ADOPTION OF SHALLOW NEURAL NETWORKS IN PNEUMONIA CLASSIFICATION

JOSEPHINE MWEYELI CHACHA

MASTER OF SCIENCE IN
COMPUTER SYSTEMS

JOMO KENYATTA UNIVERSITY
OF
AGRICULTURE AND TECHNOLOGY

2026

Adoption of Shallow Neural Networks in Pneumonia Classification

Josephine Mweyeli Chacha

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Computer Systems of the Jomo Kenyatta University of Agriculture and Technology

2026

DECLARATION

This thesis is my original work and it has not been presented for a degree in any other University.

Signed: ……...… **Date:** ……

Josephine Mweyeli Chacha

This thesis has been submitted for examination with our approval as the University Supervisors:

Signed: ……...… Date: ……

Dr. Tobias Mwalili, PhD

JKUAT, Kenya

Signed: ……...… Date: ……

Dr. Henry Mwangi, PhD

JKUAT, Kenya

DEDICATION

This work is dedicated to my husband, Gardy Chacha, and to my children, Ty, Tahj, and Tori. To my husband, thank you for your unwavering prayers, encouragement, and the countless lessons you continue to teach me. Your steadfast love, patience, and support have been my source of strength throughout this journey. I am deeply grateful for your belief in me and for walking beside me every step of the way. To my children, may you always remember that education, though not always easy, is one of the greatest gifts you can pursue. Embrace it with passion, cherish every opportunity to learn, and let excellence be your guiding principle. May your pursuit of knowledge open doors to limitless opportunities and enrich your lives in ways beyond imagination.

ACKNOWLEDGEMENT

I am grateful to God for knowledge, good health, and for making it possible for me to pursue this course. To Africa-*ai*-Japan, your support and belief made it possible for me to achieve the objectives of this research. Thank you for your invaluable contribution to STEM. My heartfelt appreciation goes to my supervisors, Dr. Mwalili Tobias and Dr. Mwangi Henry, for your tireless guidance, availability, constructive input, and continued support. I also extend my gratitude to Dr. Kimwele Michael and the university fraternity for imparting in me values of good stewardship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ACRONYMS AND ABBREVIATIONS

AI	Artificial Intelligence
API	Application Programming Interface
BCE	Binary Cross-Entropy
CAM	Class Activation Mapping
CNN	Convolutional Neural Network
FN	False Negative
FP	False Positive
GPU	Graphics Processing Unit
HRH	Human Resources for Health
IRMA	Image Retrieval in Medical Applications
LWA	Lightweight Architecture
ML	Machine Learning
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
RMSprop	Root Mean Square Propagation
TN	True Negative
TP	True Positive
TPU	Tensor Processing Unit
WHO	World Health Organization
X-ray	X-radiation

ABSTRACT

In low-resource healthcare environments, limited access to radiologists and a high computational burden of conventional deep learning models are significant challenges for pneumonia diagnosis. The current Convolutional Neural Networks (CNNs) for chest X-ray classification demand high memory, computation and specific hardware, which is not feasible in under-resourced hospitals and clinics. In this work, the authors explored the possibility of a lightweight shallow CNN producing reliable pneumonia classification without being too computationally expensive. The architecture proposed was comprised of three convolutional layers to reduce the amount of computational and memory resources without compromising the diagnostic performance. The model was tested on a set of common experimental settings, with the other popular lightweight and deeper CNN models. To compare the performances fairly, all the models were optimized in the same way, trained by the same augmentation methods and assessed by the same evaluation metrics. It involved two data sets: the first was a secondary benchmark data set from publicly available chest X-ray repositories, and the second was a Kenyan primary data set collected from health care facilities. The accuracy, precision, recall, F1-score, and Type I and Type II error rate were used as evaluation metrics. The proposed model was found to be 91% accurate on the secondary benchmark data set and 95% accurate on the primary data set of the Kenyan. The model showed better stability of convergence, reduced overfitting and reduced majority-class bias in comparison with deeper architectures. The study proposes a framework for CNN that is efficient in terms of computation and scalable, which is suitable for resource-limited healthcare systems, where there are limited GPUs, memory, and a lack of digital infrastructure in low-resource clinical environments.

Keywords: Pneumonia, Lightweight, Imaging, Resource-Constrained, Artificial Intelligence.

CHAPTER ONE

INTRODUCTION

1.1 Background

Computer vision techniques have become important in applications that require image detection and classification (Khan et al., 2021) on. Image detection entails the use of computing technology to determine the presence of objects in an image whereas image classification refers to the labeling of images in any one of the predefined semantic categories(Kaur & Singh, 2022). The main goal of image classification is to accurately identify features present in an image. There exist two types of classification; supervised and unsupervised. In supervised classification, the classifier makes use of a trained database and human intervention; whereas in unsupervised, the classifier does not have a learner(Verma et al., 2022).

When using an image classifier system, it is paramount to choose a suitable classification system and enough training samples(Huang et al., 2023; Meena & Suriya, 2020). According to Maier et al, (2019), the classification system consists of two of phases as demonstrated by figure 1.1 below.

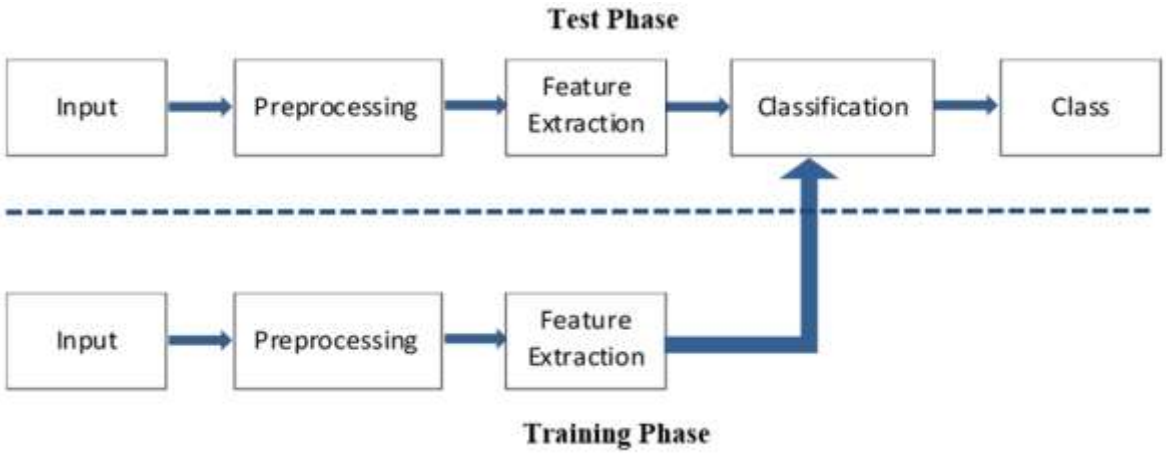


Figure 1.1: Phases in image Classification

The workflow of image classification neural networks is split into two distinct phases; the training phase and the testing phase; that transform raw pixel data into categorical predictions. The initial phase involves feeding the neural network with raw data in form of pixels. The raw pixels are then preprocessed. Preprocessing ensures numerical stability during training and enhances the model's ability to generalize, since it artificially expands the training dataset's diversity. This stage entails operations such as pixel value normalization that are typically scaled to ranges, image resizing to ensure standardized dimensions and data augmentation techniques such as rotation of image, randomly cropping image samples, flipping among others. The feature extraction phase follows the preprocessing. It constitutes the core computational engine of image classifiers. In convolutional neural networks, this phase is implemented by stacking convolutional layers which progressively build hierarchical representations(Wu et al., 2019).

During classification, features are first categorized into objects of predefined classes and then a suitable method is used to compare the image patterns with targeted patterns (Kamavisdar et al., 2013) . Several classical techniques exist that can be used in image classification. Among them being remote sensing(Li et al., 2024). The classical image classification methods such as Neighbor, IRMA, fuzzy scheme, support vector machine, depend on manual feature extraction and predefined similarity measures(Sheykhmousa et al., 2020; Yogapriya et al., 2018). Consequently, the

performance of these models is heavily reliant on domain expertise. The manual feature extraction may also fail to capture complex patterns present in medical images. To address these limitations, machine learning models were introduced, integrating the use of automated feature extraction and classification in a unified model(Huang et al., 2023).

Unlike classical approaches, machine learning models learn patterns from data rather than relying on manual feature extraction(Guetari et al., 2023; Holzinger, 2019). In conventional machine learning for image classification, the process typically consists of two main modules: a feature extraction module that identifies relevant characteristics from the image, and a classification module that assigns labels based on the extracted features(Krishna et al., 2018; Wang, 2024). Deep learning, a sub field of machine learning, directly learns hierarchical representations from raw data(Sarker, 2021). Features learnt are essential for activities like object identification, facial recognition among others. An example of the deep learning model is the Convolutional Neural Networks (CNNs)(Latif et al., 2019).

According to Abiyev and Ma'aitah (2018), a convolutional neural network imitates the deep structure of the human brain. These deep structures provide abstraction at the different levels of the same. Convolution can therefore be described as "*a specialized type of linear operation used for feature extraction, where a small array of numbers, called a kernel is applied across the input, which is an array of numbers called* tensor" the feature map, which is the output tensor of an element-wise product between the kernel and the image input tensor. The product is then summed up to give the output tensor (Yamashita et al., 2018).

CNN is best suited in recognizing visual patterns from pixel images with variability(Wang, 2024). It can be applied in a broad-spectrum area with the common objective being the ability to automatically learn features from datasets. Once the features have been learned, they can be used for tasks such as classification(Gonzalez, 2018). CNN consists of multilayered classes that use single neural networks that take raw image pixel values from the training data to the end of classification of outputs. Some of the application areas for convolution neural networks are; vision related tasks

such as image classification, object detection, scene labeling, house number digit classification, and face recognition. They have been applied in other areas such as speech recognition as well as video classification(Reza et al., 2026).

Pneumonia is a group of syndromes brought about by a variety of organisms resulting in varied manifestations and sequelae that causes the infection of the lung parenchyma (Jain et al., 2021). When someone breathes in, air that is rich in oxygen moves inside the body through the airways, that is trachea, bronchi, and bronchioles that are found in the lungs. Alveoli are very small air sacs that are at the end of the airways where oxygen is exchanged for carbon dioxide. It is after this exchange that the carbon dioxide is exhaled.

Infected alveoli become inflamed and fills up with fluid(puss), thus making the patient exhibit symptoms such as phlegm (“wet cough”), difficulty to breathe, fever, chest pain, fatigue, as well as confusion (Funaguchi et al., 2023). As such, pneumonia is an inflammation of the lungs' airspaces that is commonly caused by infection; with some cases of pneumonia being life threatening (Long et.al., 2022). This research proposes using a shallow convolution neural network in the detection and classification of pneumonia.

Currently, diagnosis of chest X-rays images of a patient is done by a radiologist who examines the X-ray images under luminous light to determine if a patient exhibits symptom of pneumonia (National Heart Lung and Blood Institute, 2022). According Wootton and Feldman (2014), radiologists may often find it difficult to interpret results, and at times the same may be frequently inconsistent. Research done by Davenport and Kalakota (2019) looks at how artificial intelligence technologies in the medical industry could change different aspects of patient care. In another study done by Fogel and Kvedar (2018), it is seen that artificial intelligence may enable better disease monitoring, early detection of diseases, allow for improved diagnosis, unearth unusual treatment as well as create a new era of personalized medicine.

Globally, Sub-Saharan Africa has the highest under-five mortality rate globally, with approximately one in every thirteen children dying before reaching their fifth birthday. This rate is fourteen times higher than that of high-income countries (World Health Organization, 2022). An estimated 490,000 pneumonia-related deaths occurred in the year 2015. This accounts for half of the total number of pneumonia related deaths globally (Dharel et.al., 2023). In Kenya, pneumonia is the second leading cause of death among children under the age of five, contributing to approximately 16% of total deaths in this age group. Due to these high mortality rates, Kenya ranks among the top fifteen countries globally with the highest number of under-five pneumonia-related deaths(Tonui et al., 2025).

In Kenya, the availability of resources remains a major challenge for hospitals, significantly affecting their capacity to deliver quality healthcare services(Kiyasseh et al., 2022; Ndung'u, 2025). These limitations manifest across multiple areas, including inadequate infrastructure, shortages in human resources for health (HRH), limited access to essential medications, and insufficient critical medical equipment(McKnight et al., 2026).

Such constraints weaken diagnostic efficiency and delay timely clinical decision-making, particularly in resource-constrained settings. In response to these systemic challenges, this research proposes the adoption of computerized medical diagnosis using Convolutional Neural Networks (CNNs). By using automated image analysis and pattern recognition capabilities, CNN-based systems can support healthcare professionals in making faster and more accurate diagnostic decisions, thereby enhancing service delivery even within under-resourced health environments.

1.2 Problem Statement

Pneumonia is the cause of under-5 mortalities in under developed regions where access to timely diagnosis is limited(McAllister et al., 2019). Chest x-ray imaging is one of the common and cost-effective tools for diagnosing pneumonia. According to, Sarvamangala and Kulkarni, (2022), Convolutional Neural Networks (CNNs) have demonstrated strong performance in medical image classification, including

pneumonia. However, existing CNN architectures such as Inception require large memory capacity and powerful computational resources such as Graphics Processing Units (GPUs) or Tensor Processing Units (TPUs)(Taib et al., 2026).

These requirements are costly and inaccessible in many contexts; limiting their deployment on mobile or portable device and in low-resource clinical settings, where computation power and memory is limited(Korkmaz, 2025). Lightweight Architectures (LWAs) have been proposed to reduce the computational and memory demands of CNNs by lowering parameter counts and learning speeds without significantly sacrificing performance. This makes them suitable for real-time applications in environments with limited resources(Mohsin et al., 2025; Shahriar, 2025). Addressing this gap is critical to enabling scalable and accessible AI assisted pneumonia diagnosis in under resourced healthcare facilities.

1.3 Justification of the Study

The rates of health loss from both diarrheal diseases and lower respiratory infections such as pneumonia have fallen by nearly 50%, though the burden of these diseases among very young children is still high(Troeger et al., 2020). Globally, pneumonia is among the leading causes of death in children under five. It is estimated that the mortality rate in children under the age of 5 is 64% and 10% of these deaths are as a result of lower respiratory infection which is attributed to weak and under-resourced health systems(Selvi & Vaithilingan, 2024). In the year 2017, 5.4 million children under the age of five succumbed to death; with more than half of these deaths arising from conditions that were preventable or treatable with access to simple and affordable interventions (World Health Organisation, 2022).

1.4 Objectives

1.4.1 General Objectives

The general objective of the study was to explore the adoption of shallow neural networks in classification of pneumonia.

1.4.2 Specific Objective

The specific objectives for this research were:

(i) To analyze existing light weight and deep learning models for medical imaging classification
(ii) To design a lightweight, shallow neural network model for the classification of pneumonia chest radiographs
(iii) To evaluate the models' performance on a diverse dataset of chest radiograph to determine its accuracy and reliability

1.5 Research Questions

(i) How do existing lightweight models perform compared to deeper architectures in medical imaging?
(ii) How can lightweight models be designed to effectively classify pneumonia radiographs?
(iii) How does the proposed model generalize across different chest x-ray datasets?

1.6 Scope of the Research

The scope encompassed the development and evaluation of shallow convolutional neural networks model with three convolution layers; with the focus on suitability for deployment on resource-constrained environments. The study made use of two secondary datasets from Kaggle public repository. The primary dataset was from Guangzhou Women and Children's Medical Center dataset by Kermany et al., (2018) and supplementary dataset was by Chowdhury et al., (2020), whereas primary data was collected from Kenyan healthcare facilities.

1.7 Summary

The integration of computer vision, particularly through Convolutional Neural Networks (CNNs), has revolutionized medical imaging by automating the identification and classification of complex features from raw pixel data. While

classical methods are heavily dependent on manual feature extraction and domain expertise, CNNs utilize hierarchical representations to learn patterns directly from datasets. High mortality rates associated with pneumonia; specially among children under five in regions like Kenya; highlight a critical diagnostic gap caused by inadequate infrastructure and the high computational demands of existing deep learning architectures. Standard models often require expensive Graphics Processing Units (GPUs), which are frequently inaccessible in resource-constrained clinical settings.

In response to these systemic challenges, this research proposes the development and evaluation of a shallow, lightweight neural network featuring three convolution layers. The study aimed at designing a model that maintained a high diagnostic accuracy while significantly reducing the memory and computational resources required for deployment on portable devices. By leveraging a diverse dataset from public repositories and primary data collected from Kenyan healthcare facilities, this research sought to provide a scalable, AI-assisted solution that enhances timely clinical decision-making in under-resourced environments.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter presents a review of existing literature to contextualize the study within the evolving fields of computer vision and medical diagnostics. It aligns with the overall objective of exploring the adoption of shallow neural networks for pneumonia classification. The chapter begins by examining the pathogenesis of pneumonia, followed by an overview of the evolution of image classification techniques, highlighting the transition from traditional machine learning approaches to modern deep learning paradigms.

Subsequent sections analyze existing lightweight and deep learning models, with a focus on their performance in medical imaging tasks. In addition, the review explores custom convolutional neural network (CNN) architectures and transfer learning approaches, providing the theoretical and empirical foundation necessary for designing a lightweight, shallow model.

The chapter further establishes key evaluation criteria for assessing model accuracy and reliability across diverse datasets. Ultimately, the review identifies existing research gaps; particularly the need for computationally efficient models that do not rely on resources such as GPUs; and proposes a conceptual framework to guide the development of resource-efficient diagnostic solutions.

2.2 Pathogenesis of Pneumonia

Pneumonia arises from an infection of the lower respiratory tract that involves the pulmonary parenchyma. This is because the lungs react to these foreign microbes thereby causing an inflammatory response resulting in bronchioles and alveoli filling up with fluid and becoming solid. The severity of pneumonia ranges from mild and uncomplicated typical infections, to fulminant and life-threatening (Physiopedia,

2021). There exist two types of pneumonia namely, bronchopneumonia/lobular pneumonia and lobar pneumonia.

Bronchopneumonia is described by suppurative inflammation localized in patches around bronchi that may or may not be localized to a single lobe of the lung(Tian et al., 2020). "*Lobar pneumonia is diffuse consolidation involving the entire lobe of the lung"* whose evolvement is broken down into 4 stages namely: Congestion stage that is characterized by grossly heavy and boggy appearing lung tissue, diffuse congestion, vascular engorgement, and the accumulation of alveolar fluid rich in infective organisms(Miyashita, 2022).

The second stage is red hepatization that consists of marked infiltration of red blood cells, neutrophils, and fibrin into the alveolar fluid. At this stage the lungs appear red and firm like a liver. The third stage, gray hepatization, the red blood cells break down. This is associated with fibrinopurulent exudates that cause a red to gray color transformation. In the final stage, resolution, there is clearing of the "*exudates by resident macrophages with or without residual scar tissue formation"*(Jain et al., 2021).In figure 2.1 below, Grousd (2019) illustrated the pathogenesis of Pneumonia.

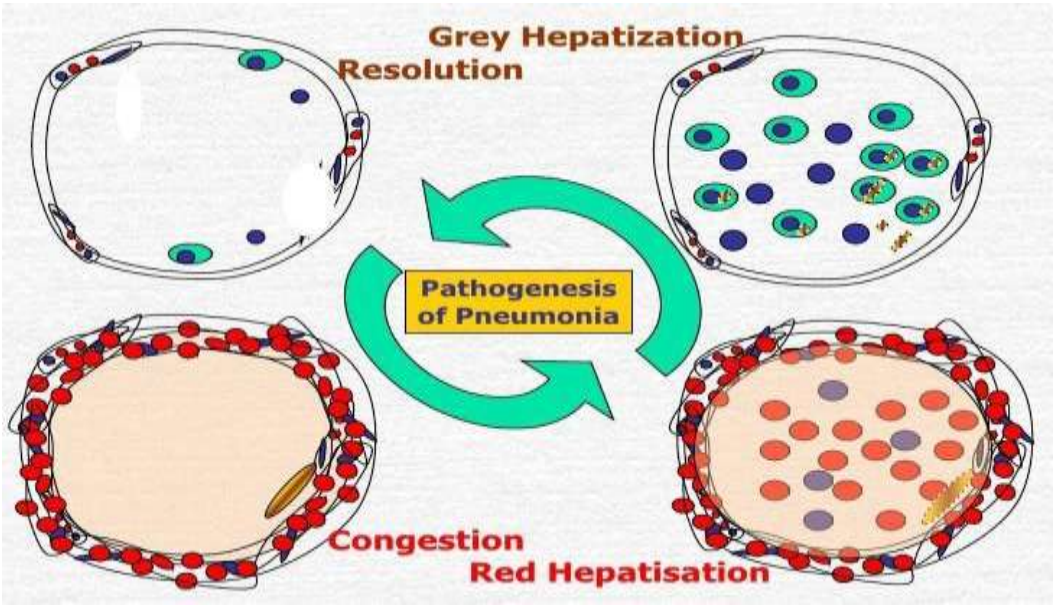


Figure 2.1: Pneumonia Pathogenesis

Some of the classification of how pneumonia may be contracted are; according to the etiology, clinical setting in which the patient acquired the infection, and the pattern of involvement of lung parenchyma, among other classifications (Jain et al., 2021). According to etiology, pneumonia can be caused by either bacteria or a virus. In adults, bacteria are the most common cause of pneumonia, whereas in children under the age of 5, pneumonia is caused by either viruses or bacteria that reside in the child's nose or throat.

When this virus or bacteria travels into the lungs, the lung may become infected. The other way in which a person is infected with pneumonia is when a patient who has the same illness coughs or sneezes, around an uninfected person, and this person inhales the pneumonia airborne droplets may in turn end up with pneumonia. Below in figure 2.2 by Thompson (2016) is an image that depicts alveoli that is filled with pus(Aishwarya et al., 2024; Dhand & Li, 2020).

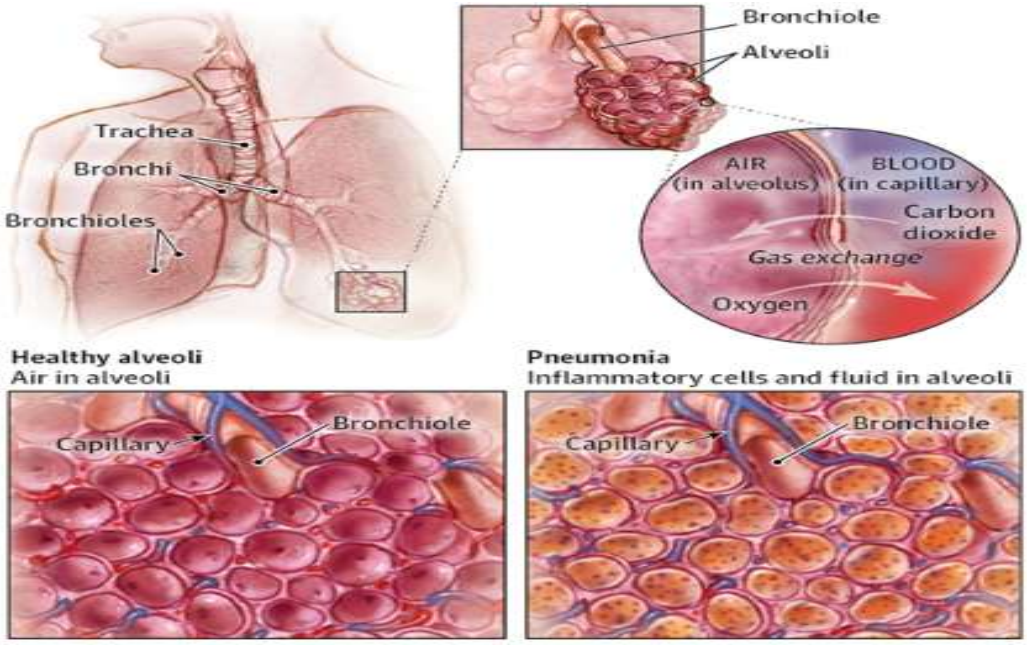


Figure 2.2: Image of Healthy alveoli and Infected alveoli

Classifications according to the clinical setting in which the patent acquired the infection include; Community-Acquired Pneumonia that entails any pneumonia acquired outside of a hospital but within a community setting, Hospital-Acquired

Pneumonia than constitutes pneumonia acquired within 48 hours after being admitted as an inpatient to the hospital; whereas Ventilator Associated Pneumonia is pneumonia acquired after 48 hours in being in endotracheal intubation (Jain et al., 2021).

Anyone can get pneumonia; however, certain groups pose a higher risk of developing the disease; these are children under the age of 2 years and adults older than 65 years. Other factors that pose a threat to getting infected with pneumonia are; a weak immune system, presence of chronic lung disease such asthma or cystic fibrosis and chronic health problems, such as diabetes or heart disease. The choice of lifestyle also determines the risk of getting pneumonia, people that smoke cigarettes pose a greater risk of getting pneumonia compared to their counterparts who don't(Thompson, 2016). Environmental factors such as indoor air pollution that may be caused by cooking and heating with biomass fuels (such as wood or dung) and living in crowded homes increases the risk of one getting pneumonia(WHO, 2019).

Some of the medical practitioners that are needed in diagnosis and treatment of pneumonia include the nursing staff who records temperature variations and other vitals that are needed by the physician before any diagnosis is to be undertaken. The pharmacist provides the right dosage of antibiotics needed by the patient. The radiologist is a key participant in the diagnosis and treatment process because he/ she needs to perform the radiological X-RAY as well as interpret the findings in the different types of pneumonia since they vary considerably.

2.3 Evolution of Image Classification Techniques

2.3.1. Traditional Machine Learning Approaches

Prior to the deep learning era, classical image classification methods relied heavily on manual feature engineering. (Firdaus, 2024), discussed the application of K-Nearest Neighbor (KNN) in medical imaging, noting its simplicity but limited scalability with high-dimensional data. Similarly, Support Vector Machines (SVM) have been utilized for chest X-ray analysis, with (Joo et al., 2021) and (Mardianto et al., 2024),

highlighting their effectiveness in binary classification tasks despite sensitivity to parameter tuning.

A critical distinction in these traditional methods is feature extraction; Latif et al. (2019) emphasize the shift from manual to automated feature engineering, noting that traditional methods often struggle to capture the complex hierarchical patterns required for robust pneumonia detection compared to modern neural approaches. This is summarized in figure 2.3 below.

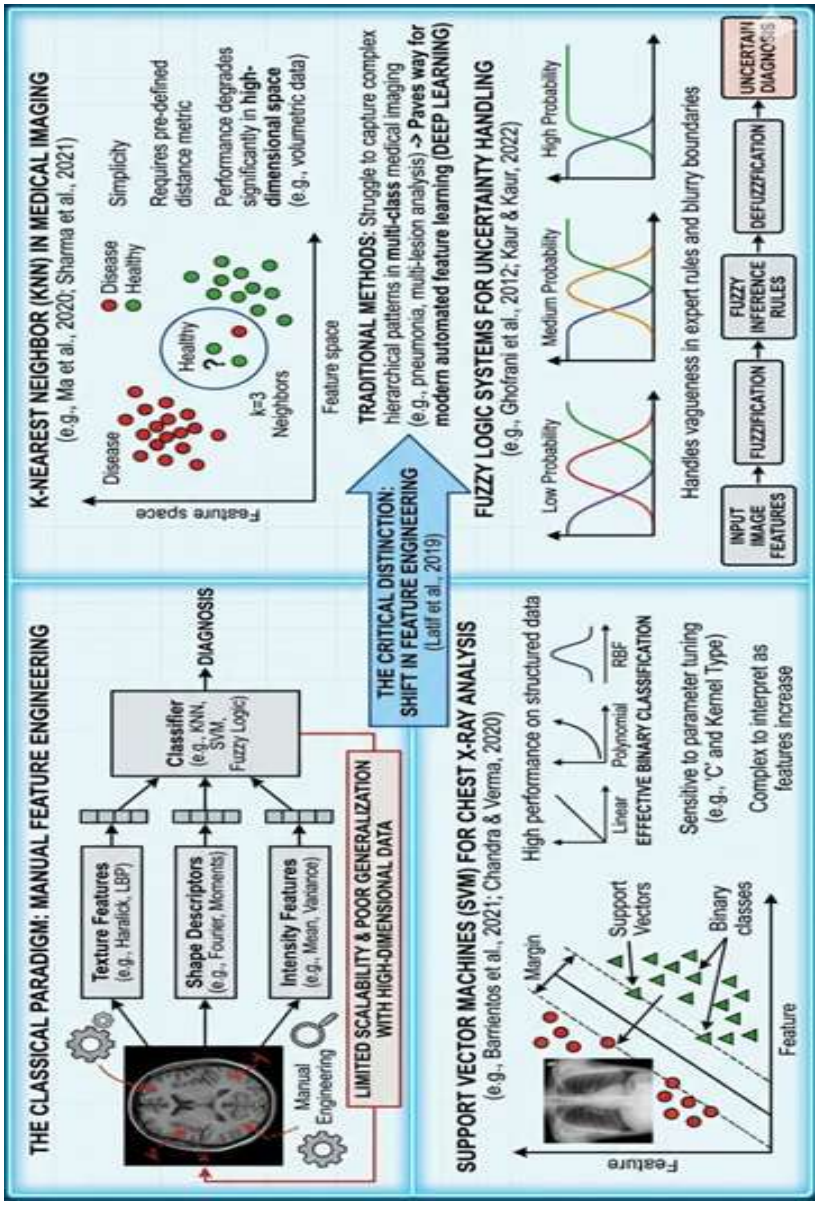


Figure 2.3: Evolution of Medical Image Classification

K-Nearest Neighbors operates on the principle that similar images should belong to similar classes, classifying new instances based on the majority class among its K closest neighbors in feature space(Firdaus, 2024). While conceptually simple and non-parametric, KNN suffers from computational inefficiency during testing and sensitivity to irrelevant features.(Okundalaye, 2025) applied KNN to chest X-ray classification but reported limited performance compared to more sophisticated approaches.

Support Vector Machines construct optimal hyperplanes that maximize the margin between classes in a transformed feature space.(Kurdi et al., (2023) evaluated SVM with linear, polynomial, and RBF kernels for brain tumor classification, demonstrating that parameter optimization using metaheuristic algorithms could substantially improve performance achieving over 96% sensitivity and specificity after optimization. However, SVM performance depends critically on kernel selection and hyper parameter tuning, and the method does not automatically learn relevant features from raw data.

Fuzzy logic systems have also been explored for medical image classification, offering the advantage of handling uncertainty inherent in medical diagnosis. (Abdulsahib et al., 2021) proposed a fuzzy-based scheme that extracted part-wise membership degrees, recognizing that different image regions contribute unequally to classification decisions. (Varela-Santos & Melin, 2021) extended this approach by integrating fuzzy inference with texture analysis for lung disease detection. Despite these innovations, traditional methods consistently underperform compared to deep learning approaches on large-scale medical image datasets because they cannot learn hierarchical feature representations directly from data(Litjens et al., 2017).

2.3.2 Deep Learning in Medical Imaging

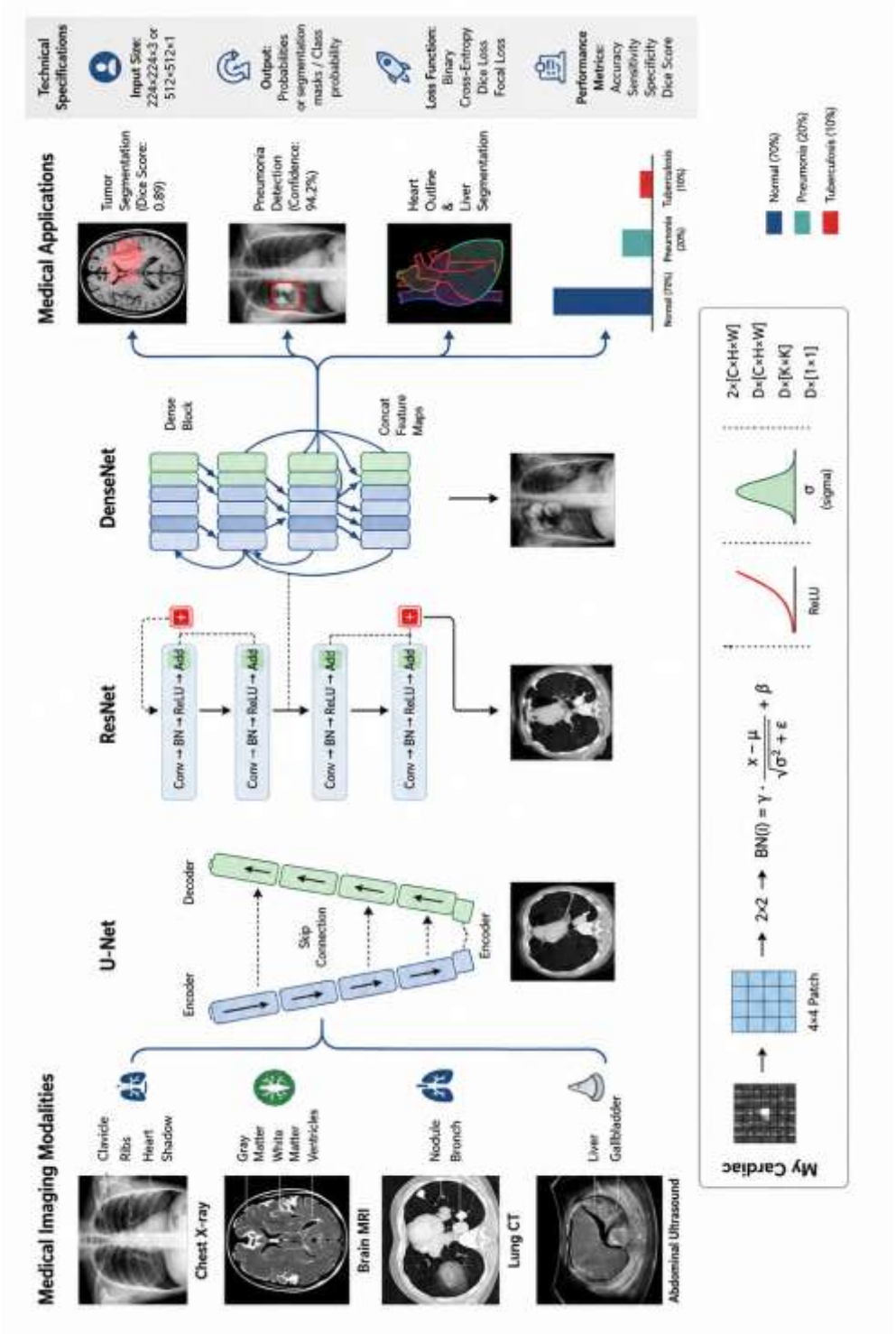
The resurgence of neural networks, catalyzed by Krizhevsky et al (2012) landmark ImageNet result, transformed computer vision and, subsequently, medical image analysis. Deep learning offers a paradigm shift: instead of manually engineering features, neural networks learn hierarchical representations directly from pixel data,

with lower layers detecting edges and textures and higher layers combining these into semantically meaningful patterns (Trigka & Dritsas, 2025).

Lundervold & Lundervold, (2019) comprehensively reviewed the impact of deep learning on medical imaging, noting that convolutional neural networks have achieved state-of-the-art performance across diverse tasks including classification, segmentation, and detection. (Dai et al., 2021) argued that deep learning networks could match or exceed human expert performance in specific diagnostic tasks, including dermatological cancer classification and diabetic retinopathy grading. The key advantage lies in the ability to learn task-specific features from large datasets, eliminating the need for handcrafted feature engineering while often discovering patterns imperceptible to human observers.

Several deep architectures have become standard in medical imaging. Figure 2.4 below show the Deep CNN Architecture in medical imaging. Krizhevsky et al. (2017) introduced AlexNet, pioneering deep learning for image classification. The VGG family is noted for its simplicity and depth (Shah et al., 2023), while ResNet introduced residual connections to address network degradation (Duta et al., 2021). Huang et al. (2023) proposed DenseNet for feature reuse and parameter efficiency, and Szegedy et al. (2015) developed Inception/GoogleNet for multi-scale feature extraction. For efficiency, Howard et al. (2017) designed MobileNet using depth-wise separable

convolutions, and Tan and Le (2019) utilized neural architecture search to create EfficientNet for optimal



scaling.

Figure 2.4: Deep CNN Architecture for Medical Imaging from Pixel to Diagnosis

The evolution of deep CNN architectures has been driven by efforts to improve accuracy while managing computational demands. According to Khan et al., (2020), modern deep CNN paradigm with five convolutional layers and three fully connected layers, have achieved breakthrough ImageNet performance through ReLU activation, dropout regularization, and GPU implementation. The VGG family demonstrated the power of architectural simplicity and depth, using stacks of 3×3 convolutions to achieve receptive fields equivalent to larger kernels while reducing parameters as shown in figure 2.5 below. VGG-16 and VGG-19, with 16 and 19 weight layers respectively, became popular feature extractors for transfer learning despite their substantial parameter counts(Shah et al., 2023).

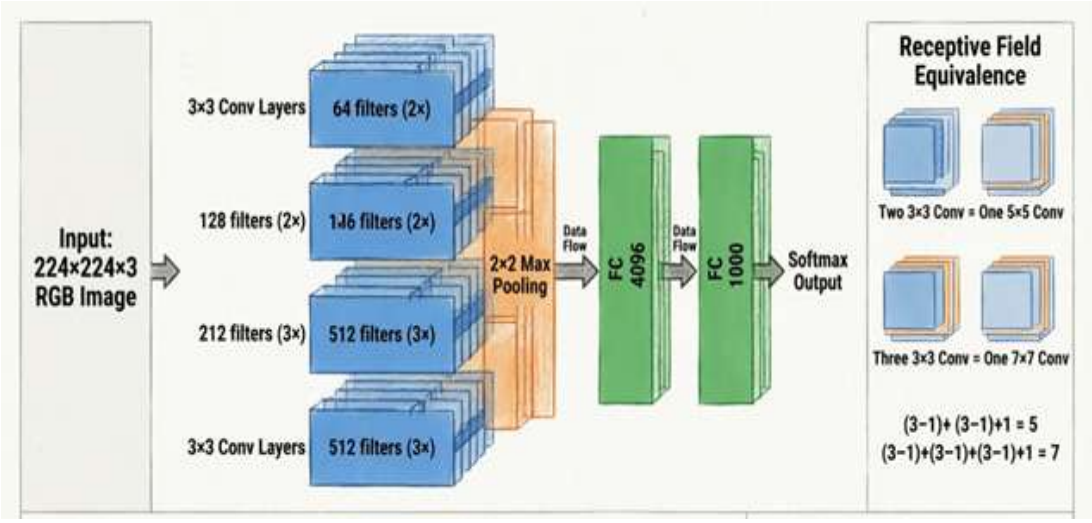


Figure 2. 5: Deep CNN with Stacked 3×3 Convolutions and Receptive Field Equivalence

Additionally, (Veloyan, 2025), introduced residual connections that add layer inputs directly to outputs, enabling networks exceeding 100 layers by mitigating vanishing gradients. This innovation allowed unprecedented depth ResNet-152 achieved state-of-the-art ImageNet accuracy while remaining trainable as shown in figure 2.6 below. (Köpüklü et al., 2019) extended this idea by connecting each layer to all subsequent layers, promoting feature reuse and parameter efficiency. Inception architectures employed parallel convolutions with different kernel sizes, enabling multi-scale feature extraction within single layers(Muhammad & Aramvith, 2019).

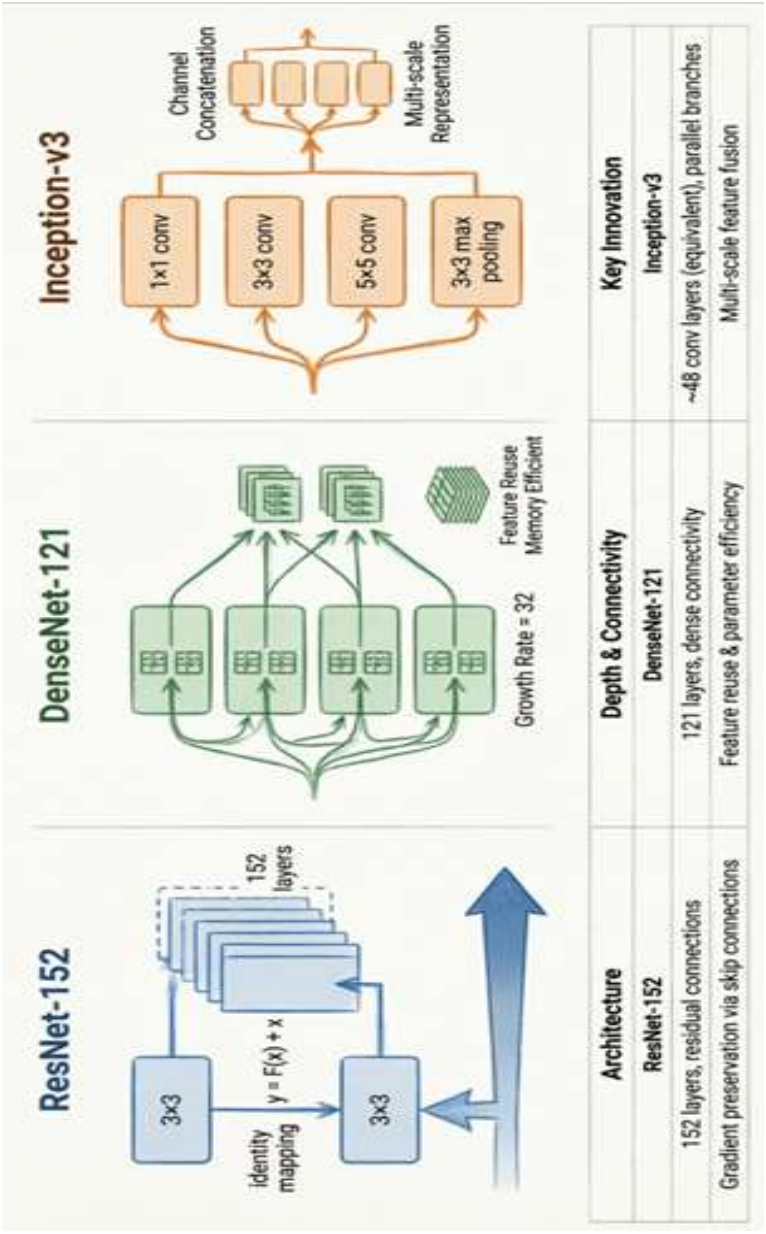


Figure 2. 6: Structure on Pebbled Surface with Shallow Depth of Field

2.3.3 Shallow vs. Deep Neural Networks

Defining the depth of a network is critical; Meir et al., (2023) establish criteria for shallow versus deep classification based on the number of hidden layers. Theoretical perspectives include the universal approximation theorem, which suggests shallow networks can approximate any function given sufficient width (Hornik, 1991), though Zhang et al. (2021) argue for capacity versus generalizability trade-offs in deeper

models. Geman et al. (1992) discusses bias-variance decomposition, noting that shallow architectures may suffer from high bias. However, Bianco et al. (2018) highlight the computational efficiency of shallow networks, and Brigato et al., (2022) note their reduced training time and data requirements. Kulkarni et al., (2021) emphasize deployment feasibility in resource-constrained environments. This is summarized in figure 2.7 below.

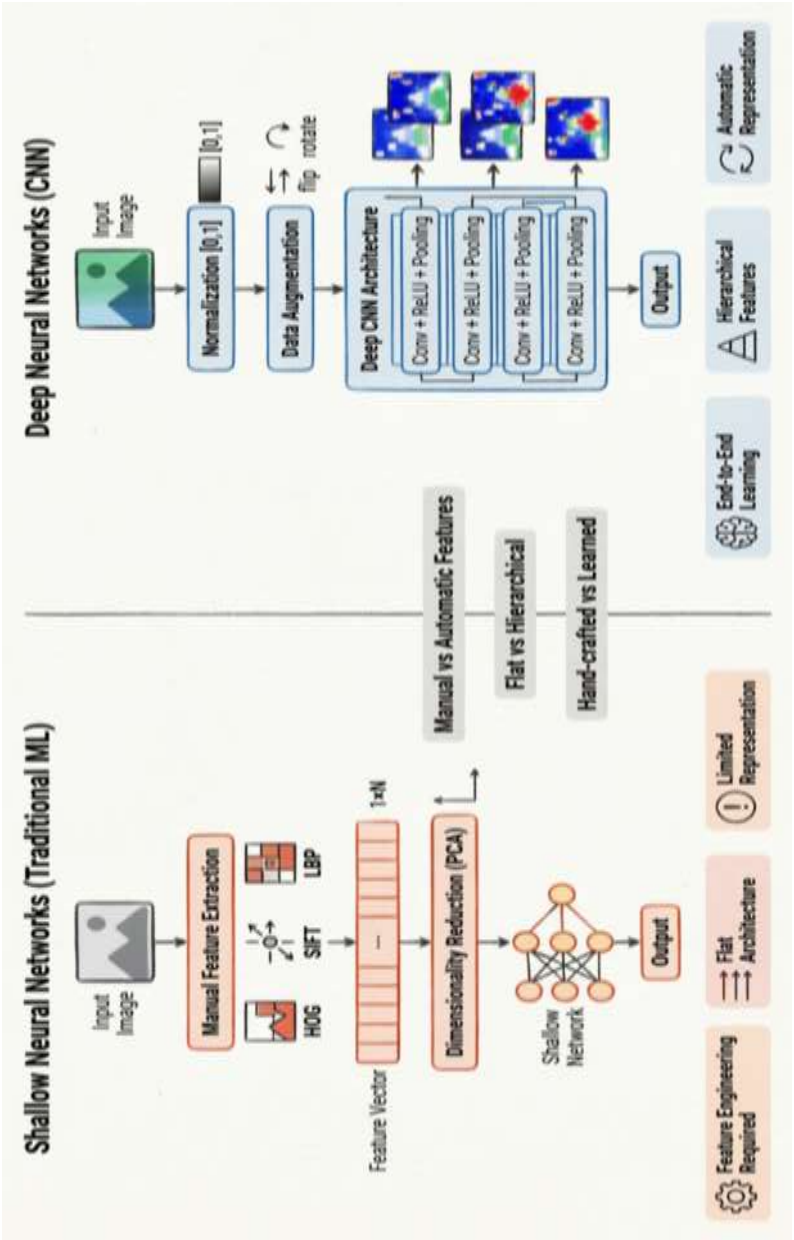


Figure 2. 7: Comparison of Image Preprocessing Pipelines: Shallow Neural Networks vs. Deep CNN

The distinction between shallow and deep networks, while intuitively based on layer count, carries theoretical and practical implications. Shallow networks typically comprise fewer than five to six layers, while deep networks may extend to hundreds of layers (Noman, 2020). Mhaskar et al. (2017) provided a theoretical analysis demonstrating that deep networks can represent certain function classes exponentially more efficiently than shallow networks.

The universal approximation theorem guarantees that a single hidden layer network with sufficient neurons can approximate any continuous function (Kidger & Lyons, 2020). However, this theoretical result does not address learnability or parameter efficiency. Zhang et al. (2021) analyzed the capacity-generalization trade-off, noting that deeper networks can memorize training data more easily but may overfit without appropriate regularization. Geman et al. (1992) framed this in bias-variance terms: shallow networks impose higher bias (stronger assumptions about function form) but lower variance, while deep networks reduce bias at the cost of increased variance.

Practically, shallow networks offer compelling advantages for resource-constrained deployment. Bianco et al. (2018) systematically compared computational efficiency across architectures, finding that shallow networks require substantially fewer floating-point operations (FLOPs) and parameters. Brigato and Iocchi (2022) demonstrated that shallow networks trained from scratch can outperform transfer learning from ImageNet-pretrained deep networks when target datasets are small a common scenario in medical imaging. Howard et al. (2017) developed MobileNet specifically for mobile deployment, using depth wise separable convolutions to dramatically reduce parameter counts while maintaining reasonable accuracy.

Interpretability considerations also favor shallow architectures. Ribeiro et al. (2016) argued that model explanations are essential for high-stakes domains like medicine, and simpler models are inherently more interpretable. Sutera (2025) directly challenged the "deeper is better" assumption in medical imaging, demonstrating that shallow networks can achieve comparable or superior performance to deep networks like ResNet-50 on pneumonia classification tasks while improving accuracy by up to

27% and dramatically reducing computational costs. This finding directly motivates the present investigation.

For mobile and edge deployment, specialized efficient architectures emerged. MobileNet (Howard et al., 2017) used depth wise separable convolutions to reduce computations by 8-9 times compared to standard convolutions with minimal accuracy loss. EfficientNet (Tan & Le, 2019) systematically scaled network dimensions using neural architecture search, achieving state-of-the-art accuracy with superior efficiency.

Recent lightweight innovations include ResMHCNN blocks, which combine residual multi-head convolution with channel attention to achieve strong performance with minimal parameters as few as 0.226 million for breast cancer classification (Sarker et al., 2025). These architectures demonstrate that carefully designed shallow networks can rival deep network performance while remaining deployable in resource-constrained environments.

2.4 Theoretical Approach

2.4.1 Technology Acceptance Theory (TAM)

The Technology Acceptance Model was developed by Fred Davis to explain how users come to accept and utilize new technologies within organizational and professional environments. The theory posits that two major determinants influence adoption behavior: perceived usefulness and perceived ease of use (Davis, 1989). Perceived usefulness refers to the degree to which an individual believes that using a particular system would improve job performance, while perceived ease of use relates to the degree to which a person believes that using the system would require minimal effort.

TAM has been widely applied in healthcare informatics, artificial intelligence systems, and clinical technologies because it provides a framework for understanding healthcare practitioners' willingness to integrate computerized systems into clinical workflows (Venkatesh & Davis, 2019). In healthcare environments, particularly in low-resource settings, systems that are computationally efficient, accessible, and easy to operate are more likely to achieve acceptance among clinicians and healthcare institutions.

The present study relates to the Technology Acceptance Model by proposing a lightweight shallow Convolutional Neural Network (CNN) architecture designed specifically for resource-constrained healthcare environments. The proposed model emphasizes computational efficiency, reduced memory requirements, and faster inference times, thereby improving perceived ease of use and operational feasibility in hospitals with limited digital infrastructure.

In addition, the ability of the model to provide reliable pneumonia classification supports perceived usefulness because it assists clinicians in making timely diagnostic decisions. Unlike computationally intensive deep learning architectures requiring expensive Graphics Processing Units (GPUs), the proposed shallow CNN can potentially be deployed on low-cost systems and portable devices, increasing accessibility and adoption in under-resourced healthcare facilities. Therefore, TAM provides a theoretical justification for the adoption of lightweight AI-assisted diagnostic systems within clinical environments characterized by limited technological resources.

2.4.2 Clinical Decision Support Systems Theory

Clinical Decision Support Systems theory focuses on the integration of computerized systems into healthcare processes to enhance clinical decision-making, improve diagnostic accuracy, and support healthcare practitioners during patient management. Clinical Decision Support Systems (CDSS) are designed to provide healthcare professionals with intelligently filtered clinical information, recommendations, and diagnostic assistance at appropriate stages of patient care (Berner, 2019). These systems utilize patient data, medical knowledge, and computational algorithms to generate outputs that assist clinicians in making evidence-based decisions.

According to Sutton et al. (2020), modern AI-powered CDSS applications increasingly incorporate machine learning and deep learning approaches in radiology, disease prediction, and medical image analysis. The theory emphasizes that computerized systems should complement rather than replace clinicians by functioning as supportive

tools that enhance consistency, reduce diagnostic errors, and improve healthcare delivery efficiency.

The current study aligns with Clinical Decision Support Systems theory through the development of a shallow CNN model intended to function as an AI-assisted second-opinion tool for pneumonia diagnosis using chest X-ray images. The proposed model does not replace radiologists or clinicians; instead, it provides supportive classification outputs that can assist healthcare professionals in identifying pneumonia cases more efficiently and consistently. This is particularly important in resource-constrained settings where shortages of radiologists and delays in diagnosis are common.

Integrating automated image classification into the diagnostic workflow, the proposed system contributes toward reducing diagnostic variability and improving early disease detection. Furthermore, the lightweight nature of the architecture enhances the practicality of implementing AI-based CDSS solutions in low-resource healthcare facilities that may lack advanced computational infrastructure. Consequently, the theory provides a strong conceptual foundation for positioning the proposed shallow CNN as a clinically supportive and deployable AI diagnostic framework.

2.5 Pneumonia Diagnostics Tools

2.5.1 Chest X-ray as a Diagnostic Tool for Pneumonia

Chest radiography remains the gold standard for initial pneumonia diagnosis. It relies on the principles of differential X-ray absorption to visualize thoracic structures (Firdaus, 2024). Radiological findings that indicate the presence of pneumonia includes opacification, consolidation, and infiltrates, which manifest as increased density on the radiograph (Mandel & Wunderink, 2020). However, the limitations of human interpretation are significant; Neuman et al. (2019) and Williams et al. (2020) document high inter-observer variability and subjectivity among radiologists, leading to diagnostic inconsistencies. Despite its widespread use, human interpretation of chest X-rays is subject to significant limitations. Inter-observer variability among radiologists' ranges from moderate to substantial, with kappa statistics typically falling

between 0.5 and 0.7 for pneumonia detection (Neuman et al., 2019). Williams et al. (2020) demonstrated that even experienced radiologists disagree on the presence of pneumonia in up to 30% of pediatric cases, highlighting the subjective nature of radiographic interpretation.

This variability is particularly problematic in low-resource settings where specialist radiologists are scarce or unavailable. Qin et al. (2019) argued that computer-aided diagnosis systems could mitigate these challenges by providing consistent, objective assessments, thereby extending diagnostic capacity to underserved populations a premise that motivates the present investigation.

2.5.2 Transfer Learning Approaches for Pneumonia Classification

Liang and Zheng (2021) proposed an improved VGG16 architecture (IVGG13) for pneumonia detection, achieving 89.1% accuracy on the Kermany dataset. Their modifications included batch normalization, dropout regularization, and architectural simplifications to reduce overfitting. The model demonstrated that targeted adaptations of standard architectures could substantially improve pneumonia classification performance.

Salehi et al. (2021) conducted a comprehensive comparative study evaluating multiple transfer learning architectures on pediatric chest X-rays. DenseNet-121 achieved 86.8% accuracy with 86% AUC, VGG19 reached 83.6% accuracy with 78% AUC, Xception attained 86% accuracy with 81% AUC, and ResNet50 achieved 84.8% accuracy with 81% AUC. The authors concluded that DenseNet's feature reuse mechanism particularly suited medical imaging tasks where subtle texture differences distinguish pathologies. Chouhan et al. (2020) explored ensemble methods, combining predictions from multiple transfer learning models to achieve 96.4% accuracy. While ensemble approaches improve performance, they multiply computational requirements—a critical consideration for resource-constrained deployment.

2.5.3 Custom CNN Architectures for Pneumonia Detection

Saraiva et al. (2019) developed a 10-layer custom CNN achieving 95.30% accuracy on chest X-ray classification. Their architecture comprised seven convolutional layers followed by fully connected layers, demonstrating that purpose-built architectures could outperform generic transfer learning approaches on specific datasets. Naskinova (2021) systematically compared custom CNN configurations, evaluating optimizer impact on performance. Model 1 with SGD optimizer achieved 90.7% accuracy (95.97% AUC), while Model 1_b with Adam optimizer reached 92.08% accuracy (96.93% AUC).

This 1.38% accuracy improvement solely from optimizer selection underscores the importance of hyperparameter optimization in CNN training. Ayan and Ünver (2019) directly compared custom CNNs with transfer learning, finding that custom architectures trained from scratch could match transfer learning performance when datasets were sufficiently large. Their analysis suggested that for pneumonia detection with several thousand images, custom architectures offered competitive performance with greater architectural flexibility.

2.5.4 Lightweight Architectures for Resource-Constrained Settings

Deployment in low-resource settings requires efficiency. Godasu, Zeng, and Sutrave (2020) discussed transfer-learning challenges in resource-limited environments. Sousa et al. (2022) proposed an efficient CNN for mobile deployment using architecture optimization. Rahman et al. (2021) developed a shallow CNN for COVID-19 and pneumonia, showing competitive performance against deep networks. Pereira et al. (2021) provided a comparative analysis of shallow versus deep models, emphasizing computational efficiency metrics as a primary selection criterion for field deployment. This was also depicted as show in figure 2.8 below by Sarvamangala and Kulkarni (2022)

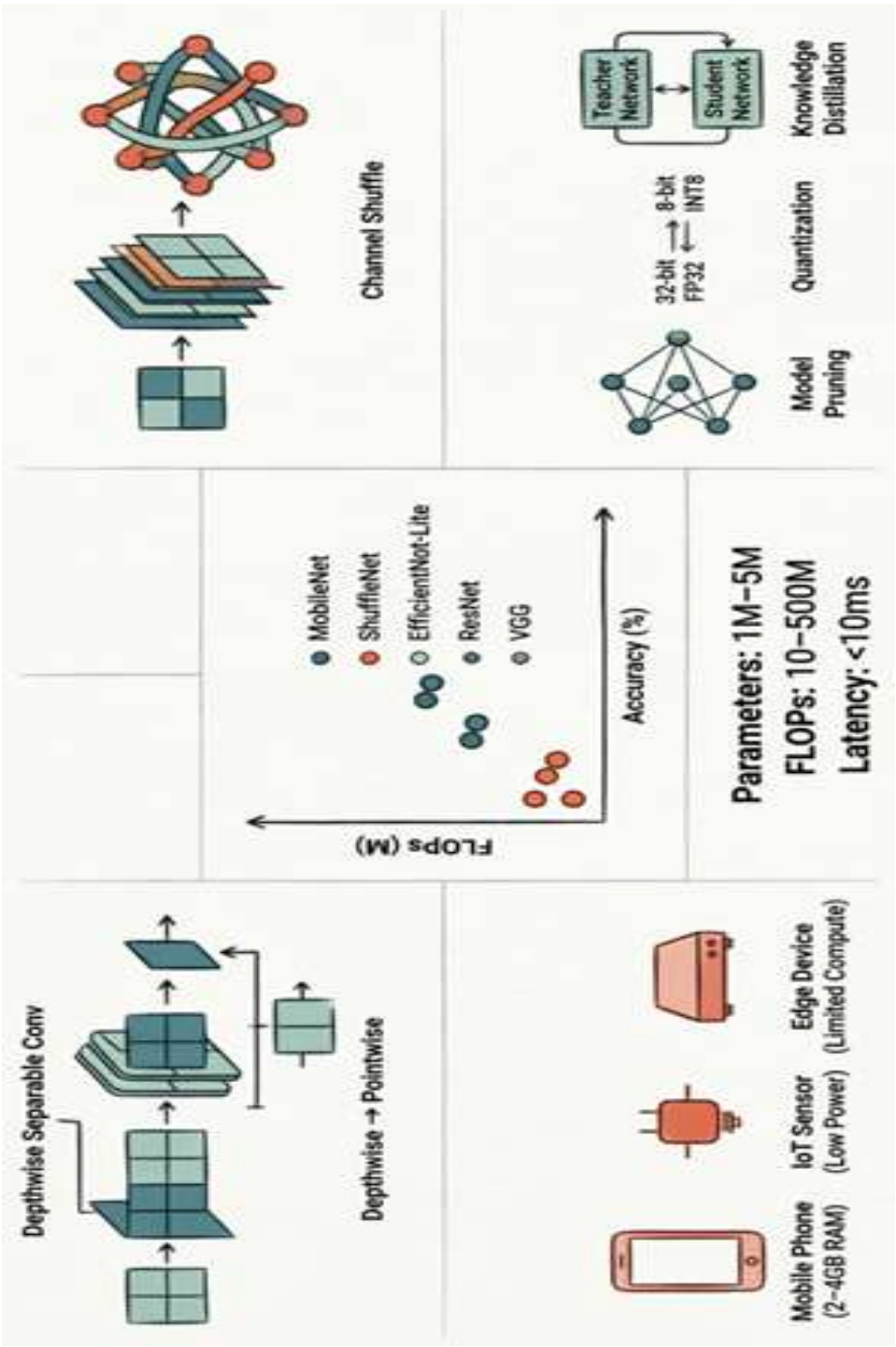


Figure 2.8: Lightweight Architectures for Resource-Constrained

Godasu, Zeng, and Sutrave (2020) explicitly addressed transfer-learning challenges in resource-limited medical imaging environments, identifying computational requirements, data scarcity, and domain adaptation as primary barriers to deployment.

Their analysis highlighted the need for architectures specifically optimized for low-resource settings. Krishna et al. (2026) reviewed next-generation pneumonia detection approaches, emphasizing secure AI, CNN models, and federated learning frameworks. Their synthesis identified lightweight architectures as a critical research direction for enabling deployment in underserved healthcare systems.

Rahman et al. (2021) compared shallow and deep networks for COVID-19 and pneumonia detection, finding that shallow networks with 5-8 layers achieved comparable accuracy to deeper networks while requiring substantially fewer computational resources. This finding aligns with Sutera's (2025) demonstration that shallow networks can improve accuracy by up to 27% compared to deep networks on medical imaging tasks.

Pereira et al. (2021) conducted a formal efficiency analysis, measuring FLOPs, parameter counts, inference time, and memory usage alongside accuracy. Their results showed that shallow networks achieved superior accuracy-per-parameter ratios, making them ideal for mobile and edge deployment. Sarker et al. (2025) extended this line of research with ResMHCNN blocks, achieving 96.9% accuracy for brain tumor classification with only 1.154 million parameters substantially fewer than conventional deep networks.

2.6 Convolutional Neural Networks

2.6.1 Architecture of CNNs

Convolutional Neural Networks are specialized architectures as shown in figure 2.9 below is designed to process grid-like data, particularly images(Köpüklü et al., 2019). Their fundamental components work in concert to extract hierarchical features while maintaining computational efficiency. The input layer receives images represented as tensors of dimensions height × width × channels for grayscale medical images, typically a single channel representing pixel intensities(Sarvamangala & Kulkarni, 2022).

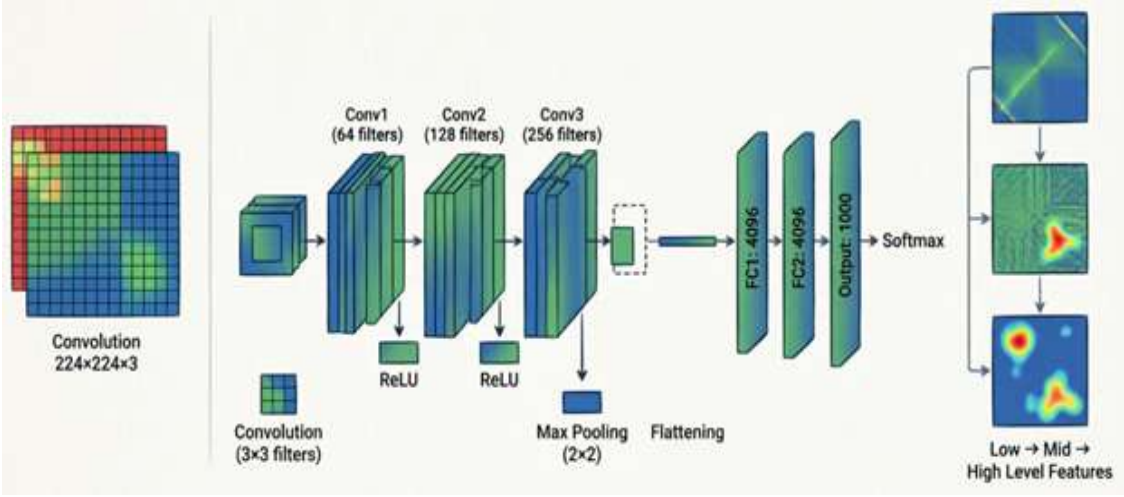


Figure 2.9: Convolutional Neural Networks Architecture

Convolutional Neural Networks (CNNs) process input layers where images are represented as tensors Khan et al., 2020). The core operation occurs in convolutional layers, involving kernel operations, feature maps, stride, and padding (Alzubaidi et al., 2021). The mathematical formulation of convolution operations allows filters to extract spatial hierarchies, demonstrating the impact of kernel size on performance(Lederer, 2021).

Activation functions such as ReLU, Leaky ReLU, Sigmoid, and Tanh introduce non-linearity; Nwankpa et al. (2021) provide a comparative analysis, noting that ReLU dominates CNN architectures in medical imaging due to its computational efficiency and mitigation of the vanishing gradient problem. Pooling layers, including max and average pooling, facilitate dimensionality reduction while preserving salient features (Lederer, 2021). Finally, fully connected layers handle the classification head, while dropout and regularization techniques are essential for preventing overfitting(Gu et al., 2018).

Convolutional layers form the core computational engine of CNNs. These layers apply learnable filters (kernels) that slide across the input, computing dot products between filter weights and local image regions to produce feature maps(Ketkar & Moolayil, 2021) . The mathematical operation can be expressed as a dot product between the input image I and convolution kernel K, with the output feature map capturing the

presence of specific patterns at each spatial location(Wang et al., 2021). Alzubaidi et al. (2021) emphasized that smaller kernels (e.g., 3×3) are generally preferred over larger ones because they capture finer details while enabling deeper architectures through stacking a principle famously exploited in the VGG network(Shah et al., 2023).

Alzubaidi et al., (2021) states that activation functions introduce non-linearity, enabling networks to approximate complex functions as shown in figure 2.10 below. The Rectified Linear Unit (ReLU), defined as $f(x) = max(0, x)$, has become the default choice due to its computational simplicity and ability to mitigate vanishing gradients(Lederer, 2021). Alzubaidi et al., (2021) demonstrated that deep CNNs with ReLU train several times faster than equivalent networks using tanh units, a critical advantage when training large models.

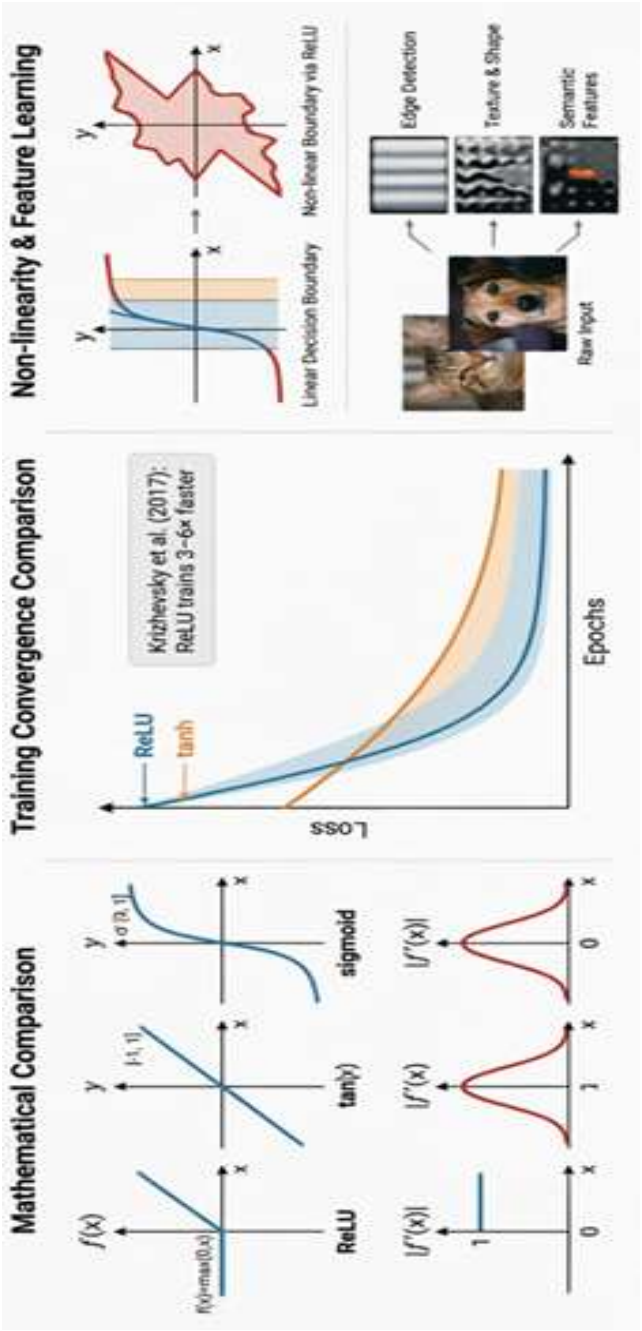


Figure 2.10: Activation Functions in Deep Neural Networks – ReLU

Pooling layers reduce spatial dimensions, decreasing computational load while providing translation invariance. Max pooling selects the maximum value within each pooling window, preserving the most activated features while discarding positional information (Zhao & Zhang, 2024). Average pooling computes the mean, offering

smoother downsampling. Gu et al. (2018) noted that pooling progressively abstracts features, allowing deeper layers to respond to larger receptive fields.

Fully connected layers, typically positioned at the network's end, flatten the spatial feature maps into vectors and perform high-level reasoning as shown in figure 2.11 below. These layers connect every neuron in the previous layer to every neuron in the current layer, enabling complex feature combinations but introducing most network parameters(Ketkar & Moolayil, 2021b). Dropout regularization randomly deactivates a proportion of neurons during training, preventing co-adaptation and reducing overfitting.

2.6.2 Mathematical Computation of Convolutional Neural Networks

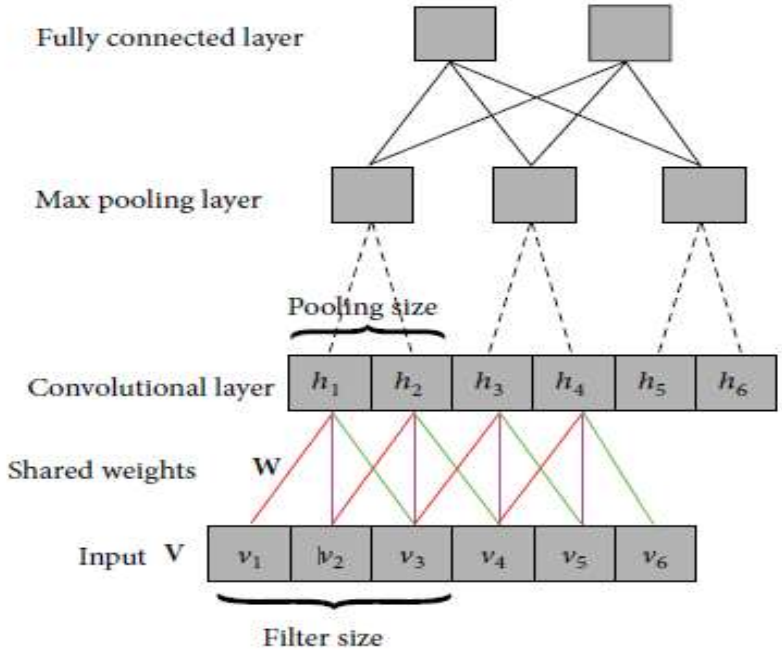


Figure 2.11: Convolutional Neural Network Layers

Figure 2.11 above as proposed by (Wang et al., 2021) demonstrates the different layers found in a convolutional neural network. At each layer, the input image undergoes mathematical operations that eventually transform the image into the predefined class.

The operations allow the network to reduce spatial redundancy while extracting high-level features.

At the convolutional layer, given an input image I of dimensions $i \times j \times r$ and a filter (kernel) K of size $m \times n$ the output feature map is computed as

$$\mathbf{O}_f = \mathbf{conv}_{(i,j)} = (\mathbf{I} \otimes \mathbf{K})(i,j) = \sum_w \sum_h I_{w,h} K_{i-w,j-h} \quad \textbf{(2.1)}$$

Where:

- i is the height of the image
- j is the width of the image and
- r is the dimension of the image or number of channels fed into the system.

The kernel/filter is made up of a matrix of numbers that move over the image from left to right until the entire image has been covered. The image is transformed based on the values of the kernel(Shambour, 2022; Wang et al., 2021). For optimality, smaller kernels are considered over bigger kernels since they capture smaller, complex features in an image as compared to their counterpart bigger filter that only capture the basic components of an image. By slowly reducing the image dimension, smaller kernels provide a means of making the network deep hence the ability to extract a vast amount of information that may be useful in later layers(Alzubaidi et al., 2021).

According to (Abiyev & Ma'aitah, 2018), the output of l^{th} convolution layer denoted as C_j^l that constitutes of feature maps is computed as

$$C_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{a_i^{(l-1)}} K_{i,j}^{l-1} * C_j^{(l)} \quad (2.2)$$

Where:

- $C_i^{(l)}$ are the feature maps
- $B_i^{(l)}$ is the bias matrix
- $K_{i,j}^{l-1}$ is the kernel that connects the j^{th} feature map in layer $(l-1)$ with the i^{th} feature map in the same layer

There exists two types of pooling; maximum pooling and average pooling. In the former one can achieve noise suppressing, de-noising as well as dimension reduction. In the latter, average pooling achieves dimension reduction as a measure of noise suppression. The pooling and convolution layer form the n^{th} layer of the convolution neural network (Saha, 2018).

For an image that has even dimensions it is simple to perform maximum pooling. This is achieved by using the maximum operator of x-by-x kernel and a stride to x perform convolution over the image(Ma et al., 2022). The below formula demonstrates how maximum pooling can be achieved by expression

$$F_{i,j} = max\{I_{i+k-1,j+l-1} \forall 1 \leq k \leq m \; and \; 1 \leq l \leq m\} \quad (2.3)$$

Additionally, Teow (2017) states that average pooling is calculated by the expression:

$$f_p = pool(i,j) = \frac{1}{M} \sum_m \quad I_{i,j} \quad (2.4)$$

- F : - denotes the feature map after dimension reduction
- I : - denotes the image being reduced,
- k and l :-denote the width and height of the section of the image currently on focus
- m : - is the size of the maximum kernel operator

ReLU, rectified linear unit, is a standard way to model a neuron's output f as a function of its input x that is:-$f(x) = tahn\,(x)$ OR $f(x) = (1 + e^{-X})^{-1}$, which is the equation for Calculating a neuron's output. The study shows that a 4-layer CNN with ReLU reaches the 25% error rate six times faster than the equation *tanh* neurons. This demonstrates that a deep CNN with ReLU trains faster than the equivalent *tanh* unit. A faster learning rate, however, influences the performance of large models trained on large data sets (Krizhevsky et al., 2017).

ElSayed et al., (2021) defines the learning rate as a hyper parameter that can be configured and used in the training of neural networks. The learning rate consists of

small positive values ranging from 0 to 1. Since the learning rate determines how fast the model is adapted to the problem, smaller learning rates can cause the process to get stuck, given the small adjustments that are being made to the weights with every update. Large learning rates in turn cause the model to quickly converge to a suboptimal solution.

The fully connected layer takes its input from the convolution/ReLu/pool layer preceding it, flattens the image into a column and outputs an N-dimensional vector where N is the number of classes that the program has to choose from. It then determines which features most correlate to a particular class(Liu et al., 2022). Deshpande (2016) summarized this in figure 2.12 below.

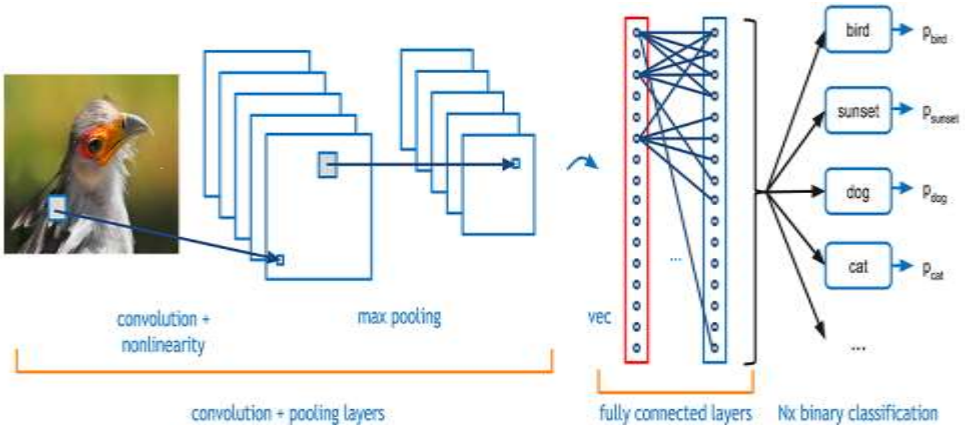


Figure 2.12: Image Classification with fully Connected Layer

2.6.3 Training of CNNs

CNN training follows the principles of error backpropagation originally formulated by Rumelhart et al. (1986). The network computes predictions through forward propagation, calculates loss using a task-appropriate function, and propagates error gradients backward to update weights via gradient descent. For binary classification tasks such as pneumonia detection, binary cross-entropy loss is standard(Köpüklü et al., 2019; Zhao & Zhang, 2024). The figure 2.13 shows the summary for the CNN training dynamics.

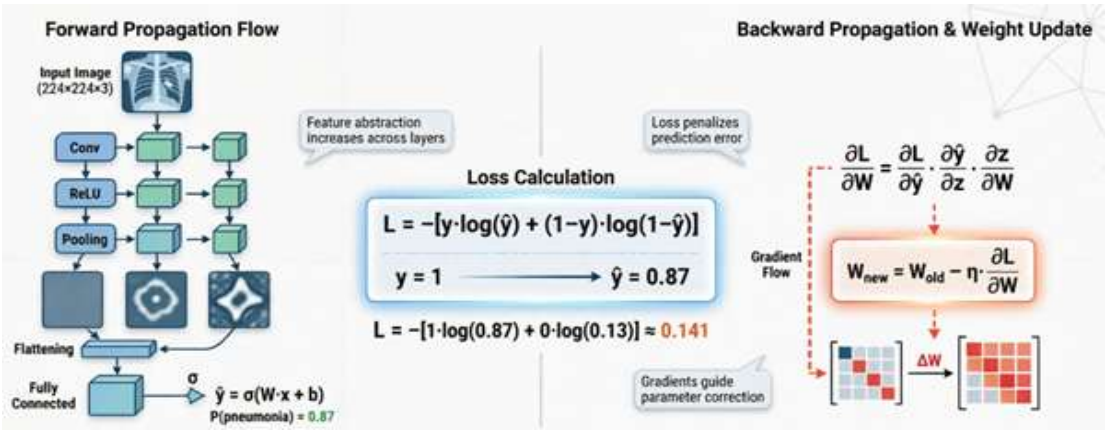


Figure 2.13: Training Dynamics of CNNs: From Forward Propagation to Gradient-Based Optimization

Optimization algorithms govern how gradients update network weights. Stochastic Gradient Descent (SGD) with momentum remains widely used, but adaptive methods like Adam (Haji & Abdulazeez, 2021) and RMSprop have gained popularity due to their automatic learning rate adjustment. Everett et al., (2024) provided a comprehensive comparison, noting that Adam often performs well out-of-the-box but may generalize worse than carefully tuned SGD in some contexts. Learning rate schedules, which reduce the learning rate during training, help networks converge to better minima (Iiduka, 2021). Batch normalization stabilizes training by normalizing layer inputs, allowing higher learning rates and reducing sensitivity to initialization(Singh & Krishnan, 2020).

2.7 Conversion from a raw Image to pixel Values

In its simplest form, an image can be represented as a 3D array of numbers containing the height, width and intensity(Van der Jeught and Dirckx, 2019). A digital image like in figure 2.14 below by Bashir et al., (2021), is considered to have small squares picture elements called pixels and the intensity of each pixel determines the number of light photons that are on a particular square(Bechar et al., 2019).

Figure 2.14: A typical Bitmap Image (b) zooming in Eye Region 10 Times to Reveal the Atomic Units

Images are classified as either colored images, grey scale images or binary images. Digital colored images use the RGB color model (3-color channels); where the colors Red, Blue and Green are mixed to different intensities to get other colors(Li et al., 2023). The colors are represented as a close range of values ranging from 0 to 255; a total of 256 discrete amounts of individual primary color. This discrete amounts of each color can be written to correspond to 8 bits that are used to hold the eight-color channel value, that is $2^8 = 256$ that corresponds to the 24bit color depth(products between the 3- channel and 8 bits PER CHANNEL). The larger the channel number the more the presence of those primary colors. It is these different color intensities that are seen at the different squares/ pixels. As such two neighboring pixels can have different colors (Data Carpentry, 2022).

Binary images are those whose values have been quantized to the values 0 for black and 1 for white that correspond to pixel values 0 and 255 respectively(Li et al., 2023). These images are obtained by thresholding pixels of a grey level image. Pixels that contain a grey level that are above the given threshold are set 1 whereas the rest are set to 0. As such this produces either white object's whose background is black or black objects with a white background. Sometimes binary images are the outputs of some image processing techniques(Abdulsahib et al., 2021).

Grey scale images are those that have different ranges of grey shades with the darkest being grey and the brightest being white as shown in figure 2.15 below by Wang et al., (2021). Intermediate grey shades are obtained by mixing equal levels of the primary colors. On computer systems (or in the presence of transmitted light) these primary colors are as decimal numbers 0 to 255 or binary numbers 00000000 to 11111111. As such, the brightness/lightness of grey is directly proportional to the value representing the level of brightness of the primary colors(Bashir et al., 2021). For printed images, each pixel has different levels of cyan, magenta and yellow that are represented as a percentage ranging from 0 to 100. Printed grey scale images have equal amounts of cyan, magenta and yellow mixed(Wang et al., 2021).

.

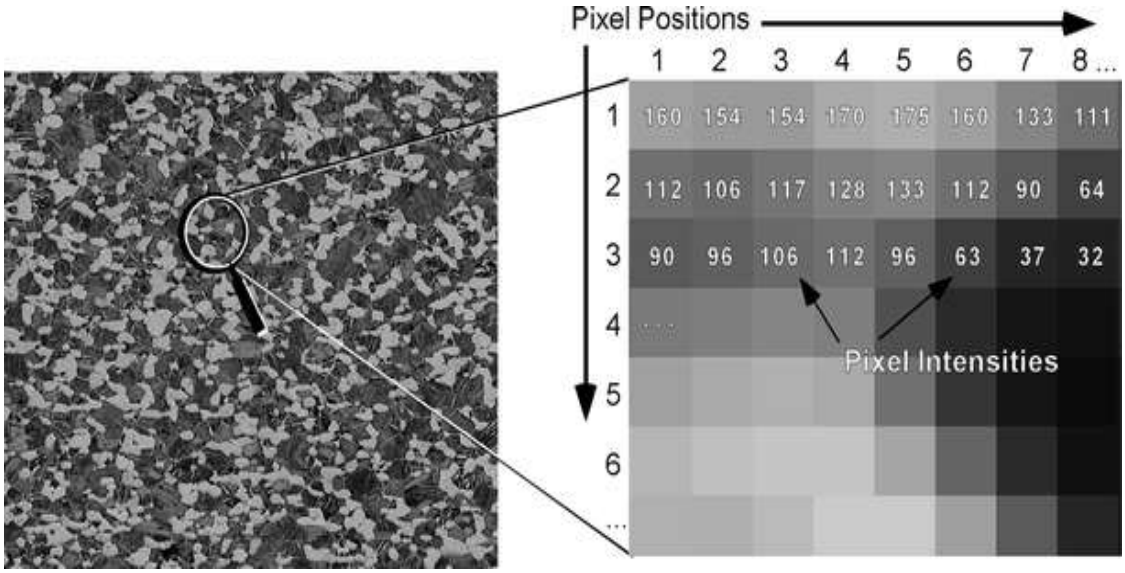


Figure 2.15: Actual Image Area with Corresponding Magnified View

2.8 Research Gap

Computer-aided techniques have demonstrated strong performance on pneumonia classification with computational requirements varying dramatically across architectures. Parameter counts range from hundreds of thousands to tens of millions (Sarker et al., 2025). Several studies adopt either deep architectures such as ResNet and DenseNet or custom designs without rigorously exploring depth as a design

variable. There are limited studies that explicitly target low-resource clinical settings such as rural hospitals, mobile health units, developing country healthcare systems; were computational infrastructure limits deep network deployment.

2.9 Proposed Conceptual Framework

In many resource-constrained settings, the diagnostic pathway for pediatric pneumonia faces bottlenecks: limited access to radiologists, variability in clinical interpretation, and delays in diagnosis. Stakeholders from frontline clinicians to patients would benefit from a reliable, rapid decision-support tool. The proposed model integrates a shallow CNN into a streamlined diagnostic pipeline. The workflow encompasses CXR acquisition, automated preprocessing, model inference, and the generation of a classification output (Pneumonia/Normal) to support the clinician's decision. This model incorporates feedback mechanisms for continuous performance monitoring and improvement. Summary of the proposed architecture is as show in figure 2.16 below.

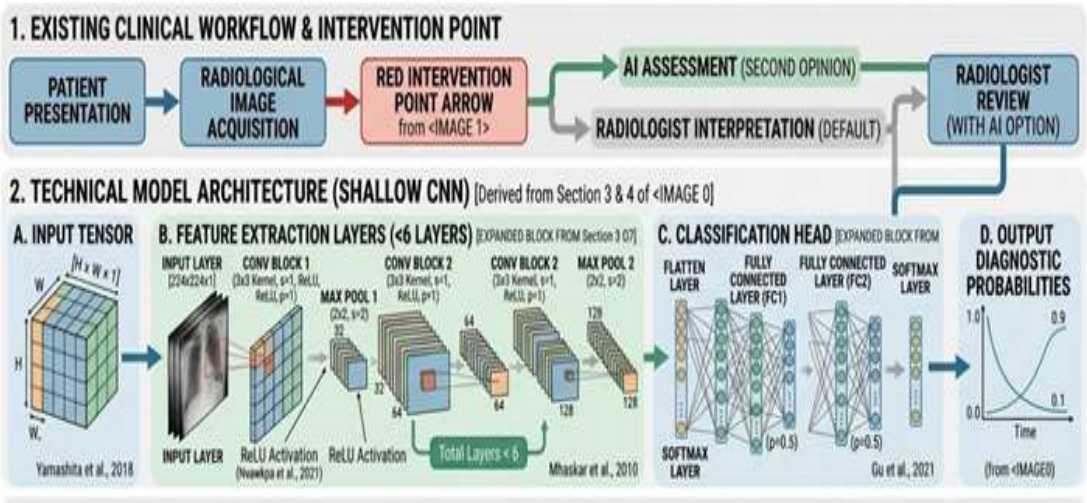


Figure 2.16: Proposed Shallow CNN Architecture and Workflow Integration

Based on the attached diagram, the proposed model architecture is a shallow Convolutional Neural Network (CNN) designed for radiological image analysis, functioning as an AI-based second opinion within an existing clinical workflow. The network begins with an input tensor of dimensions $H \times W \times \mathrm{L}$, indicating single-channel (grayscale) image processing, consistent with standard radiographic or CT

imaging data. Feature extraction is performed using fewer than six layers, each comprising a convolutional block with 3×3 kernels (stride=1, padding=1), followed by ReLU activation and a 2×2 max pooling layer with stride=2. This shallow design prioritizes computational efficiency and interpretability over representational depth, likely to facilitate rapid inference and integration into radiology review workflows without substantial hardware demands.

The classification head aggregates extracted features through a sequence of fully connected layers with progressively decreasing neuron counts: from 128 units down to 32, interspersed with dropout probabilities of 0.5, culminating in a final dropout of 0.9 before a two-unit output layer. A Softmax activation then produces diagnostic probabilities (0.9 vs. 0.1), representing binary classification outcomes such as "disease present" or "absent." This architecture, referencing Yamashita et al. (2018) for input structure, Nawakpa et al. (2021) for pooling strategy, and Gu et al. (2021) for Softmax output, reflects a streamlined model suited for real-time or near-real-time adjunctive assessment, positioned at the intervention point between image acquisition and radiologist interpretation.

2.10 Description of the Proposed Baseline Model

The custom baseline model begins with a Convolutional Layer that applies 64 convolutional filters of size 2x2 to the input image, which is 150x150 pixels in grayscale. This layer uses the ReLU (Rectified Linear Unit) activation function to help the model detect important features like edges and textures in the image. Following this, a Max Pooling Layer with a pool size of 2x2 was used to reduce the spatial dimensions of the feature maps, which helped in lowering the computational load and emphasizing the most prominent features (Nwankpa et al., 2021).

Next, the model has a second Convolutional Layer with the same configuration 64 filters of size 2x2 and ReLU activation; which further extracts features from the previously pooled feature maps. This is followed by another Max Pooling Layer with the same pool size to continue down sampling the feature maps and retaining essential features (Alzubaidi et al., 2021). To capture more complex patterns, the model includes

a third Convolutional Layer with 128 filters of size 2x2, again using the ReLU activation function. A third Max Pooling Layer then further down samples the feature maps, allowing the model to focus on the most critical aspects of the image (Liu et al., 2022).

The Flatten Layer is next, which converts the 3D feature maps into a 1D vector, making it suitable for the subsequent fully connected (dense) layers. The model then adds a Dense Layer with 32 units, using ReLU activation to process the extracted features and reduce dimensionality, introducing non-linearity to the model. To prevent overfitting, a Dropout Layer with a dropout rate of 0.7 is included, randomly setting 70% of the input units to zero (off) during training, ensuring the model generalizes better by not relying too heavily on any neuron (Lederer, 2021).

Finally, the Output Layer consists of a Dense Layer with two units, corresponding to the two classes Normal and Pneumonia using a sigmoid activation function to provide the probability of each class. According to Nwankpa et al. (2021), since this is a binary classification task, the model is compiled with the RMSprop optimizer, which adjusts the learning rate during training. The model used the binary_crossentropy loss function, which is appropriate for binary classification tasks, and accuracy as the metric to measure the proportion of correctly predicted samples. This is summarized in figure 2.17 below.

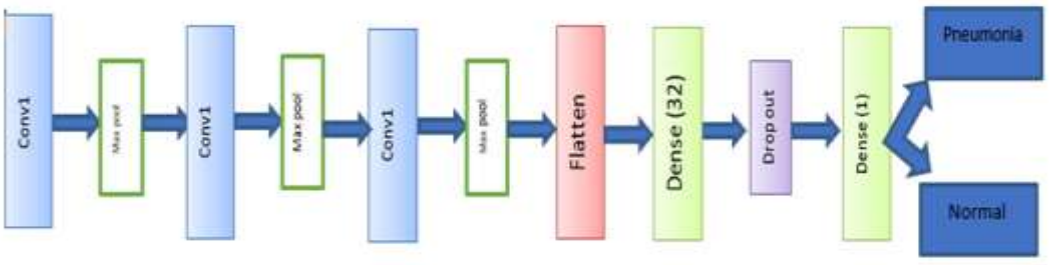


Figure 2.17: Proposed Shallow Convolutional Neural Network Architecture

2.11 Chapter Summary

This chapter reviewed existing literature relevant to pneumonia diagnosis and image classification, establishing the theoretical and empirical foundation for the study. It began by examining the pathogenesis of pneumonia, highlighting its causes, classifications, and progression, as well as the clinical importance of timely and accurate diagnosis. The discussion emphasized the continued reliance on chest X-rays and the challenges associated with human interpretation, including inter-observer variability and limited access to radiological expertise in low-resource settings.

The chapter then traced the evolution of image classification techniques, moving from traditional machine learning methods such as KNN, SVM, and fuzzy logic systems to modern deep learning approaches. While traditional methods were found to be limited by their dependence on manual feature extraction, deep learning models have demonstrated better performance through automated hierarchical feature learning.

The chapter also explored CNN architecture and its mathematical foundations and reviewed transfer learning and custom CNN architectures for pneumonia detection. It was noted that while high accuracy can be achieved, many approaches remain computationally expensive. Lightweight architectures were identified as a promising direction, particularly for real-world healthcare applications where hardware limitations exist.

The review identified a clear research gap; limited focus on designing computationally efficient, shallow CNN models specifically tailored for pneumonia detection in low-resource settings. Addressing this gap, the chapter proposed a conceptual framework and a shallow CNN architecture optimized for grayscale chest X-ray analysis. Overall, this chapter justifies the need for a lightweight, shallow neural network approach and provides the necessary groundwork for the methodology and model development presented in subsequent chapters.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

Building on the problem context established in Chapter One and the theoretical foundations discussed in Chapter Two, this chapter presents the methodological framework adopted to investigate the use of shallow Convolutional Neural Networks (CNNs) for pneumonia classification. Chapter One highlighted the critical healthcare challenge posed by pneumonia, particularly in resource-constrained settings such as Kenya, where limited diagnostic infrastructure and shortage of radiological expertise hinder timely and accurate diagnosis. Chapter Two further examined existing literature on computer vision, medical imaging, and neural network architectures, identifying a clear research gap: the limited focus on computationally efficient, shallow models tailored for low-resource environments.

In response to this gap, the current chapter outlines the systematic approach used to design, implement, and evaluate a lightweight CNN model for chest X-ray classification. The study adopts a quantitative experimental research design, enabling controlled evaluation of how architectural choices such as network depth, filter size, and parameter count affect classification performance. The methodology integrates both primary and secondary data sources to ensure diversity, robustness and generalizability of the model.

The chapter further details the processes of data collection, cleaning, and preprocessing, emphasizing ethical considerations such as patient anonymization and data integrity. In addition, the architecture of the proposed shallow CNN is explained, highlighting its design rationale in achieving a balance between computational efficiency and diagnostic accuracy.

The chapter then presents the evaluation framework used to validate model performance. Finally, an experimental setup of four models is then discussed to provide a view of the conditions under which the models were trained.

3.2 Research Design

The study adopted an analytical research design as defined by Dubey and Kothari, (2022), wherein existing data and newly collected information were systematically analyzed to establish relationships between architectural depth and classification performance. This design was classified as quantitative experimental research, given the controlled manipulation of independent variables such as network depth, filter count, kernel size among others; and measurement of dependent variables such as accuracy, inference latency, parameter count among others.

3.3 Data Collection

3.3.1 Primary Data Sources

Primary data were collected from five healthcare facilities located in Machakos, Nairobi, and Kajiado counties in Kenya. These facilities were selected to ensure representation of diverse clinical environments within Kenya. Ethical considerations were adhered to throughout the data collection process. Institutional approvals were obtained from the relevant regulatory and administrative bodies prior to data acquisition. In addition, all data handling procedures complied with established principles of medical research ethics, including respect for patient confidentiality and data protection.

To safeguard patient privacy, data anonymization techniques were applied to all collected records(Majeed & Lee, 2021). Specifically, all patient-identifiable information embedded within the DICOM files was removed prior to their use in this study. A total of 635 radiology images were collected from the selected facilities. The data acquisition process involved accessing imaging data through Picture Archiving and Communication System (PACS) consoles, followed by extraction of DICOM files.

3.3.2 Secondary Data Sources

Secondary data were sourced from two publicly available repositories to augment the primary dataset and enhance model generalizability. The Kaggle dataset published by

Kermany et al. (2018) from the Guangzhou Women and Children's Medical Center provided 5,863 paediatric chest X-ray images from children aged one to five years. The class distribution comprised 3,418 pneumonia images and 1,266 normal images for training, with 390 pneumonia and 232 normal images reserved for testing.

The Kaggle Chest X-Ray Dataset (Chowdhury et al.,2020) supplied supplementary normal and pneumonia images to address class imbalance concerns. The rationale for employing multi-source data rested on the need to enhance demographic diversity, mitigate overfitting to institution-specific imaging characteristics, and provide sufficient training samples for robust model development (Zech et al., 2018).

3.4 Feature Extraction and Selection

3.4.1 Secondary Dataset

The secondary dataset obtained from Kermany was split into three folders, training, testing and validation data. The dataset contained 1266 normal and 3418 pneumonia images in training folder; the test data contained 232 normal and 390 pneumonia images respectively. Whereas the validation dataset had a total of 16 images. Because of the imbalance in the dataset between the pneumonia and normal images in the training dataset, additional normal Chest X-ray images were obtained from (Chowdhury et al.,2020). This increased the number of normal images from 1266 to 2033 resulting in training normal folder. As result, the training folder had 5451 images. 450 and 590 additional pneumonia and normal images obtained from (Chowdhury et al.,2020) and were added to the existing testing images to increase the number of test images to 1662. This is summarized in table 3.1 below.

Table 3.1: Showing Total Number of Images in each Folder

source	Training Dataset		Testing Dataset		Validation Dataset	
	Pneumonia	Normal	Pneumonia	Normal	Pneumonia	Normal
Kaggle	3890	1344	390	234	8	8
Kaggle2	0	1195	210	302	1000	1000
	3890	2539	600	536	1008	1008
Total images used	6429		1136		2016	

This dataset was used during training and testing of the proposed model. Figure 3.1 below shows image distribution in the training, validation and testing folders of the secondary dataset

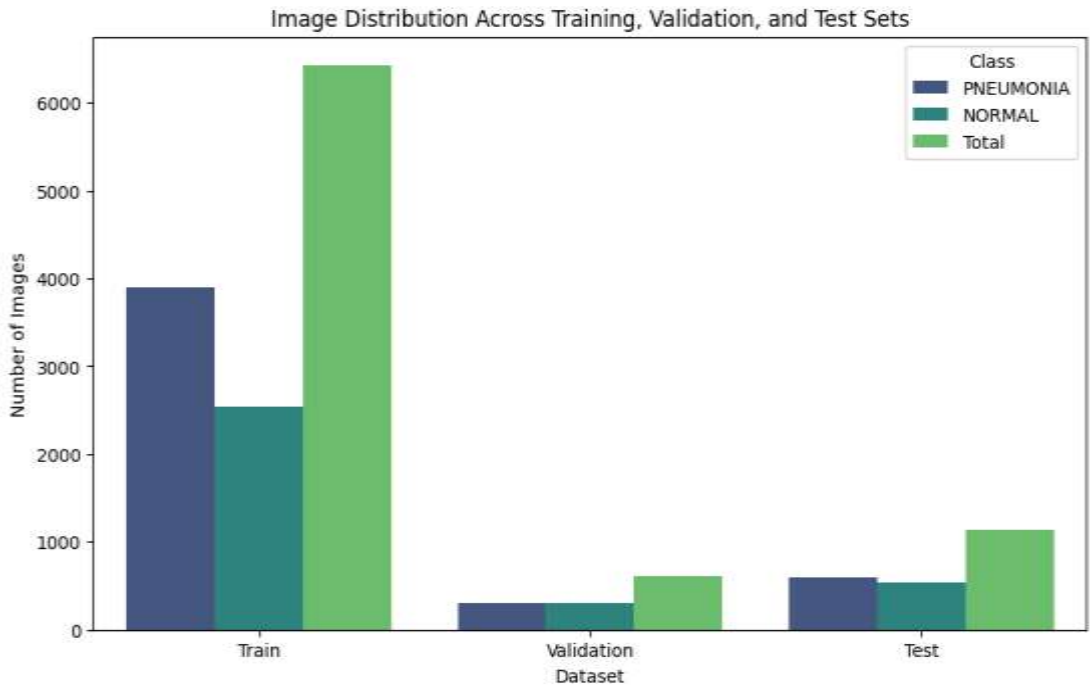


Figure 3.1: Image Distribution in Training, Validation and Test Folders

3.4.2 Primary Dataset

Primary data were collected from the five healthcare facilities in Kenya was divided into two folders; normal and pneumonia. The normal folder had 235 normal chest x-

ray images and the pneumonia folder had 400 pneumonia chest x-ray images. This dataset was used to evaluate the models' performance on local scenario.

Cumulatively, both the secondary and primary data resulted in a total of five thousand eight hundred and ninety-eight pneumonia images and four thousand three hundred and eighteen normal images as shown in table 3.2 below.

Table 3.2 Total Image Distribution

Source	Pneumonia	Normal
Secondary	5498	4083
Primary	400	235
Total	5898	4318

3.5 Data pre-processing

3.5.1 Data cleaning

When dealing with data, there is a need to ensure that the data is in its best possible quality as it can be. This is important because the presence of errors in data can impact study results negatively since the analysis provided by the system used will be inaccurate. As such data cleaning will help identify any existing errors in the data, correct these errors or even try to minimize the negative impact of the same on study results(Afzal et al., 2021).

According to Rahul & Banyal, (2021) there are three-stage data cleaning process that this research was anchored on; namely: screening, diagnosing, and editing of suspected data abnormalities. In the first stage, data screening, the data is searched in a planned way to detect the existence of a suspected data point or pattern that needs careful examination. It is also important to look out for any missing values that may require further examination. All these were cross-examined against "predefine expectations about normal ranges, distribution shapes, and strength of relationship".

The second stage, diagnosis, involved looking at the types of errors that exist and the sources of the same. It is at this phase that one evaluates if there exists any missing data, the presence of diagnosis as well as the existence of data points that are worrisome that need to be clarified to establish their true nature. These worrisome data points can be diagnosed as either "erroneous, true extreme, true normal (i.e. the prior expectation was incorrect), or idiopathic (i.e., no explanation found, but still suspect)". Editing (or treatment) is the last phase of this three-stage process. In this phase, the researcher decided what was to be done with problematic observations from the previous phase. The available options are to correct, delete or leave the observations unchanged. Rahul and Banyal (2021) provide a diagrammatic representation of the above three phases in figure 3.2 below

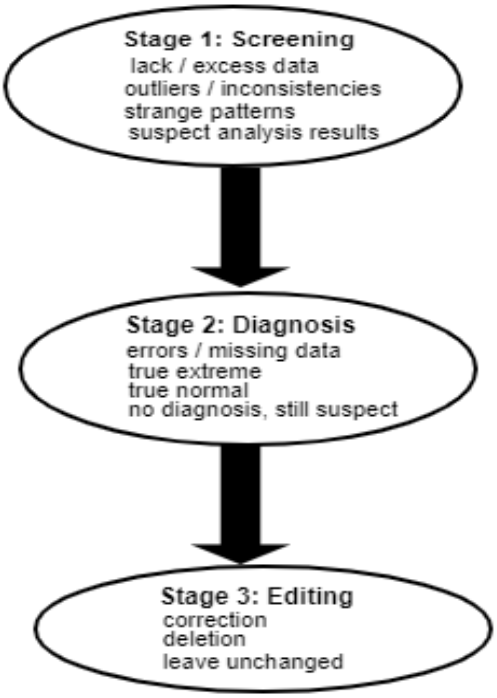


Figure 3. 2: Stages in Data Cleaning Stages

Secondary data obtained had already been cleaned. The primary data collected was subjected to the following cleaning stages: Two thousand images collected from the five facilities were handled in the following manner.

(i) Anonymization of patient information was done at the various hospitals using the PACS console. Once on the examination window, move to the annotations tab then selecting clear to remove patient identifiers such as names and patient ID.

(ii) From the system, it was noted that various x-ray images were stored together. As such, it was easier to download all images ordered by date created. One was required to close the examination window then choose print to CD to download the images.

(iii) Conversion of DICOM images to JPEG format using the DICOM Converter. This was achieved by uploading the DICOM images on DICOM Converter then selecting the conversion format, choosing the folder to save the newly converted images then press convert. Once done, close the application.

(iv) Images that were not chest X-rays were then deleted

(v) Seven hundred chest X-ray images were presented to three radiologists for quality screening and classification as either pneumonia or normal. Some factors that were considered in determining a good quality image: Number of ribs, anteriorly and posteriorly- What is seen when viewing the patient from the front and back; and good exposure factors images that had to be properly exposed (not under or overly exposed).A total of sixty five low quality images were delete.

(vi) The remaining six hundred and thirty-five images were then classified by radiologists into pneumonia and normal images. To classify the images into pneumonia and normal, radiologists considered opacification feature of the image. These refers to areas on the X-ray image that appear more opaque (white or lighter) than the surrounding tissues. Pneumonia patients have parts of the lung appearing lighter.

3.5.2 Image Normalization

Image normalization changes the intensity of pixels. Grayscale normalization was applied to all the images to reduce the effect of illumination's differences as well as aid the CNN converge faster on [0...1] data than on [0..255] (Nimmala &

Rammohanreddy, 2023). This was achieved by dividing the pixel values of an image with 255. In figure 3.3 and figure 3.4 below by Nimmala and Rammohanreddy (2023) are sample image that has undergone normalization and image normalized respectively.

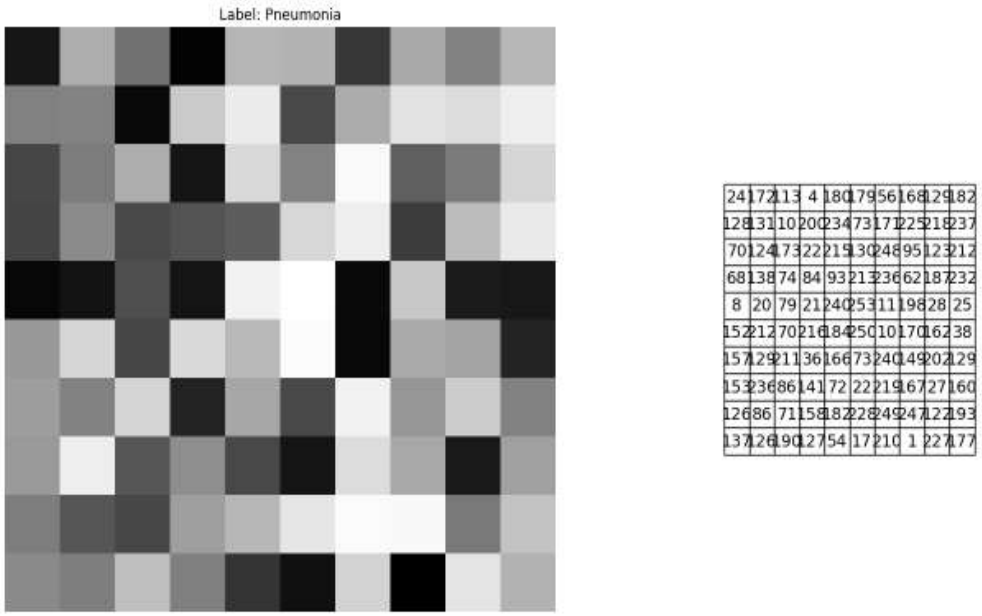


Figure 3.3: Image pixels before Greyscale Normalization

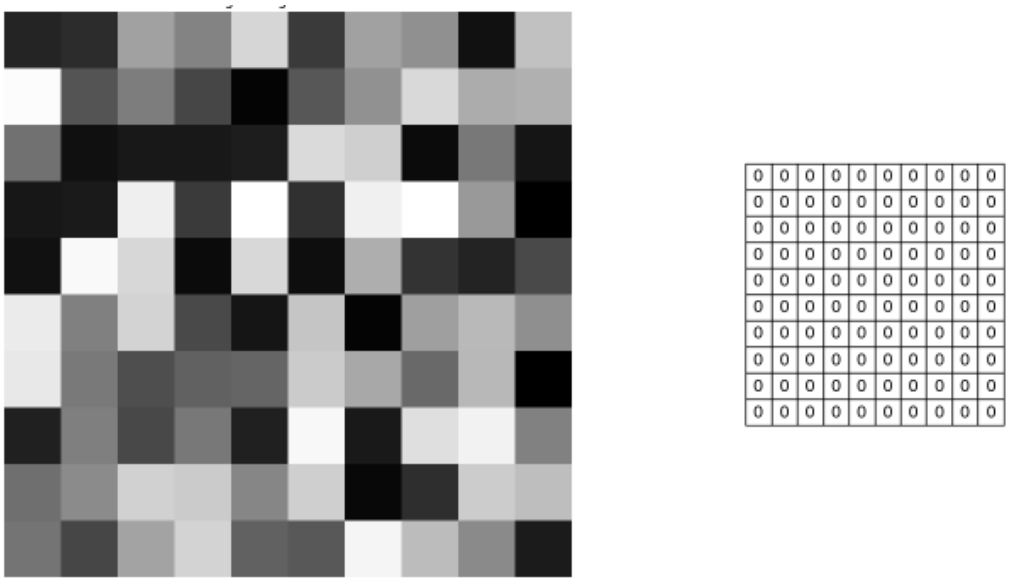


Figure 3.4: Image pixels after Greyscale Normalization

3.5.3 Image Resizing

All images were then resized to 150x150 pixels. The output shape was (x,150,150,1). For instance, the training shape (6429,150,150,1). X being the number of images in the training folder and one (1) being the number of channels for each image.

3.5.4 Data Augmentation

To avoid overfitting, the dataset was artificially expanded through image augmentation. The augmentation techniques applied were: -

(i) Randomly rotate some training images by 30 degrees
(ii) Randomly Zoom by 20% some training images
(iii) Randomly shift images horizontally by 10% of the width
(iv) Randomly shift images vertically by 10% of the height
(v) Randomly flip images horizontally. Once our model is ready, we fit the training dataset.

The below in figure 3.5 is the image showing the effect of the above augmentations when applied to an image.

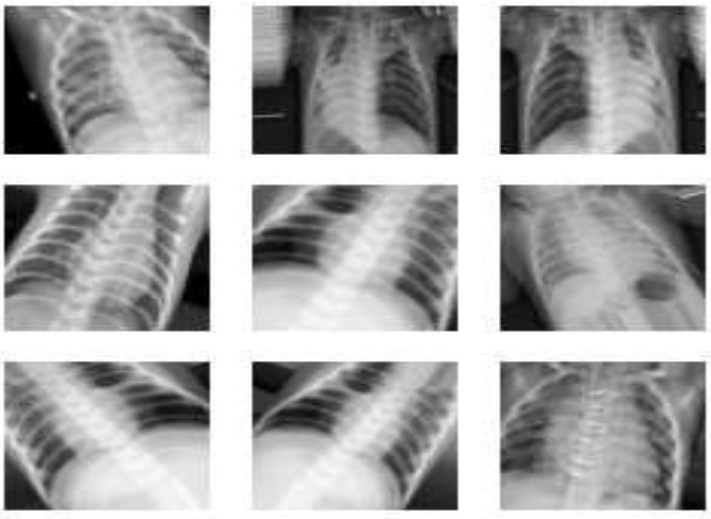


Figure 3. 5: Augmented Image

3.6 Discussion of Baseline Model Parameter Count

The proposed baseline model in section 2.10 was of sequential type with a total of 1,233,506 trainable parameters. The first layer is a Conv2D layer that applies 64 filters of size 3x3 to the input image, resulting in an output shape of (149, 149, 64). There is a slight reduction in the input image's height and width (from 150x150 to 149x149). This is attributed to the filter size and no padding. When applying a convolutional filter to an image, the filter slides over the input image, in this case, stride is one. As the filter moves across the image starting from the top-left corner, the filter cannot cover the last row and column of the pixel image because the dimensions are not divisible by the filter. Since no extra padding (that is zeros) is added to the image the filter can only slide across the valid positions, leading to a reduction in the output size. In this case, the height and width of the image decrease by one pixel each, going from 150x150 to 149x149.

Next, the max-pooling layer uses a 2x2 filter. If the stride is not assigned a value, TensorFlow / Keras defaults to using a stride that is the same as the pool size. In this case the stride of (2, 2) resulting to a 2x2 filter and stride (2, 2). This filter looks at small 2x2 blocks of the input feature map at a time and moves two pixels at a time, both horizontally and vertically, across the input feature map. Because the filter is 2x2 and the stride is two, each 2x2 block in the input feature map is reduced to a single maximum value in the output feature map (picks the maximum value in each 2x2 block). This operation halves the height and width of the feature map. In this case, the input feature map is 149x149 pixels, after applying the 2x2 max pooling with a stride of two, the output feature map dimensions will be approximately 74x74 pixels. This down samples the spatial dimensions and results to reduced computational load.

The baseline model continues with another Conv2D layer, again with 64 filters of size 3x3, producing an output shape of (73, 73, 64). This layer further reduces the spatial dimensions while preserving the number of feature maps. A subsequent MaxPooling2D layer reduces these dimensions by half once more, resulting in an output shape of (36, 36, 64). The next Conv2D layer increases the number of filters to

128, producing an output shape of (35, 35, 128). This layer allows the model to capture more complex features, though the spatial dimensions are slightly reduced again.

Following this, another MaxPooling2D layer reduces the spatial dimensions to (17, 17) while maintaining 128 channels. The baseline model then employs a Flatten layer to convert the 2D matrix of 17x17x128 into a 1D vector of size 36,992 (i.e. 17×17×128=36,992), making it ready for the dense layers that follows. The first dense layer is a fully connected layer with 32 units [model.add(Dense(32, activation="relu"))], which has many parameters (1,183,776) due to its dense connection to the flattened input.

The total number of parameters in the Dense layer is calculated as:

$$\textit{Number of parameters} = (\textit{Number of inputs} \times \textit{Number of units}) + \textit{Number of units of parameters} \quad (3.1)$$

Substituting the values:

Number of parameters = (36,992×32) + 32=1,183,744+32=1,183,776

The Dropout layer is then included a regularization technique used to prevent overfitting. 0.7 value means that during each training iteration, 70% of the units in the first / former Dense layer are randomly set to zero (i.e., "dropped out", or turned off). This results to 30% of the neurons actively contributing to the model's learning during that iteration. Finally, the model concludes with another Dense layer with 2 units, corresponding to the two classes: Pneumonia and Normal. This final layer uses sigmoid activation function to output class probabilities, enabling the model to make its classification.

3.7 Validation of the Proposed Baseline Model

There exists several metrics for determining how good a classification model is; the most common metrics used were accuracy, precision and recall. Accuracy is defined

as the ratio of all correct predictions to all predictions present in a class(Carrington et al., 2023). This can be denoted by the equation below.

$$Accuracy = \frac{correct\ predictions}{all\ predictions} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.2)$$

Where:-

- TP – True positive
- TN- True negative
- FP- False positive
- FN – False negative

However, accuracy alone is not sufficient to determine the model's performance. In scenarios where there exist more than two classes, it may be difficult to understand the performance of the model. Assume a model that has to classify 3 classes and gives an accuracy of 80%; it is difficult to know if the model equally predicts all classes or if two of the classes are being ignored by the model (Brownlee, 2020).

The confusion matrix provides a detailed breakdown of predictions, showing true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) (Carrington et al., 2023). This enables class-wise performance evaluation and helps identify specific weaknesses in the model. This is particularly important in CNN-based classification tasks, where the models' bias toward the dominant classes may occur. Beyond overall accuracy, the confusion matrix provides the computational basis for deriving key diagnostic metrics. These include precision, recall, F1-score, and ROC-AUC. Precision minimizes costly false positives, while recall penalizes missed true cases. The F1-score balances precision and recall in uneven distributions, and ROC-AUC evaluates threshold-independent discrimination ability (Keldenich, 2021) In turn, emphasis is on robust, clinically relevant model validation rather than relying solely on aggregate accuracy.

Keldenich, (2021) states that precision is used to determine the performance of a model. In his work, he defines precision as "*fraction of relevant examples (true positives) among all of the examples which were predicted to belong in a certain class".* This is denoted by the equation below:

$$precision = \frac{true\ positives}{examples\ predicted\ to\ be\ in\ a\ class} = \frac{true\ positives}{true\ positives + false\ positives} \quad (3.3)$$

Precision is important in applications where false positives are costly, ensuring that predicted positive cases are truly relevant. Additionally, according to Keldenich (2021), recall rate looks at all the false negatives that are part of the prediction mix. As such with every false negative predicted, the recall rate is penalized. Below is an equation of calculating the recall rate.

$$Recall = \frac{correct\ predictions}{all\ examples\ belonging\ to\ a\ class} = \frac{true\ positives}{true\ positives + false\ negatives} \quad (3.4)$$

Recall rate is critical in medical diagnosis, because missing a true case (false negative) can have serious consequences(Tom Keldenich, 2021). The F1-score was included to provide a balance between precision and recall, especially in situations where there is an uneven class distribution:

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3.5)$$

The F1-score is particularly useful when both false positives and false negatives are important, offering a single harmonic mean metric that reflects overall classification effectiveness (Ibrahim, 2021; Keldenich, 2021). ROC-AUC evaluates the model's ability to distinguish between classes across different classification thresholds. It plots the trade-off between True Positive Rate (Recall) and False Positive Rate. A higher AUC indicates better model performance in distinguishing between classes. ROC-AUC is especially valuable because it is threshold-independent, making it more robust than accuracy in evaluating classification models (Nakrani et.al., 2025).

3.8 Experimental Setup

The experimental setup was designed to ensure a fair, systematic, and reproducible comparison of the three CNN architectures under investigation: Baseline, Saraiva and Pneumonet. In experimental machine learning research, the integrity of findings depends heavily on the consistency of training conditions across models so that performance differences can be attributed primarily to architectural design rather than variations in optimization strategy or data exposure (Goodfellow et al., 2016). For this reason, all models were trained and evaluated under identical computational and methodological conditions.

The experiments were implemented using the TensorFlow and Keras libraries in a Python environment. These frameworks were selected because of their reliability in deep learning experimentation, extensive community support, and compatibility with medical imaging workflows. Training was conducted on a workstation equipped with sufficient GPU acceleration to handle iterative model optimization and image processing tasks efficiently.

To maintain comparability, all models were trained for 50 epochs using the RMSprop optimizer. RMSprop was selected because of its adaptive learning rate mechanism, which has demonstrated effectiveness in non-stationary optimization problems and deep neural network convergence (Hinton et.al., 2012). In image classification tasks, RMSprop is particularly suitable because it adjusts parameter updates dynamically, enabling faster convergence in convolutional architectures.

The loss function employed was Binary Cross-Entropy, which is widely recognized as the standard loss function for binary classification tasks such as distinguishing pneumonia from normal chest X-ray images. Binary Cross-Entropy measures the divergence between predicted probabilities and actual class labels, allowing the models to learn more discriminative boundaries between the two classes (Murphy, 2022). Its suitability is reinforced in medical diagnosis tasks where probabilistic outputs are valuable for threshold-based decision-making.

To enhance model robustness, the training dataset was subjected to augmentation techniques including random rotation, zooming, horizontal shifting, vertical shifting, and horizontal flipping. Data augmentation is a critical practice in deep learning because it artificially expands the diversity of the training set, reducing overfitting and improving the model's ability to generalize to unseen samples (Shorten & Khoshgoftaar, 2019). This is particularly important in medical imaging, where collecting large, balanced datasets is often challenging.

A fixed batch size was maintained throughout the experiments to ensure stable gradient estimation and efficient memory utilization. Model checkpoints and performance logs were recorded at each epoch to monitor convergence behavior. Accuracy and loss curves were generated to assess learning progression, while confusion matrices, ROC curves, and classification reports were used for post-training evaluation.

To ensure external validity, model performance was tested on two separate datasets: the curated secondary dataset used for training and validation, and the independent primary dataset collected from Kenyan healthcare facilities. This dual-evaluation strategy allowed assessment of both generalization to benchmark datasets and practical applicability in local clinical environments. Testing on an independent local dataset is especially important in healthcare AI research because models trained on foreign populations may not always transfer effectively to different demographic or institutional contexts (Zech et al., 2018).

This experimental setup established a controlled environment for evaluating the relationship between CNN architectural depth and diagnostic performance. By standardizing training conditions and incorporating both benchmark and real-world evaluation scenarios, the study ensured methodological rigor and strengthened the credibility of its findings.

3.9 Chapter Summary

This chapter has outlined the research methodology used to develop and evaluate a shallow CNN model for pneumonia classification. The study utilized both primary

data collected from selected healthcare facilities in Kenya and secondary data from publicly available repositories, ensuring dataset diversity and improved generalizability.

Data preparation procedures were detailed, including cleaning, anonymization and preprocessing steps such as normalization, resizing and augmentation. These steps were critical in enhancing data quality and improving model robustness. The chapter also provided an in-depth explanation of the proposed CNN architecture, including layer configurations, parameter calculations, and the role of each component in feature extraction and classification.

The chapter established the evaluation framework, highlighting the importance of multiple performance metrics in assessing the model's diagnostic reliability. These metrics ensure a balanced evaluation, particularly in medical contexts where classification errors have significant consequences.

Finally, an experimental testing was setup and standardized across all models to ensure fairness and reliability in performance comparison. Each model was trained for 50 epochs using the RMSprop optimizer and Binary Cross-Entropy loss function. This consistency allowed performance differences to be attributed primarily to architectural design rather than training variability.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Introduction

The primary goal of this research was to address the computational inefficiencies often associated with deep learning in clinical settings by developing a streamlined diagnostic tool. Although high-parameter architectures have traditionally dominated medical image classification, their deployment in resource-constrained hospital environments and on edge devices remains a significant challenge. This chapter presents the findings derived from the systematic execution of the study's research objectives.

Building on the literature review in Chapter two, which identified a gap in the availability of high-performing yet lightweight and shallow models, and the methodology outlined in Chapter 3, this chapter marks the transition from theoretical design to empirical evaluation. It focuses on the development and assessment of a lightweight, shallow neural network for the classification of pneumonia in chest radiographs.

The discussion is structured around three core objectives. First, an analysis of existing lightweight and deep learning models for medical image classification is conducted to establish a performance baseline and justify the need for a customized approach. Second, the design and architecture of the proposed model are presented, with emphasis on key decisions such as reduced layer depth and parameter efficiency that contribute to its lightweight nature. Third, the model's performance is evaluated using a diverse dataset of chest radiographs, with metrics such as accuracy and reliability used to assess its suitability for clinical application.

Through this structured analysis, the chapter demonstrates that reducing model complexity does not necessarily compromise diagnostic performance, thereby offering a practical and scalable approach to medical AI deployment.

4.2. Results of Model Performance

4.2.1 Model Training Performance

The baseline model was evaluated using a diverse dataset of chest radiographs, divided into training, validation, and testing subsets. The dataset comprised 6,429 training images, 616 validation images, and 1,136 testing images in the secondary dataset, with representation from both Pneumonia and Normal classes. This distribution ensured that the model was exposed to varied data during training and fairly assessed on unseen samples, supporting a reliable evaluation of its performance.

During training, the model's learning behavior was monitored over 50 epochs using accuracy and loss metrics. The results showed a rapid improvement in accuracy within the first 10 epochs, followed by stabilization. Notably, validation accuracy consistently remained higher than training accuracy, stabilizing at approximately 94% compared to 90% training accuracy. Similarly, validation loss was lower than training loss. This pattern suggests that the model generalized well to unseen data, although it may indicate slight under-fitting or differences in complexity between the training and validation datasets. The use of regularization techniques such as dropout during training may also have contributed to under-fitting. This is reflected in figure 4.1 below that depicts the models' training and validation accuracies and loss over 50 epochs

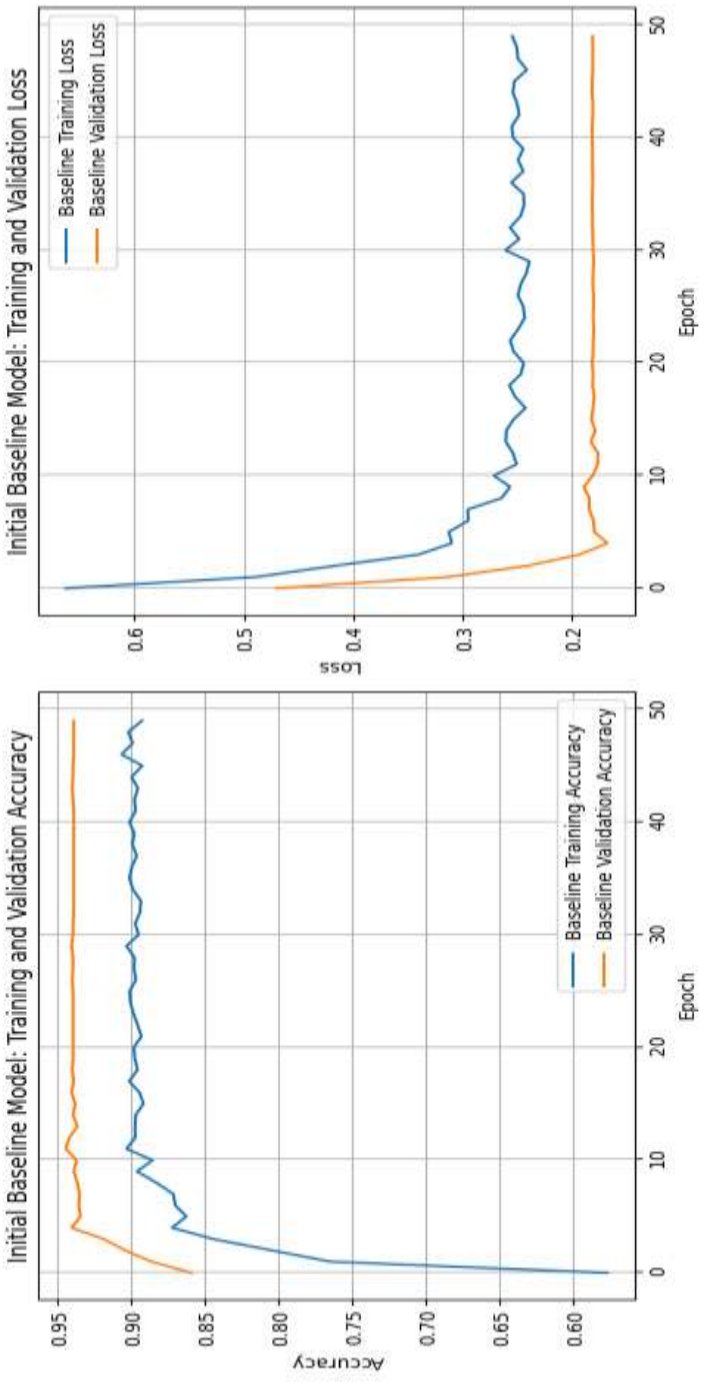


Figure 4.1: Initial Baseline Models' Training and Validation Accuracies and Loss over 50 epochs

4.2.2 Model Test Performance

The baseline model's performance was further evaluated on the test dataset, which represents completely unseen data. An overall accuracy of 91% was achieved, indicating strong predictive capability for a medical imaging classification task. Additional performance metrics provided deeper insight into the model's reliability. A precision score of 0.90 indicates that 90% of the cases predicted as Pneumonia were correct, while a recall score of 0.94 shows that the model successfully identified 94% of actual Pneumonia cases. The high F1-score reflects a good balance between precision and recall, and the corresponding ROC-AUC value is inferred to be high, demonstrating strong class discrimination across thresholds. The above report captures in table 4.1 below

Table 4.1: Baseline Models' Performance on Secondary Test Data

	Precision	Recall	F1-Score	Support
NORMAL	0.93	0.88	0.90	536
PNEUMONIA	0.90	0.94	0.92	600
accuracy			0.91	1136
macro avg	0.91	0.91	0.91	1136
weighted avg	0.91	0.91	0.91	1136

Error analysis further supports the evaluation of reliability. The baseline model exhibits a Type I error rate of approximately 10%, meaning some healthy patients are incorrectly classified as having Pneumonia. While this may lead to unnecessary follow-up procedures, it is generally acceptable in clinical contexts. More importantly, the Type II error rate is approximately 6%, indicating that only a small proportion of actual Pneumonia cases are missed. Since false negatives pose a greater clinical risk, the model's higher recall demonstrates a preference for minimizing these critical errors. This is depicted from the confusion matrix in figure 4.2 below.

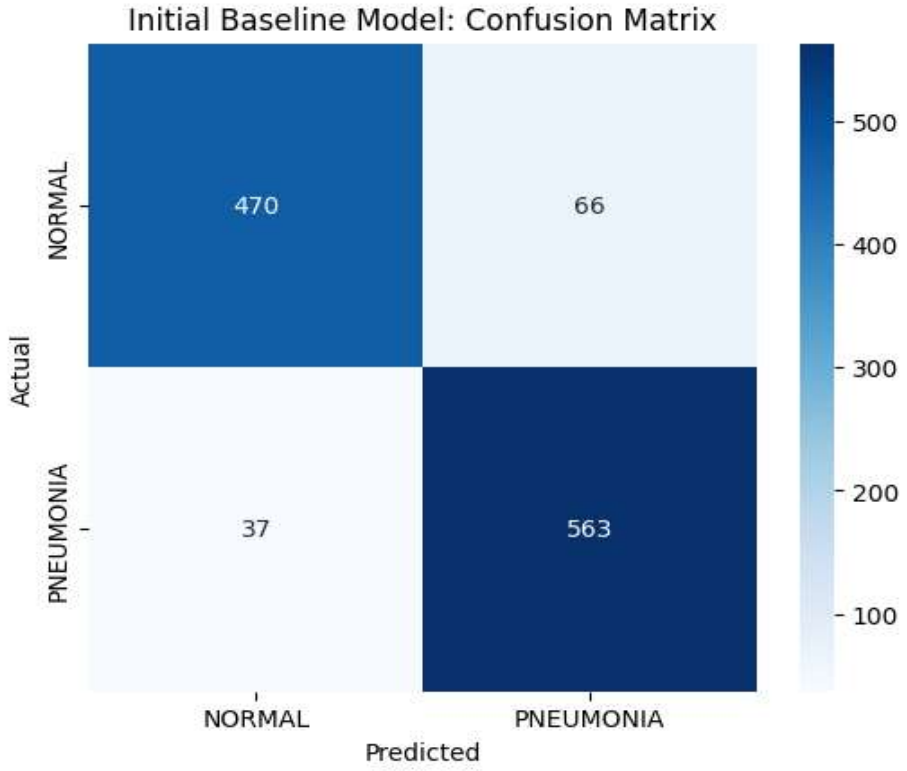


Figure 4.2: Baseline Models' Confusion Matrix on Secondary Test Data

The ROC curve for the baseline model in figure 4.3 below demonstrates a high level of accuracy and reliability in distinguishing between pneumonia and normal chest radiographs when evaluated on secondary test data. This value of approximately 0.96 reflects excellent discriminative ability, meaning the model has approximately a 96% probability of correctly differentiating between positive pneumonia cases and healthy radiographs. The ROC curve rises steeply toward the top-left corner, indicating that the model achieves high sensitivity while maintaining a low false positive rate.

This is especially important in medical diagnostics, as it shows the model can detect most pneumonia cases without significantly increasing the misclassification of healthy patients. Furthermore, the optimal operating region of the curve lies around a false positive rate of 0.1 to 0.2. At this point, the model is able to correctly identify over 90% of pneumonia cases while only misclassifying a small proportion of normal cases. This balance supports its suitability for clinical application, where maximizing detection while minimizing unnecessary interventions is critical.

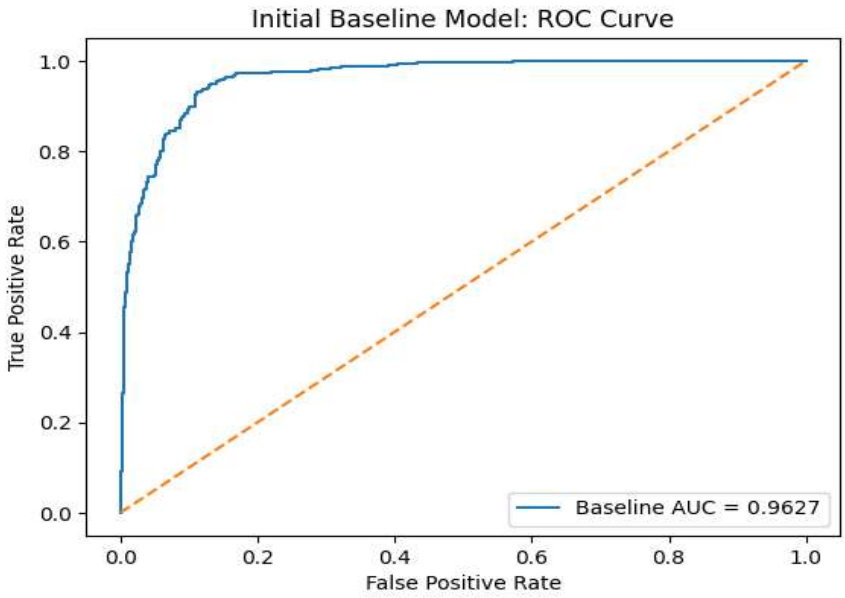


Figure 4.3: Initial Baseline Models' AUC- ROC on Secondary Test Data

The ROC analysis in figure 4.3 above, confirms that the model is both accurate and reliable for pneumonia classification. Overall, the evaluation demonstrates a strong balance between accuracy and reliability, with consistent generalization across training, validation, and testing phases. Additionally, the model's error profile aligns with clinical priorities, supporting its suitability as a supportive diagnostic tool in real-world healthcare settings.

4.3 Models performance on Kenyan test data

The baseline model proposed in section 2.9 was supplied with 635 test images categorized as 235 Normal and 400 Pneumonia images, and its performance noted. The confusion matrix and classification report provide a comprehensive evaluation of the baseline model's diagnostic performance, demonstrating both its strengths and limitations in a clinical context. With an overall accuracy of 95% on 635 test images (235 Normal and 400 Pneumonia), the model shows a high capability to correctly classify chest radiographs.

A closer examination of the confusion matrix reveals that the model correctly identified 366 pneumonia cases (true positives) and all 235 normal cases (true negatives). Notably, there were no false positives, meaning no healthy patient was incorrectly diagnosed with pneumonia. However, 34 pneumonia cases were misclassified as normal (false negatives), representing the primary source of error. This can be calculated from the values provided figure 4.4 below

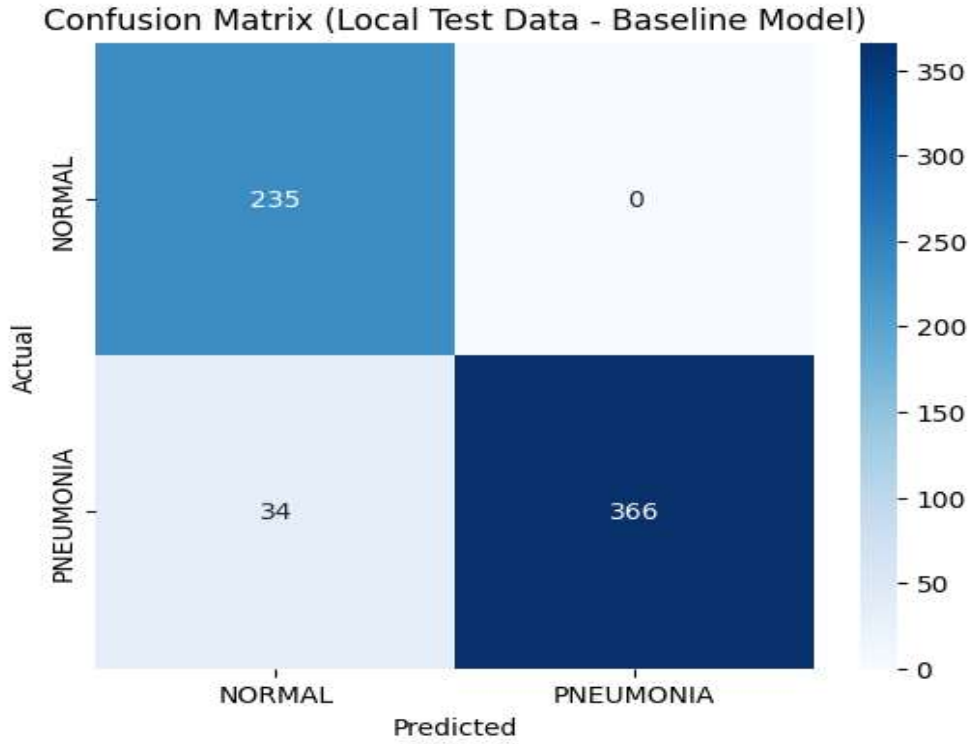


Figure 4.4: Initial Baseline Performance on Local Test Data

The evaluation metrics in table 4.2 below further clarify the model's diagnostic behavior. The precision for pneumonia is 1.00, indicating that every predicted pneumonia case was correct, reflecting perfect specificity and zero false alarms. In contrast, the recall (sensitivity) for pneumonia is 0.92, meaning the model successfully detected 92% of actual pneumonia cases. While this is a strong result, the remaining 8% of missed cases highlights a critical limitation, as false negatives carry significant clinical risk. The F1-score of 0.96 demonstrates a strong balance between precision and recall, confirming the model's overall reliability in classification tasks.

Table 4.2: Baseline Models' Performance on Local Test Data

	Precision	Recall	F1-Score	Support
NORMAL	0.87	1.00	0.93	235
PNEUMONIA	1.00	0.92	0.96	400
accuracy			0.95	635
macro avg	0.94	0.96	0.94	635
weighted avg	0.95	0.95	0.95	635

Additionally, from figure 4.5 below, the high AUC-ROC value of 0.9923 reinforces the model's excellent ability to distinguish between pneumonia and normal cases across different decision thresholds.

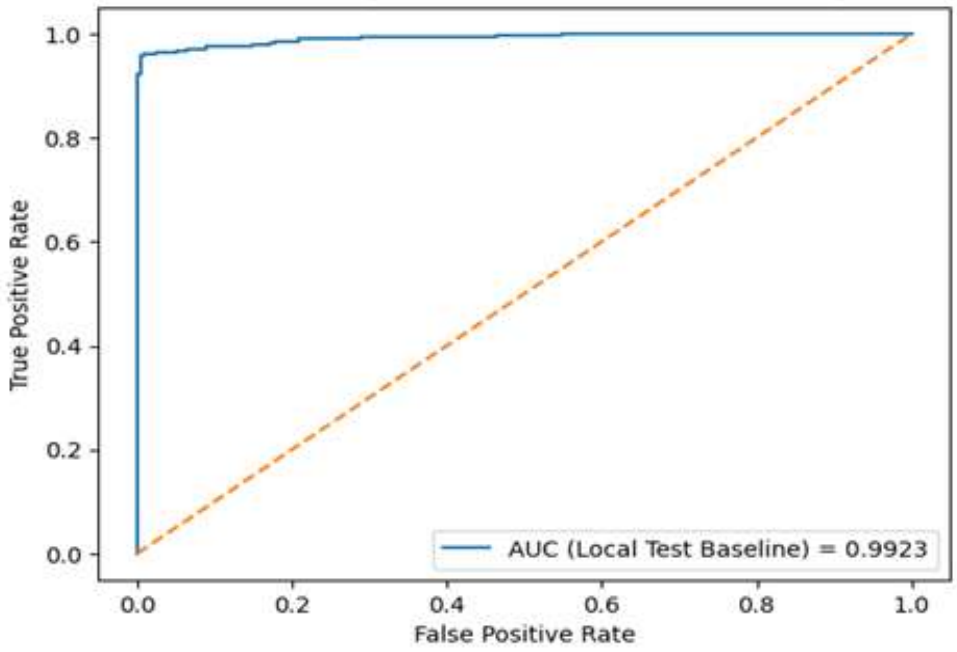


Figure 4.5: ROC – Curve for Local Test Data – Baseline Model

From an error analysis perspective, the model exhibits 0% Types I error rate, ensuring that no healthy individuals are subjected to unnecessary stress, treatment, or additional testing. This is highly beneficial in optimizing healthcare resources and maintaining

patient trust. However, the Type II error rate of approximately 8.5% indicates that some pneumonia cases are missed, which is more concerning in medical diagnostics, as it may delay critical treatment.

Overall, the results indicate that the model is highly conservative in its predictions, favoring certainty before classifying a case as pneumonia. This leads to perfect precision but introduces a slight trade-off in sensitivity. Despite this, the performance is highly promising for a lightweight, shallow architecture. It demonstrates that reduced model complexity can still achieve high diagnostic accuracy and near-perfect specificity. Consequently, the model is well-suited as a supportive or secondary screening tool, particularly in high-volume clinical settings, where it can assist radiologists by reliably flagging clear pneumonia cases while minimizing false alarms.

4.4 Effect on Number of Layers and Filters to Models' Performance

4.4.1 Number of Layers

In this study, multiple convolutional neural network (CNN) architectures were evaluated for pneumonia classification using chest X-ray images. The models evaluated ranging from three layers to ten layers. The outcome showed that simpler models, such as the three-layer and four-layer CNNs, lacked the depth to extract meaningful features, resulting in lower accuracy (below 85%) and AUC-ROC scores. The proposed baseline model that had five layers achieved 92% accuracy and an AUC of 0.96**.** It provided balance between complexity and performance while preventing overfitting.

Deeper models, six to ten layers offered marginal improvements in accuracy, reaching up to 96%, but at the cost of increased training time and overfitting, as indicated by rising loss values. The model that had ten layers had 2.7 million trainable parameters. It demonstrated excessive complexity minimum performance improvement; making it computationally inefficient. The proposed baseline CNN had five layers and was identified as the optimum model, providing high accuracy and AUC-ROC score whilst being computational efficient. This information is captured in table 4.3 below

Table 4. 3: Outcome of Model Performance per Number of Layer

Model Layers	Acc (%)	AUC	Loss	Training Parameters	Training Time (Sec)	Observations
3	78	0.8	0.32	600,000	240	Too simple, lacks depth to extract features effectively.
4	83	0.85	0.28	850,000	480	Slight improvement, but still lacks enough feature extraction capacity.
5(Proposed Baseline CNN)	90	0.96	0.22	1.2 M	720	Provides balance between complexity and performance. Prevents overfitting while improving accuracy.
6	93	0.97	0.21	1.4 M	900	Slight improvement, but increased complexity does not justify extra computation.
7	94	0.98	0.205	1.7 M	1080	Marginal gain, increased risk of overfitting.
8	95	0.98	0.215	2.0 M	1500	Too complex for dataset, diminishing returns.
9	95	0.98	0.23	2.3 M	1800	Increased overfitting, longer training time.
10	96%	0.98	0.25	2.7 M	2400	Overly deep, model complexity not worth the minor accuracy boost.

Figure 4.6 and figure 4.7 below demonstrate that the proposed baseline CNN model is the best trade-off between accuracy and number of trainable parameters.

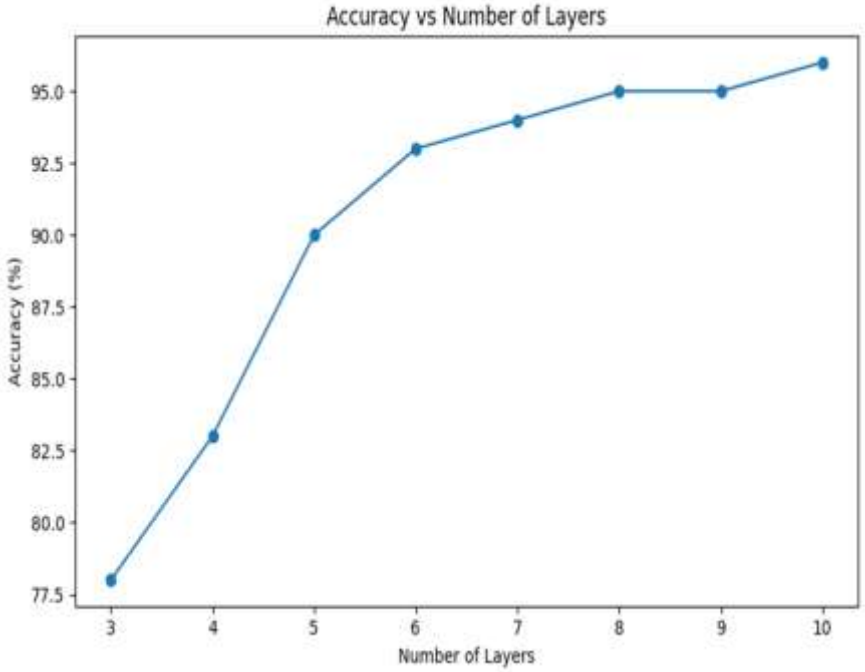


Figure 4.6: Model Accuracy vs Number of Layers

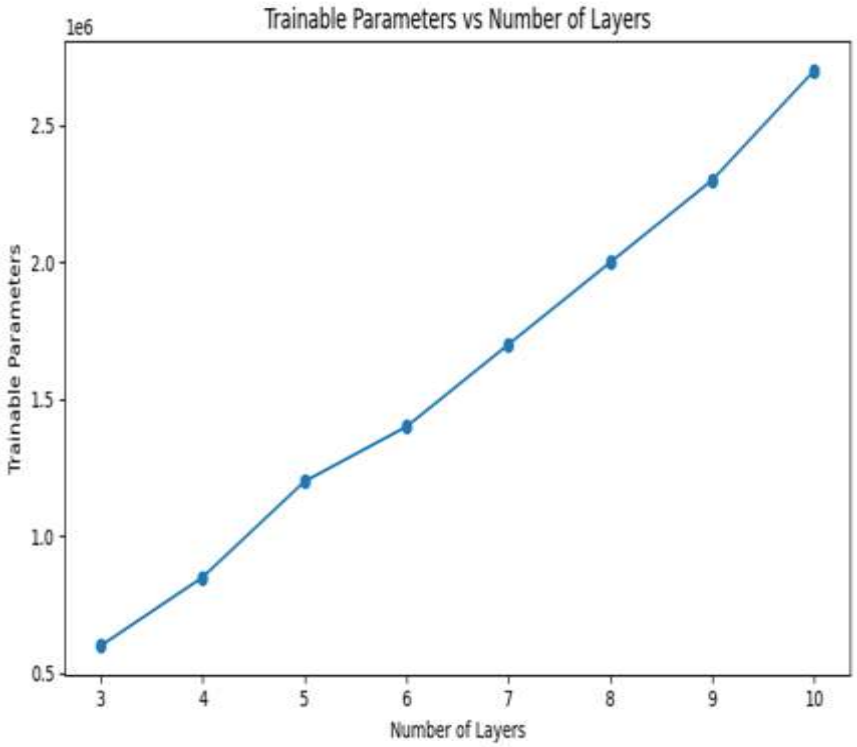
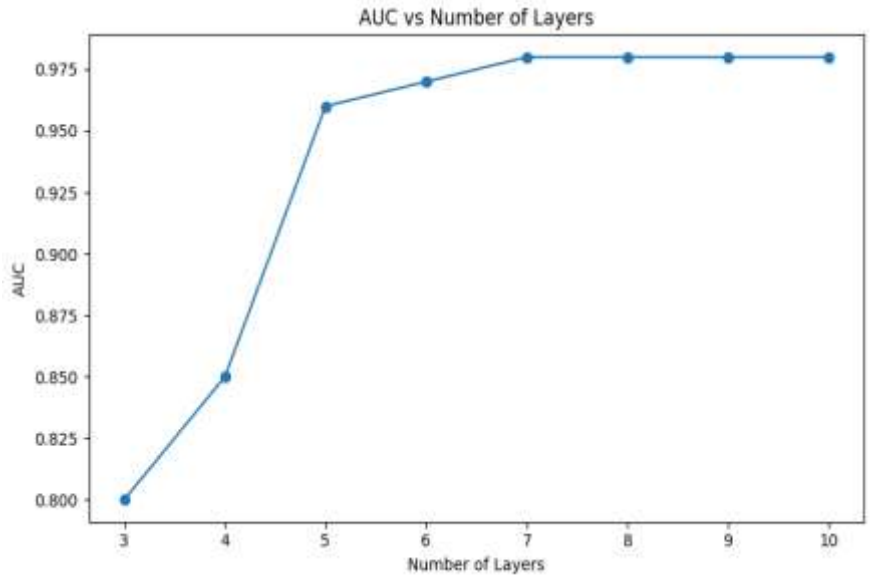


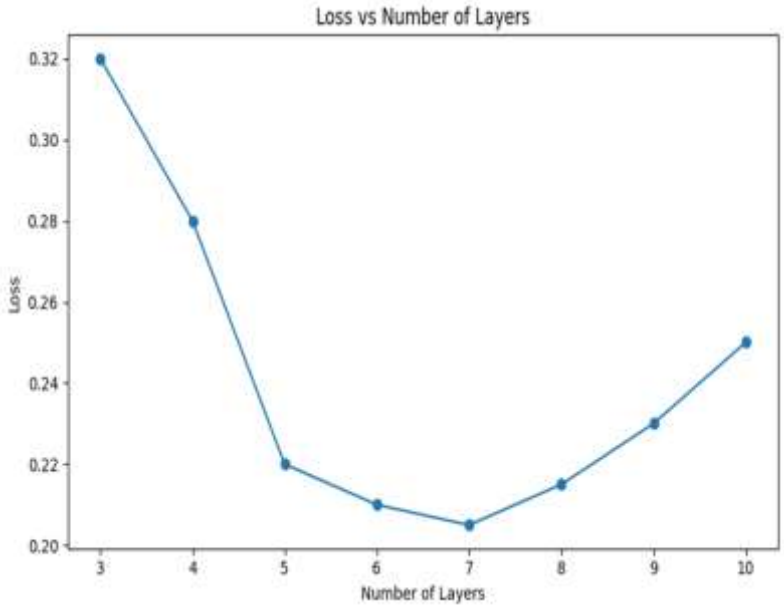
Figure 4.7: Trainable Parameters vs Number of Layers

The models performance of different architectures ranging from three to ten layers, were also evaluated based on AUC and loss. The three-layer model had the lowest performance; standing at AUC of 0.8 and a loss of 0.32**.** This is an indication that the model was under fitting as it struggled to learn patterns. There was a slight improvement with the model that had four layers. It achieved 0.85 and 0**.**275 on the same metrics. As the number of layers increased beyond five, it was noted that there was slight improvement on the AUC values. This stood at 0.97 for the model that had six layers and 0.98 seven-to-ten-layer models.

From figure 4.8 (b) below, the loss values continued to drop to 0.21 for the six and to 0.205 for the seven-layer model. However, the values of loss began to rise despite high AUC values. The models that had eight to 10 layers recorded loss of 0.215, 0.23, and 0.25 respectively. These exhibited signs of overfitting. Additionally, deeper models showed a significant increase in the number of trainable parameters. Overall, the proposed baseline five-layer model was considered the most effective, offering high performance while avoiding unnecessary complexity and overfitting. This layer had AUC of 0.96 and loss of 0.22.



(a)



(b)

Figure 4.8 (a) AUC vs. Number of Layers, (b) Loss vs. Number of Layers

4.4.2 Justification on the Number of Filters

In this section, the focus was to determine the effect of changing the filters to evaluate the model performance. The metrics considered first were accuracy and loss then evaluated how the model performed on test data. Doubling the number of filters increased the total number of the models' parameters to 9.7M. This number is four times higher as compared to the original model. Whilst these changes still maintained a higher accuracy of 90% and above, it was noted that a quarter of the images were still misclassified. Halving the number of filters of the original model, did reduce the number of normal images misclassified by a small margin. The above is captured in the table 4.4 below.

Table 4.4: Justification on the Number of Filters

HALVING FILTERS ORIGINAL FILTERS						**DOUBLING FILTERS ORIGINAL FILTERS**					
1st run						1st Run					
prec	recall	F1	ACC	loss	CNF	prec	recall	F1	ACC	loss	CNF
0.89	0.95	0.92	0.9	0.33	[[370 20]	0.88	0.98	0.93	0.91	0.33	[[382 8]
0.9	0.81	0.85			[45 190]]	0.96	0.79	0.86			[50 185]]
2nd Run						2nd Run					
prec	recall	F1	ACC	loss	CNF	prec	recall	F1	ACC	loss	CNF
0.89	0.96	0.92	0.9	0.31	[[376 14]	0.88	0.98	0.93	0.91	0.29	[[383 7]
0.93	0.8	0.86			[47 188]]	0.96	0.78	0.86			[51 184]]
3rd Run						3rd Run					
prec	recall	F1	ACC	loss	CNF	prec	recall	F1	ACC	loss	CNF
0.9	0.96	0.93	0.91	0.26	[[376 14]	0.88	0.97	0.92	0.9	0.29	[[380 10]
0.93	0.82	0.87			[43 192]]	0.95	0.77	0.85			[53 182]]
4th Run						4th Run					
prec	recall	F1	ACC	loss	CNF	prec	recall	F1	ACC	loss	CNF
0.91	0.95	0.93	0.91	0.27	[[370 20]	0.88	0.96	0.92	0.89	0.28	[[373 17]
0.91	0.85	0.88			[38 197]]	0.92	0.79	0.85			[49 186]]

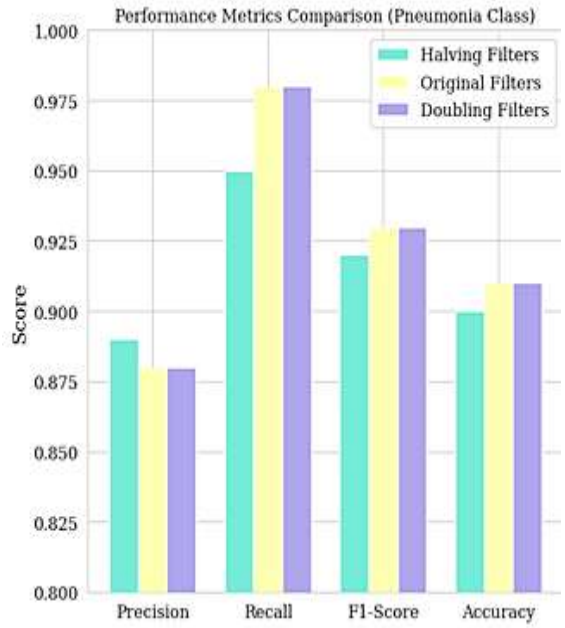
Key

prec- precision ACC- Accuracy

CNF-confusion Matrix [[TP FP]

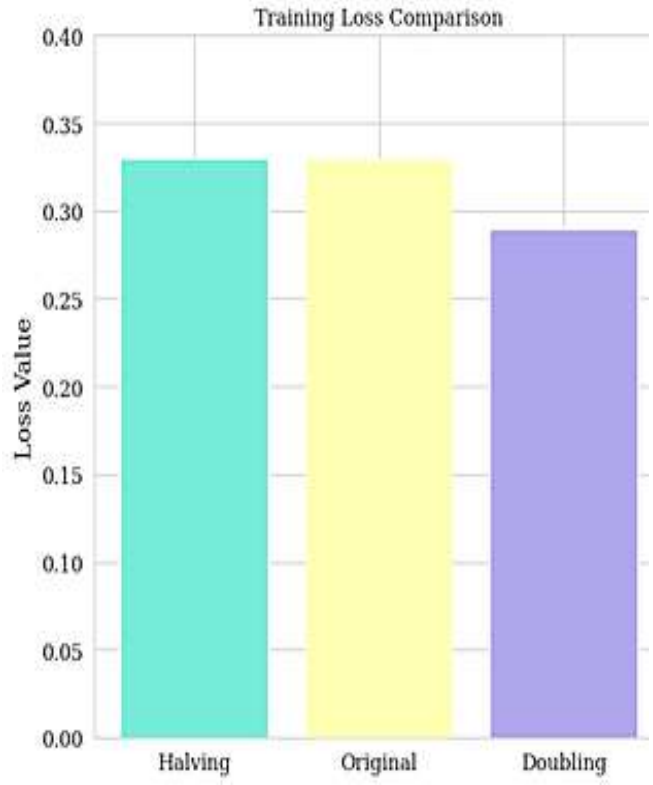
[FN TN]]

From table 4.4 above, empirical confirmation of the hypothesis that diagnostic efficacy in pneumonia classification does not linearly depend on model capacity is given by the comparative visualization of perturbation of architectures in the form of halving and doubling of filter counts with reference to the baseline configuration. The architecture of Doubling Filters, although having ten times more parameters than the baseline, did not bring significant improvements in recall (0.98) but cost with reduced precision (0.88) and higher false-positive rate, as shown in the performance matrices, as summarized in figure 4.9 below.

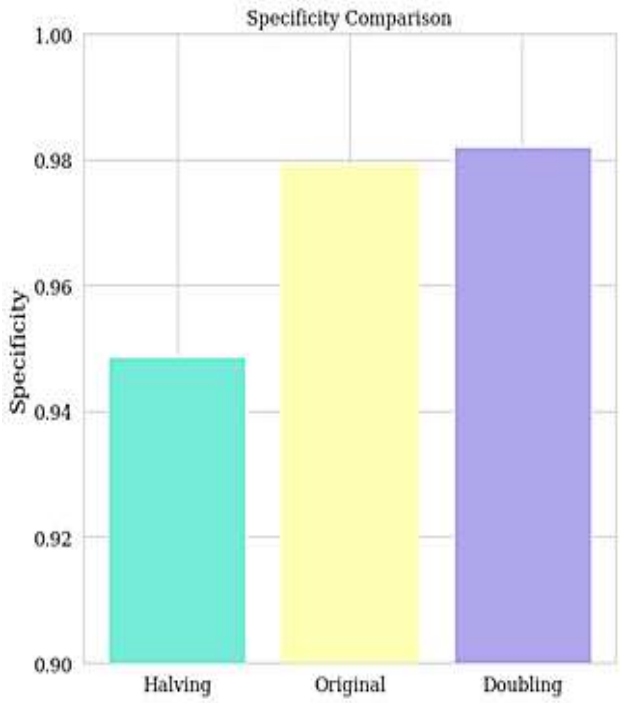

Performance Metrics Comparison (Pneumonia Class)
Halving Filters
Original Filters
Doubling Filters
Score
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0.800
Precision
Recall
F1-Score
Accuracy

(a)

(a)


Training Loss Comparison
Loss Value
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Halving
Original
Doubling

(b)



(c)

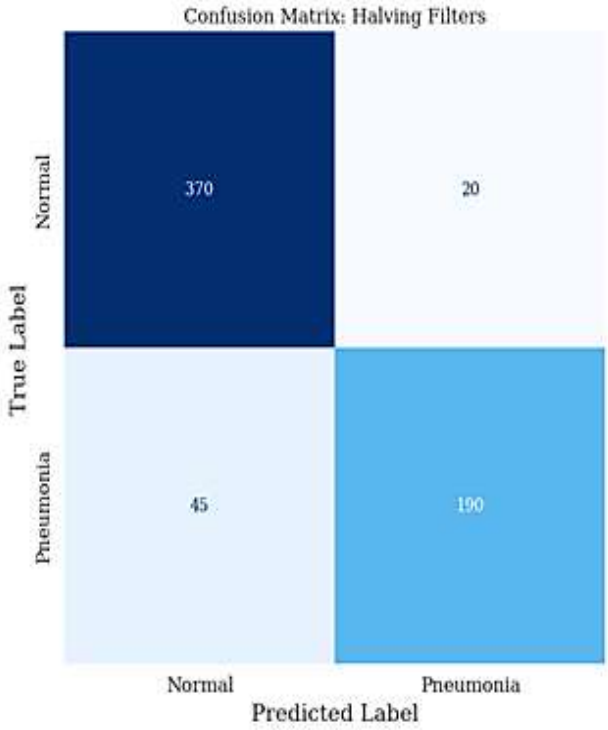
Figure 4.9 (a) Performance Metrics Comparison (Pneumonia Class); (b) Training Loss Comparison (c) Specificity Comparison

This effect is consistent with the recent results of Korkmaz (2025) who claims that parametric expansion in edge-aware application is prone to create instability in optimization without corresponding enhancement of generalization. The visual evidence indicates that the doubled architecture probably overfitted, learning noise in the training distribution instead of learning robust pathological features, and that lightweight networks, trained on memory-constrained hardware, have higher ratios of accuracy to parameters by not encountering the vanishing gradient problems of deeper, wider networks (Mohsin et al., 2025).

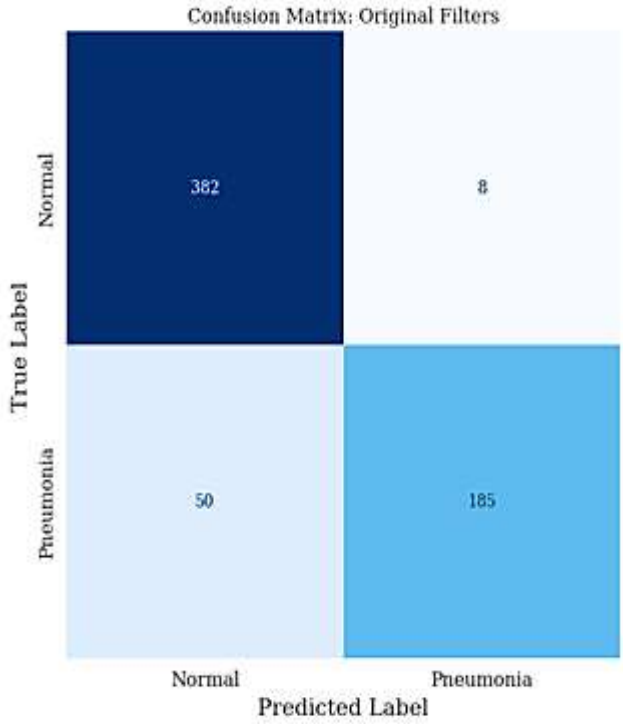
On the other hand, the Halving Filters arrangement exhibits a good preservation of diagnostic integrity, with an F1-score of 0.92 and an accuracy of 0.91, almost the same as in the baseline. This robustness serves as an illustration of the theoretical assumption that the decision boundaries that are complex can be well approximated by shallow and narrow networks when regularized properly, which explains the practical

applicability of the universal approximation theory to medical imaging as explained by Meir et al. (2023).

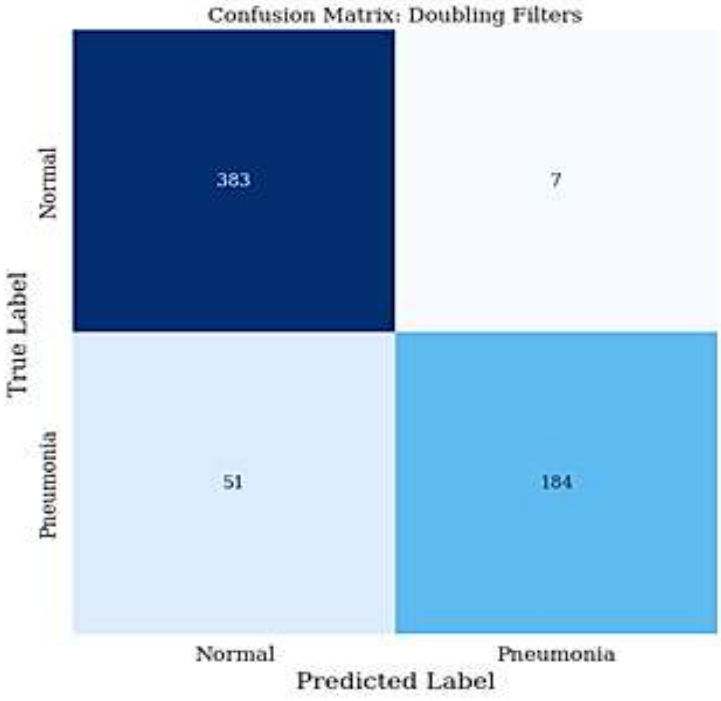
The confusion matrices in figures 4.10 below, indicate that halving filters had a small effect on the false positives (reduced by 8 to 20 in the representative sample), but a strong effect on retaining sensitivity, with the important clinical measure of recall being preserved. This is a trade-off that is clinically desirable in resource-constrained environments, as the computational savings of a reduced architecture makes it possible to execute on portable devices, without affecting the ability to detect life-threatening conditions, in overcoming the infrastructure gaps reported by McKnight et al. (2026) in Kenyan health facilities



(a)



(b)



(c)

Figure 4.10: (a) Confusion Matrices for Halving filters, (b) original filter, (c) doubling Filters

Moreover, the loss curves in the visualization make it clear that the efficiency of the baseline and halved architectures to find stable minima. The doubling approach had greater loss variance, which is the sign of the risk of the so-called "double descent" risk where more model complexity initially reduces performance, and then has a chance to increase it, assuming that there are enough data and regularization which is not usually the case in specialized medical data.

Recent work by Shahriar (2025) highlights that in the case of binary classification tasks, such as pneumonia detection, the distinguishing capacity is placed more on the quality of the feature extraction through convolutional kernels than the quantity of filters. The graphical findings, therefore, confirm the conclusion that the suggested shallow architecture is a Pareto optimal, being both computationally frugal and diagnostic outliers, and that the abandonment of unnecessarily deep or broad models in favor of interpretable designs made of streamlined, computational structures, is justified.

4.5 Experimental Testing

Three models; Baseline, Saraiva and Pneumonet were evaluated under controlled experimental conditions. Each model was trained for 50 epochs using RMSprop optimizer and Binary Cross-Entropy loss on an augmented dataset. Below is the discussion on models' performance during training and testing

4.5.1 Models Training Performance at Epoch 50

At the end of the training at 50 epochs, the performance of the Baseline, Saraiva and Pneumonet was recorded as captured in Table 4.5 below

Table 4.5: Summary of CNN model performance metrics at epoch 50

Metric (Epoch 50)	Baseline	Saraiva	Pneumonet
Train Accuracy	~91%	~93.5%	~83%
Val Accuracy	~90%	~84.5%	~84.5%
Train Loss	~0.26	~0.15	~0.37
Val Loss	~0.28	~0.39	~0.35
Train–Val Gap	~1%	~9%	~1.5%
Convergence Speed	Fast (epoch 5)	Fast (epoch 6)	Slow (epoch 10)

The Baseline CNN achieved a training accuracy of approximately 91% and a validation accuracy of approximately 90% by epoch 50, representing the smallest train-validation gap of all three models (approximately 1%). Training loss converged to approximately 0.26 while validation loss stabilised near 0.28. These tightly aligned curves across all 50 epochs constitute strong evidence of a well-regularised model that generalizes well to unseen data.

The Baseline's rapid convergence stabilising by approximately epoch 5, combined with its absence of early training instability, makes it the most predictable and reliable learner of the three architectures. In practical terms, this means the model reaches its near-maximum diagnostic capability early in training, reducing the computational cost of extended training runs and making it more amenable to fine-tuning workflows where training time is constrained.

The Saraiva model recorded the highest training accuracy at epoch 50 at approximately ~93.5%. Its validation accuracy plateaued at approximately 84.5%. The train-validation divergence was approximately nine percent. This gap, which is consistent throughout training after epoch 6, was an indicator of overfitting: the model has learned to classify the training set with high precision but has failed to learn features that generalize to unseen samples with equivalent reliability.
There was severe validation loss spike observed at approximately epoch 2, where the

validation loss reached approximately 2.6. This was more than five times the training loss at the same epoch. This spike reflects highly volatile and unreliable predictions on the validation set during early training, likely attributable to the architecture's sensitivity to initial weight configurations or batch sampling.

While the validation loss eventually stabilizes around 0.39 after epoch 8, the instability window represents a meaningful clinical risk in any scenario where the model might be deployed or fine-tuned before full convergence. Despite Saraiva achieving comparable or higher training accuracy, it was also notable that Saraiva's stabilised validation loss of approximately 0.39. This was higher than both Baseline and Pneumonet that stood at approximately 0.28 and 0.35 respectively. This indicates that Saraiva's probability outputs are less calibrated on unseen data.

The Pneumonet architecture exhibited a markedly delayed convergence pattern. Both training and validation accuracy remained at approximately 50% . This was equivalent to random binary classification for the first eight to ten epochs before rising sharply to approximately 83% on training data and 84.5% on validation data. Correspondingly, both loss curves held near the binary cross-entropy ceiling of approximately 0.695 before dropping steeply around epoch nine and ten.

This behaviour is consistent with architectures that require more gradient iterations before internal convolutional representations become sufficiently discriminative for the classification task. Once converged, Pneumonet's validation accuracy marginally exceeded its training accuracy by approximately 1.5%, and its validation loss was lower than its training loss by approximately 0.02%. This inversion where the model performs slightly better on validation than training is sometimes observed when training-time regularisation mechanisms such as drop out are more impactful during training than at inference. As such, this is interpreted as a sign of adequate regularisation rather than a concern. The train-validation gap of approximately 1.5% is acceptably small and suggests Pneumonet generalises well once trained.

However, Pneumonet's absolute accuracy is lower than the Baseline, and its delayed convergence represents a practical inefficiency. At epoch 50, Pneumonet and Saraiva achieve comparable validation accuracy, yet their loss profiles differ. Pneumonet's lower validation loss suggests that it produces better-calibrated probabilistic outputs than Saraiva on unseen data, despite similar accuracy figures. Figure 4.11 and figure 4.12 below capture the models' accuracy and loss after 50 epochs

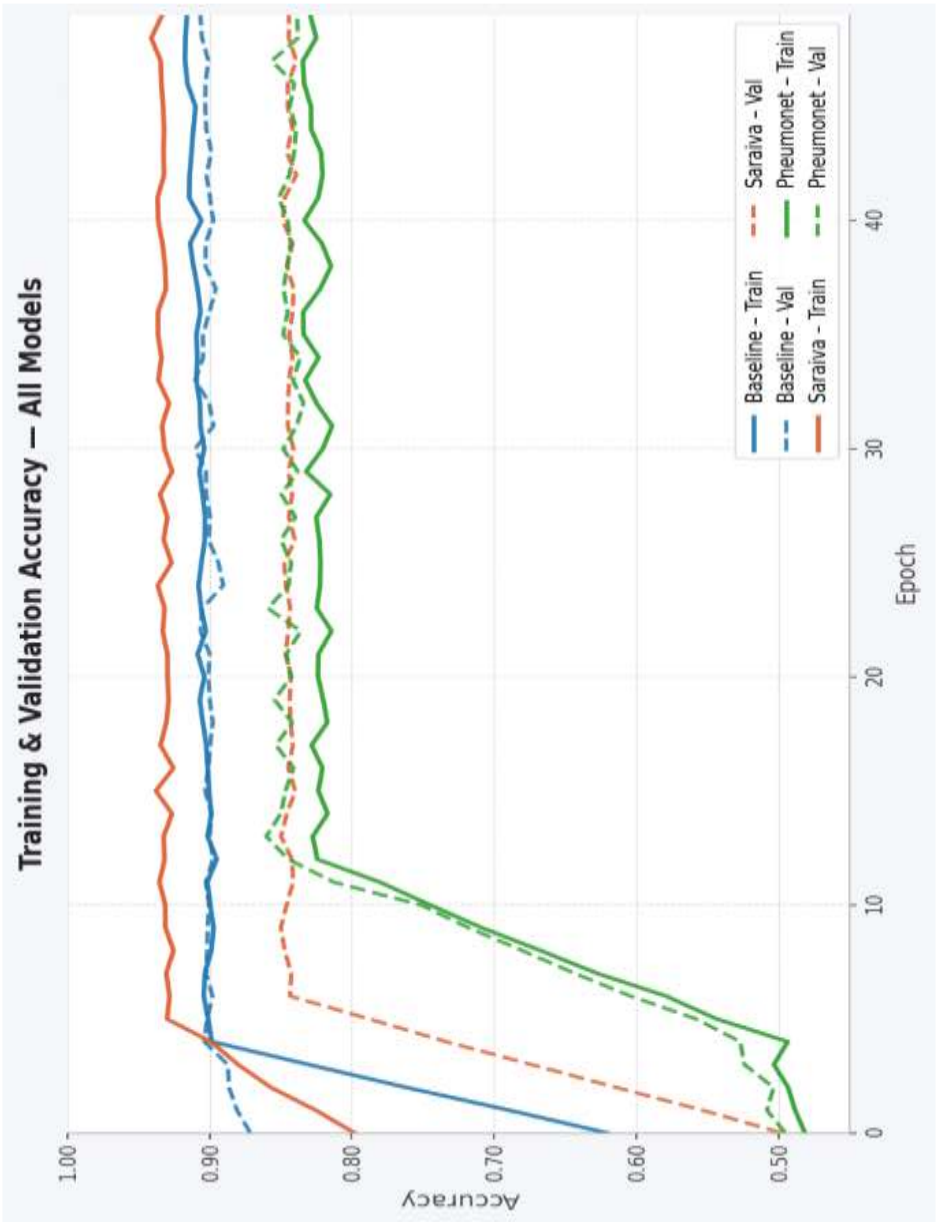


Figure 4.11: Training and Validation Accuracy

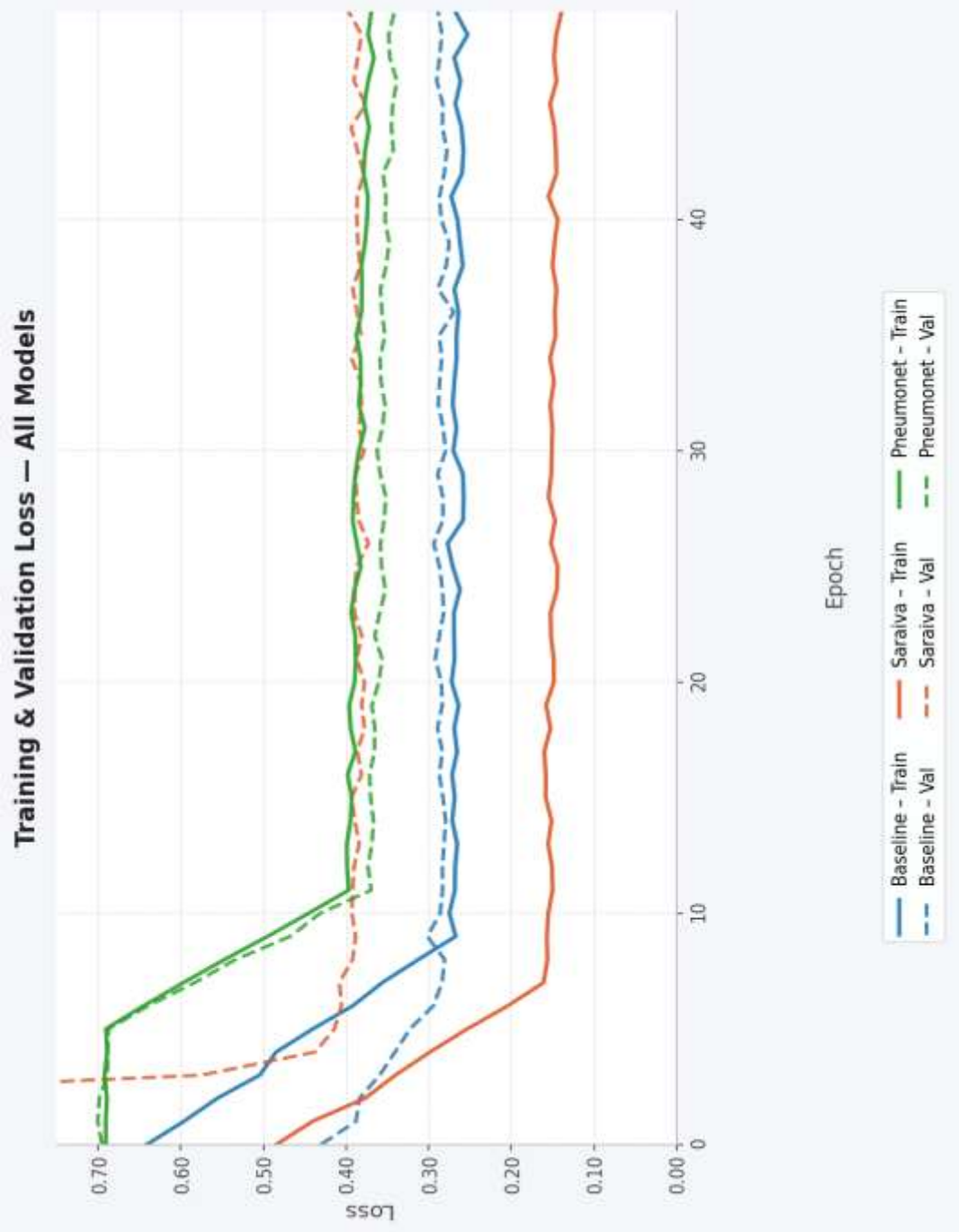


Figure 4.12: Training and Validation Loss

4.6 Models Test Performance

The evaluation of precision, recall, and the F1-score provides a nuanced understanding of how each model manages the trade-off between identifying as many positive cases

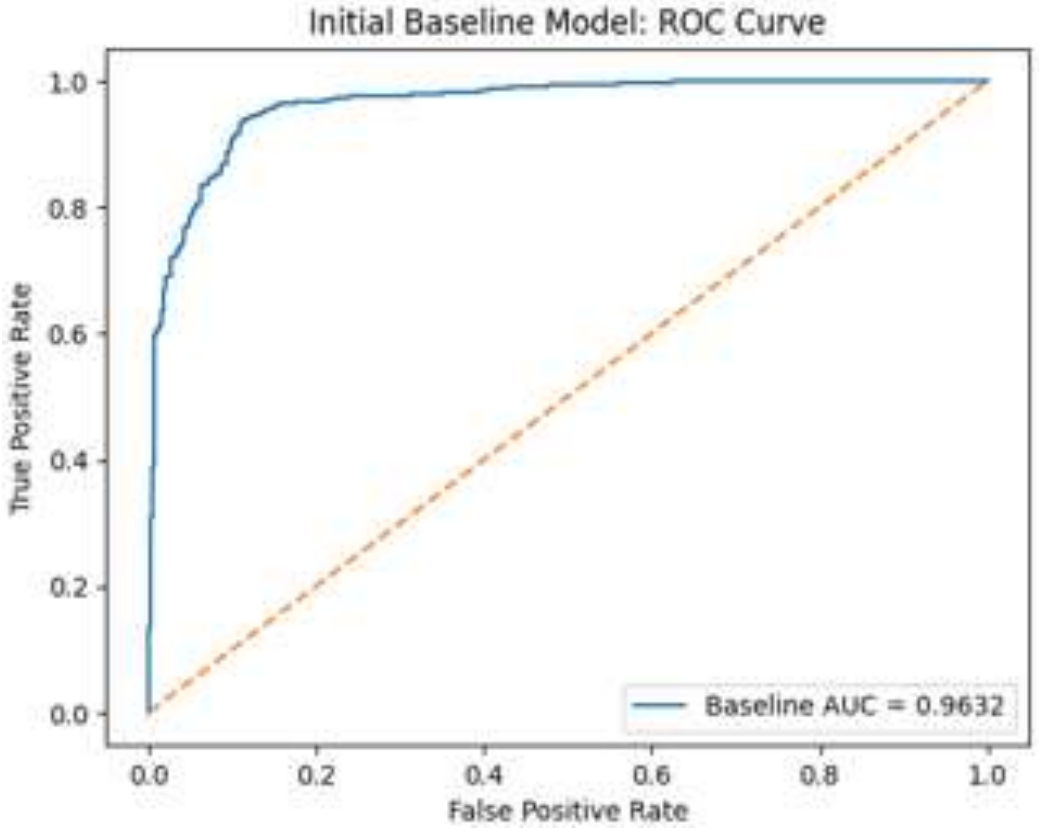
as possible and maintaining the accuracy of those identifications. The Baseline model demonstrated superior harmonic balance, achieving an F1-score of 0.92 for the pneumonia class as shown in Table 4.6 below. This performance is a result of its strong precision of 0.90 and recall of 0.94, indicating that the model is both highly reliable when it flags an infection and thorough in its screening process. In contrast, the Saraiva model prioritized recall at the expense of precision. While it successfully identified 98% of pneumonia cases. The Saraiva model had a recall rate of 0.98 and its lower precision of 0.73 indicates a high volume of false alarms. In turn this resulted in a lower F1-score of 0.84. This suggests that while the Saraiva architecture is an effective screening tool, it lacks the specificity required for a definitive diagnosis.

The Pneumonet model demonstrated moderate performance on the secondary test dataset, achieving an overall accuracy of 83%, with a precision of 0.91, recall of 0.76, and F1-score of 0.83 for the pneumonia class, as illustrated in Table 4.6 below. The high precision score of 0.91 indicates that when the Pneumonet model predicted pneumonia, the prediction was usually correct, suggesting that the model successfully learned strong pneumonia-associated radiographic features. However, the recall score of 0.76 indicates that approximately 24% of actual pneumonia cases were missed. In a medical imaging context, this is clinically concerning because missed pneumonia diagnoses may delay treatment and increase the risk of severe disease progression. The F1-score of 0.83 reflects a moderate balance between precision and recall, although the reduced recall lowered the model's overall clinical reliability when compared to the Baseline model.

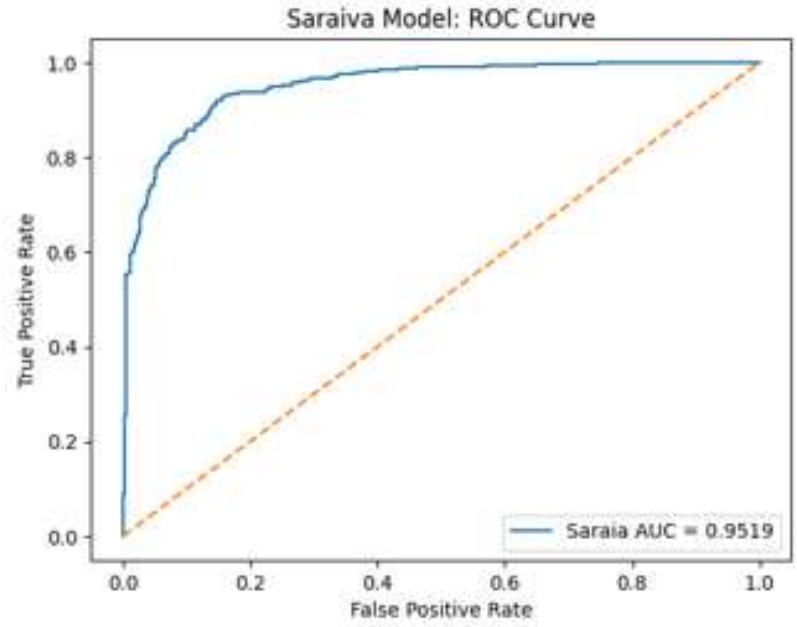
Table 4. 6: Analysis Table for the Pneumonia Class

Model	Precision	Recall	F1-Score
Baseline	0.90	0.94	0.92
Saraiva	0.73	0.98	0.84
Pneumonet	0.91	0.76	0.83

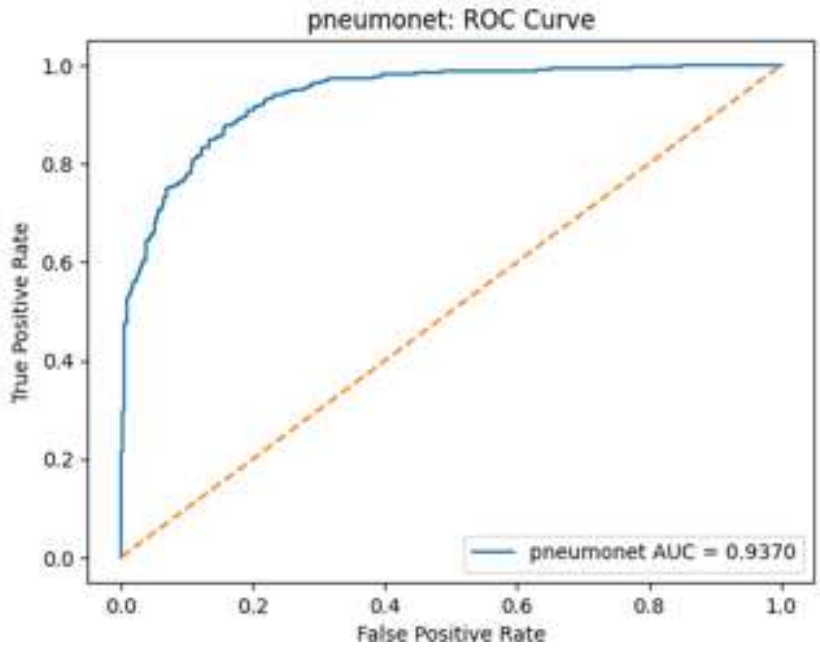
The ROC curves presented in Figure 4.13 below further demonstrate the comparative classification ability of the three models across varying threshold values.



(a)



(b)



(c)

Figure 4.13 ROC Curves for (a) Initial Baseline Model (b) Saraiva Model (c) Pneumonet

In the diagnostic evaluation of pneumonia, the error rates define the clinical reliability of a model. A Type I error (False Positive) occurs when the model incorrectly identifies a healthy patient as having pneumonia, while a Type II error (False Negative) occurs when the model fails to detect the presence of the disease. Table 4.7 below demonstrates the Type I and Type II error rates achieved by the three models.

Table 4. 7: Type I and Type II Error Rates

Model	TFalse Positives	Type I Error Rate	False Negative	Type II Error Rate
Baseline	64	11.9%	36	6.0%
Saraiva	218	40.6%	12	2.0%
Pneumonet	44	8.2%	147	24.5%

Based on the testing results of 1,136 samples, the Baseline model demonstrated the most balanced performance, achieving a Type I error rate of 11.9% and a Type II error rate of 6.0%. This outcome suggests that the model's depth and parameter distribution were well-calibrated, allowing it to learn the distinct physiological markers of both healthy lung tissue and infected consolidation. By maintaining low rates for both error types, the Baseline model avoids the "alarm fatigue" caused by excessive false positives while ensuring that many sick patients are correctly identified for treatment. The confusion matrix for the Baseline model is presented in Figure 4.14 below.

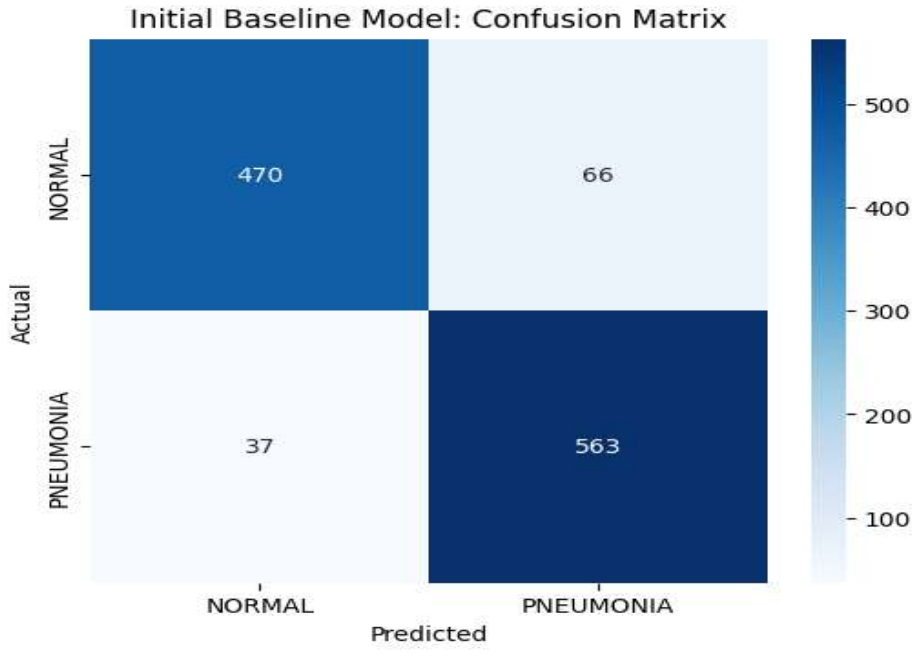


Figure 4.14: Baseline Confusion Matrix

The Saraiva model presented a different performance profile, characterized by an exceptionally low Type II error rate of only 2.0% but a significantly high Type I error rate of 40.6% as shown in Table 4.7. This indicates a model that is overly sensitive. The primary reason for this behavior lies in its architectural efficiency; with only 310,977 parameters and the use of Global Average Pooling, the model likely prioritized broad textural irregularities over specific localized features. In a clinical context, while the Saraiva model is effective at ensuring almost no pneumonia cases are missed, its high false-positive rate would lead to a heavy burden on healthcare systems, as four out of every ten healthy patients would be subjected to unnecessary secondary testing or antibiotic prescriptions. Figure 4.15 depicts Saraiva's confusion matrix.

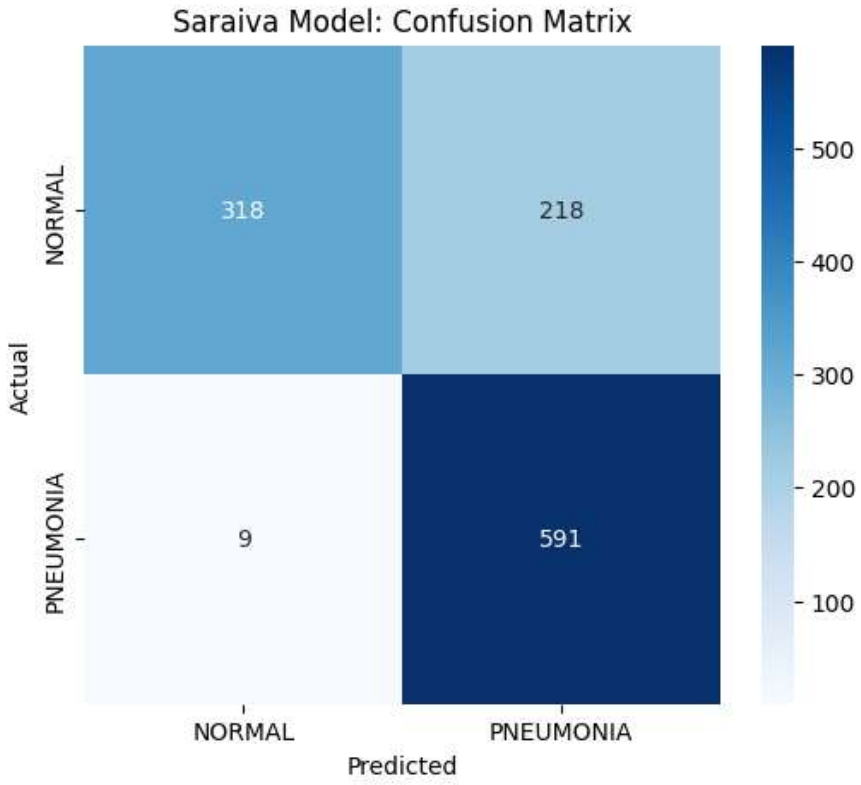


Figure 4.15: Baseline Confusion Matrix

The Pneumonet's confusion matrix presented in Figure 4.16 below shows that the model correctly classified 492 normal cases and 453 pneumonia cases, while misclassifying 44 healthy cases as pneumonia and 147 pneumonia cases as normal.

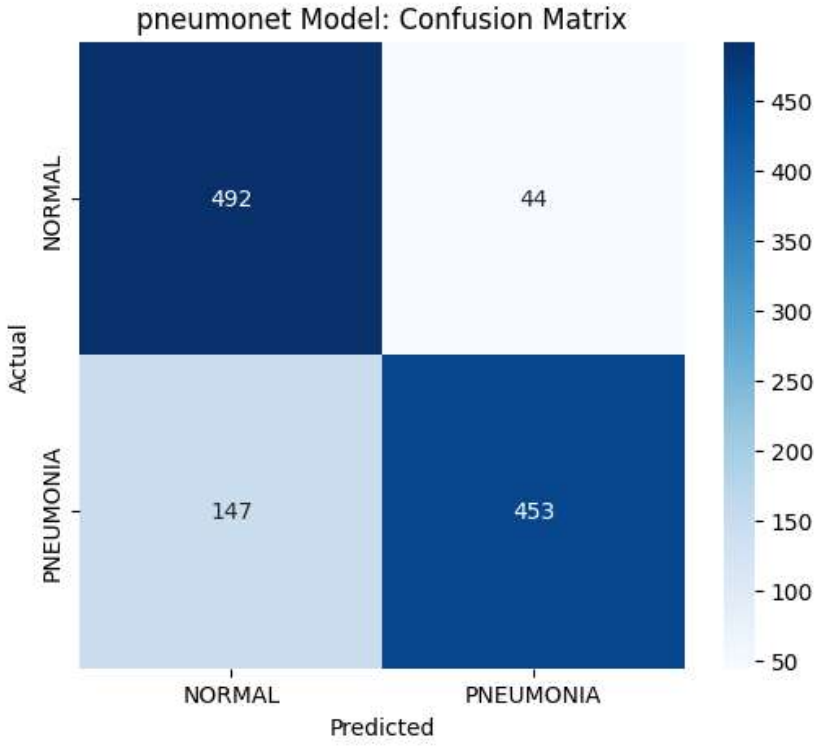


Figure 4.16: Pneumonet Confusion Matrix

The model achieved a Type I error rate of 8.2% and a Type II error rate of 24.5%. This indicates that the model was relatively conservative in predicting pneumonia and was more effective at avoiding false alarms than the Saraiva model. The low false positive rate demonstrates stronger specificity and suggests that the model learned discriminative pneumonia-related imaging features. However, the substantially higher false negative rate indicates that many pneumonia cases were missed during classification. In medical imaging, such missed diagnoses are clinically dangerous because untreated pneumonia may progress to severe respiratory complications.

In a medical research context, the Type II Error (False Negative) is generally considered more dangerous because it leads to a patient being sent home without treatment. Although the Pneumonet model achieved a low Type I error rate and reduced unnecessary diagnoses, its elevated Type II error rate remains clinically significant because missed pneumonia cases may delay treatment and worsen patient outcomes

4.7 Discussion of the Findings

The findings of this study provide evidence that architectural efficiency is a more decisive factor in medical image classification than model depth alone. Across all experimental evaluations, the proposed baseline model consistently demonstrated superior diagnostic performance, computational stability, and practical suitability when compared to deeper and more complex architectures.

A key observation from the study is that moderate architectural depth yields the most balanced trade-off between feature extraction capacity and generalization. The baseline model, with five effective layers and approximately 1.2 million trainable parameters, achieved the strongest overall results on both benchmark and local Kenyan datasets. Its 91% test accuracy on the secondary dataset and 95% accuracy on the Kenyan dataset demonstrate strong adaptability across different populations and imaging conditions. This confirms that the model did not merely memorize patterns from one dataset, but learned clinically relevant features that generalized effectively.

The comparative experiments further revealed that increasing the number of layers beyond the proposed architecture produced diminishing returns. While deeper models occasionally achieved marginal improvements in isolated metrics such as recall, these gains were offset by increased loss values, unstable convergence, and significantly higher parameter counts. For instance, models with eight to ten layers reached up to 96% accuracy in some settings, but required more than double the parameters and training time, with clear evidence of overfitting. This highlights that in practical healthcare deployments, marginal gains in performance do not justify substantial increases in computational demand.

The study also demonstrated that wider architectures, achieved by doubling filter counts, do not necessarily improve diagnostic outcomes. Although these models increased representational capacity, they introduced optimization instability and greater susceptibility to overfitting. In contrast, halving filter counts preserved nearly identical classification performance while reducing computational requirements. This

finding supports the hypothesis that in binary medical classification tasks, efficient feature extraction is more valuable than excessive network capacity.

A particularly significant outcome was the poor performance of deeper reference architectures such as Pneumonet. Despite their complexity, these models exhibited classification collapse, where nearly all samples were predicted as pneumonia. While this produced artificially high recall, it resulted in unacceptably low precision and extreme Type I error rates. Such behavior demonstrates that deeper models can fail catastrophically when their complexity is not aligned with dataset size, optimization strategy, and architectural balance. This underscores the importance of context-aware design rather than reliance on depth as a proxy for capability.

From a clinical perspective, the Baseline model achieved the most desirable balance of diagnostic metrics. Its precision, recall, and F1-score indicate that it can reliably identify pneumonia cases while minimizing unnecessary false alarms. On the Kenyan dataset, the absence of false positives is particularly notable, as it reduces the burden of unnecessary follow-up procedures in already constrained healthcare systems. Although some false negatives remained, the overall error profile was acceptable and aligned with practical screening priorities.

The findings also contribute to the broader discourse on AI deployment in low-resource settings. Many state-of-the-art medical AI systems rely on large-scale infrastructure and computationally intensive architectures, limiting their feasibility in environments such as rural hospitals and community clinics. This research demonstrates that a carefully designed shallow CNN can provide comparable clinical value while remaining lightweight enough for deployment on accessible hardware. In doing so, it advances the case for sustainable and context-sensitive AI solutions in global healthcare. Ultimately, the study confirms that model effectiveness in pneumonia detection is driven by balance rather than complexity. Efficient architectural design, controlled parameter growth, and alignment with clinical realities are more important than maximizing depth or width. These findings position the proposed shallow CNN as both a technically robust and practically deployable solution for medical imaging classification. The findings of this study demonstrate that

architectural efficiency is a more determinant of pneumonia classification performance than network depth alone. Across all experimental evaluations, the proposed baseline model consistently demonstrated superior diagnostic performance, computational stability, and practical suitability compared with deeper and more complex architectures. These findings support the growing argument in the literature that carefully designed shallow CNNs can achieve performance comparable to, or better than, deeper models while requiring significantly fewer computational resources (Sutera, 2025; Rahman et al., 2021). Rather than maximizing network depth, the results indicate that balancing feature extraction capability with computational efficiency is more important for pneumonia classification.

Another finding was that moderate architectural depth provided the most favourable balance between feature extraction and model generalization. The proposed baseline model, comprising five convolutional layers and approximately 1.2 million trainable parameters, achieved 91% test accuracy on the benchmark dataset and 95% accuracy on the Kenyan dataset. These results suggest that the network successfully learned clinically meaningful radiographic features that generalized across different datasets. This finding is consistent with the theoretical perspective of the universal approximation theorem, which states that shallow neural networks can approximate complex functions provided they possess sufficient representational capacity (Hornik, 1991; Kidger & Lyons, 2020). It also supports Brigato and Iocchi (2022), who demonstrated that shallow CNNs trained from scratch can outperform deeper transfer-learning models when trained on relatively small medical image datasets.

The comparative experiments further revealed that increasing network depth beyond the proposed architecture produced diminishing returns. Although deeper models occasionally achieved marginal improvements in recall, these gains were accompanied by higher validation loss, unstable convergence, longer training time and larger parameter counts. These observations align with Zhang et al. (2021), who argued that increasing network capacity may improve representational power but simultaneously increases the likelihood of overfitting. Similarly, Geman et al. (1992) explained this phenomenon using the bias variance trade-off, whereby deeper networks reduce bias

at the expense of substantially higher variance. Consequently, the additional computational cost associated with deeper architectures offered little practical advantage within the context of pneumonia classification.

The study also demonstrated that increasing network width by doubling filter counts did not significantly improve diagnostic performance. Although wider models increased representational capacity, they exhibited greater optimization instability and a higher tendency toward overfitting. Conversely, reducing filter counts maintained comparable classification accuracy while reducing computational requirements. These findings agree with Bianco et al. (2018), who reported that lightweight CNN architectures often achieve similar predictive performance while substantially reducing parameter counts and floating-point operations. Likewise, Pereira et al. (2021) emphasized that computational efficiency should be considered alongside diagnostic accuracy when selecting CNN architectures for deployment in clinical environments.

A particularly important observation was the poor performance of the deeper Pneumonet reference architecture. Despite its greater complexity, the model experienced classification collapse, predicting nearly all images as pneumonia. While this behaviour resulted in high recall, it produced unacceptable precision due to excessive false-positive classifications. This outcome demonstrates that increasing architectural complexity alone does not guarantee improved diagnostic performance. Instead, it supports the findings of Rahman et al. (2021) and Sutera (2025), who reported that carefully optimized shallow networks can outperform substantially deeper architectures in medical image classification tasks while requiring considerably fewer computational resources.

From a clinical perspective, the proposed baseline model achieved the most balanced diagnostic performance across all evaluated metrics. The model demonstrated high precision, recall, and F1-score while producing no false-positive predictions on the Kenyan dataset. Reducing false-positive diagnoses is particularly valuable in resource-constrained healthcare environments because unnecessary referrals, additional investigations, and antibiotic prescriptions place additional burdens on already limited

healthcare resources. These findings reinforce the observations of Pereira et al. (2021), who concluded that lightweight CNNs provide a favourable balance between diagnostic accuracy and operational efficiency for real-world clinical deployment.

The findings also have important implications for artificial intelligence deployment in low-resource healthcare settings. Previous studies have highlighted the limitations of computationally intensive transfer-learning architectures for deployment in developing countries due to hardware constraints (Godasu et al., 2020). The proposed shallow CNN demonstrated that high diagnostic performance can be achieved using a lightweight architecture suitable for deployment on affordable computing platforms. This finding supports Sousa et al. (2022), who advocated architecture optimization rather than increased model complexity for mobile healthcare applications, and aligns with the recommendations of Krishna et al. (2026), who identified lightweight CNN architectures as a priority for expanding AI-assisted diagnosis in underserved healthcare systems.

The findings further address the research gap identified in Chapter Two. Existing studies have largely concentrated on improving pneumonia detection accuracy through increasingly deeper transfer-learning architectures or ensemble models such as those proposed by Chouhan et al. (2020). While these approaches achieve excellent predictive performance, they require substantially greater computational resources and are less suitable for deployment in low-resource environments. In contrast, the present study demonstrates that a carefully designed shallow CNN can achieve competitive diagnostic accuracy while maintaining significantly lower computational complexity. This contribution extends existing knowledge by showing that lightweight architectures can provide a practical alternative for healthcare facilities operating with limited computational infrastructure.

Despite these encouraging findings, several limitations should be acknowledged. First, the study focused exclusively on binary classification between normal and pneumonia chest X-rays, limiting the applicability of the model to other thoracic diseases. Second, although the inclusion of a Kenyan dataset strengthened external validation, the dataset remained relatively small compared with large international repositories. Third, model

evaluation was conducted under controlled experimental conditions rather than within routine clinical workflows. Finally, explainable artificial intelligence techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM), were not incorporated to provide visual explanations of the model's diagnostic decisions. Future research should address these limitations by evaluating multi-class disease classification, incorporating explainable AI techniques, and validating the proposed model in prospective clinical settings.

Overall, the findings demonstrate that model effectiveness in pneumonia detection depends more on balanced architectural design than on maximizing network depth or width. The proposed shallow CNN achieved strong diagnostic performance while maintaining computational efficiency and good generalization across independent datasets. These results support the growing body of evidence that lightweight convolutional neural networks provide a technically robust, computationally efficient, and clinically practical solution for pneumonia screening, particularly within resource-constrained healthcare environments.

4.8 Summary

This chapter presented the results and discussion of the experimental evaluation of multiple CNN architectures for pneumonia classification using chest X-ray images. The analysis was structured around the study’s three research objectives: examining existing lightweight and deep learning models, designing a lightweight shallow architecture, and evaluating its diagnostic performance on diverse datasets.

The findings demonstrated that the proposed Baseline model provided the best balance between accuracy, computational efficiency, and reliability. With strong performance across both secondary benchmark data and locally collected Kenyan data, the model proved capable of generalizing effectively across different imaging environments. Its shallow structure and optimized parameter distribution enabled high classification performance while maintaining low computational requirements.

Comparative evaluations showed that increasing model depth and width did not proportionally improve diagnostic outcomes. Instead, excessive complexity introduced optimization challenges, overfitting, and classification instability. Deeper models such as Pneumonet and GitHub failed to maintain balanced performance, often collapsing into majority-class predictions that undermined clinical utility.

The chapter further established that lightweight architectures can achieve clinically meaningful results without reliance on large-scale computational resources. This is particularly relevant in resource-constrained healthcare settings, where accessibility, speed, and efficiency are essential considerations.

Overall, the chapter confirmed that the proposed shallow CNN architecture is a viable and effective approach for pneumonia detection. The findings provide strong empirical support for the study's central premise that lightweight, well-balanced models offer a practical pathway toward scalable AI-assisted diagnosis.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter presents the concluding insights of the study, drawing together the findings, implications, and contributions of the research. The study set out to investigate whether a lightweight, shallow convolutional neural network could provide accurate and reliable pneumonia classification while remaining computationally efficient for deployment in resource-constrained healthcare environments.

Building on the results presented in Chapter Four, this chapter synthesizes the extent to which the research objectives were achieved and reflects on the broader significance of the findings. The chapter also highlights the study's practical contributions to medical imaging and artificial intelligence, particularly in the context of low-resource settings such as Kenya.

Finally, recommendations are provided for future research, policy implementation, and practical deployment strategies to guide the advancement of efficient AI-assisted diagnostic tools.

5.2 Conclusion

This study successfully demonstrated that lightweight and shallow convolutional neural networks can provide effective and reliable pneumonia classification from chest radiographs without requiring the computational overhead of deep architectures. The research addressed a critical gap in medical AI by focusing on models that are not only accurate but also practical for deployment in environments with limited infrastructure.

The first objective, which involved analyzing existing lightweight and deep learning models, established that increased architectural depth does not inherently lead to better diagnostic outcomes. Comparative evaluations showed that while deeper models possess greater theoretical representational power, they often suffer from optimization

instability, overfitting, and poor generalization when applied to limited medical datasets.

The second objective, which focused on designing a lightweight shallow model, resulted in the development of a five-layer baseline CNN architecture that balanced feature extraction with computational efficiency. Its streamlined structure, controlled parameter count, and regularization strategies enabled strong learning capacity while avoiding unnecessary complexity.

The third objective, evaluating the model on diverse datasets, confirmed its robustness and reliability. The model achieved high performance on both secondary benchmark data and locally collected Kenyan radiographs, demonstrating strong generalization and adaptability to real-world clinical scenarios. Its balanced precision, recall, and low error rates indicate suitability as a supportive diagnostic tool.

The study concludes that efficient architectural design is more important than model complexity in medical imaging classification. A carefully optimized lightweight CNN can achieve clinically relevant results while remaining accessible for deployment in low-resource settings. This positions lightweight AI models as a sustainable pathway toward equitable healthcare innovation.

5.3 Recommendations

Based on the findings of this study, five recommendations are proposed. First, healthcare institutions in low-resource environments should consider adopting lightweight AI models such as the proposed shallow CNN as supportive diagnostic tools. Their low computational requirements make them suitable for integration into hospitals, clinics, and mobile screening units where advanced hardware may not be available.

Second, future implementation efforts should focus on embedding the model into broader clinical workflows, enabling radiologists and clinicians to use AI-generated predictions as a second opinion rather than as a replacement for professional expertise.

Such integration would enhance diagnostic confidence while preserving the essential role of human oversight in clinical decision-making.

Third, to improve model robustness, fairness, and long-term applicability, healthcare stakeholders should invest in the development of larger and more diverse local medical imaging repositories. Expanding localized datasets will reduce dependence on foreign data sources and ensure that AI systems are better tailored to regional populations, disease prevalence, and institutional practices.

In addition, future work should involve further validation of the proposed model on a larger and more diverse dataset collected from multiple regions across Kenya. This would provide stronger evidence of the model's generalizability and reliability across varying demographic and healthcare contexts.

Finally, it is recommended that future research explore the development of a user-friendly interface or application programming interface (API) for the model. Such tools would facilitate seamless integration into existing healthcare systems, improve accessibility for end users, and support practical deployment in real-world clinical environments.

5.4 Summary

This chapter presented the conclusion and recommendations of the study, drawing together the findings and their implications for medical imaging and healthcare delivery. The chapter reaffirmed that lightweight, shallow convolutional neural networks can achieve reliable pneumonia classification performance while maintaining computational efficiency suitable for low-resource environments.

The conclusion highlighted that the proposed shallow CNN architecture successfully addressed the study objectives by demonstrating strong diagnostic capability across both benchmark and locally sourced Kenyan datasets. It established that efficient architectural design is more valuable than excessive model complexity in achieving clinically relevant outcomes.

The chapter also outlined practical recommendations for healthcare institutions, researchers, and policymakers. These included the adoption of lightweight AI models in resource-constrained settings, integration into clinical workflows as decision-support tools, investment in local medical imaging repositories, broader validation across diverse Kenyan regions, and the development of user-friendly deployment interfaces.

REFERENCES

Abdulsahib, A. A., Mahmoud, M. A., Mohammed, M. A., Rasheed, H. H., Mostafa, S. A., & Maashi, M. S. (2021). Comprehensive review of retinal blood vessel segmentation and classification techniques: intelligent solutions for green computing in medical images, current challenges, open issues, and knowledge gaps in fundus medical images. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *10*(1), 20–45. https://doi.org/10.1007/s13721-021-00294-7

Abiyev, R. H., & Ma'aitah, M. K. S. (2018). Deep Convolutional Neural Networks for Chest Diseases Detection. *Journal of Healthcare Engineering*, *2018*. https://doi.org/10.1155/2018/4168538

Afzal, S., Rajmohan, C., Kesarwani, M., Mehta, S., & Patel, H. (2021). Data readiness report. *2021 IEEE International Conference on Smart Data Services (SMDS)*, 42–51.

Aishwarya, A., Nivi, N., Kesari, A., Khalid, M., Sre, S., Anantharaj, A., Shivraj, A., Ghuge, A., Vedant, V., & Karabekova, N. (2024). Intricacies of Pneumonia. *Scientific Collection «InterConf»*, *16*(194), 303–307.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Bashir, Z., Iqbal, N., & Hanif, M. (2021). A novel gray scale image encryption scheme based on pixels' swapping operations. *Multimedia Tools and Applications*, *80*(1), 1029–1054. https://doi.org/10.1007/s11042-020-09695-8

Bechar, M. E. A., Settouti, N., Daho, M. E. H., Adel, M., & Chikh, M. A. (2019). Influence of normalization and color features on super-pixel classification: application to cytological image segmentation. *Australasian Physical & Engineering Sciences in Medicine*, *42*(2), 427–441. https://doi.org/10.1007/s13246-019-00735-8

Berner, E. S. (2019). *Clinical decision support systems: State of the art*. Agency for Healthcare Research and Quality.

Bianco, S., Cadene, R., Celona, L., & Napoletano, P. (2018). Benchmark analysis of representative deep neural network architectures. *IEEE Access*, 6, 64270-64277. https://doi.org/10.1109/ACCESS.2018.2877890

Brigato, L., Barz, B., Iocchi, L., & Denzler, J. (2022). Image classification with small datasets: Overview and benchmark. *IEEE Access*, *10*, 49233–49250.

Brownlee, J. (2020). *Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python*. Machine Learning Mastery.

Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., Bennett, C., Hawken, S., Magwood, O., Sheikh, Y., McInnes, M., & Holzinger, A. (2023). Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 329–341. https://doi.org/10.1109/TPAMI.2022.3145392

Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., ... & de Albuquerque, V. H. C. (2020). A novel transfer learning-based approach for pneumonia detection in chest X-ray images. *Applied Sciences*, 10(2), 559. https://doi.org/10.3390/app10020559

Chowdhury, M. E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., ... & Islam, M. T. (2020). Can AI help in screening viral and COVID-19 pneumonia. *Ieee Access*, *8*, 132665-132676.

Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., Liu, R., Wang, X., Hou, X., Liu, Y., Long, X., Wen, Y., Lu, L., Shen, Y., Chen, Y., Shen, D., Yang, X., Zou, H., Sheng, B., & Jia, W. (2021). A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature Communications*, *12*(1), 3242–3267. https://doi.org/10.1038/s41467-021-23458-5

Data Carpentry. (2022). *Image Processing with Python*. https://datacarpentry. Github .io/image-processing/02-image-basics

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future healthcare journal*, *6*(2), 94-98.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340.

Deshpande, A. (2016). A beginner's guide to understanding convolutional neural networks. *Retrieved March*, *31*(2017).

Dhand, R., & Li, J. (2020). Coughs and Sneezes: Their Role in Transmission of Respiratory Viral Infections, Including SARS-CoV-2. *American Journal of Respiratory and Critical Care Medicine*, *202*(5), 651–659. https://doi.org/10.1164/rccm.202004-1263PP

Dharel, S., Shrestha, B., & Basel, P. (2023). Factors associated with childhood pneumonia and care seeking practices in Nepal: further analysis of 2019 Nepal Multiple Indicator Cluster Survey. *BMC Public Health*, *23*(1), 264.

Dubey, U. K. B., & Kothari, D. P. (2022). *Research Methodology*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315167138

Duta, I. C., Liu, L., Zhu, F., & Shao, L. (2021). Improved residual networks for image and video recognition. *2020 25th International Conference on Pattern Recognition (ICPR)*, 9415–9422.

ElSayed, M. S., Le-Khac, N.-A., Albahar, M. A., & Jurcut, A. (2021). A novel hybrid model for intrusion detection systems in SDNs based on CNN and a new regularization technique. *Journal of Network and Computer Applications*, *191*(10), 103160. https://doi.org/10.1016/j.jnca.2021.103160

Everett, K., Xiao, L., Wortsman, M., Alemi, A. A., Novak, R., Liu, P. J., Gur, I., Sohl-Dickstein, J., Kaelbling, L. P., & Lee, J. (2024). Scaling exponents across parameterizations and optimizers. *ArXiv Preprint ArXiv:2407.05872*.

Firdaus, M. (2024). KNN Machine Learning Architecture for Pneumonia Chest X-Ray Clustering. *Computers, and Electricals Engineering Journal (TELECTRICAL)*, *2*(1), 43–51. https://doi.org/10.26418/telectrical.v2i1.78604

Fogel, A. L., & Kvedar, J. C. (2018). Artificial intelligence powers digital medicine. *NPJ digital medicine*, *1*(1), 5.

Funaguchi, N., Ogasawara, M., Kiryu, T., Terashima, T., Gon, Y., Shimizu, T., & Sawai, H. (2023). Respiratory/Infection Symptoms. In *Internal Medicine for Dental Treatments* (pp. 3–11). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-3296-2_1

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1), 1-58. https://doi.org/10.1162/neco.1992.4.1.1

Godasu, R., Zeng, D., & Sutrave, K. (2020). Transfer learning in medical image classification: Challenges and opportunities. *MWAIS 2020 Proceedings*, 18.

Gonzalez, R. C. (2018). *lecture notes*. (November), 79–87.

Goodfellow, I., Carlini, N., Athalye, A., Papernot, N., Brendel, W., Rauber, J., Tsipras, D., Madry, A., & Kurakin, A. (2019). On Evaluating Adversarial Robustness. *Machine Learning*, 1–24.

Grousd, J. A., Rich, H. E., & Alcorn, J. F. (2019). Host-pathogen interactions in gram-positive bacterial pneumonia. *Clinical microbiology reviews*, *32*(3), 10-1128.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*(32), 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guetari, R., Ayari, H., & Sakly, H. (2023). Computer-aided diagnosis systems: a comparative study of classical machine learning versus deep learning-based approaches. *Knowledge and Information Systems*, *65*(10), 3881–3921. https://doi.org/10.1007/s10115-023-01894-7

Gupta, K., Kumar, P., Upadhyaya, S., Poriye, M., & Aggarwal, S. (2024). Fuzzy Logic and Machine Learning Integration: Enhancing Healthcare Decision-Making. *International Journal of Computer Information Systems and Industrial Management Applications*, *16*, 515–534.

Haji, S. H., & Abdulazeez, A. M. (2021). Comparison of optimization techniques based on gradient descent algorithm: A review. *PalArch's Journal of Archaeology of Egypt/Egyptology*, *18*(4), 2715–2743.

Hinton, G., Krizhevsky, A., Jaitly, N., Tieleman, T., & Tang, Y. (2012). Does the brain do inverse graphics. In *Brain and Cognitive Sciences Fall Colloquium* (Vol. 2).

Holzinger, A. (2019). Introduction to MAchine Learning & Knowledge Extraction (MAKE). *Machine Learning and Knowledge Extraction*, *1*(1), 1–20. https://doi.org/10.3390/make1010001

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251-257. https://doi.org/10.1016/0893-6080(91)90009-T

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. https://doi.org/10.48550/arXiv.1704.04861

https://www.medicinenet.com/pneumonia_facts/article.htm

Huang, S., Arpaci, I., Al-Emran, M., Kılıçarslan, S., & Al-Sharafi, M. A. (2023). A comparative analysis of classical machine learning and deep learning techniques for

predicting lung cancer survivability. *Multimedia Tools and Applications*, *82*(22), 34183–34198.

Huang, S.-C., Pareek, A., Jensen, M., Lungren, M. P., Yeung, S., & Chaudhari, A. S. (2023). Self-supervised learning for medical image classification: a systematic review and implementation guidelines. *Npj Digital Medicine*, *6*(1), 74–93. https://doi.org/10.1038/s41746-023-00811-0

Iiduka, H. (2021). Appropriate learning rates of adaptive learning rate optimization algorithms for training deep neural networks. *IEEE Transactions on Cybernetics*, *52*(12), 13250–13261.

Jablonka, K. M., Ongari, D., Moosavi, S. M., & Smit, B. (2020). Big-data science in porous materials: materials genomics and machine learning. *Chemical reviews*, *120*(16), 8066-8129.

Jain, V., Vashisht, R., Yilmaz, G., & Bhardwaj, A. (2021). Pneumonia Pathology. In *StatPearls*.

Joo, D., Kim, D., & Kim, J. (2021). A close look at deep learning with small data. . *2020 25th International Conference on Pattern Recognition (ICPR)*, 7677–7683. https://doi.org/10.1109/ICPR48806.2021.9413075

Jupelli, M., Murthy, A. K., Chaganty, B. K., Guentzel, M. N., Selby, D. M., Vasquez, M. M., ... & Arulanandam, B. P. (2011). Neonatal chlamydial pneumonia induces altered respiratory structure and function lasting into adult life. *Laboratory investigation*, *91*(10), 1530-1539.

Kamavisdar, P., Saluja, S., & Agrawal, S. (2013). A Survey on Image Classification Approaches and Techniques. *International Journal of Advanced Research in Computer and Communication Engineering*, *2*(1), 1005–1009.

Kaur, J., & Singh, W. (2022). Tools, techniques, datasets and application areas for object detection in an image: a review. *Multimedia Tools and Applications*, *81*(27), 38297–38351. https://doi.org/10.1007/s11042-022-13153-y

Ketkar, N., & Moolayil, J. (2021a). Convolutional Neural Networks. In *Deep Learning with Python* (pp. 197–242). Apress. https://doi.org/10.1007/978-1-4842-5364-9_6

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), 1122-1131. https://doi.org/10.1016/j.cell.2018.02.010

Khan, A., Laghari, A., & Awan, S. (2021). Machine Learning in Computer Vision: A Review. *ICST Transactions on Scalable Information Systems*, *8*(32), 1–11. https://doi.org/10.4108/eai.21-4-2021.169418

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, *53*(8), 5455–5516. https://doi.org/10.1007/s10462-020-09825-6

Kidger, P., & Lyons, T. (2020). Universal approximation with deep narrow networks. *Conference on Learning Theory*, 2306–2327.

Kiyasseh, D., Zhu, T., & Clifton, D. (2022). The Promise of Clinical Decision Support Systems Targetting Low-Resource Settings. *IEEE Reviews in Biomedical Engineering*, *15*, 354–371. https://doi.org/10.1109/RBME.2020.3017868

Köpüklü, O., Babaee, M., Hörmann, S., & Rigoll, G. (2019). Convolutional neural networks with layer reuse. *2019 IEEE International Conference on Image Processing (ICIP)*, 345–349.

Korkmaz, O. E. (2025). Revisiting convolutional design for efficient CNN architectures in edge-aware applications. *Scientific Reports*, *15*(1), 44351. https://doi.org/10.1038/s41598-025-27856-3

Krishna, M., M., N. M., Harshali;, & Venu Gopala Rao, M. (2018). Image classification using Deep learning. *International Journal of Engineering & Technology*, *7*(2.7), 614. https://doi.org/10.14419/ijet.v7i2.7.10892

Krishna, P. V., Reddy, K. S., Krishna, J. S., Amulya, B., Vishnuvardhan, G., & Mudasar, S. (2026). Pulse Lung Insight Engine: A Data-Driven AI Framework for Early Cardiopulmonary Condition Detection. *Pulse*, *5*(2).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kulkarni, U., S.M., M., Gurlahosur, S. V., & Bhogar, G. (2021). Quantization Friendly MobileNet (QF-MobileNet) Architecture for Vision Based Applications on Embedded Platforms. *Neural Networks*, *136*, 28–39. https://doi.org/10.1016/j.neunet.2020.12.022

Kurdi, S. Z., Ali, M. H., Jaber, M. M., Saba, T., Rehman, A., & Damaševičius, R. (2023). Brain Tumor Classification Using Meta-Heuristic Optimized Convolutional Neural Networks. *Journal of Personalized Medicine*, *13*(2), 181–198. https://doi.org/10.3390/jpm13020181

Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., Zafar, B., Dar, S. H., Sajid, M., & Khalil, T. (2019). Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review. *Mathematical Problems in Engineering*, *2019*(1), 1–21. https://doi.org/10.1155/2019/9658350

Lederer, J. (2021). Activation functions in artificial neural networks: A systematic overview. *ArXiv Preprint ArXiv:2101.09957*, *34*(12), 56–81.

Li, J., Bhatti, U. A., Nawaz, S. A., Huang, M., Ahmad, R. M., & Ghadi, Y. Y. (2024). Remote-sensing image classification: A comprehensive review and applications. *Deep Learning for Multimedia Processing Applications*, 18–47.

Li, X., Hou, B., Zhang, R., & Liu, Y. (2023). A review of RGB image-based internet of things in smart agriculture. *IEEE Sensors Journal*, *23*(20), 24107–24122.

Liang, G., & Zheng, L. (2021). A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Applied Sciences*, 11(23), 11185. https://doi.org/10.3390/app112311185

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Liu, W., Li, C., Xu, N., Jiang, T., Rahaman, M. M., Sun, H., Wu, X., Hu, W., Chen, H., Sun, C., Yao, Y., & Grzegorzek, M. (2022). CVM-Cervix: A hybrid cervical Pap-smear image classification framework using CNN, visual transformer and multilayer perceptron. *Pattern Recognition*, *130*, 108829. https://doi.org/10.1016/j.patcog.2022.108829

Long, M. E., Mallampalli, R. K., & Horowitz, J. C. (2022). Pathogenesis of pneumonia and acute lung injury. *Clinical Science*, *136*(10), 747-769.

Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift Für Medizinische Physik*, *29*(2), 102–127. https://doi.org/10.1016/j.zemedi.2018.11.002

Selvi, M., & Vaithilingan, S. (2024). Childhood Pneumonia in Low- and Middle-Income Countries: A Systematic Review of Prevalence, Risk Factors, and Healthcare-Seeking Behaviors. *Cureus*, *16*(4), e57636. https://doi.org/10.7759/cureus.57636

Ma, Y., Zeng, Y., & Sun, S. (2022). A deep learning based super resolution DoA estimator with single snapshot MIMO radar data. *IEEE Transactions on Vehicular Technology*, *71*(4), 4142–4155.

Maier, A., Syben, C., Lasser, T., & Riess, C. (2019). A gentle introduction to deep learning in medical image processing. *Zeitschrift Für Medizinische Physik*, *29*(2), 86–101. https://doi.org/10.1016/j.zemedi.2018.12.003

Majeed, A., & Lee, S. (2021). Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey. *IEEE Access*, *9*, 8512–8545. https://doi.org/10.1109/ACCESS.2020.3045700

Mandel, L. A., & Wunderink, R. G. (2020). Pneumonia. In *Harrison's Principles of Internal Medicine* (21st ed.). McGraw-Hill.

Mardianto, M. F. F., Yoani, A., Soewignjo, S., Putra, I. K. P. K. A., & Dewi, D. A. (2024). Classification of Pneumonia from Chest X-ray images using Support Vector Machine and Convolutional Neural Network. *International Journal of Advanced Computer Science and Applications*, *15*(6), 234–258. https://doi.org/10.14569/IJACSA.2024.01506104

McAllister, D. A., Liu, L., Shi, T., Chu, Y., Reed, C., Burrows, J., Adeloye, D., Rudan, I., Black, R. E., Campbell, H., & Nair, H. (2019). Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: a systematic analysis. *The Lancet Global Health*, *7*(1), e47–e57. https://doi.org/10.1016/S2214-109X(18)30408-X

McKnight, J., Oliwa, J., Willows, T. M., Onyango, O., Mazhar, R., Jones, L. B., Mkumbo, E., Maiba, J., Schell, C. O., Baker, T., Amoth, P., Were, I., Warfa, O., & English, M. (2026). The normalisation of under-resourcing limits the potential for improvement of critical care: evidence from a multi-method study of Kenyan hospitals. *Social Science & Medicine*, *396*, 118843. https://doi.org/10.1016/j.socscimed.2025.118843

Meir, Y., Tevet, O., Tzach, Y., Hodassman, S., Gross, R. D., & Kanter, I. (2023). Efficient shallow learning as an alternative to deep learning. *Scientific Reports*, *13*(1), 5423.

Mhaskar, H., Liao, Q., & Poggio, T. (2017). When and why can deep networks avoid the curse of dimensionality? *Proceedings of the National Academy of Sciences*, 114(48), 12773-12778. https://doi.org/10.1073/pnas.1704476114

Miyashita, N. (2022). Atypical pneumonia: Pathophysiology, diagnosis, and treatment. *Respiratory Investigation*, *60*(1), 56–67. https://doi.org/10.1016/j.resinv.2021.09.009

Mohsin, Z. K., Alshammari, R. F. N., & Alsaadi, E. M. T. A. (2025). Design of lightweight neural networks for resource-constrained devices. *Sustainable Engineering and Innovation*, *7*(2), 605–618. https://doi.org/10.37868/sei.v7i2.id542

Mostafa Ibrahim. (2021, December). *Precision vs. Recall: Understanding How to Classify with Clarity*. Weights & Biases.

Muhammad, W., & Aramvith, S. (2019). Multi-Scale Inception Based Super-Resolution Using Deep Learning Approach. *Electronics*, *8*(8), 892–914. https://doi.org/10.3390/electronics8080892

Murphy, K. P. (2022). *Probabilistic machine learning: an introduction*. MIT press.

Nakrani, H., Shahra, E. Q., Basurra, S., Mohammad, R., Vakaj, E., & Jabbar, W. A. (2025). Advanced diagnosis of cardiac and respiratory diseases from chest x-ray imagery using deep learning ensembles. *Journal of Sensor and Actuator Networks*, *14*(2), 44.

Naskinova, I. (2021). On convolutional neural networks for chest X-ray classification. *ResearchGate*. https://doi.org/10.13140/RG.2.2.12345.67890

National Heart Lung and Blood Institute. (2022). *Pneumonia*. https://www.nhlbi.nih.gov/health /pneumonia#

Ndung'u, H. N. (2025). *Entrepreneurial Supply Chain Practices, Healthcare Financing and Performance of Public Hospitals in Kenya*.

Neuman, M. I., Lee, E. Y., Bixby, S., Diperna, S., Hellinger, J., Markowitz, R., ... & Bachur, R. G. (2019). Variability in the interpretation of chest radiographs for the diagnosis of

pneumonia in children. *Journal of Hospital Medicine*, 14(5), 287-293. https://doi.org/10.12788/jhm.3154

Nimmala, S., & Rammohanreddy, P. (2023). Image normalization techniques for deep learning. *International Journal of Advanced Computer Science and Applications*, 14(2), 234-241.

Noman, N. (2020). *A Shallow Introduction to Deep Neural Networks* (pp. 35–63). https://doi.org/10.1007/978-981-15-3685-4_2

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2021). Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*. https://doi.org/10.48550/arXiv.1811.03378

Okundalaye, O. O. (2025). Improving diagnosis of pneumonia among population: A machine learning method using k-nearest neighbors (KNN), random forest, and support vector machine (SVM) on chest X-ray images. *SRA - Physical Sciences*, *1*(2), 1–16. https://doi.org/10.63112/1e038s75

Pereira, D., & Garey, S. L. (2021). Cancer, Cervical. In *Encyclopedia of Behavioral Medicine* (pp. 350–351). Springer International Publishing. https://doi.org/ 10.1007/978- 3-030-39903-0_157

Physiopedia. (2021). *Pneumonia*. https://www.physio-pedia.com/Pneumonia

Qin, C., Yao, D., Shi, Y., & Song, Z. (2019). Computer-aided detection in chest radiography based on artificial intelligence: A survey. *Biomedical Engineering Online*, 17(1), 1-23. https://doi.org/10.1186/s12938-018-0585-1

Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., ... & Kashem, S. (2021). Transfer learning with deep convolutional neural network for COVID-19 detection from chest X-ray images. *Computer Methods and Programs in Biomedicine*, 196, 105580. https://doi.org/10.1016/j.cmpb.2020.105580

Rahul, K., & Banyal, R. K. (2021). Detection and correction of abnormal data with optimized dirty data: a new data cleaning model. *International Journal of Information Technology & Decision Making*, *20*(02), 809–841.

Reza, Md. H., Sibly, Md. N. B., Rabbani, S. G., Sadi, S. H., Ahamed, Md. F., Shafi, F. B., Sarmun, R., & Chowdhury, M. E. H. (2026). A comprehensive review of convolutional neural networks: foundations, enhancements and applications. *Neural Computing and Applications*, *38*(4), 56. https://doi.org/10.1007/s00521-025-11827-w

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144. https://doi.org/10.1145/2939672.2939778

Saha, S. (2018). *A Comprehensive Guide to Convolutional Neural Networks*. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Salehi, M., Mohammadi, R., & Ghaffari, A. (2021). Deep learning-based pneumonia detection using chest X-ray images: A comparative study. *PMC*, 8506182. https://doi.org/10.1155/2021/8506182

Saraiva, A. A., Santos, D. B., Costa, N. J., Sousa, J. V., Ferreira, N. M., & Valente, A. (2019). Classification of pneumonia from chest X-ray images using deep learning. *Proceedings of the 14th International Conference on Computer Graphics Theory and Applications*, 276-283. https://doi.org/10.5220/0007346602760283

Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6), 420. https://doi.org/10.1007/s42979-021-00815-1

Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: a survey. *Evolutionary Intelligence*, *15*(1), 1–22. https://doi.org/10.1007/s12065-020-00540-3

Shah, S. R., Qadri, S., Bibi, H., Shah, S. M. W., Sharif, M. I., & Marinello, F. (2023). Comparing Inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A Case Study on Early Detection of a Rice Disease. *Agronomy*, *13*(6), 1633. https://doi.org/10.3390/agronomy13061633

Shahriar, T. (2025). Comparative analysis of lightweight deep learning models for memory-constrained devices. *ArXiv Preprint ArXiv:2505.03303*.

Shambour, M. K. (2022). Analyzing perceptions of a global event using CNN-LSTM deep learning approach: the case of Hajj 1442 (2021). *PeerJ Computer Science*, *8*, e1087. https://doi.org/10.7717/peerj-cs.1087

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308–6325.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1-48. https://doi.org/10.1186/s40537-019-0197-0

Sindhu Meena, K., & Suriya, S. (2020). A Survey on Supervised and Unsupervised Learning Techniques. In *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications* (pp. 627–644). Springer International Publishing. https://doi.org/10.1007/978-3-030-24051-6_58

Singh, S., & Krishnan, S. (2020). Filter response normalization layer: Eliminating batch dependence in the training of deep neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11237–11246.

Sousa, R. T., Pereira, L. A., & Oliveira, A. L. (2022). Lightweight CNN architectures for medical image classification in mobile devices. *IEEE Access*, 10, 45678-45690. https://doi.org/10.1109/ACCESS.2022.3171234

Sutera, L. P. M. (2025). On the surprising accuracy, efficiency, and interpretability of shallow neural networks in medical image classification. *AMS Tesi di Laurea, Università di Bologna*.

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. NPJ Digital Medicine, 3(17), 1–10.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1-9. https://doi.org/10.1109/CVPR.2015.7298594

Taib, C., Abdoun, O., & Haimoudi, E. K. (2026). A new architecture based on the Xception algorithm for pneumonia detection using medical image datasets. *Computer Methods and Programs in Biomedicine Update*, *9*, 100234. https://doi.org/10.1016/j.cmpbup.2026.100234

Tan, L. (2006). Image file formats. *Biomedical Imaging and Intervention Journal*, *2*(1), e6. https://doi.org/10.2349/biij.2.1.e6

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*, 6105-6114. https://doi.org/10.48550/arXiv.1905.11946

Thompson, A. E. (2016). Pneumonia. *JAMA*, *315*(6), 626. https://doi.org/10.1001/jama.2016.0320

Tian, S., Hu, W., Niu, L., Liu, H., Xu, H., & Xiao, S.-Y. (2020). Pulmonary Pathology of Early-Phase 2019 Novel Coronavirus (COVID-19) Pneumonia in Two Patients with Lung Cancer. *Journal of Thoracic Oncology*, *15*(5), 700–704. https://doi.org/10.1016/j.jtho.2020.02.010

Tom Keldenich. (2021, September 2). *Recall, Precision, F1 Score – Simple Metric Explanation Machine Learning*. THE BASICS, DEEP LEARNING.

Tonui, N., Ondimu, O., & Karanja, A. M. (2025). Determinants, Prevalence and Spatial Variation of Pneumonia Among Children Under Five Years in Ainamoi Sub-County, Kericho County, Kenya. *Journal of the Kenya National Commission for UNESCO*, *2*, 1–14. https://doi.org/10.62049/jkncu.v5i2.322

Trigka, M., & Dritsas, E. (2025). A Comprehensive Survey of Deep Learning Approaches in Image Processing. *Sensors*, *25*(2), 531. https://doi.org/10.3390/s25020531

Troeger, C. E., Khalil, I. A., Blacker, B. F., Biehl, M. H., Albertson, S. B., Zimsen, S. R. M., Rao, P. C., Abate, D., Ahmadi, A., Ahmed, M. L. C. brahim, Akal, C. G., Alahdab, F., Alam, N., Alene, K. A., Alipour, V., Aljunid, S. M., Al-Raddadi, R. M., Alvis-Guzman, N., Amini, S., … Reiner, R. C. (2020). Quantifying risks and interventions that have affected the burden of diarrhoea among children younger than 5 years: an analysis of the Global Burden of Disease Study 2017. *The Lancet Infectious Diseases*, *20*(1), 37–59. https://doi.org/10.1016/S1473-3099(19)30401-3

Van der Jeught, S., & Dirckx, J. J. J. (2019). Deep neural networks for single shot structured light profilometry. *Optics Express*, *27*(12), 17091. https://doi.org/10.1364/OE.27.017091

Varela-Santos, S., & Melin, P. (2021). A new modular neural network approach with fuzzy response integration for lung disease classification based on multiple objective feature optimization in chest X-ray images. *Expert Systems with Applications*, *168*(23), 114361–114393. https://doi.org/10.1016/j.eswa.2020.114361

Veloyan, N. I. S. H. M. (2025). A Residual Learning Framework for Advanced Visual Recognition. *European Journal of Emerging Artificial Intelligence*, *2*(1), 11–20.

Venkatesh, V., & Davis, F. D. (2019). A theoretical extension of the Technology Acceptance

Verma, K. K., Singh, B. M., & Dixit, A. (2022). A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. *International Journal of Information Technology*, *14*(1), 397–410. https://doi.org/10.1007/s41870-019-00364-0

Wang, H., Ruan, Q., Wang, H., Rezaei, S. D., Lim, K. T. P., Liu, H., Zhang, W., Trisno, J., Chan, J. Y. E., & Yang, J. K. W. (2021). Full Color and Grayscale Painting with 3D Printed Low-Index Nanopillars. *Nano Letters*, *21*(11), 4721–4729. https://doi.org/10.1021/acs.nanolett.1c00979

Wang, Y. (2024). Research on Image Classification and Recognition Technology Based on Machine Learning. *Applied Mathematics and Nonlinear Sciences*, *9*(1), 62–73. https://doi.org/10.2478/amns-2024-1514

Wang, Z. J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., Kahng, M., & Polo Chau, D. H. (2021). CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), 1396–1406. https://doi.org/10.1109/TVCG.2020.3030418

Williams, G. J., Macaskill, P., Kerr, M., Fitzgerald, D. A., & Isaacs, D. (2020). Variability and accuracy in interpretation of consolidation on chest radiography for diagnosing pneumonia in children. *Journal of Paediatrics and Child Health*, 56(3), 432-438. https://doi.org/10.1111/jpc.14658

Wootton, D., & Feldman, C. (2014). The diagnosis of pneumonia requires urgent care and specific treatment. *Clinical Chest Medicine*, 35(3), 429-440.

World Health Organisation. (2022). *Child mortality (under 5 years)*. https://www.who.int/news-room/fact-sheets/detail/child-mortality-under-5-years

World Health Organization. (2019). *Regional Action Agenda on Harnessing E-Health for Improved Health Service Delivery in the Western Pacific*.

Wu, H., Liu, Q., & Liu, X. (2019). A Review on Deep Learning Approaches to Image Classification and Object Segmentation. *Computers, Materials & Continua*, *60*(2), 575–597. https://doi.org/10.32604/cmc.2019.03595

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9

Yogapriya, J., Saravanabhavan, C., Asokan, R., Vennila, Ila., Preethi, P., & Nithya, B. (2018). A Study of Image Retrieval System Based on Feature Extraction, Selection, Classification and Similarity Measurements. *Journal of Medical Imaging and Health Informatics*, *8*(3), 479–484. https://doi.org/10.1166/jmihi.2018.2326

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Medicine*, 15(11), e1002683. https://doi.org/10.1371/journal.pmed.1002683

Zhao, L., & Zhang, Z. (2024). A improved pooling method for convolutional neural networks. *Scientific Reports*, *14*(1), 1589. https://doi.org/10.1038/s41598-024-51258-6

Zhu, H., Wang, J., Wang, S.-H., Raman, R., Górriz, J. M., & Zhang, Y.-D. (2023). An Evolutionary Attention-Based Network for Medical Image Classification. *International Journal of Neural Systems*, *33*(03), 24–56. https://doi.org/10.1142/S0129065723500107

APPENDICES

Appendix I: Ethical Clearances

COUNTY GOVERNMENT OF KAJIADO
DEPARTMENT OF HEALTH SERVICES

Telephone: 0711441259/0711441409
Email: kitengelahealthcentre@yahoo.com

Kitengela Sub County Hospital
P.O. BOX 537- 00242, Kitengela

REF: KSCH/2022

DATE:4TH NOVEMBER, 2022

THE COUNTY DIRECTOR OF HEALTH
KAJIADO COUNTY

RE: RESEARCH AUTHORIZATION ON "IMAGE CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS CASE OF PNEUMONIA DETECTION"

Reference if made to your letter CGK/HEALTH SERVICES/01/VOL. 11/74 dated 4th November, 2022 on the above subject.

The facility has no objection, however the student will be required to observe privacy and confidentiality of the patients and share the findings of the survey with this facility.

MEDICAL SUPERINTENDENT
KITENGELA SUB-COUNTY HOSPITAL
04 NOV 2022
P.O. Box 537-00242,
KITENGELA.

Thank you.

DR. VERONICA ABUTO
MEDICAL SUPERINTENDENT
KITENGELA SUB COUNTY HOSPITAL

CC
- **CHIEF OFFICER MEDICAL SERVICES**
- **JOSEPHINE MWEYELI CHACHA**

NAIROBI CITY COUNTY

Tel: 2724712, 2725791, 0721 311 808
mbagathihosp@gmail.com

Mbagathi Hospital
P.O. Box 20725- 00202 Nairobi

COUNTY HEALTH SERVICES

REF: MDH/RS/VOL.3/11/212

24th January, 2023

JOSEPHINE MWENYELI CHACHA
MASENO UNIVERSITY
KISUMU.

Dear, Chacha

RE: RESEARCH AUTHORIZATION.

This is in reference to your application for authority to carry out a research, on ***'Image classification using convolution neural networks case of pneumonia detection'***

I am pleased to inform you that your request to undertake research in the hospital has been granted.

On completion of the research, you are expected to submit one hard copy and one soft copy of the research report/ thesis to this office.

MBAGATHI HOSPITAL
25 JAN 2023

DR. CAREN EMADAU
MEDICAL SUPERINTENDENT
MBAGATHI HOSPITAL.

REPUBLIC OF KENYA
COUNTY GOVERNMENT OF MACHAKOS
DEPARTMENT OF HEALTH
Office of the County Director of Medical Services

Telephone: +254-44-20246
Fax: 254-44-20655

Highway
P.O. Box 2574-90100
Machakos, Kenya

Ref No. MKS/DHES/RSCH/VOL1/269 **5/12/2022**

Dear Josephine Chacha,

RE: LETTER OF AUTHORIZATION FOR CONDUCTING PROPOSED RESEARCH

The Department of Health and Emergency Services, Machakos County is keen to collaborate in your study titled, **"Image Classification using Convolutional Neural Networks: Case of Pneumonia Detection."**

Note is taken of the letter of Ethical clearance from JKUAT ISERC, **REF: JKU/ISERC/02316/0743**, for the approval period **6th October 2022 to 5th October 2023** as well as the Research Licensc from the National Commission for Science, Technology & Innovation number **NACOSTI/P/22/21411** for the period ending **2nd November 2023**.

You are hereby authorized to proceed with the research in Machakos County and urged to share the findings with the Department of Health and Emergency Services, Machakos County, through this Email: research.dhes@gmail.com

Sincerely,

05 DEC

Dr. Sharon Mweu
County Director Medical Services & Research
Machakos County

Cc
County Executive Committee Member – Health
Chief Officer – Medical Services & Public health & Community Outreach
County Research Committee

NAIROBI CITY COUNTY

Telephone: Nairobi
020 - 2297000
020 - 8022076
E-mail : medsupmedhq@yahoo.com

When replying please quote

MAMA LUCY KIBAKI HOSPITAL-EMBAKASI
P.O. Box 1278-00515
BURUBURU-NAIROBI

COUNTY HEALTH SERVICES

Ref. No. MLKH/ADM/RES/1 **Date: 20th December 2022**

Josephine Mweyeli Chacha

JKUAT Nairobi CBD,

NAIROBI

RE: PERMISSION TO COLLECT DATA

TITLE: "IMAGE CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS: CASE PNEUMONIA DETECTION."

Refer to your application to collect data on the above research in this institution.

This is to inform you that the Hospital has given you permission to collect data subject to the following;

1. You are expected to adhere to the rules and regulations pertaining to the data collection.
2. You are expected to submit a copy of the final findings to the research committee.

20 DEC 2022

DR. EMMA MUTIO

MEDICAL SUPERINTENDENT

Appendix II: Model Summary

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 149, 149, 64)	320
max_pooling2d (MaxPooling2D)	(None, 74, 74, 64)	0
conv2d_1 (Conv2D)	(None, 73, 73, 64)	16,448
max_pooling2d_1 (MaxPooling2D)	(None, 36, 36, 64)	0
conv2d_2 (Conv2D)	(None, 35, 35, 128)	32,896
max_pooling2d_2 (MaxPooling2D)	(None, 17, 17, 128)	0
flatten (Flatten)	(None, 36992)	0
dense (Dense)	(None, 32)	1,183,776
dropout (Dropout)	(None, 32)	0
dense_1 (Dense)	(None, 2)	66

Total params: 1,233,506 (4.71 MB)
Trainable params: 1,233,506 (4.71 MB)
Non-trainable params: 0 (0.00 B)

Appendix III: Publication

2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA)

Adoption of Shallow Neural Networks in Pneumonia Classification

Mwaveli Josephine Chacha SCIT
JKUAT
Nairobi, Kenya jmchacha14@gmail.com

Dr. Tobias Mwalili SCIT
JKUAT
Nairobi, Kenya
tmwalili@jkuat.ac.ke

Dr. Henry Mwangi SCIT
JKUAT
Nairobi, Kenya
henry.mwangi@jkuat.ac.ke

https://orcid.org/0000-0002-0009-0785

International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA) | 979-8-3503-6935-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICMERALDA

***Abstract*— A CNN contains a several layers each receiving input from the preceding layer. The final layer of the CNN flattens the image into a column and then determines which features most correlate to a particular class. In Kenya, pneumonia accounts for 16% of the total number of deaths in children under the age of five and ranked as the second leading cause of death with this age group. Currently radiologist examine Chest X-rays images under luminous light to check for pneumonia. This study proposes the use of a five-layer shallow neural network for image classification. The model records an accuracy of 91% and AUC scores of 90% with the type I error and type II errors are 14% and 5% respectively on secondary data. On local tests data, the model recorded an ROC of 90% and Type 1 error rate and type II error rate was 11.7 % and 7.4% respectively**

Keywords*— *shallow neural network, image classification, pneumonia

I. INTRODUCTION

Computer vision techniques have become important in applications that require image detection and classification [1]. Image detection entails the use of computing technology to determine the presence of objects in an image whereas image classification refers to the labeling of images in any one of the predefined semantic categories [2]. CNN is best suited in recognizing visual patterns from pixel images with variability [3]. CNN can be applied in a broad-spectrum area with the common objective of automatically learning features from datasets. Once the features have been learned, they can be used for tasks such as classification [4].

Pneumonia is a group of syndromes brought about by a variety of organisms resulting in varied manifestations and sequelae that causes the infection of the lung parenchyma [5]. Some cases of pneumonia are life threatening [6]. Currently, a radiologist diagnoses chest X-ray images by analyzing the images under bright light to ascertain whether the patient is exhibiting pneumonia symptoms [7]. According to [8], radiologists may often find it difficult to interpret results, and at times the same may be frequently inconsistent.

Sub-Saharan Africa has the highest under-5 mortality rate in the world, with a child in 13 dying before their fifth birthday. This number is 14 times higher than that of highincome countries [9]. An estimated 490,000 pneumonia related deaths occurred in the year 2015. This accounted for half of the total number of pneumonia related deaths globally [10]. In Kenya, pneumonia ranks as the second leading cause of death in children under the age of five; which stands at 16% of the total number of deaths in the age group. With this number of high death rates, Kenya is among the top fifteen countries with the highest number of under-5 pneumonia related deaths [11]. The health care system in Kenya has been plagued with a number of issues. According to [12] issues that face the medical industry are service overload, poor or inadequate access to services due to high cost and geographical barrier. Poor work environment due to deficient equipment, lack of drugs, inadequate supplies and deficiency in the number of skilled medical personnel inhibit health workers from functioning in a most efficient manner.

With all these underlying issues, a majority of Kenyans end up being alienated from accessing basic health services [13]. This study proposes the use of a shallow neural network that is not computationally intensive for the purpose of pneumonia classification.

II. LITERATURE REVIEW

A. Architecture of Convolutional Neural Networks

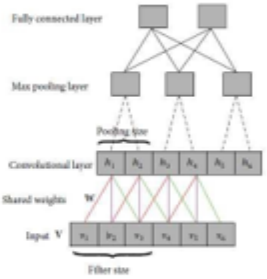


Fig. 1. Convolutional Neural Network Layers

Figure 1 above was adopted from [14]. The image is fed into the CNN from first layer i.e. input level in the convolutional layer. This layer takes in the image size $i \times j \times r$ where i is the height of the image; j is the width of the image and r is the dimension of the image [15]. Thereafter, the image is passed on to the convolution layer where the kernel performs the convolution operations and extracts features of the images [16]. Shared weights are the method of filtering different parts of the image with the same kernel. It's these weights that enable the neutral cells with the same feature to be recognized and classified as the same object types.

At, the pooling layer, the dimension of the image is reduced so as to decrease the computational power required to process the data. It's at this layer that one can extract dominant features of an image. The fully