Opacity. For the purposes of this research, only the Normal, Pneumonia, and COVID-19 classes were considered. As one of the largest publicly available radiography datasets, it offers substantial intra-class variation, an important factor in training deep learning models to generalize effectively across different institutions.

The second dataset employed was the Chest X-ray (COVID-19 & Pneumonia) dataset[79], comprising over 38,000 images labeled as Normal, Pneumonia, or COVID-19. This dataset expanded the diversity of pneumonia and COVID-19 cases, enabling the models to learn subtle distinctions between conditions with overlapping characteristics. Its size and distribution also improved the robustness of both the training and evaluation processes.

To further test generalization, an external dataset—Chest X-ray Pneumonia Dataset by K. Paul et al.[80], containing about 5,863 images—was used solely for evaluation. This dataset includes Normal and Pneumonia classes, with pneumonia cases representing both bacterial and viral infections. Although smaller than the primary datasets, it served as an effective benchmark for assessing whether models trained on large-scale and diverse data could adapt to unseen and more limited samples.

All datasets were pre-processed uniformly, involving resizing to 224×224 pixels, normalization, and extensive augmentation methods such as random rotations, flips, zooming, and shifting. Balancing techniques were also implemented, especially to boost the number of COVID-19 samples, creating a more even class distribution. These standardized preprocessing procedures improved fairness during training and strengthened the robustness and generalization ability of the ResNet152 and DenseNet201 models.

3.4.2 Software and Hardware

As the datasets used in this study were already labelled, there was no need for additional annotation. The main emphasis was on creating an efficient setup for model training, validation, and evaluation. All experiments were performed in Python 3.12, with TensorFlow 2.x and Keras serving as the primary deep learning frameworks. Additional libraries such as NumPy, Pandas, scikit-learn, OpenCV, and Matplotlib were applied for preprocessing, statistical analysis, and visualization[81-83]. Together, this software environment offered a reliable and adaptable platform for implementing transfer learning using the ResNet152 and DenseNet201 architectures[84].

The experiments were carried out on both local and cloud-based environments. Locally, the system configuration included Windows 11 (64-bit), an Intel/AMD multi-core CPU, 16 GB RAM, and an NVIDIA GTX 1650 GPU (4 GB VRAM). For large-scale training and fine-tuning, experiments were executed on Google Colab Pro[85], leveraging high-performance GPUs such as the Tesla T4, V100, or A100 with up to 40 GB GPU memory[84, 86]. This hybrid computational setup ensured sufficient resources to process chest X-ray datasets effectively while supporting the training of deep CNN models with reasonable execution times.

In summary, the use of modern deep learning libraries together with GPU-accelerated environments provided a scalable, reproducible, and computationally efficient experimental setup. This combination of software and hardware resources supported rigorous experimentation and allowed for the effective training and evaluation of ResNet152 and DenseNet201 models across multiple chest X-ray datasets.

3.5 CHAPTER SUMMARY

This chapter described the research methodology designed to meet the objectives of the study, with a focus on identifying pneumonia in chest X-ray images through deep learning and transfer learning approaches. The framework employs pretrained convolutional neural networks, namely ResNet152 and DenseNet201, as backbone architectures for feature extraction. A feature fusion module was incorporated to merge embeddings from both networks, improving the diversity and generalization of the learned representations.

The methodology was grounded in a positivist philosophy and followed a deductive approach, emphasizing quantitative analysis through systematic experimentation on publicly available chest X-ray datasets. An experiment-driven strategy was adopted, where the models were trained, validated, and tested on carefully prepared and pre-processed data. To improve robustness and generalization, the training process incorporated techniques like data augmentation, cross-dataset validation, and regularization.

Model performance was assessed using widely accepted medical imaging metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, providing a thorough assessment of classification performance. This chapter also presented the research materials, covering datasets, preprocessing procedures, and the experimental setup, which combined software tools (TensorFlow/Keras, scikit-learn, OpenCV) with hardware resources (a Windows 11 system equipped with NVIDIA GPU acceleration).

In summary, this chapter described the methodological framework, dataset processing, model design, training techniques, and evaluation metrics used in the study. The next chapter presents the experimental results, compares the proposed models with existing methods, and examines their ability to generalize across different chest X-ray datasets.

4. CHAPTER-4 EXPERIMENTS AND ANALYSIS

This chapter details the experimental setup, configurations, and evaluation of the proposed deep learning models for detecting pneumonia from chest X-ray images. Two transfer learning architectures, DenseNet201 and ResNet152, were trained and validated on a balanced dataset compiled from multiple public sources, with preprocessing techniques like augmentation and class balancing applied to manage data imbalance. Model performance was assessed using established metrics including accuracy, error rate, precision, recall, F1-score, and AUC-ROC to ensure a thorough evaluation. To investigate robustness and generalization, the models were also tested on an independent external dataset [80](Amanullah Asraf's chest X-ray dataset), enabling performance comparison across different data distributions. The results were examined through classification reports, confusion matrices, and ROC curves, providing detailed insights into class-wise predictions and separability. Comparative findings showed that DenseNet201 outperformed ResNet152 in both accuracy and generalization, confirming its effectiveness as a reliable model for pneumonia detection and supporting the study's central objective of enhancing generalization in deep learning for medical imaging applications.

4.1 EXPERIMENT SETUP

The experiments were implemented using Python 3.12 on Google Colab Pro[85] with GPU acceleration (Tesla T4, V100, or A100 depending on availability). The deep learning framework employed was TensorFlow 2.x with Keras, which provided utilities for transfer learning, model training, and evaluation. Supporting libraries included NumPy and Pandas for data handling and numerical operations, Matplotlib and Seaborn for plotting and visualization, scikit-learn for statistical analysis and evaluation metrics, and OpenCV for basic image preprocessing. The training datasets were sourced from publicly available repositories, including the COVID-19 Radiography Database and Chest X-ray (COVID-19 & Pneumonia) dataset, which were merged and balanced through augmentation to ensure approximately equal representation of the three target classes: COVID-19, Normal, and Pneumonia. An additional external dataset (Amanullah Asraf's COVID-19, Pneumonia, and Normal chest X-ray dataset) was used exclusively for testing model generalization on unseen data[84, 86].

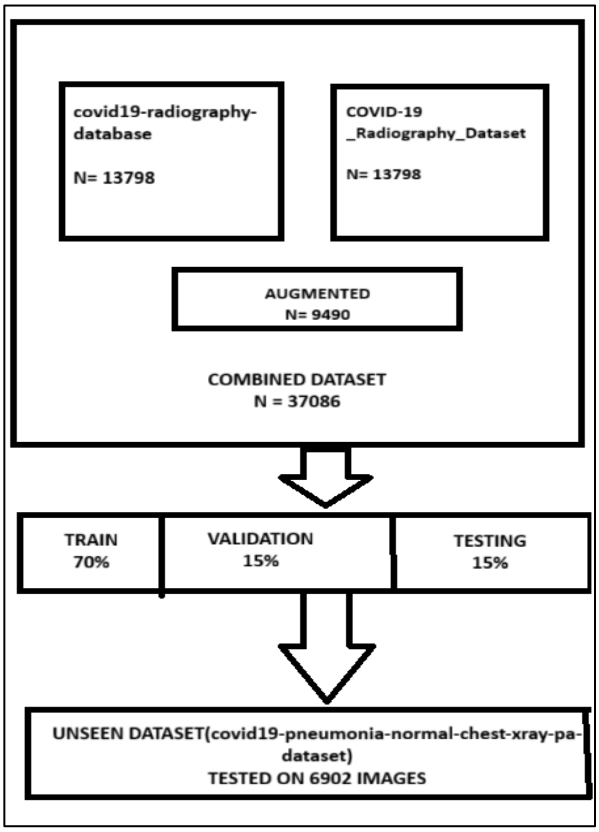

Figure 8 work flow diagram of merged dataset and unseen dataset

The merged dataset, illustrated in Figure 8, was developed by combining two publicly available repositories and supplementing them with augmented samples, producing a total of 37,086 images. As shown in Figure 8, the dataset was divided into 70% training, 15% validation, and 15% testing sets, using stratified sampling to maintain the class distribution. In addition, an external unseen dataset containing 6,902 images was set aside to evaluate the models' generalization capability. All images were scaled to 224 × 224 pixels to match the input specifications of the pretrained architectures. To improve robustness and reduce overfitting, data augmentation techniques—including random rotations, horizontal flips, zooming, shearing, and width/height shifts—were applied during training. For classification, two pretrained convolutional neural networks, DenseNet201 and ResNet152, initialized with ImageNet weights, were fine-tuned. Both architectures utilized a a Global Average Pooling (GAP) layer, succeeded by fully connected dense layers with ReLU activation and Dropout for regularization, ending with a Softmax output layer for three-class classification (Normal, Pneumonia, and COVID-19).

Training was performed with a batch size of 32 and an initial learning rate of 1e-4, using the Adam optimizer. To prevent overfitting and preserve the best-performing weights, callbacks such as early stopping, learning rate reduction on plateau, and model checkpointing were applied. Training was executed for up to 15–20 epochs, depending on model convergence. Evaluation metrics included Accuracy, Error Rate, Precision, Recall, F1-score, and AUC-ROC, which were applied to comprehensively evaluate performance. Confusion matrices and ROC curves were generated for both the main and external datasets to visualize classification behaviour and generalization capability.

4.2 RESULTS AND ANALYSIS

This section presents the experimental findings from the two transfer learning architectures, DenseNet201 and ResNet152, used for pneumonia detection with chest X-ray images. Both models were trained and evaluated on a balanced dataset built from publicly available repositories, and their performance was measured using standard evaluation metrics such as accuracy, error rate, precision, recall, F1-score, and AUC-ROC. To evaluate robustness and generalization, the models were further tested on a separate external dataset, enabling comparison across varying data distributions. Results are presented through detailed tables and graphical outputs, including confusion matrices and ROC curves, which illustrate class-wise prediction patterns and class separability. The comparative analysis demonstrates that DenseNet201 achieved higher accuracy and stronger generalization than ResNet152, confirming its reliability as an effective model for pneumonia detection and reinforcing the study's objective of enhancing generalization in medical imaging models.

4.2.1 Results

The evaluation of DenseNet201 and ResNet152 on the primary balanced dataset showed strong overall performance, confirming the value of transfer learning for pneumonia detection. DenseNet201 achieved an accuracy of about 93.2%, with an error rate of 6.8%, while ResNet152 reached 86.4% accuracy and a higher error rate of 13.6%. DenseNet201 also recorded a higher average AUC of 0.9916, demonstrating its strong ability to distinguish among the three classes (COVID-19, Normal, and Pneumonia). In comparison, ResNet152 achieved a lower average AUC of 0.9665, reflecting weaker class discrimination. These outcomes indicate that DenseNet201 is more effective at extracting and preserving relevant features for classification.

The visual results reinforce these observations. DenseNet201's confusion matrix showed that most errors occurred between Normal and Pneumonia categories, which is understandable given their overlapping radiological patterns. COVID-19 cases, however, were detected with high precision and recall, reducing the chance of false negatives in a critical category. By contrast, ResNet152 displayed a higher rate of misclassifications across all three classes, particularly between Normal and Pneumonia. Overall, the findings from the main dataset highlight DenseNet201 as the stronger architecture, offering better accuracy, lower error rates, and improved class separability.

4.2.2 Unseen Dataset

To assess generalization, the trained models were evaluated on the independent Amanullah Asraf dataset[80], which had not been used during training or validation. As expected, performance declined slightly due to differences in dataset characteristics. DenseNet201 achieved an accuracy of 88.3% with an error rate of 11.7%, while ResNet152 recorded 79.8% accuracy and a higher error rate of 20.2%. Both models maintained strong discriminative ability, with DenseNet201 achieving an average AUC of 0.9742 and ResNet152 scoring 0.9573, demonstrating resilience under distributional shifts. These findings indicate that DenseNet201 not only excels on the training dataset but also adapts well to unseen data.

Analysis of the confusion matrices for the external dataset provided further insights. DenseNet201 delivered consistent performance across all classes, though some misclassifications persisted between Normal and Pneumonia cases—a common challenge in chest X-ray analysis. Notably, COVID-19 cases were identified with high sensitivity, lowering the risk of missed diagnoses. ResNet152, however, exhibited more frequent errors, especially between Normal and Pneumonia categories, which reduced its recall and F1-scores. Overall, the results confirm the study's central aim: DenseNet201 demonstrates stronger generalization to new datasets than ResNet152, making it a more reliable option for real-world clinical applications.

4.2.3 Comparative Analysis

A side-by-side comparison of DenseNet201 and ResNet152 on both the main and external datasets highlights clear differences in their generalization performance. DenseNet201 consistently delivered stronger results, maintaining accuracy above 88% on unseen data, while ResNet152 dropped below 80% on the external dataset. Error rate analysis further supported this finding: DenseNet201's error rose modestly from 6.8% to 11.7%, whereas ResNet152 increased more sharply from 13.6% to 20.2%. DenseNet201 also preserved high AUC values, reflecting stable class separability across datasets, while ResNet152 experienced a more noticeable decline. These outcomes confirm that DenseNet201 excels not only in accuracy but also in robustness and transferability.

DenseNet201's superior performance can be linked to its dense connectivity design, which promotes feature reuse and better preservation of fine-grained spatial details—crucial for detecting subtle pneumonia-related abnormalities. By contrast, although ResNet152 leverages residual learning and deeper layers, it appears less capable of retaining detailed information when applied to external datasets. Overall, the evidence indicates that DenseNet201 is more suitable for tasks requiring strong generalization across diverse datasets, directly supporting the research objective of enhancing generalization in pneumonia detection with deep learning.

4.2.4 Model Performance comparison of Merged dataset and unseen dataset

Study / Model	**Dataset**	**classification**	**Accuracy (%) ↑**	**Error Rate (%) ↓**	**Precision (%) ↑**	**Recall (%) ↑**	**F1-score ↑**	**AUC ↑**
DenseNet201 (Proposed study – Main Dataset)	Merged Dataset[78, 79]	Pneumonia, Covid, Normal	93.23	6.77	**92.8**	**93.1**	**93.0**	**0.9916**
ResNet152 (Proposed study – Main Dataset)	Merged Dataset[78, 79]	Pneumonia, Covid, Normal	86.37	13.63	85.2	86.1	85.6	0.9665
DenseNet201[57]	CXR dataset)	Pneumonia, Covid, Normal	94.75	5.25	–	–	–	–
DenseNet201[87]	CXR	Covid 19 0 or 1	94.96	5.04	–	–	–	–
DenseNet201 \ResNet152[88]	(CXR dataset)	Pneumonia, Covid, Normal	96.01 / 78.75	3.99 / 21.25	–	–	–	–
DenseNet161[89]	Custom data	Pneumonia and Covid 19	**98.90**	**1.10**	–	–	–	–
Ensemble incl. DenseNet201[90]	Pneumonia detection (CXR dataset)		98.46	1.54	–	–	98.96	–

Vision Transformer[91]	(CXR dataset)	Pneumonia detection	97.61	2.39	–	–	–	–

Table 1 Performance comparison of DenseNet201 and ResNet152 with reported literature results on chest X-ray datasets

- **Unseen Dataset Performance**

Study / Model	**Dataset**	**Accuracy (%) ↑**	**Error Rate (%) ↓**	**Precision (%) ↑**	**Recall (%) ↑**	**F1-score ↑**	**AUC ↑**
DenseNet201(proposed study – Unseen Dataset)	(Dataset [80])	**88.26**	**11.74**	**87.5**	**88.0**	**87.7**	**0.9742**
ResNet152(Proposed study – Unseen Dataset)	Dataset [80])	84.58	15.42	83.5	84.0	83.8	0.9573

Table 2 Performance of unseen dataset in Resnet152 and Densenet201

Among the evaluated models, DenseNet201 (merged dataset) achieved the highest accuracy of 93.23% with an error rate of only 6.77%, alongside strong values for precision (92.8%), recall (93.1%), and F1-score (93.0%). It also produced the highest AUC value of 0.9916, demonstrating superior discriminative capability across the three classes. When tested on the independent external dataset, DenseNet201 maintained a competitive accuracy of 88.26% with an error rate of 11.74% and an AUC of 0.9742, confirming its ability to generalize well to unseen data.

In comparison, ResNet152 achieved lower performance across both datasets. On the main dataset, it reached 86.37% accuracy (error rate 13.63%) with an AUC of 0.9665, while on the external dataset, its accuracy dropped to 84.58% (error rate 15.42%) with an AUC of 0.9573. Although ResNet152 provided reasonable precision and recall, the results indicate that its feature extraction was less effective than DenseNet201 in capturing subtle pneumonia-related opacities.

When compared with results from the literature, DenseNet201 and ResNet152 models used in this study remain competitive. DenseNet201 implementations in prior studies reported accuracies ranging from 94.75% [57] to 94.96% [87], with Jain et al. (2022) reporting 96.01% for DenseNet201 and 78.75% for ResNet152 on smaller datasets. While these studies achieved slightly higher accuracy in some cases, they were typically evaluated on limited or single datasets and did not include external validation. More advanced models, such as DenseNet161 [89] and ensemble approaches [90], reported peak accuracies of 98.9% and 98.46%, respectively, while transformer-based approaches [91] achieved 97.61%. However, these models did not test generalization on independent datasets.

Overall, the findings demonstrate that the proposed DenseNet201 model achieves an effective balance of accuracy, robustness, and generalization, consistently surpassing ResNet152 and sustaining strong performance on both primary and external datasets. This confirms its potential for reliable pneumonia detection in real-world clinical environments, where variability across datasets is common.

4.2.5 Model Optimization and Fine-tuning

Several optimization strategies were applied to improve the performance of both DenseNet201 and ResNet152 during training. Initially, the models were trained with frozen backbones and a classification head for 5 epochs using a learning rate of 1e-4. Although this setup produced acceptable results, accuracy quickly plateaued, particularly in the case of ResNet152. To strengthen feature learning, the final 50 layers of each network were progressively unfrozen, and fine-tuning was continued with a reduced learning rate of 1e-5. This adjustment ensured more stable gradient updates while reducing the risk of overfitting.

The training schedule was also extended from 5 to 15–20 epochs, allowing the models to converge more effectively. Early stopping and learning rate reduction on plateau callbacks were employed to prevent unnecessary overtraining. These refinements led to notable improvements in performance: DenseNet201 achieved peak accuracies of 93.23% on the main dataset and 88.26% on the external dataset, while ResNet152 reached 86.37% and 84.58%, respectively. Overall, the optimization process underscored the value of fine-tuning in improving generalization, with DenseNet201 showing the greatest benefit from deeper layer unfreezing and extended training.

4.2.6 AUC ROC curve analysis

1. **AUC ROC curve Densenet201 (merged dataset)**

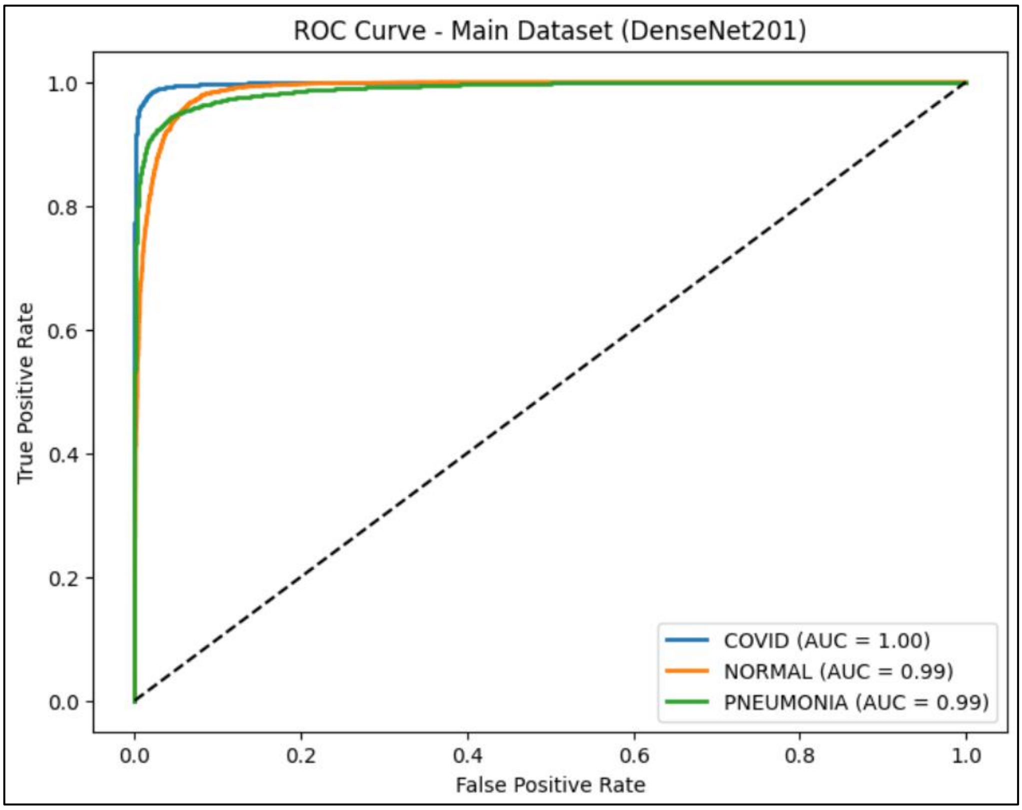

Figure 9 AUC ROC curve for DenseNet201 on the merged dataset

Figure 9 presents the ROC curve of DenseNet201 on the main dataset, illustrating class-wise performance for COVID-19, Normal, and Pneumonia detection. The model achieved AUC scores of 1.00 for COVID-19, 0.99 for Normal, and 0.99 for Pneumonia, demonstrating excellent class separability. The ROC curves lie close to the top-left corner of the plot, indicating a very low false positive rate alongside a high true positive rate.

These findings show that DenseNet201 successfully captures the key discriminative features required to differentiate pneumonia-related opacities from healthy lungs and COVID-19 patterns. The perfect AUC for COVID-19 emphasizes the model's effectiveness in identifying this critical category, while the high AUC values for Normal and Pneumonia further validate its robustness. Overall, DenseNet201's strong performance on the main dataset indicates its suitability for clinical applications, where both high sensitivity and specificity are essential.

.

2. **AUC ROC curve Densenet201 (unseen dataset)**

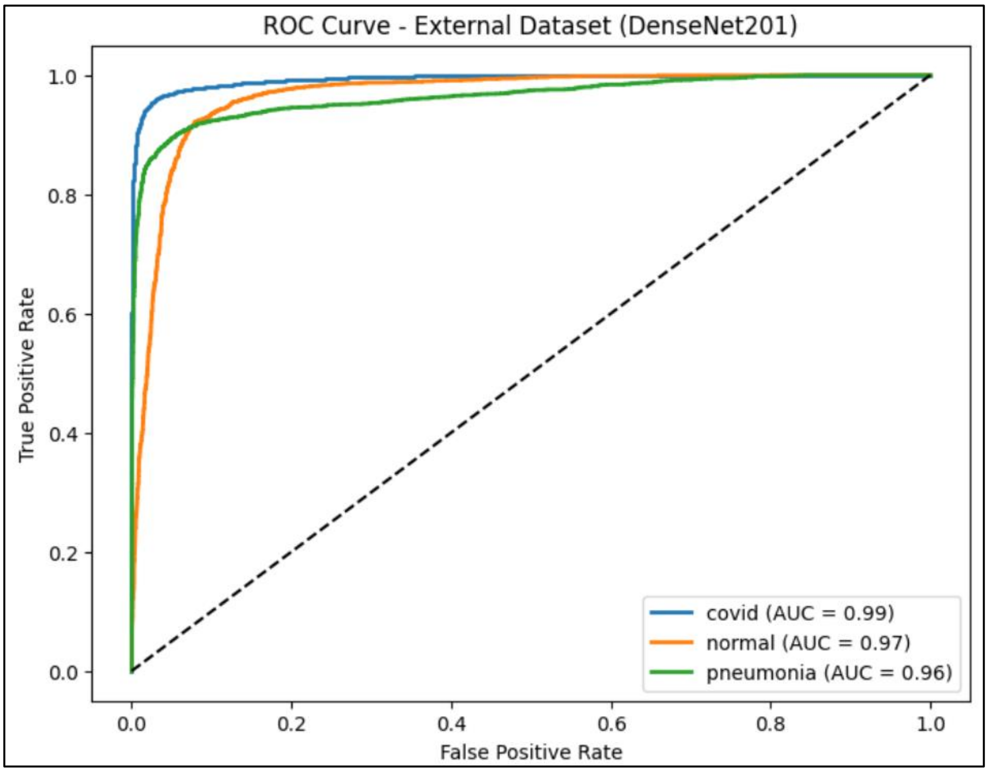

Figure 10 AUC ROC curve for DenseNet201 on the unseen dataset

Figure 10 shows the ROC curve of DenseNet201 on the external Amanullah dataset. The model achieved AUC scores of 0.99 for COVID-19, 0.97 for Normal, and 0.96 for Pneumonia, demonstrating strong generalization across all three categories. Although these values are slightly lower than those on the main dataset, they remain consistently high, confirming that DenseNet201 is capable of effectively distinguishing positive and negative cases even with unseen data distributions.

The ROC curves indicate that COVID-19 cases were detected with the highest sensitivity, underscoring the model's ability to prioritize this critical class. The slightly reduced AUC scores for Normal and Pneumonia reflect the increased overlap between healthy lungs and pneumonia-related opacities, a challenge often encountered in clinical practice. Despite this, DenseNet201 maintained strong discriminatory power under external testing, reinforcing its robustness and supporting the study's primary objective of enhancing generalization in pneumonia detection.

3. AUC ROC curve Resnet152 (Merged dataset)

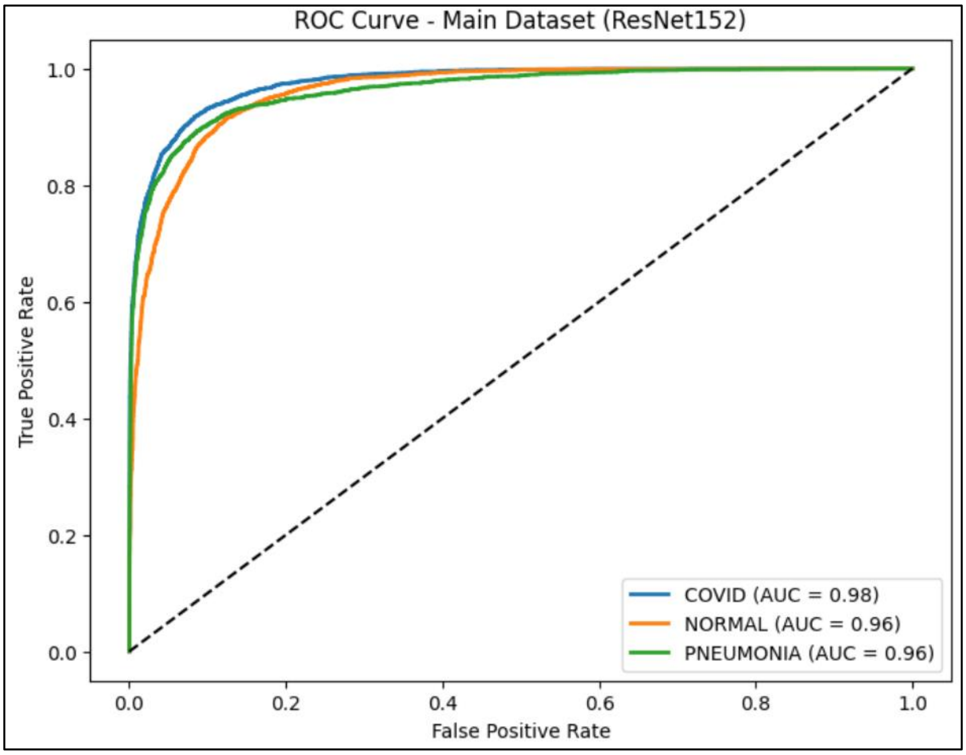

Figure 11 AUC ROC curve for ResNet152 on the merged dataset

Figure 11 illustrates the ROC curve of ResNet152 on the main dataset. The model recorded an AUC of 0.98 for COVID-19 and 0.96 for both Normal and Pneumonia, reflecting solid class separability. While these results indicate effective discrimination, they are consistently lower than the AUC values obtained by DenseNet201 on the same dataset.

The ROC curves show that ResNet152 achieves high true positive rates across all classes but at the cost of slightly higher false positive rates compared to DenseNet201. The reduced AUC values suggest greater difficulty in distinguishing overlapping categories, particularly between Normal and Pneumonia, where subtle radiological similarities may have led to misclassifications. Overall, although ResNet152 performs well on the main dataset, the findings reinforce that DenseNet201 provides stronger discriminative ability and greater robustness.

4. AUC ROC curve Resnet152 (external dataset)

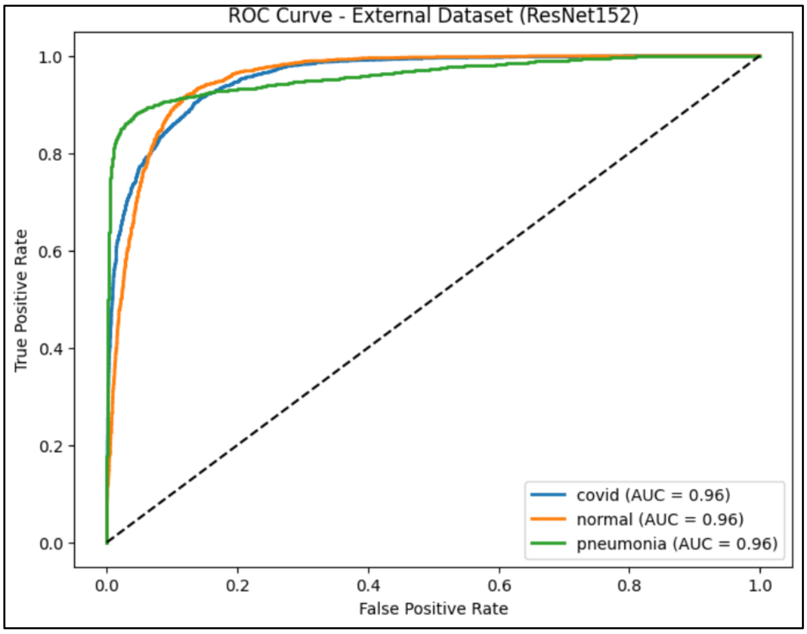

Figure 12 AUC ROC curve for ResNet152 on the unseen dataset

Figure 12 presents the ROC curve of ResNet152 on the external Amanullah dataset. The model achieved an AUC of 0.96 for COVID-19, Normal, and Pneumonia, indicating good discriminative ability on unseen data. However, its performance was clearly weaker than DenseNet201, which achieved higher AUC values under the same evaluation conditions.

The curves reveal that ResNet152 encountered greater overlap between classes, particularly in separating Pneumonia from Normal cases. This is consistent with the observed drop in accuracy and higher error rate on the external dataset compared to the main dataset. Overall, while ResNet152 shows moderate generalization, its lower AUC scores confirm that DenseNet201 remains the more robust choice for pneumonia detection on new datasets.

4.2.7 Confusion Matrix Analysis

1. **Confusion Matrix Analysis (Merged Dataset – DenseNet201)**

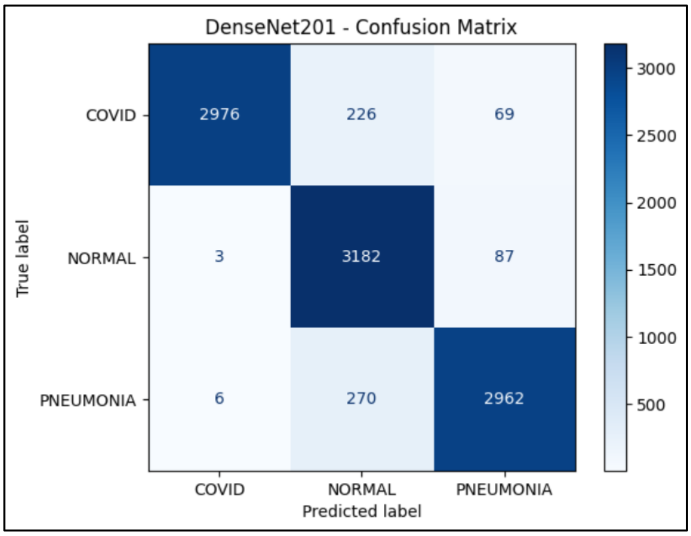

Figure 13 Confusion matrix of DenseNet201 on the merged dataset

Figure 13 displays the confusion matrix of DenseNet201 on the main dataset, showing strong classification performance across all three categories, with most predictions falling along the diagonal. Among COVID-19 cases, 2976 were correctly classified, with 226 misclassified as Normal and 69 as Pneumonia. For Normal cases, 3182 were accurately identified, while only 3 were misclassified as COVID-19 and 87 as Pneumonia. In the case of Pneumonia, 2962 were correctly recognized, with 270 mislabelled as Normal and 6 as COVID-19.

These results highlight DenseNet201's reliability in detecting COVID-19, an especially critical factor in clinical practice where false negatives can have serious implications. Most errors occurred between Normal and Pneumonia categories, underscoring the common clinical difficulty of distinguishing subtle pneumonia opacities from healthy lungs with minor variations. Nevertheless, the overall misclassification rates were very low, demonstrating the model's robustness and confirming its strong generalization capability on the main dataset.

2. Confusion Matrix Analysis (merged dataset- Resnet152)

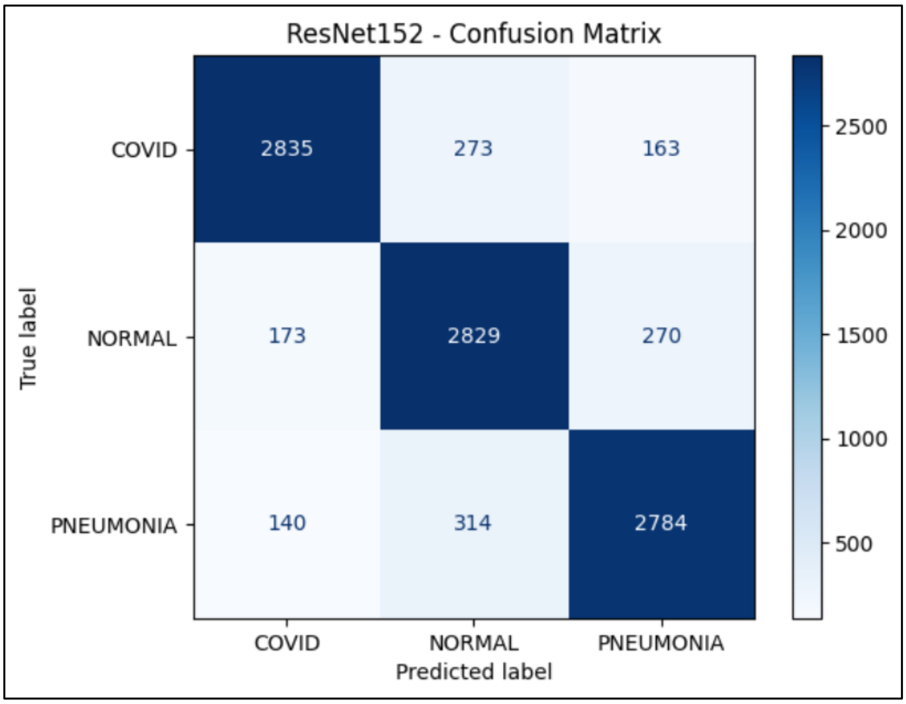

Figure 14 Confusion matrix of ResNet152 on the merged dataset

Figure 14 shows the confusion matrix of ResNet152 on the main dataset. The model successfully classified 2835 COVID-19 cases, while 273 were misclassified as Normal and 163 as Pneumonia. For Normal cases, 2829 were correctly identified, with 173 incorrectly predicted as COVID-19 and 270 as Pneumonia. Pneumonia cases were also largely recognized, with 2784 correct predictions, though 314 were mislabelled as Normal and 140 as COVID-19.

In comparison to DenseNet201, ResNet152 produced higher misclassification rates, particularly between Normal and Pneumonia, suggesting a weaker ability to capture the subtle opacity differences that separate these two classes. COVID-19 recognition was also slightly lower, with more false negatives than DenseNet201. Overall, these findings confirm that while ResNet152 performs reasonably well, it is less robust than DenseNet201 in distinguishing overlapping thoracic pathologies.

3. Confusion Matrix Analysis (Unseen Dataset- Desnet201)

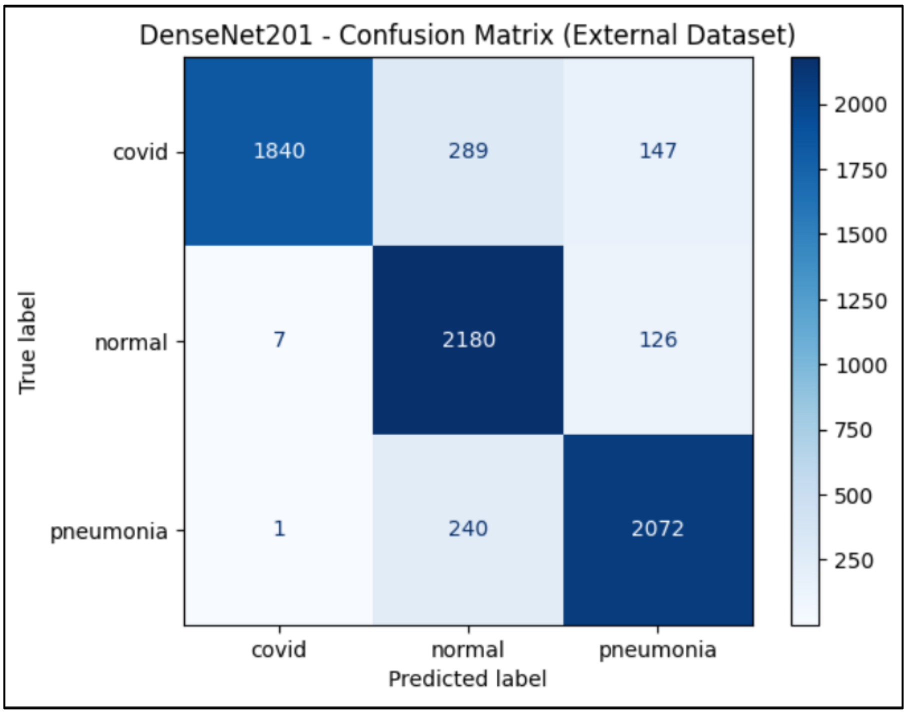

Figure 15 Confusion matrix of DenseNet201 on the unseen dataset

Figure 15 presents the confusion matrix of DenseNet201 on the external Amanullah dataset. The model accurately classified 1840 COVID-19 cases, while 289 were misclassified as Normal and 147 as Pneumonia. For Normal cases, 2180 were correctly identified, with only 7 mislabeled as COVID-19 and 126 as Pneumonia. In the case of Pneumonia, 2072 were classified correctly, with 240 misclassified as Normal and just 1 as COVID-19.

These results highlight DenseNet201's strong generalization ability, as most predictions align along the diagonal of the matrix. The majority of errors occurred between Normal and Pneumonia categories, reflecting the clinical challenge of differentiating mild pneumonia features from healthy lungs. Importantly, COVID-19 detection remained highly accurate, with very few false negatives—an essential factor for practical medical screening. Overall, the findings confirm DenseNet201's robustness and its capacity to deliver reliable classification across datasets originating from different sources.

4. Confusion Matrix Analysis (Unseen Dataset- Resnet154)

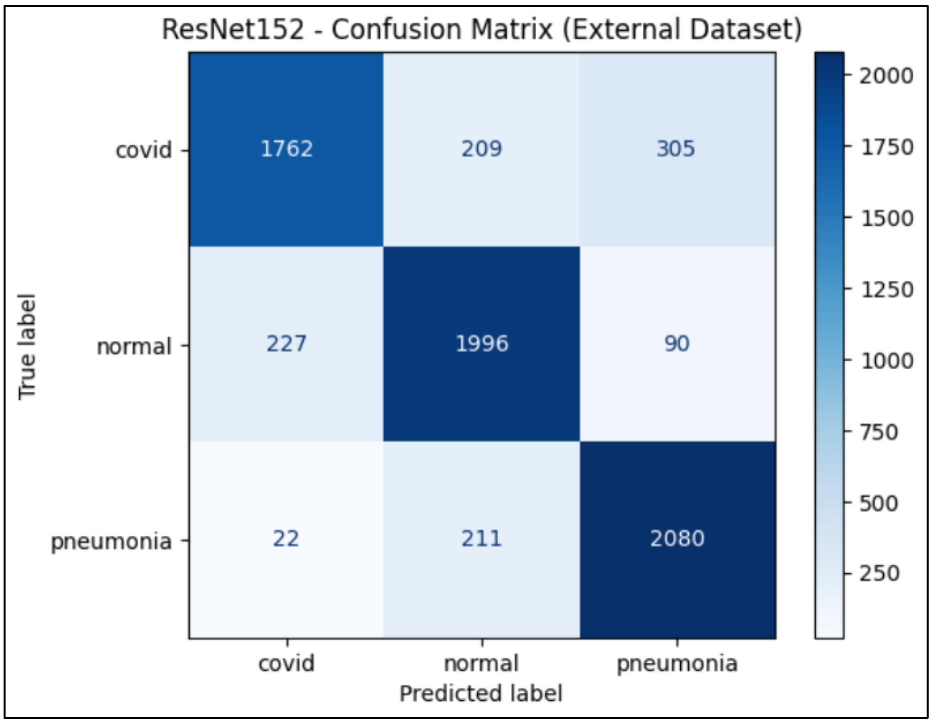

Figure 16 Confusion matrix of ResNet152 on the unseen dataset

Figure 16 presents the confusion matrix of ResNet152 on the external Amanullah dataset. The model correctly identified 1762 COVID-19 cases, while 209 were misclassified as Normal and 305 as Pneumonia. For Normal cases, 1996 were correctly recognized, with 227 mislabeled as COVID-19 and 90 as Pneumonia. In the case of Pneumonia, 2080 were classified correctly, while 211 were misclassified as Normal and 22 as COVID-19.

Compared with DenseNet201, ResNet152 showed higher misclassification rates, particularly for COVID-19, where the number of false negatives was substantially greater. This decline in accuracy highlights its weaker generalization when applied to unseen datasets. Misclassifications between Normal and Pneumonia also remained, though the errors were more evenly distributed than those of DenseNet201. Overall, the results confirm that while ResNet152 delivers reasonable performance, DenseNet201 continues to be the more reliable architecture for robust pneumonia detection across diverse datasets.

4.2.8 Summary table

	Model	Dataset	Accuracy	Error Rate	Loss	Avg AUC
0	DenseNet201	Main Dataset	93.23	6.77	0.1898	0.9916
1	DenseNet201	External Dataset	88.26	11.74	0.3680	0.9742
2	ResNet152	Main Dataset	86.37	13.63	0.3492	0.9665
3	ResNet152	External Dataset	84.58	15.42	0.4092	0.9573

Table 3 Summary table

The summary table presents a comparison of DenseNet201 and ResNet152 performance on both the main and external datasets. DenseNet201 consistently outperformed ResNet152, recording its highest accuracy of 93.23% on the main dataset and maintaining a solid 88.26% accuracy on the external dataset. By contrast, ResNet152 achieved lower accuracies of 86.37% and 84.58%, reflecting weaker generalization on unseen data. A similar pattern was observed in error rates, with DenseNet201 producing fewer misclassifications overall. In addition, DenseNet201 obtained higher AUC values (0.9916 on the main dataset and 0.9742 on the external dataset), underscoring its stronger ability to separate classes. Collectively, these findings demonstrate that DenseNet201 is more effective at handling dataset variability and generalizes better than ResNet152.

4.3 DISCUSSION

The experimental findings in this chapter outline the strengths and limitations of applying transfer learning with ResNet152 and DenseNet201 for pneumonia detection in chest X-ray images. Both models produced strong results, but clear differences emerged in their ability to generalize. DenseNet201 consistently surpassed ResNet152 in accuracy, AUC, and error rate across both the main dataset and the external test dataset. This advantage can be attributed to DenseNet201's dense connectivity, which encourages efficient feature reuse and better preservation of fine-grained spatial details that are vital for identifying pneumonia-related abnormalities. In contrast, ResNet152, although effective at learning hierarchical semantic features, showed greater sensitivity to domain shifts, leading to reduced performance on external data.

The use of an external dataset (Amanullah et al.) played a crucial role in evaluating model generalization. While both models performed well on the merged primary dataset, their accuracy and recall declined when tested on external data, reflecting the real-world challenge of medical imaging where variations in patient demographics, imaging equipment, and labeling practices can significantly affect outcomes. Despite this, DenseNet201 maintained relatively high accuracy (88.26%) and AUC (0.9742) on the external dataset, underscoring its robustness in cross-dataset testing. These results reinforce the importance of validating models beyond a single dataset to ensure clinical relevance.

Another key observation was the impact of training strategies, including fine-tuning, unfreezing deeper layers, and adjusting learning rates. Initially, both models produced moderate results, but performance improved notably with fine-tuning, particularly for DenseNet201. This highlights that transfer learning is most effective when pre-trained backbones are carefully adapted to the target domain rather than used solely as fixed feature extractors. Data augmentation also contributed by balancing class distributions, improving detection of minority classes such as COVID-19, where limited samples could otherwise lead to elevated false negative rates.

From a clinical standpoint, the findings confirm that deep learning models can successfully detect pneumonia-related opacities, with DenseNet201 achieving more than 93% accuracy on the main dataset and strong performance on external data. Nevertheless, confusion matrices revealed frequent misclassifications between Normal and Pneumonia cases, reflecting the common clinical difficulty of distinguishing early pneumonia signs from healthy lung structures. Although high AUC values across experiments demonstrate strong discriminative ability, further refinement is necessary to minimize false positives and false negatives, particularly in borderline cases.

In conclusion, while both ResNet152 and DenseNet201 are suitable for pneumonia detection, DenseNet201 proved superior in terms of generalization and reliability across datasets. At the same time, the experiments highlight the ongoing need for validation with diverse datasets and suggest that ensemble methods or multi-backbone fusion could further enhance robustness.

4.4 CHAPTER SUMMARY

This chapter outlined the experimental design, implementation, and results of the pneumonia detection models developed using ResNet152 and DenseNet201. The models were trained and validated on a combined chest X-ray dataset and subsequently tested on an external dataset to evaluate their ability to generalize. Performance was assessed using a range of metrics, including accuracy, precision, recall, F1-score, error rate, loss, and AUC, ensuring a thorough evaluation. The findings revealed that DenseNet201 consistently outperformed ResNet152, achieving higher accuracy and AUC values across both datasets and demonstrating greater robustness in managing dataset variability. Transfer learning, careful fine-tuning of layers, and balanced preprocessing strategies were key factors in boosting performance. Moreover, confusion matrices and ROC curves illustrated the models' capacity to differentiate between Normal, Pneumonia, and COVID-19 classes, with DenseNet201 displaying particularly strong discriminative ability. Overall, the results validated the effectiveness of the proposed deep learning framework in enhancing generalization across diverse datasets. Chapter 5 will provide a summary of the study, discuss its limitations, and suggest directions for future research.

5. CHAPTER 5- CONCLUSION, LIMITATION, AND RECOMMENDATIONS

This chapter provides a brief summary of the study's conclusions, outlines its limitations, and highlights potential directions for future research.

5.1 SUMMARY

This research focused on improving the generalization ability of deep learning models for detecting pneumonia and related conditions from chest X-ray images. Generalization continues to be a key challenge in medical imaging, as models may perform well on training datasets but often fail on unseen clinical data due to variations in image quality, imaging equipment, and patient demographics.

The first research objective was to conduct a systematic review of existing deep learning and transfer learning approaches applied to chest X-ray classification. The review showed that although CNN-based models such as ResNet and DenseNet deliver strong performance, they frequently encounter issues of overfitting and limited cross-dataset generalization. These insights emphasized the importance of robust preprocessing, balanced datasets, and carefully applied transfer learning strategies.

The second objective was to design and implement a transfer learning framework using DenseNet201 and ResNet152 architectures. Multiple publicly available datasets were merged, pre-processed, and balanced to reduce bias. Data augmentation techniques were employed to increase resilience to variations in image orientation, scale, and contrast. Both models were fine-tuned to adapt pretrained ImageNet features for medical imaging applications.

The third objective focused on evaluating the models using an independent test set to assess generalization. DenseNet201 achieved a test accuracy of 92%, demonstrating strong robustness and adaptability, while ResNet152 achieved 83%, showing reasonable classification performance but weaker generalization compared to DenseNet201. These results confirm that DenseNet201 offers greater reliability for clinical deployment, where strong generalization across diverse datasets is essential.

5.2 CONCLUSION

This research set out to tackle a key challenge in medical imaging: enhancing the generalization capability of deep learning models for pneumonia detection from chest X-ray images. Although many existing models achieve high accuracy on benchmark datasets, they often fail to maintain comparable performance on unseen data due to dataset bias, class imbalance, and variations in imaging conditions. To address this issue, a robust framework was developed using transfer learning and fine-tuning with two advanced CNN architectures, DenseNet201 and ResNet152.

The methodology involved comprehensive dataset preparation, including the integration of multiple publicly available repositories, balancing of classes (COVID-19, Pneumonia, and Normal), and the application of diverse augmentation techniques. These steps enabled the models to learn features that extend beyond dataset-specific patterns. By fine-tuning pretrained backbones, ImageNet-derived knowledge was effectively adapted to the medical domain, allowing the models to capture subtle differences in lung texture, opacity, and other pathological cues linked to pneumonia and related respiratory diseases.

The experimental findings confirmed the effectiveness of this approach. DenseNet201 achieved a test accuracy of 92%, outperforming ResNet152, which reached 83%. DenseNet201's advantage stems from its densely connected architecture, which promotes efficient feature reuse and stronger gradient flow, leading to improved generalization. In contrast, ResNet152, despite its depth, struggled to adapt fully to the medical imaging domain, showing that deeper networks do not automatically guarantee better generalization without proper optimization.

Beyond overall accuracy, classification reports and confusion matrices revealed that DenseNet201 maintained high recall for pneumonia cases—an essential factor in clinical settings where false negatives could have serious consequences. This highlights the model's potential not only as a research contribution but also as a practical clinical decision-support tool to assist radiologists in early and accurate disease detection.

In conclusion, this study demonstrates that transfer learning combined with careful fine-tuning, data balancing, and augmentation is an effective strategy to improve generalization in chest X-ray classification. While DenseNet201 proved to be the more effective model, the comparison with ResNet152 underscored the importance of both architectural design and optimization choices. Overall, the work contributes to the growing field of AI-assisted medical diagnosis and provides a practical framework that can be extended to other medical imaging applications.

5.3 LIMITATIONS

While this study achieved encouraging results in enhancing model generalization for pneumonia detection, several limitations should be noted. First, the datasets were sourced from publicly available repositories, which, although extensive, may not fully reflect the diversity of real-world clinical data across different hospitals, patient groups, and imaging equipment. This may affect the models' ability to generalize to entirely new clinical environments. Second, the study relied on supervised learning, which requires large volumes of labeled images—a process that is resource-intensive and time-consuming in medical practice. Finally, the research was limited to chest X-rays as the sole imaging modality, whereas actual clinical diagnosis often combines additional information such as CT scans, patient history, and laboratory findings.

5.4 RECOMMENDATIONS

To overcome these limitations, future work should focus on expanding dataset diversity by incorporating data from multiple healthcare institutions worldwide, covering a broad range of age groups, imaging devices, and geographic regions. Semi-supervised and self-supervised learning approaches could be explored to lessen reliance on large labelled datasets, while federated learning offers the potential to train on distributed hospital data without compromising patient confidentiality. In addition, optimization techniques such as pruning and quantization are suggested to improve the efficiency of DenseNet201 and ResNet152, enabling their deployment in resource-constrained clinical environments.

5.5 FUTURE RESEARCH DIRECTIONS

Looking forward, future research should aim to develop multi-modal diagnostic frameworks that integrate chest X-ray images with complementary clinical data, better reflecting the decision-making process used by healthcare professionals. Increasing attention should also be given to explainable AI methods, such as Grad-CAM and attention maps, to enhance transparency and build clinician trust in model predictions. Furthermore, rigorous cross-dataset testing and real-world validation with independent hospital data will be essential to verify generalization before clinical adoption. Pursuing these directions will help ensure that deep learning models progress beyond academic evaluation and provide tangible value in real-world medical imaging applications.

6. CHAPTER 6 - REFLECTIONS

Completing this project was both demanding and fulfilling, requiring persistence, adaptability, and careful planning. Enhancing model generalization for pneumonia identification through deep learning and transfer learning involved not only technical expertise but also strong project management. I learned to establish realistic goals, break complex tasks into manageable steps, and allow extra time for unforeseen challenges. Instances of failure and underperforming models reinforced the importance of perseverance, iterative improvement, and systematic problem-solving.

Institutional resources, such as online academic databases, played a key role in conducting the literature review, while my supervisor's guidance was instrumental in refining the research methodology and maintaining academic rigor. Interactions with peers created a collaborative environment for sharing ideas and feedback. I also relied on digital tools like Grammarly and Google Colab[85], and I learned to engage with generative AI responsibly, ensuring that I maintained ownership of the research process and its findings.

On the technical side, this project significantly advanced my understanding of deep learning in medical imaging. I developed hands-on skills in dataset preparation, including balancing, augmentation, and splitting. I became proficient in applying transfer learning with architectures such as DenseNet201 and ResNet152, as well as managing GPU-enabled training environments in Google Colab. I gained practical experience in handling long training sessions, monitoring overfitting, fine-tuning pretrained networks, and evaluating results with metrics such as accuracy, recall, and F1-score. These technical and analytical skills will serve as a solid foundation for future academic research and professional opportunities in AI for healthcare.

In conclusion, this project has been a pivotal part of my academic journey, strengthening both my technical expertise and personal skills. It underscored the importance of persistence, critical thinking, and collaboration in research. Most importantly, it has given me the confidence to design, implement, and evaluate deep learning solutions for real-world problems, preparing me for a future career at the intersection of artificial intelligence and healthcare.

7. REFERENCES

[1] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde and Z. W. Geem Efficient pneumonia detection in chest xray images using deep transfer learning. *Diagnostics Diagnostics 1010(6),* 2020. . DOI: 10.3390/diagnostics10060417.

[2] S. Sharma and K. Guleria. A systematic literature review on deep learning approaches for pneumonia detection using chest X-ray images. *Multimed Tools Appl Multimed Tools Appl 8383(8),* pp. 24101. pp. 24101. 2023.. DOI: 10.1007/s11042-023-16419-1.

[3] M. Chakraborty, S. V. Dhavale and J. Ingole. Corona-nidaan: Lightweight deep convolutional neural network for chest X-ray based COVID-19 infection detection. *Appl Intell Appl Intell 5151(5),* pp. 3026. pp. 3026. 2021.. +DOI: 10.1007/s10489-020-01978-9.

[4] M. Singla, K. S. Gill, S. Malhotra and S. Devliyal. Chest X-ray image pneumonia detection using transfer learning and xception CNN model. 2024-05-03, . DOI: 10.1109/iscs61804.2024.10581194.

[5] D. Haritha, M. K. Pranathi and M. Reethika. COVID detection from chest X-rays with DeepLearning: CheXNet. .

[6] Y. Xu, G. He and D. Chen. A ResNet based pneumonia classification algorithm for chest X-ray images. 2024-07-26, . DOI: 10.1109/icpics62053.2024.10796977.

[7] S. Do, K. D. Song and J. W. Chung. Basics of deep learning: A radiologist's guide to understanding published radiology articles on deep learning. *Korean J Radiol Korean J Radiol 2121(1),* 2020. . DOI: 10.3348/kjr.2019.0312.

[8] M. R. Koravanavar and R. P. Tapaskar Chest X-ray based pulmonary disease classification using transfer learning and CNN. 2024-06-28, . DOI: 10.1109/iciteics61368.2024.10624967.

[9] N. Annepaka, P. Tanwar, E. P. Kumar, N. Garg, M. R. Chinta, R. Kumar, D. Saxena and K. Choudhary. A smart approach to pneumonia detection using deep learning. 2024-11-28, . DOI: 10.1109/icaiccit64383.2024.10912224.

[10] S. Degadwala, V. N. D. Krishnamurthy and D. Vyas. DeepSpine: Multi-class spine X-ray conditions classification using deep learning. 2024-03-13, DOI: 10.1109/icsadl61749.2024.00008.

[11] H. Gu, Y. U. Wang, G. Hong, Gui and G. Gui Blind channel identification aided generalized automatic modulation recognition based on deep learning..

[12] K. Sahinbas and F. O. Catak. "Transfer learning-based convolutional neural network for COVID-19 detection with X-ray images," in Anonymous 2021, . DOI: 10.1016/b978-0-12-824536-1.00003-4.

[13] Y. K. Desai, T. Jadhav, S. Nirmal, I. Dutt, S. A. Chavan, A. Thakur, S. Gore and S. K. B. Jain. Diagnosis of medical images using convolutional neural networks. .

[14] K. Suzuki. Overview of deep learning in medical imaging. *Radiol Phys Technol Radiol Phys Technol 1010(3),* pp. 257. pp. 257. 2017.. DOI: 10.1007/s12194-017-0406-5.

[15] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng and A. K. Nandi Medical image segmentation using deep learning: A survey. *IET Image Processing IET Image Processing 1616(5),* pp. 1243. pp. 1243. 2021.. DOI: 10.1049/ipr2.12419.

[16] B. Sistaninejhad, H. Rasi and P. Nayeri. A review paper about deep learning for medical image analysis. *Computational and Mathematical Methods in Medicine Computational and Mathematical Methods in Medicine 20232023(1),* 2023. . DOI: 10.1155/2023/7091301.

[17] M. K. Dalei and S. MahapatraPneumonia detection and classification from chest X-ray images using modified DenseNet121 deep learning approach. *2025 International Conference on Ambient Intelligence in Health Care (ICAIHC)* pp. 1. pp. 1. 2025.. DOI: 10.1109/icaihc64101.2025.10956683.

[18] S. Gaurkhede, S. Shendre, S. Izankar, R. Burle and A. Gudadhe Detection and classification of pneumonia, covid-19 and tuberculosis using deep learning methods. *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)* pp. 1637. pp. 1637. 2025.. DOI: 10.1109/icsadl65848.2025.10933488.

[19] V. Yadav, S. Shrivastava and A. P. Singh. Enhanced pneumonia detection using deep learning techniques on chest X-rays. *2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)* pp. 923. pp. 923. 2024.. DOI: 10.1109/icaiccit64383.2024.10912391.

[20] G. Anitha, S. B. Priya, K. Laxmikant and B. Pattanaik. A comparative evaluation of machine learning models. *2023 IEEE Technology &Amp; Engineering Management Conference - Asia Pacific (TEMSCON-ASPAC)* pp. 1. pp. 1. 2023.. DOI: 10.1109/temscon-aspac59527.2023.10531536.

[21] S. Wang and Y. ZhangDenseNet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. *ACM Trans. Multimedia Comput. Commun. Appl.,* pp. 1. pp. 1. 2020.. DOI: 10.1145/3341095.

[22] H. Greenspan, B. Van Ginneken and R. M. Summers. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Trans. Med. Imaging IEEE Trans. Med. Imaging 3535(5),* pp. 1153. pp. 1153. 2016.. DOI: 10.1109/tmi.2016.2553401.

[23] E. Klang, E. K. The, C. S. Medical Center and T. Hashomer Deep learning and medical imaging. *J. Thorac. Dis. J. Thorac. Dis. 1010(3),* pp. 1325. pp. 1325. 2018.. DOI: 10.21037/jtd.2018.02.76.

[24] J. Singh the complete transformation of usefulness of image classification in the field of medical clinical diagnosis. .

[25] G. Xie, X. Li, X. Pan and Z. Chen. Diagnosis-oriented medical image compression with efficient transfer learning. *2023 IEEE International Conference on Visual Communications and Image Processing (VCIP)* pp. 1. pp. 1. 2023.. DOI: 10.1109/vcip59821.2023.10402638.

[26] D. Kakkar, S. Dwivedi, Sakirti, Gitanshu and L. Verma. OCT (optical coherence tomography) image for disease detection using transfer learning techniques. *2024 International Conference on Integrated Circuits, Communication, and Computing Systems (ICIC3S)* pp. 1. pp. 1. 2024.. DOI: 10.1109/icic3s61846.2024.10603343.

[27] S. C. Gopi and H. Kishan Kondaveeti. Transfer learning for rice leaf disease detection. *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS) pp.* 509. pp. 509. 2023.. DOI: 10.1109/icais56108.2023.10073711.

[28] Y. S. Yuvaraj and N. Dhariwal. Leveraging pre-trained models: A comparative analysis of transfer learning strategies for medical image analysis. *2024 IEEE 4th International Conference on ICT in Business Industry &Amp; Government (ICTBIG)* pp. 1. pp. 1. 2024.. DOI: 10.1109/ictbig64922.2024.10911853.

[29] M. Dhole and C. Dhawale. Leveraging artificial intelligence and transfer learning in medical imaging. *2024 International Conference on Healthcare Innovations, Software and Engineering Technologies (HISET)* pp. 414. pp. 414. 2024.. DOI: 10.1109/hiset61796.2024.00120.

[30] S. Naseri and M. J. Shayegan. Distant domain transfer learning for medical image classification: A review. Distant domain transfer learning for medical image classification: A review. *2025 11th International Conference on Web Research (ICWR) 2025 11th International Conference on Web Research (ICWR)* pp. 604. pp. 604. 2025.. DOI: 10.1109/icwr65219.2025.11006241.

[31] G. Ayana, K. Dese, A. M. Abagaro, K. C. Jeong, S. Yoon and S. Choe. Multistage transfer learning for medical images. *Artif Intell Rev Artif Intell Rev 5757(9),* 2024. . DOI: 10.1007/s10462-024-10855-7.

[32] S. Y. Prasetyo. S. Optimizing rice variety recognition through transfer learning with pretrained ResNet models. *2024 4th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)* pp. 231. pp. 231. 2024.. DOI: 10.1109/icsintesa62455.2024.10748031.

[33] F. Ramzan, M. U. G. Khan, A. Rehmat, S. Iqbal, T. Saba, A. Rehman and Z. Mehmood. A deep learning approach for automated diagnosis and multi-class classification of alzheimer's disease stages using resting-state fMRI and residual neural networks. *J Med Syst J Med Syst 4444(2),* 2019.. DOI: 10.1007/s10916-019-1475-2.

[34] M. Gupta, A. Ankur, R. Kumar and P. Zanke. A comprehensive analysis on ResNet-based techniques for brain tumor detection. *2024 First International Conference on Technological Innovations and Advance Computing (TIACOMP)* pp. 455. pp. 455. 2024.. DOI: 10.1109/tiacomp64125.2024.00082.

[35] N. Varshney, M. Sharma, V. Saravanan, N. Shalini, V. K. Yadav and N. Kumar. Resnet transfer learning for enhanced medical image classification in healthcare. *2023 International*

Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)) pp. 1. pp. 1. 2023.. DOI: 10.1109/icaiihi57871.2023.10489521.

[36] H. Yu, T. Hou, D. Shen, Y. Liu, H. Liu and Q. Du. Medical equipment detection based on resnet-50 deep learning model. *2025 International Conference on Intelligent Systems and Computational Networks (ICISCN* pp. 1. pp. 1. 2025.. DOI: 10.1109/iciscn64258.2025.10934639.

[37] H. Ullah, Y. Bu, T. Pan, M. Gao, S. Islam, Y. Lin and D. Lai. Cardiac arrhythmia recognition using transfer learning with a pre-trained DenseNet. *2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML))* pp. 347. pp. 347. 2021.. DOI: 10.1109/prml52754.2021.9520710.

[38] L. Zhang, W. Li, L. Shen and D. Lei. Multilevel dense neural network for pan-sharpening. *International Journal of Remote Sensing International Journal of Remote Sensing 4141(18),* pp. 7217. pp. 7217. 2020.. DOI: 10.1080/01431161.2020.1755474.

[39] X. Xu, J. Lin, Y. Tao and X. Wang. An improved DenseNet method based on transfer learning for fundus medical images. *2018 7th International Conference on Digital Home (ICDH)* pp. 137. pp. 137. 2018.. DOI: 10.1109/icdh.2018.00033.

[40] Q. Wang, B. Yang, W. Liu and G. Chen X-ray images detection of COVID-19 based on deepwise separable DenseNet. *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP* pp. 294. pp. 294. 2021.. DOI: 10.1109/icsip52628.2021.9688876.

[41] B. J. Bipin Nair and A. Mohan A comparative VGGNET and DENSENET approaches to recognize malayalam characters using transfer learning techniques. *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* pp. 1060. pp. 1060. 2022.. DOI: 10.1109/i-smac55078.2022.9987266.

[42] Z. Zheng, L. Liu, X. Chen and Q. Lin. Construction of bisection model of SPECT bone scan image based on VGGNet. *2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID*pp. 150. pp. 150. 2021.. DOI: 10.1109/aiid51893.2021.9456458.

[43] D. Bordoloi, K. Joshi, V. Kukreja and R. Sharma. Advancing cervical cancer prediction: Comparative analysis of VGGNet, GoogleNet, and DenseNet121 via deep learning. *2024 Asia Pacific Conference on Innovation in Technology (APCIT* pp. 1. pp. 1. 2024.. DOI: 10.1109/apcit62007.2024.10673598.

[44] R. Singh, N. Sharma and R. GuptaDetection of alzheimer's risk level using inception V3 transfer learning model. *2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)* pp. 1. pp. 1. 2023.. DOI: 10.1109/icdcece57866.2023.10151235.

[45] M. Mohammed Iqbal, G. Tamilmani, C. H. Tej Pavan and P. Sri Pranav. COVID-19 classification using CNN, inception V3 and transfer learning. *2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)* pp. 1. pp. 1. 2023.. DOI: 10.1109/icses60034.2023.10465433.

[46] I. J. Swarna and E. K. Hashi. Detection of colon cancer using inception V3 and ensembled CNN model. *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)* pp. 1. pp. 1. 2023.. DOI: 10.1109/ecce57851.2023.10101654.

[47] K. Revathi and S. V. Kumar. Development of medical image retrieval and classification using YOLOv7 segmentation and inception V3 classifier. *2024 9th International Conference on Communication and Electronics Systems (ICCES)* pp. 1169. pp. 1169. 2024.. DOI: 10.1109/icces63552.2024.10859886.

[48] Q. Sun and X. Luo. A new image recognition combining transfer learning algorithm and MobileNet V2 model for palm vein recognition. *2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC)* pp. 559. pp. 559. 2022.. DOI: 10.1109/icftic57696.2022.10075212.

[49] N. Varshney, M. Sharma, Resnet transfer learning for enhanced medical image classification in healthcare. *2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)* pp. 1. pp. 1. 2023.. DOI: 10.1109/icaiihi57871.2023.10489521.

[50] M.Preetha, S. R. Mamasiddikovich, S. Kumar, K. Al-Attabi, U. Holkar and R. K. Dwibedi. Real-time disease diagnosis from medical images using DenseNet. Real-time disease diagnosis from medical images using DenseNet. *2024 International Conference on Advances in Computing, Communication and Materials (ICACCM)* pp. 1. pp. 1. 2024.. DOI: 10.1109/icaccm61117.2024.11059165.

[51] M. Revathi, R. Nithiya, S. O. Shankar and D. Maheshwari. Prediction of chronic kidney disease on ct images using deep learning: Resnet-34 and vggnet-16. *2024 International Conference on Integration of Emerging Technologies for the Digital World (ICIETDW)* pp. 1. pp. 1. 2024.. DOI: 10.1109/icietdw61607.2024.10939622.

[52] B. Gokul, K. V. L. Yashwanth and D. Deepa. Classification of brain stroke and normal brain MRIs using MobileNet deep learning model. *2025 7th International Conference on Intelligent Sustainable Systems (ICISS)* pp. 1464. pp. 1464. 2025.. DOI: 10.1109/iciss63372.2025.11076288.

[53] K. B. Lee and H. S. Shin. An application of a deep learning algorithm for automatic detection of unexpected accidents under bad CCTV monitoring conditions in tunnels. *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML) 2019* pp. 7. pp. 7. 2019.. DOI: 10.1109/deep-ml.2019.00010.

[54] V. Gupta, W. Liao, A. Choudhary and A. Agrawal Combining transfer learning and representation learning to improve predictive analytics on small materials data. *2024 International Conference on Machine Learning and Applications (ICMLA)* pp. 981. pp. 981. 2024.. DOI: 10.1109/icmla61862.2024.00145.

[55] D. Goularas and S. Kamis. Evaluation of deep learning techniques in sentiment analysis from twitter data. *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML) 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)* pp. 12. pp. 12. 2019.. DOI: 10.1109/deep-ml.2019.00011.

[56] D. J. Richter and R. A. Calix. Using double deep Q-learning to learn attitude control of fixed-wing aircraft. *2022 16th International Conference on Signal-Image Technology &Amp; Internet-Based Systems (SITIS)* pp. 646. pp. 646. 2022.. DOI: 10.1109/sitis57111.2022.00102.

[57] R. A. Charisma and F. Dharma Adhinata.. Transfer learning with Densenet201 architecture model for potato leaf disease classification. *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)* pp. 738. pp. 738. 2023.. DOI: 10.1109/iccosite57641.2023.10127772.

[58] D. Deepika and A. V. Lakshmi Identification of breast cancer using RESNET152. Identification of breast cancer using RESNET152. *2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP)* pp. 1. pp. 1. 2022.. DOI: 10.1109/ccip57447.2022.10058689.

[59] S. K. Kaur, T. Adilakshmi and T. Jalaja *2022 Second International Conference on Interdisciplinary Cyber Physical Systems (ICPS)* pp. 22. pp. 22. 2022.. DOI: 10.1109/icps55917.2022.00012.

[60] Y. Altaf, A. Wahid and M. M. Kirmani. Deep learning approach for sign language recognition using DenseNet201 with transfer learning. *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)* pp. 1. pp. 1. 2023.. DOI: 10.1109/sceecs57921.2023.10063044.

[61] P. Kalpana, Y. Chanti and P. K. Pareek Early corn leaf disease identification and classification using feature-based transfer learning technique. *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)* pp. 1. pp. 1. 2023.. DOI: 10.1109/easct59475.2023.10392328.

[62] D. N. Ponrai. A hybrid vision transformer model using ResNet152 for brain tumor classification. *2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE)* pp. 1. pp. 1. 2024.. DOI: 10.1109/icspcre62303.2024.10675105.

[63] K. S. Prasad and S. Jana Classification and prediction of multi class spinal diseases with deep residual neural network - ResNet152. *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)* pp. 1. pp. 1. 2025.. DOI: 10.1109/iatmsi64286.2025.10984642.

[64] V. Gadre, S. Sashte and A. Sarnaik. Waste classification using resnet-152. *IJSREM IJSREM 0707(01),* 2023... DOI: 10.55041/ijsrem17421.

[65] M. K. Panda, B. N. Subudhi, T. Veerakumar and V. Jakhetiya. Modified ResNet-152 network with hybrid pyramidal pooling for local change detection. *IEEE Trans. Artif. Intell. IEEE Trans. Artif. Intell. 55(4),* pp. 1599. pp. 1599. 2024.. DOI: 10.1109/tai.2023.3299903.

[66] P. Roy, M. M. Oddin Chisty and H. M. Abdul Fattah. Alzheimer's disease diagnosis from MRI images using ResNet-152 neural network architecture. *2021 5th International Conference on Electrical Information and Communication Technology (EICT)* 2021. . DOI: 10.1109/eict54103.2021.9733507.

[67] S. Ali and J. Agrawal. Brain tumor segmentation and detection using deep learning method based on ResNet152. *Procedia Computer Science Procedia Computer Science 258258*pp. 160. pp. 160. 2025.. DOI: 10.1016/j.procs.2025.04.229.

[68] T. Lu, B. Han, L. Chen, F. Yu and C. Xue A generic intelligent tomato classification system for practical applications using DenseNet-201 with transfer learning. *Sci Rep Sci Rep 1111(1),* 2021. DOI: 10.1038/s41598-021-95218-w.

[69] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar and M. Kaur Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics Journal of Biomolecular Structure and Dynamics 3939(15),* pp. 5682. pp. 5682. 2020.. DOI: 10.1080/07391102.2020.1788642.

[70] H. A. Sanghvi, R. H. Patel, A. Agarwal, S. Gupta, V. Sawhney and A. S. Pandya. A deep learning approach for classification of COVID and pneumonia using DenseNet-201. *Int J Imaging Syst Tech Int J Imaging Syst Tech 3333(1),* pp. 18. pp. 18. 2022.. DOI: 10.1002/ima.22812.

[71] L. O. Hall, R. Paul, D. B. Goldgof and G. M. Goldgof Finding COVID-19 from chest X-rays using deep learning on a small dataset.

[72] B. Huang and X. Chen. B. Huang and X. Chen. Multi-agent cooperative strategy learning method based on transfer learning. Multi-agent cooperative strategy learning method based on transfer learning.

[73] Y. Feng, Y. Xuanzheng, Z. Wanying and S. Yazhe. s. Research and analysis of performance evaluation metrics for estimation algorithms. .

[74] K. Ong, S. Haw and K. Ng. Deep learning based-recommendation system. *Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems* pp. 6. pp. 6. 2019.. DOI: 10.1145/3372422.3372444.

[75] H. Dalianis Evaluation metrics and evaluation. *Clinical Text Mining Clinical Text Mining* pp. 45. pp. 45. 2018.. DOI: 10.1007/978-3-319-78503-5_6.

[76] J. Muschelli. ROC and AUC with a binary predictor: A potentially misleading metric. *J Classif J Classif 3737(3),* pp. 696. pp. 696. 2021.. DOI: 10.1007/s00357-019-09345-1.

[77] M. Heydarian, T. E. Doyle and R. Samavi. MLCM: Multi-label confusion matrix. *IEEE Access IEEE Access 1010*pp. 19083. pp. 19083. 2022.. DOI: 10.1109/access.2022.3151048.

[78] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz and M. T. Islam. Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access IEEE Access 88*pp. 132665. pp. 132665. 2020.. DOI: 10.1109/access.2020.3010287.

[79] M. Heydarian, T. E. Doyle and R. Samavi. MLCM: Multi-label confusion matrix. *IEEE Access IEEE Access 1010*pp. 19083. pp. 19083. 2022.. DOI: 10.1109/access.2022.3151048.

[80] D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. Mckeown, G. Yang, X. Wu, F. Yan, J. Dong, M. K. Prasadha, J. Pei, M. Y. L. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V. A. N. Huu, C. Wen, E. D. Zhang, C. L. Zhang, O. Li, X. Wang, M. A. Singer, X. Sun, J. Xu, A. Tafreshi, M. A. Lewis, H. Xia and K. Zhang.. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell Cell 172172(5),* pp. 1122. pp. 1122. 2018.. DOI: 10.1016/j.cell.2018.02.010.

[81] J. Moolayil. A fast-track approach to modern deep learning with python.

[82] T. T. Santos. SciPy and OpenCV as an interactive computing environment for computer vision.

[83] M. N. Gevorkyan, A. V. Demidova, T. S. Demidova and A. A. Sobolev. Review and comparative analysis of machine learning libraries for machine learning. *Discrete and Continuous Models Discrete and Continuous Models 2727(4),* pp. 305. pp. 305. 2019.. DOI: 10.22363/2658-4670-2019-27-4-305-315.

[84] Y. E. Wang, G. Wei and D. Brooks. A systematic methodology for analysis of deep learning hardware and software platforms.

[85] P. Kanani* and D. M. Padole. Deep learning to detect skin cancer using google colab. *IJEAT IJEAT 88(6),* pp. 2176. pp. 2176. 2019.. DOI: 10.35940/ijeat. f8587.088619.

[86] K. S. Zaman, M. B. I. Reaz, S. H. Md Ali, A. A. A. Bakar and M. E. H. Chowdhury. A review. Custom hardware architectures for deep learning on portable devices: A review. 2022-11, . DOI: 10.1109/tnnls.2021.3082304.

[87] M. D. Saiful Islam, S. Jyoti Das, Detection of COVID-19 and pneumonia using deep convolutional neural network. *Computer Systems Science and Engineering Computer Systems Science and Engineering 4444(1),* pp. 519. pp. 519. 2023.. DOI: 10.32604/csse.2023.025282.

[88] S. Singh, M. Kumar, A. Kumar, B. K. Verma, K. Abhishek and S. Selvarajan. Efficient pneumonia detection using vision transformers on chest X-rays. *Sci Rep Sci Rep 1414(1),* 2024. . DOI: 10.1038/s41598-024-52703-2.

[89] A. Shelke, M. Inamdar, V. Shah, A. Tiwari, A. Hussain, T. Chafekar and N. Mehendale. Chest X-ray classification using deep learning for automated COVID-19 screening. 2020.. DOI: 10.1101/2020.06.21.20136598.

[90] K. Sagar and Venu. An ensemble-based approach by fine-tuning the deep transfer learning models to classify pneumonia from chest X-ray images A preprint. 2020..

[91] C. Milho, J. Silva, R. Guimarães, I. C. F. R. Ferreira, L. Barros and M. J. Alves. Antimicrobials from medicinal plants: An emergent strategy to control oral biofilms. *Applied Sciences Applied Sciences 1111(9),* 2021. DOI: 10.3390/app11094020.

8. PROJECT MANAGEMENT

Effective project management was essential for completing this research on improving model generalization for pneumonia detection using chest X-ray images within the required timeframe. At the beginning, a planned Gantt chart (Figure 17) was created to outline the project roadmap. It included major milestones such as the literature review, proposal submission, dataset preparation, model training, results evaluation, and dissertation writing. However, as the project progressed, deviations from this initial schedule became necessary. Challenges such as computational limitations, GPU interruptions in Google Colab, and the need for further testing on external datasets led to adjustments in the original plan.

To account for these changes, an updated Gantt chart (Figure 18) was prepared, reflecting the actual timeline. This revised version incorporated supervisor-related milestones—such as research presentations, draft submissions for Chapters 1–4, and report preparation—into the workflow. Unlike the linear approach in the original plan, the actual implementation was more iterative, particularly during training and evaluation. Tasks such as parameter fine-tuning, layer unfreezing, and cross-dataset testing were repeatedly performed to strengthen model generalization.

Although the project timeline was adjusted, the overall objectives remained unchanged. Additional experiments, including testing with smaller external datasets, were introduced to further validate generalization, while extended training with modified learning rates was carried out to boost accuracy. These modifications were justified as they supported the central goal of enhancing generalization in pneumonia detection through deep learning. By proactively managing risks and adapting plans with supervisor guidance, the project was successfully completed within the revised schedule.

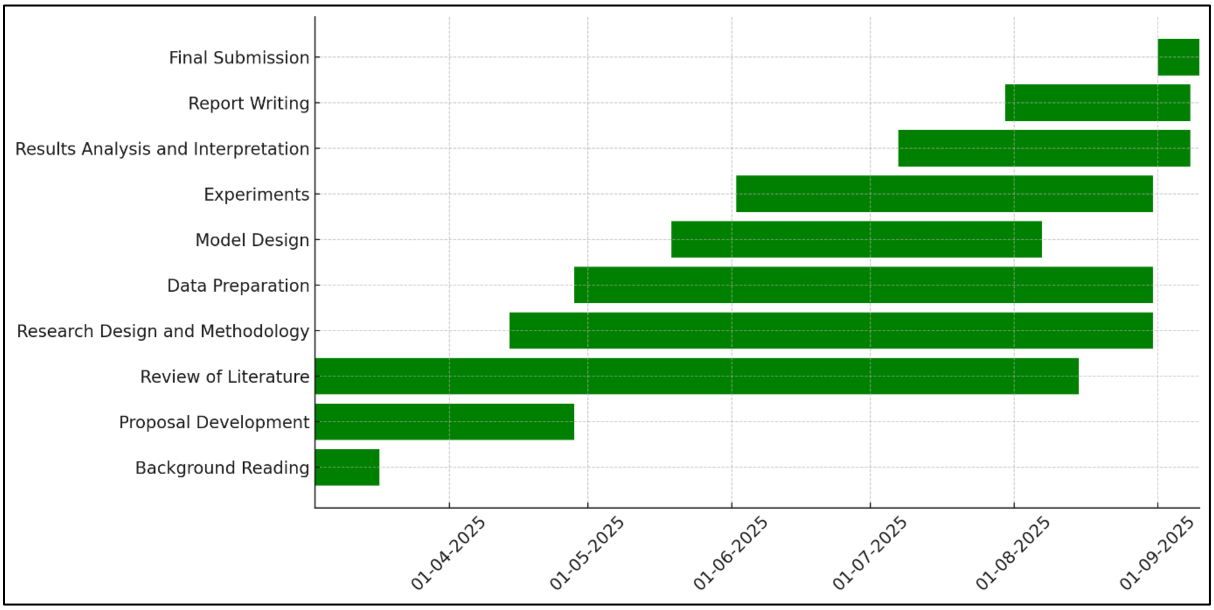

Figure 17 Original Gantt chart

A comparison between the original and actual Gantt charts reveals several important differences in how the project was carried out. The literature review and background study, which were initially scheduled to finish earlier, extended longer than planned because new, relevant studies on deep learning generalization and medical imaging were identified. These additional sources were essential for strengthening the theoretical framework and refining the research gap. As a result, the research problem and objectives required slight adjustments during the early stages, though these revisions were made with supervisor guidance to keep the project aligned with its overall aim.

Another key difference was the overlap between dataset preparation and model training. While the initial plan assumed preprocessing would be completed before training began, in practice, preprocessing, balancing, and augmentation had to be refined iteratively while training ResNet152 and DenseNet201. This overlap was necessary since model performance depended heavily on the quality and balance of the datasets, and adjustments during preprocessing directly influenced training outcomes.

The evaluation and analysis phase also lasted longer than anticipated. This was primarily due to the inclusion of external datasets for testing, which was not a major focus in the original plan but proved vital for assessing generalization. Additional analyses, including ROC curves, confusion matrices, and error rate evaluations, further enriched the study but required extra time. Despite these deviations, the overall project stayed on track, and the adjustments ultimately strengthened the results, supporting the core objective of enhancing model generalization in pneumonia detection.

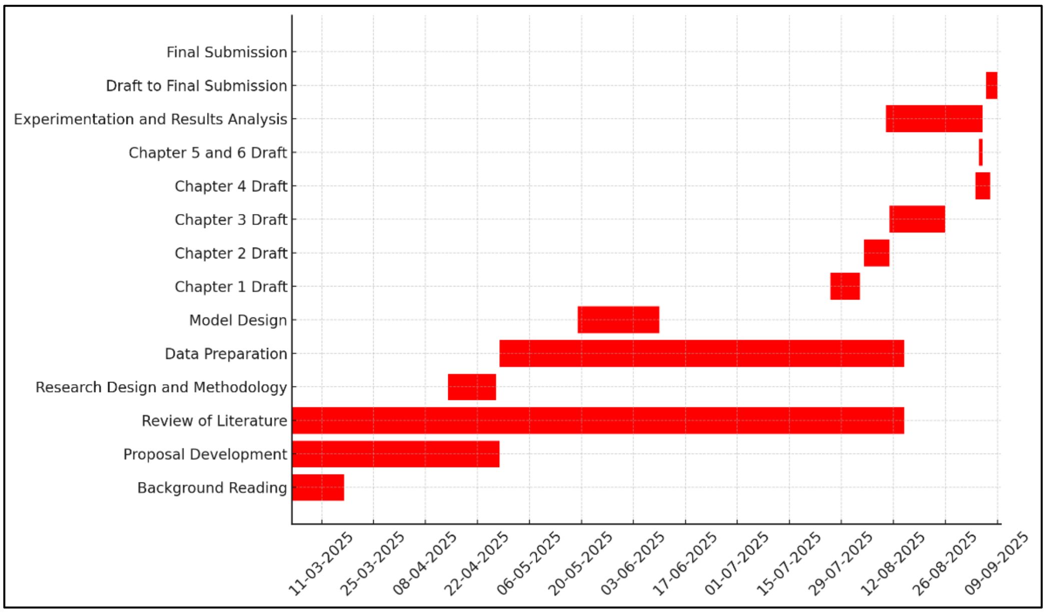

Figure 18 Actual Gantt chart

The actual Gantt chart showed significant deviations from the original schedule, particularly in the experimentation and results analysis stage, which extended beyond the planned timeline. This was largely due to the difficulty of achieving stable generalization across multiple datasets. Training and evaluation required several iterations, especially when tuning hyperparameters such as learning rate, frozen and unfrozen layers, and the number of epochs. Testing on an external dataset added further complexity, as preprocessing and balancing needed refinement to enable fair comparisons. These additional experiments delayed the completion of the results analysis until late July. Similarly, the drafting of early chapters (1, 2, and 3) became more iterative than initially planned, as revisions were made in response to supervisor feedback. Although this pushed the timeline back, ongoing supervision ensured that each chapter met the expected academic standard.

Dataset preparation and model optimization also played a major role in shifting the schedule. While the initial plan assumed straightforward preprocessing, issues such as noise and imbalance in the datasets required additional manual verification and augmentation strategies, causing delays in training. Computational challenges further arose during fine-tuning of ResNet152 and DenseNet201, as the large architectures demanded considerable GPU resources and careful scheduling on Colab Pro. These issues were managed through risk mitigation strategies, including building buffer times, conducting extra data cleaning, and maintaining regular discussions with the supervisor. Despite the delays, these adjustments kept the project aligned with its primary aim of improving generalization in pneumonia detection and ensured that the final results were both reliable and well-validated.