Arab American University

Faculty of Graduate Studies

Deep Learning for Chest X-ray Analysis: Pneumonia Detection

By

Inshirah Hussein Yousef Odeh

Supervisor

Dr. Khalid S. Rabayah

This thesis was submitted in partial fulfillment of the requirements for the Master's degree in Computer Science.

3 / 2025

Thesis Approval

Deep Learning for Chest X-ray Analysis: Pneumonia Detection

By

Inshirah Hussein Yousef Odeh

This thesis was defended successfully on 08 /03/2025 and approved by:

Committee members	Signature
1. Dr. Khalid Rabayah: Supervisor	
2. Dr. Amjad Rattrout: Internal Examiner	
3. Prof. Yousef Daraghma: External Examiner	

Declaration

I acknowledge that this thesis entitled " Deep Learning for Chest X-ray Analysis: Pneumonia Detection" results from my research and work, and I write it alone. It is devoid of any other research work except for the reference and has not been submitted to any other scientific or scientific degree.

The Name of The Student: Inshirah Hussein Yousef Odeh

ID: 201720328

Signature: 

Date: 25.8.2025

Acknowledgments

I want to thank my supervisor (Dr. Khalid S. Rabayah), who made this work possible. His guidance and advice carried me through all the stages of writing my thesis. A debt of gratitude is also owed to my university, AAUP, and my respected teachers for their invaluable help in preparing this research.

I would also like to thank my parents, husband, children, brothers, and family for their continuous support and understanding while researching and writing my thesis. Your prayer for me was what sustained me this far.

I humbly thank all concerned persons who cooperated with me in this regard.

Finally, I would like to thank God for letting me through all the difficulties. I have experienced your guidance day by day. You are the one who let me finish my degree, and I will keep on trusting you for my future.

Abstract

Pneumonia, which ranks among the primary causes of death on a global scale, poses a substantial threat to the health of children everywhere, especially those below the age of five. An accurate and prompt diagnosis is essential for effective therapy and patient outcomes. This study aims to analyze chest X-rays in the best method, which will provide a definitive method for pneumonia assessment.

The investigation employed a comprehensive, global dataset acquired from the Kaggle platform. The dataset contains 5856 chest X-ray images. Data preprocessing is used in training deep learning models to improve the ability to predict pneumonia in X-ray images. After that, several deep-learning models, namely ResNet, ResNet50, VGG16, and GoogLeNet, enhanced techniques based on the CNN model, have been used to detect pneumonia in X-rays. These models have been subjected to rigorous testing to guarantee their accuracy and performance.

The results of the DenseNet techniques are highly effective, and the accuracy is very good. On the other hand, GoogLeNet and other techniques could not achieve the desired degree of accuracy. Considering the notable variations in performance, the DenseNet algorithm appears well-suited for this dataset type.

Further research could investigate various algorithmic modifications to improve performance further. Additional examination of its composition and unique characteristics is required to ascertain its efficacy in this domain.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANNs	Artificial Neural Networks
ES	Ensemble
FN	False-Negative
FP	False-Positive
LM	Levenberg-Marquardt
LMS	Learning Management System
ML	Machine Learning
MLP	Multi-Layer Perceptron
MLT	Machine Learning Techniques
MSE	Mean Squared Error
NNs	Neural Networks
ROC Area	Receiver Operator Characteristic Curves
TN	True Negative
TP	True Positive

Chapter 1

Introduction

1.1 Introduction

Respiratory disorders are a prominent contributor to mortality worldwide, as stated by the World Health Organization, according to (Pham et al., 2020). These include conditions such as lower respiratory tract infections (LRTI), asthma, pneumonia, chronic obstructive pulmonary disease (COPD), lung cancer, and tuberculosis, As noted by (Khan et al., 2021). In the study conducted by (Ervural & Ceylan , 2021), Pneumonia is a medical illness marked by the inflammation of the alveoli in either one or both of the lungs. Symptoms of fluid or pus accumulation in the air sacs include a productive cough with phlegm or purulent material, increased body temperature, chills, and difficulty breathing, as described by (Bhatt et al., 2023) . Pneumonia can be attributed to many species, including bacteria, viruses, and fungi. (Pal & Das 2011) highlight that pneumonia can exhibit a spectrum of severity, ranging from a benign ailment to a life-threatening condition. As (Gabruseva et al., 2020) mentions, the population cohorts most vulnerable to the gravity of this ailment include infants and young children, those aged 65 and older, as well as individuals with pre-existing medical illnesses or weakened immune systems, as found in (Rajpurkar et al., 2017) . Timely detection of this illness can significantly enhance the individual's likelihood of rapid recuperation and survival, as the sickness can be fatal which is revealed by (Barhoom & Abu-Naser, 2022) . Pneumonia outbreaks can have significant implications for public health, as demonstrated by infections like COVID-19. (Perna Tagarelli & , 2019) points out that deploying Artificial Intelligence (AI) driven early detection systems can expedite the recognition of potential epidemics and enable swift measures to curb their spread .

The work of (Ignat, A., & Găină, R. A. (2023) illustrates that artificial Intelligence (AI) is a broad discipline in computer science that focuses on the ability of machines to exhibit rational behavior in response to external inputs. The goal of AI is to create computers that can perform tasks that

would normally need human intelligence. AI has numerous benefits in practical applications, such as minimizing human errors across various domains, including medicine and education, executing tasks flawlessly in any profession, and handling monotonous and repetitive work which aligns with existing research (Surden, 2019). AI is apparent in multiple facets of everyday life, such as the employment of virtual assistants, the adoption of search prediction technologies, and the provision of ride-hailing services. Deep learning, machine learning, and other AI approaches are widely utilized across several domains to assist humans in challenging jobs like predicting natural phenomena and diagnosing intricate diseases which has been demonstrated in previous research (Mustapha et al., 2020).

As (Dong et al., 2021) mentioned, deep Learning (DL) technology is highly esteemed in machine learning, artificial intelligence, data science, and analytics. This is mostly related to its capacity to learn information from the given data. Several prominent firms, such as Google, Microsoft, and Nokia, actively investigate this topic, as it promises to produce significant results in different classification and regression tasks involving distinct datasets. As described by (Tain et al., 2020), Within professional applications, DL is often regarded as a subset of machine learning (ML) and artificial intelligence (AI) . DL can be defined as an artificial intelligence technique replicating the human brain's cognitive data-processing skills. The worldwide significance of "DL" is progressively increasing, as demonstrated by examining past data acquired from Google trends. Deep learning technology employs a hierarchical architecture with multiple layers to efficiently capture and depict the intricate abstractions available in data, thus enabling the creation of robust computer models. Research by (Dong et al., 2021) indicates that Training deep learning models is time-consuming because of their thorough parameterization. Unlike traditional machine learning methods, these models demonstrate comparatively lower execution times during testing.

Convolutional neural networks (CNNs), including ResNet, ResNet50, VGG16, and GoogLeNet, are extensively used in healthcare systems because they can extract features and classify data.

Deep learning approaches have been a popular option for distinguishing between categories, such as positive and negative, infected and healthy, carcinogenic and non-cancerous reflecting the results of prior research (Ibrahim et al., 2021).

Pneumonia detection utilizing deep learning techniques involves using sophisticated medical technology to identify and diagnose a respiratory ailment known as pneumonia. Healthcare practitioners employ many techniques, including X-ray imaging, computed tomography (CT) scans, and artificial intelligence (AI) algorithms, to meticulously analyze lung images and detect indications of infection as many researchers have found (Gabruseva et al., 2020). The prompt and precise diagnosis of pneumonia by healthcare professionals enables timely implementation of treatment, hence improving the effectiveness of patient recovery. The primary objective of this study field is to maximize the precision and efficiency of diagnostic techniques, resulting in improved patient outcomes and overall excellence in healthcare. According to (Barhoom & Abu-Naser, 2022) , pneumonia detection is a compelling study area for various reasons: 1) Prompt and accurate identification of a problem is crucial for administering suitable treatment interventions and improving the patient's overall prognosis. 2) Pneumonia presents a wide range of symptoms that differ in severity, making it difficult to diagnose. Automated detection systems can augment healthcare professionals' diagnostic capacities by analyzing various data sources, such as medical imaging and clinical information.

1.2 Research Motivation and Problem Description

The study seeks to find which of the well-known deep learning models produces the greatest classification accuracy on a certain dataset. Also, the research challenge concerns the substantial

variation in the proficiency of medical practitioners within a hospital to categorize chest X-ray images associated with pneumonia precisely. The variation arises from disparities in the qualifications, experiences, and levels of expertise among individual doctors. Chest X-rays present a distinctive difficulty in clinical diagnosis compared to other imaging techniques like CT scans or MRIs, owing to their intrinsically intricate and subtle characteristics. Hence, this discrepancy in diagnostic expertise gives rise to apprehensions regarding the precision and dependability of pneumonia diagnoses made using chest X-ray pictures in hospital environments.

1.3 Research Questions

The main research questions for this research are:

1. Which deep learning algorithms are most effective in detecting pulmonary abnormalities in chest X-ray images?
2. Which data augmentation techniques most significantly enhance the performance of deep learning models in chest X-ray analysis?
3. What preprocessing methods yield the best diagnostic accuracy when applied to chest X-ray images in deep learning workflows?
4. How do deep learning algorithms distinguish between normal and abnormal pulmonary conditions in chest X-ray images?

1.4 Research Objectives

The main objective of this research is to improve the accuracy of determining whether a person has pneumonia or not pneumonia in chest X-ray images using artificial intelligence techniques applied to datasets. In other words, choosing the appropriate mechanism for analyzing chest X-ray images and finding a more accurate way to evaluate pneumonia is the primary goal of this study. Early detection and accurate diagnosis of pneumonia using artificial intelligence technology

contribute to the treatment plan, in addition to avoiding complications that occur to the person due to this infection. Particularly, this thesis addresses the following objectives:

- Applying the Deep Learning (DL) model using the VGG-16 model is based on the Convolutional Neural Networks (CNN) technique to classify and detect Pneumonia through an X-ray images dataset.
- Assessing the performance of deep learning models using metrics like accuracy and lost. In addition, validate their robustness and generalizability across various clinical settings.
- Optimizing the performance of the deep learning model through preprocessing and data augmentation. On the other hand, creating a Python-based Graphical User Interface for efficient image classification.

1.5 Contribution

Our original addition to the field is a comprehensive analysis of all current deep learning methods, with an emphasis on their performance on global dataset. So, the use of datasets dedicated to the task of automating the diagnosis of pneumonia from chest X-ray pictures is a direct result of the ongoing research into this area. Then, to reliably detect pneumonia cases, these datasets are processed using deep learning algorithms. Well known Deep learning algorithm have been on mentioned dataset. implemented All these algorithms have different desirable features about precision, speed, and applicability in different settings in medicine which is why deep learning is a very useful tool in the struggle against pneumonia.

1.6 Outline of the Thesis

The thesis is organized as follows:

Chapter 2 briefly describes the previous work and literature survey about pneumonia detection and reviews the related work on providing Deep Learning Models in pneumonia detection. Chapter 3 describes the research methodology employed in this study and explains the dataset used, the pre-processing phase, data augmentation techniques, and deep learning techniques. Chapter 4 discusses experimental findings for each technique. In contrast, Chapter 5 presents the main investigation's outcomes, numerical results, and suggestions for further research to improve chest X-ray image classification using deep learning algorithms. Chapter 6 presents the thesis's limitations and recommendations and concludes the thesis work with a discussion on possible future extensions to the research work.

Chapter 2

Literature Review

2.1 Introduction

In this section, we will explore crucial inquiries regarding the identification procedure of respiratory problems in chest X-ray pictures. Also, we will investigate cutting-edge deep learning techniques to determine the most efficient ones for this specific purpose. In addition, we will examine the most effective data augmentation strategies and preprocessing techniques that have shown the ability to achieve more accuracy in detection. Furthermore, we will analyze how these deep learning algorithms are formulated to differentiate between regular and irregular chest X-ray images. We will provide insight into the inventive techniques and methodologies utilized in medical image analysis.[16]

2.2 Pneumonia Detection

Many researchers in have proposed a set of models and used different machine learning and deep learning techniques to predict the presence or absence of pneumonia. These techniques have been applied to different data sets to identify cases of pneumonia, whether through X-ray images or symptoms appearing on the sick person. On the other hand, some results were accurate, and others were inaccurate. This review will present the most relevant studies for our research as noted in the literature (Jaiswal et al., 2019).

2.2.1 Machine Learning pneumonia systems

Deep learning, a subdivision of artificial intelligence, has transformed the domain of medical imaging by offering robust tools for automated disease detection and diagnosis. Deep learning has become more commonly employed in the detection of pneumonia, a severe lung infection that can result in substantial illness and death, particularly among children and the elderly. Pneumonia has been mainly diagnosed clinically and by using radiological investigations, with a special

inclination towards chest roentgenograms. However, the assessment of these pictures requires considerable experience, and the results are bound to differ from one radiologist to another. Deep learning models, particularly convolutional neural networks, have shown a lot of promise in accurately diagnosing pneumonia through big data, especially chest X-ray images. Such capability results in the minimization of diagnostic mistakes and the optimization of the clinical execution.

Different techniques in deep learning have been explored to enhance the accuracy and robustness of pneumonia detection. Sequences incorporated into CNN that allow obtaining the spatial hierarchies of features from X-ray images in real-time are provided. DenseNet architecture is a type of deeper and more intricate network architecture than the traditional one, similar to ResNet. With the help of transfer learning techniques such as fine-tuning VGG-16 and Inception models, pneumonia detection performs well when the dataset is small and the training time is short. Also, concerning the development of more advanced solutions, researchers have designed CNNs that took components of attention processes and ensemble techniques. These models increase detection rates and are appropriate for several types of patients. In the case of pneumonia, deep learning is flexible in combating the disease in that each of those methods provides different levels of accuracy, computational throughput, and fungibility for various diseases.

Delineated in the work (Singh et al., 2023) is the novelty in the network plan: Quaternion Channel-Spatial Attention. This architecture has a pooling mechanism in the 2D domain to bear spatial and channel attention. Moreover, it employs a Quaternion residual network to classify chest x-ray images and pneumonia. Thus, higher accuracy may be obtained. The dataset is organized into three primary directories: train, test, and validation. Every directory consists of a subdirectory, [Pneumonia and Normal], which is the name of an image category. The line-up has 5,845 digitalized chest X-rays (CXR), all converted to JPEG files. The images are separated into the two

categories of: good images (P) and evil images (N). We participated in a retrospective study at Guangzhou Women and Children's Medical Centre and used CXR images from cohorts of infants 1-5 years old. Chest X-rays (CXR) were regularly conducted as part of the patient's treatment plan. Two skilled physicians thoroughly evaluated the images before using the photographs to train the AI system. The evaluation attempt is fortified by a subject expert who has the power to facilitate the grading process and fix any grading-based issues that may rise. The composing training dataset contains 5136 pictures. Again, the test dataset was independently set to 700 images. The proposed scheme apportioned a classification precision of 94. By the sensitivity of 53% and AUC, under the distance from the furthest receiver operating characteristic curve (AUC). 89. Aside from this, one also needs to mention that it is known that introducing the attention mechanism in QCNN increases performance.

However, research by (Rāčić et al , 2021) . applied the same techniques to chest X-ray images to make the diagnosis with the deep learning algorithms more accurate. The study strongly emphasizes deploying the deep learning technique, a convolutional neural network, in the model's design. With this model, we plan to classify digital human chest X-rays to consider whether pulmonary alteration due to pneumonia is present. The data handling was implemented by exploiting a chest X-ray image library containing 5,856 JPEG files. The data is partitioned into three directories: train and val set, training and value sets, and test set for testing the dataset. This team has confirmed that the Convolutional Neural Network (CNN) achieved an accuracy of more than 88 percent. Within the algorithm, the machine is 90% prescribed.

(Chandra , Verma, 2020) primarily focused on analyzing the pixels within the segmented Region of Interest (ROI) in the lungs. The main objective is to determine the importance of these pixels in detecting pneumonia, particularly in comparison to the surrounding regions. As a result, the

characteristics of the small region within the divided region of interest (ROI) of the lungs are obtained. The proposed methodology has been assessed using five widely recognized classifiers: Multilayer Perceptron, Random Forest, Sequential Minimal Optimization (SMO), Logistic Regression, and Classification via Regression. The tests employed a dataset including 412 chest X-ray pictures, with an equal distribution of 206 normal cases and 206 cases with pneumonia. The images were obtained from the ChestX-ray14 collection. The performance evaluation of the proposed approach is carried out by comparing it with the previous method using benchmarked classifiers. The experimental results suggest that the strategy proposed in this study demonstrated superior performance compared to the current method. More precisely, the suggested approach attained a significantly higher accuracy rate of 95.63% when using the Logistic Regression classifier and 95.39% when applying the Multilayer Perceptron.

(Darici et al., 2020) have used a multi-class classification framework to differentiate between bacterial pneumonia, viral pneumonia, and healthy patients. This categorization exercise aimed to assign examples to one of three unique classes, depending on whether they exhibited bacterial pneumonia, viral pneumonia, or no pneumonia. To solve the issue of uneven class distribution in the dataset, the SMOTE (Synthetic Minority Over-sampling Technique) method has been used. The design of Convolutional Neural Networks (CNN) models and models in Ensemble Learning has necessitated the creation of novel architectures instead of depending on pre-established networks. Two deep learning algorithms, Convolutional Neural Network (CNN) and ensemble learning, have been utilized for each classification task. The models attained a mean accuracy of 95% for binary classification and 78% and 75% for multi-class classification.

Machine learning techniques were utilized by (Sirazitdinov et al., 2019) to create automated approaches for diagnosing and pinpointing the location of pneumonia in chest x-ray images. The

authors proposed a fusion of two convolutional neural networks, RetinaNet, and Mask R-CNN, for pneumonia detection and localization. The study authors validated their proposed approach by utilizing a dataset of 26,684 photos acquired from the Kaggle Pneumonia Detection Challenge. The efficacy of their method was assessed using a recall score of 0.793. (Varshni et al., 2019) assessed the effectiveness of pre-trained convolutional neural network (CNN) models as feature extractors combined with different classifiers to classify abnormal and normal chest X-Rays. The models included are XCeption, DenseNet-169, VGG-16, VGG-19, and ResNet-50. The image collection utilized comprises 112,120 frontal chest X-ray pictures obtained from a total of 30,085 patients. They methodically identify the optimal convolutional neural network (CNN) model for the specified aim. It has been demonstrated that the DenseNet-169 model achieved the highest performance with an accuracy of 80.0%,

2.2.2 Deep learning pneumonia systems

(Yi et al., 2023) presented a new deep convolutional neural network (DCNN) that can be expanded in size and easily understood. The main goal is to precisely identify instances of pneumonia by examining chest X-ray images. The revised DCNN model first conducts feature extraction on the images and then classifies them into normal and pneumonia categories. The researchers employed a dataset comprising chest X-ray and CT scans, which may be obtained from the UCI Kaggle databases. The dataset comprises 5856 photos, classified into pneumonia and normal images. The dataset consists of 1583 photos labeled as normal and 4273 images labeled as pneumonia. The dataset exclusively comprises images in the JPEG format. The dataset is partitioned into two subsets, specifically the training and testing sets. The training set constitutes 70% of the data and is employed to train the model, while the testing set encompasses the remaining 30% of the data and is used for testing and verifying the model. 30% of the dataset is designated for special reasons.

More precisely, 20% of the data is allocated for testing the model, while the remaining 10% is used for validating the model. The system in question has been trained and tested using a dataset including chest X-ray pictures. The stability and effectiveness of the suggested model have been evaluated using a variety of performance metrics. The testing results showed that the performance of the suggested model exceeds that of current cutting-edge methods used for pneumonia detection, achieving an accuracy of 98.02.

On the other hand, (Bhatt & Shah, 2023) developed a model that is both efficient and can be easily implemented while keeping a high level of precision to aid in diagnosing Pneumonia. A Convolutional Neural Network (CNN) consisting of three different models with varying kernel sizes was built. The outcomes of these models were combined using a novel weighted ensemble technique that proposes an adjustable threshold value to customize the diagnostic capabilities of the model as required. By employing a flexible threshold value, it becomes possible to modify the weighting allocated to the output of each model. This allows for adjusting the classification conclusion based on the individual conditions. The model's performance was evaluated using multiple metrics, including accuracy, recall, precision, and f1-score. The dataset used in this work has been obtained from Guangzhou Women and Children's Medical Center, Guangzhou, and is publicly available on Kaggle. The dataset has 5863 JPEG images, classified into two unique categories: Pneumonia and Normal. The experimental results demonstrated that the combined convolutional neural network (CNN), which incorporated the predictions of three separate CNN models, attained an accuracy rate of 84.12%.

Furthermore, (Yang et al., 2022) Proposed a deep learning methodology that considers multiple factors associated with the image background. In addition, they assessed the efficacy of the proposed approach by utilizing explainable deep learning approaches, namely ResNet50 and

VGG16, to improve interpretability. The dataset containing chest X-ray pictures of pneumonia patients and normal participants in children aged 1 to 5 was acquired from Guangzhou Women's and Children's Medical Centre. The collection comprises 5840 photos, with 4265 X-ray images illustrating individuals diagnosed with pneumonia and 1575 X-ray images exhibiting normal cases. An equal number of X-ray pictures were selected from persons diagnosed with pneumonia and those in good condition to maintain the integrity of the model training and testing results.

The training set comprises 80% randomly chosen samples, whereas the testing set encompasses 20% of the dataset. This division enables the incorporation of chest X-ray images with diverse backgrounds. Before starting the training and testing process, it is usual to adjust the size of the image to a uniform dimension, often 224 × 224. The main goal is to remove the background from the image, improve the precision of pneumonia detection, and utilize the Grad-CAM technique to obtain an interpretable deep-learning model for identifying pneumonia. The suggested approach encompassed multiple sequential procedures. Initially, rudimentary deep learning models are developed to detect pneumonia in X-ray pictures, disregarding the backdrop.

The training set comprises 80% randomly chosen samples, whereas the testing set encompasses 20% of the dataset. This division enables the incorporation of chest X-ray images with diverse backgrounds. Before starting the training and testing process, it is usual to adjust the size of the image to a uniform dimension, often 224 × 224. The main goal is to remove the background from the image, improve the precision of pneumonia detection, and utilize the Grad-CAM technique to obtain an interpretable deep-learning model for identifying pneumonia. The suggested approach encompassed multiple sequential procedures. Initially, rudimentary deep learning models are developed to detect pneumonia in X-ray pictures, disregarding the backdrop.

Furthermore, deep learning models are constructed to consider the context to improve the precision of pneumonia detection. Finally, the Grad-CAM technique is employed to assess the interpretability of the models. The proposed methodology improves the accuracy of pneumonia identification, with the VGG16 model having a maximum accuracy of 95.6%. The proposed methodology can be applied in pneumonia diagnostics to identify and intervene promptly.

Moreover, (Yang & Mei , 2022) Employed five deep-learning models to differentiate pneumonia X-ray pictures and subsequently performed a comparison study under different conditions. By employing this procedure, we could determine the optimal model for each patient, resulting in improved efficiency in detecting pneumonia. The suggested structure comprises numerous sequential stages. Initially, they acquired chest X-ray images of persons diagnosed with pneumonia and individuals who did not have the ailment from the Guangzhou Women's and Children's Medical Centre. They used a dataset that was publicly accessible. To improve the accuracy of model training and testing, the researchers manually filtered the dataset, removing data that showed significant degrees of inconsistency and duplication. Therefore, only pneumonia X-ray images of exceptional quality were selected for further study. In addition, due to the imbalanced distribution of pneumonia and normal X-ray images in the dataset used for this investigation, a subset of the pneumonia X-ray photos was subsequently removed. Ultimately, an equal number of photos portraying persons with pneumonia and those illustrating healthy participants were selected. 80% of the data were randomly selected for the training set, while the remaining 20% was allocated to the testing set. Before executing the training and testing phases, it was imperative to standardize the dimensions of each image by resizing them to a uniform size, such as 227×227.

Furthermore, five deep learning models, namely LeNet5, AlexNet, ResNet18, MobileNet, and Vision Transformer, are developed specifically for pneumonia recognition. Furthermore, a

thorough evaluation is conducted among the five models, and the optimal model is selected based on the individual circumstances. The study revealed that the LeNet5 and AlexNet models effectively detected pneumonia when utilized with limited datasets.

In contrast, the MobileNet and ResNet18 models showed a higher aptitude for pneumonia detection when working with extensive datasets. An in-depth analysis and comparison of each model in different settings can provide a deeper understanding of the effectiveness of each model in detecting pneumonia. As a result, this makes it easier to use and select deep-learning models for identifying pneumonia.

Kundu et al. (Kundu, Das, Geem, Han, & Sarkar, 2021) developed a computer-aided diagnostic technique to automatically detect pneumonia by analyzing chest X-ray pictures. Deep transfer learning was employed to overcome the scarcity of available data. A collection of three convolutional neural network models, specifically GoogLeNet, ResNet-18, and DenseNet-121, was generated. The study utilized a weighted average ensemble technique, where the weights assigned to the base learners were determined using a novel approach. A weight vector is generated by combining four common assessment metrics: precision, recall, f1-score, and the area under the curve. The weight vector in current research is frequently obtained by experimental methods, which are prone to potential mistakes. The proposed methodology was assessed on two pneumonia X-ray datasets, which were publicly accessible and contributed by Kermany et al. and the Radiological Society of North America (RSNA). The assessment utilized a five-fold cross-validation approach. The presented technique exhibited accuracy rates of 98.81% and 86.85%, together with sensitivity rates of 98.80% and 87.02%, when assessed on the Kermany and RSNA datasets, respectively. The results demonstrated superior performance in comparison to current cutting-edge procedures, with their approach surpassing commonly used ensemble techniques. The

datasets underwent statistical analysis using McNemar's and ANOVA tests, revealing the robustness of the approach.

(Manickam et al., 2021) Proposed a novel deep-learning approach for the automated detection of pneumonia. This approach utilizes advanced transfer learning methods to optimize the detection process, improving accuracy. The provided chest X-ray pictures underwent preprocessing to identify the existence of pneumonia. The U-Net architecture was employed for segmentation, enabling the classification of pneumonia as normal or abnormal (particularly bacterial or viral).

Pre-trained models like ResNet50, InceptionV3, and InceptionResNetV2 underwent training on the ImageNet dataset to achieve this categorization.

Furthermore, two optimizers, specifically Adam and Stochastic Gradient Descent (SGD), were utilized to improve the precision of pre-trained models and extract efficient characteristics. The optimizers' performance was assessed using batch sizes of 16 and 32. The acquired values are employed to examine and compare the performances of the utilized pre-trained models. The models in question encompass DenseNet-169+SVM, VGG16, RetinaNet + Mask RCNN, VGG16, Xception, fully connected RCNN, and more. The comparison is performed with many metrics. The results demonstrate that the ResNet50 model achieved an accuracy of 93.06%, a precision rate of 88.97%, a recall rate of 96.78%, and an F1-score rate of 92.71%. These values exceed those produced by the previously mentioned models.

The study by (Chouhan et al., 2020) introduced a novel deep-learning architecture that employed transfer learning to improve the detection of pneumonia. This approach entailed extracting visual characteristics by employing diverse neural network models pre-trained on the ImageNet dataset. The collected features are then sent into a classifier to enable prediction. An examination was

conducted on the performance of five separate created models. They subsequently introduced a pioneering ensemble model that combines the results of multiple pre-trained models. The ensemble model exhibited superior performance compared to the individual models, attaining the present maximum degree of accuracy in the domain of pneumonia identification. When tested using undisclosed data from the Guangzhou Women and Children's Medical Centre dataset, the ensemble model had a classification accuracy of 96.4% and a recall rate of 99.62%.

(Jain et al., 2020) proposed using Convolutional Neural Network (CNN) models to detect pneumonia by analyzing x-ray pictures. Several Convolutional Neural Networks (CNNs) were trained to classify x-ray pictures into two distinct categories: pneumonia and non-pneumonia. This was accomplished by adjusting various parameters, hyperparameters, and the number of convolutional layers. The study has cited a total of six models. The first model consists of two convolutional layers, while the subsequent model consists of three convolutional layers. The remaining four models are pre-trained: VGG16, VGG19, ResNet50, and Inception-v3. The initial and subsequent models achieve validation accuracies of 85.26% and 92.31%, respectively. The accuracies of VGG16, VGG19, ResNet50, and Inception-v3 are 87.28%, 88.46%, 77.56%, and 70.99% respectively. Furthermore, The diagnosis of pneumonia was conducted (Ayan & Ünver, 2019) using two well-known convolutional neural network models, specifically Xception and Vgg16. The training step involved the utilization of transfer learning and fine-tuning approaches. The dataset used in this study consists of 5856 frontal chest X-ray pictures. The collection has photographs with various resolutions, including proportions such as 712x439 and 2338x2025. The dataset has 1583 photos illustrating normal cases and 4273 images illustrating instances of pneumonia. The test results demonstrate that the Vgg16 network exhibited superior performance

than the Xception network, achieving an accuracy rate of 0.87% compared to 0.82% for the Xception network.

Conversely, the Xception network has shown greater effectiveness in identifying patients with pneumonia. As a result, it was noted that each network has unique skills when used on the same dataset. The table below shows a summary of all previous studies that have been mentioned.

Table 1: Summery of pneumonia intelligent systems

Ref	**Year**	**Objectives**	**Algorithm**	**Dataset**	**Results**	
Singh, Kumar, Kumar, Verma, & Shitharth, 2023	2023	They are detecting pneumonia	QCSA	5,856 chest X-ray	QCSA accuracy of 94.53%	
Yi et al., 2023	2023	They identified pneumonia	DCNN	5856 images	Accuracy = 98.02.	
Bhatt & Shah, 2023	2023	They identified pneumonia.	CNN	5863 images in the JPEG	CNN achieved an accuracy rate of 84.12%.	
Yang, Mei, Piccialli, & Bioinformatics, 2022	2022	They identified Pneumonia	ResNet50, and VGG16	5840 images	VGG16 accuracy of 95.6%.	
Yang & Mei, 2022	2022	They are detecting pneumonia.	LeNet5, AlexNet, ResNet18, MobileNet, and Vision Transformer	Guangzhou Women's and Children's Medical Centre	LeNet5 and AlexNet models had superior performance in small datasets. While the MobileNet and ResNet18 models exhibited greater suitability in large datasets	
Račić, Popović, & Šandi, 2021	2021	They are selecting the appropriate diagnosis for detecting pneumonia.	CNN	5,856 images of chest X-rays	They have shown that CNN gave 88.90% accuracy in the classification process.	
Kundu, Das, Geem, Han, & Sarkar, 2021	2021	They automatically detect pneumonia.	GoogLeNet, ResNet-18, and DenseNet-121,	two pneumonia X-ray datasets	The proposed model gave accuracy rates of 98.81% and 86.85%	
Manickam et al., 2021	2021	automated identification of pneumonia.	ResNet50, InceptionV3, and InceptionResNetV2	chest X-ray	The ResNet50 model, as proposed, attained an accuracy of 93.06%	

Chandra & Verma, 2020	2020	examining their significance in detecting pneumonia compared to the surrounding regions.	Multilayer Perceptron, Random Forest, Sequential Minimal Optimization (SMO), Logistic Regression, and Classification through Regression	412 chest X-ray images	The proposed method achieved a notably better accuracy rate of 95.63% when employing the Logistic Regression classifier and 95.39% while utilizing the Multilayer Perceptron.	
Darici, Dokur, Olmez, & Engineering, 2020	2020	They distinguish between bacterial pneumonia, viral pneumonia, and healthy individuals.	CNN and ensemble learning	Chest X-ray dataset	The models achieved an average accuracy of 95% for binary classification and 78% and 75%, respectively, for multi-class classification.	
Chouhan et al., 2020	2020	They identification of pneumonia	Deep Learning Framework	ImageNet dataset	The ensemble model achieved a classification accuracy of 96.4%	
Jain, Nagrath, Kataria, Kaushik, & Hemanth, 2020	2020	They identified pneumonia	VGG16, VGG19, ResNet50, and Inception-v3	x-ray images	The respective accuracies of VGG16, VGG19, ResNet50, and Inception-v3 are 87.28%, 88.46%, 77.56%, and 70.99%.	
Ayan & Ünver, 2019	2019	The diagnosis of pneumonia	Xception and Vgg16	5856 frontal chest X-ray images	The test findings indicate that the Vgg16 network outperformed the Xception network in terms of accuracy, with a higher accuracy rate of 0.87% compared to 0.82% for the Xception network.	
Sirazitdinov et al., 2019	2019	They develop automated methods for the diagnosis and localization of pneumonia in chest x-ray images.	RetinaNet and Mask R-CNN	dataset with 26,684 images	The performance of their approach was evaluated using a recall score of 0.793	
(Varshni, Thakral, Agarwal, Nijhawan, & Mittal, 2019	2019	They develop CNN methods for the diagnosis of pneumonia in chest x-ray images.	XCeption, DenseNet-169, VGG-16, VGG-19, and ResNet-50	112,120 frontal chest X-ray images	They have shown that the DenseNet-169 gave the best results of 80.0 compared with other models.	

Various machine learning and deep learning techniques were employed in multiple previously summarized works to tackle the issue of pneumonia diagnosis. Nevertheless, there are obstacles such as:

1) A restricted chest X-ray dataset was utilized, which was inadequate for generating a meaningful model to detect pneumonia.

2) Due to the omission of data augmentation approaches, the absence of essential pretreatment processes, and the failure to utilize huge datasets, the models implemented in previous articles yield suboptimal results in the detection phase.

Based on our observations in previous studies, two crucial factors need to be considered: the dataset's origin and the approach used. The most reliable source can be found at the hospital or other medical facility. Obtaining this source is a time-consuming process requiring significant work and many permissions from the university and the hospital. This is a significant problem; we can address it using a preexisting data repository called Kaggle. The second matter pertains to identifying the most optimal deep learning models for detecting. The CNN is considered the superior model, featuring many state-of-the-art variants such as ResNet-18, DenseNet-121, ResNet50, InceptionV3, VGG16, and VGG19.

2.3 Dataset Description

The dataset utilized in this thesis was acquired from the Kaggle website1 and pertains to pneumonia. The dataset comprises 5,856 chest X-Ray images in JPEG format, categorized into

[1] https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

Pneumonia and Normal. The study employed anterior-posterior chest X-ray images acquired from a retrospective cohort of pediatric patients aged one to five, obtained from Guangzhou Women and Children's Medical Centre in Guangzhou. The chest X-ray imaging was performed as part of the patients' routine clinical care. A preliminary screening method was implemented to ensure quality control when analyzing chest x-ray pictures. This entailed the elimination of all scans that were considered to have substandard quality or were unreadable. Two skilled physicians assessed and awarded diagnostic scores to the photos before they were approved for training in the artificial intelligence (AI) system. A third specialist subjected the evaluation to an additional assessment to minimize the possibility of grading inaccuracies.

The dataset was partitioned into three files: training, validation, and testing. Each individual possesses chest X-rays that are categorized into two groups: Pneumonia and Normal. The table and the figure below display the number of photos in the dataset.

Table 2: Distribution of Pneumonia Dataset

	Normal	Pneumonia
Training	1,341	3,875
Validation	8	8
Testing	234	390

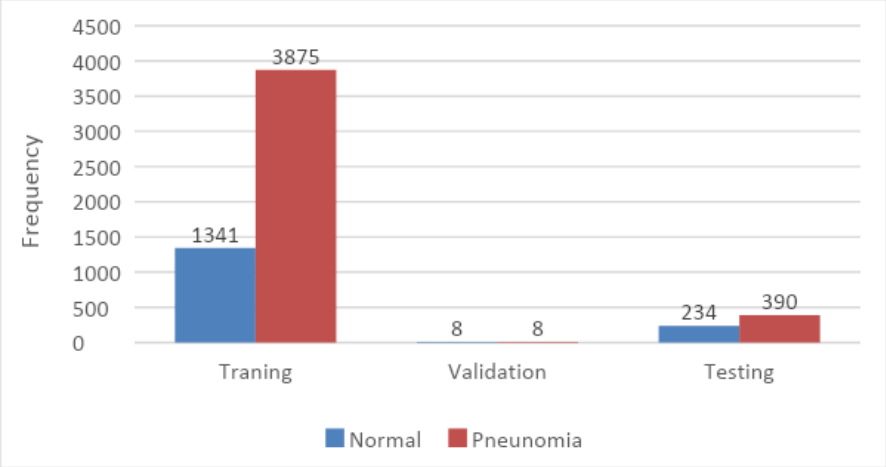


Figure 1 Distribution of Pneumonia Dataset

2.4 Summary

This chapter aimed to present a literature review of the pneumonia detection process. Moreover, this chapter talked about current pneumonia detection systems, techniques used by pneumonia detection frameworks, and the obstacles of these systems. Finally, the machine learning and deep learning algorithms and how they are applied in the pneumonia detection domain as an intelligent system were discussed to improve the efficiency of the detection process in this environment.

Chapter 3

Methodology

3.1 Introduction

This chapter outlines the proposed deep learning (ML) models intended to be used and applied to datasets, enabling a comparative study of classification techniques. Our model aims to determine the most effective techniques for classifying pneumonia disease with high accuracy. The investigation begins by utilizing the worldwide chest X-ray dataset for pneumonia detection, downloaded from the Kaggle dataset.

3.2 Proposed Model

The proposed methodology aims to increase the accuracy of diagnosis in chest X-ray images and involves many phases. The X-ray dataset is downloaded from the Kaggle platform first. The x-ray dataset has many pictures showing chest images, which are normal and pneumonia as showing in figure 2.

The next important stage is data annotation, which involves classifying each gathered image according to the type of sickness (normal, pneumonia). Supervised learning requires providing the model with the ground truth labels, and it needs to be trained as shown in recent studies (Jaiswal et al., 2019)

Research by (Garstka & Strzelecki , 2020) indicates that data augmentation is then used to increase the diversity of the dataset. After that, the images are subjected to the preprocessing phase, and the image generator function is used to generate new training instances by applying rotation, scaling, flipping, etc., to the images.

(Saul et al., 2019) emphasizes that preparing the data to be uniformly distributed prepares the images for input into the model. The input data is usually improved at this point by scaling the images to a uniform size, leveling the pixel values, and making other adjustments.

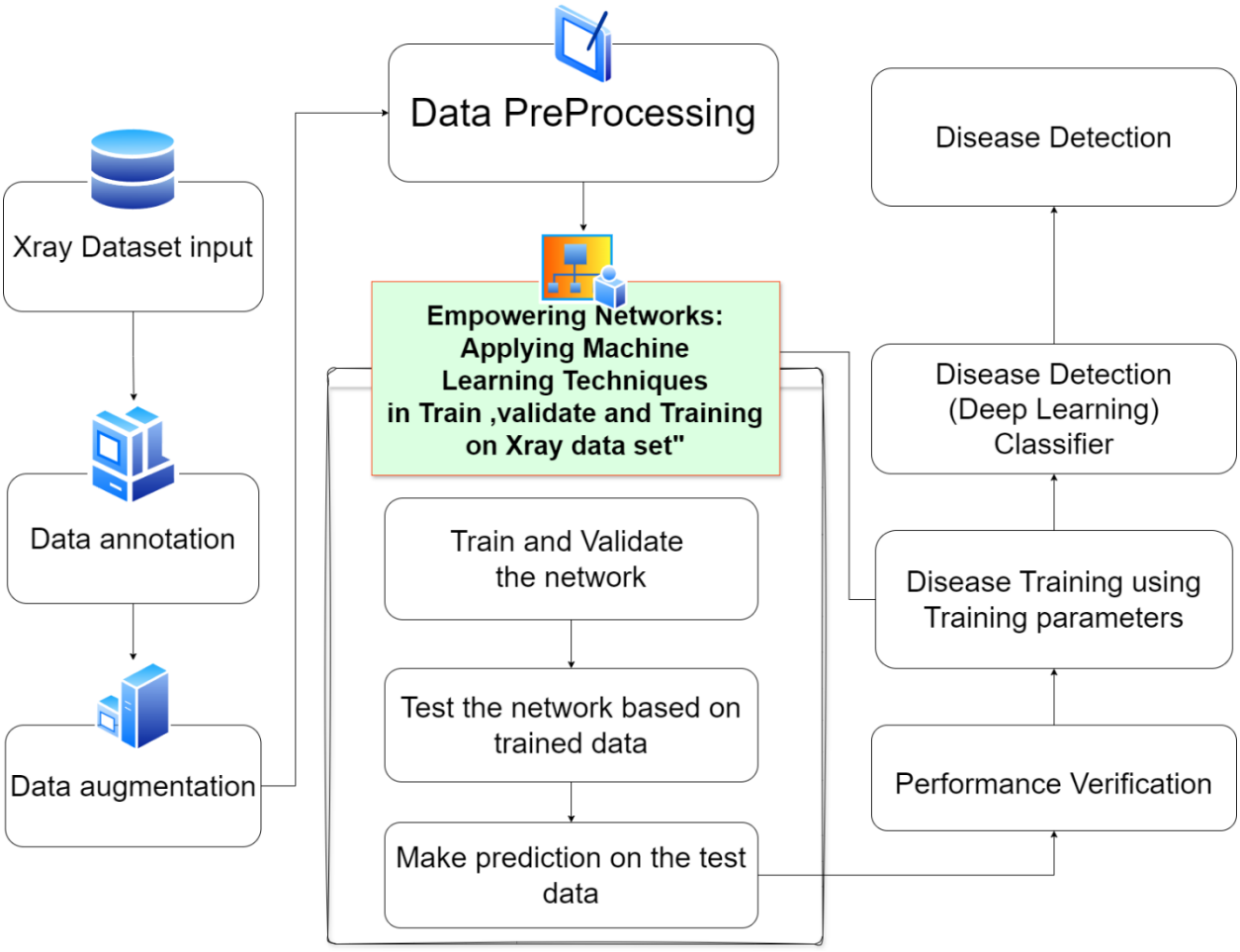


Figure 2 The proposed model

The proposed model uses a deep learning classifier trained on chest X-ray images for polynomial identification. This classifier's training parameters are fine-tuned by repeatedly training and validating the model on the labeled photos in the dataset until it can differentiate between various chest states with high accuracy and minimum loss.

After that, the trained model is tested on a test dataset to predict the model's accuracy, and after that, the model can be deployed and used by healthcare specialists to diagnose pneumonia.

3.3 Dataset Input

Pneumonia is a severe illness that causes inflammation in the air sacs of the lungs, commonly affecting either one or both lungs. The children often suffer more from this disease than other contagious diseases do each year and have the highest death rate. The diagnosis of the condition usually depends on chest X-rays, which play a crucial role in verifying its presence.

The evaluated dataset consists of 5,856 verified chest X-ray images, divided into separate training and testing sets containing samples from different patients. Each image's labeling consists of the disease category (NORMAL, BACTERIA, or VIRUS), a random patient ID, and the patient's corresponding image number. The chest images were mostly obtained from retrospective cohorts of pediatric patients aged one to five years from the Guangzhou Women and Children's Medical Center in Guangzhou, as described in the accompanying publication as described in (Ervural & Ceylan , 2021).

3.4 Dataset preparation

The dataset was downloaded from the Kaggle platform, after which we can clean it to make it ready. This refers to sizing, replacing and correcting the orientations, and removing noises, which is valuable for training the algorithms.

3.4.1 Data annotation

Regards as the sources of labels, which are the underpinnings of training deep learning algorithm that is applied in chest X-rays for pneumonia diagnosis. Pneumonia may manifest as density patches on radiographic imaging in opacities, infiltrates, and consolidations that the delicate surrounding areas can swamp. Specialists who analyze these characteristics thoroughly ensure they can correctly distinguish between regular and irregular X-rays. Research by (Jaiswal et al., 2019)

indicates that annotated datasets provide the progress, evaluation, and enhancement of algorithms, increasing the algorithm's efficiencies and reliability. Additionally, annotated data provides data standardization, which contributes to evaluating and comparing algorithms used for pneumonia diagnosis, helping facilitate advancement in this field.

3.4.2 Data augmentation

One primary factor that X-ray machines for pneumonia diagnosis exist is the issue of mistaken diagnosis. It is a common scenario that the appearances of X-rays of pneumonia vary from the point of illumination and location orientation, which are crucial factors in the AI model's prediction performance which has been demonstrated in previous research (Garstka & Strzelecki , 2020). Data augmentation mechanisms such as rotation, flipping, and scaling can help eliminate the whole dataset's variance by creating more training samples. The model can then learn to handle different kinds of situations by this process. The above aside, pneumonia datasets might be of relatively small dimension, which could cause overfitting if no component is added to prevent it. Data augmentation is a technique that allows data growth, thus diminishing the overfitting probability and improving the model's generalizing capability.

Data augmentation enables models to acquire the ability to recognize major pieces of the X-ray pictures and features of pneumonia cases, making a model do a better job in pneumonia detection.

3.4.3 Data preprocessing:

Its ability to detect pneumonia from X-ray images explains why AI algorithms are crucial for proper diagnosis. For machine-learning models, noise, distortion, and position variations in X-ray pictures are among their foes because they make their efficiency lower. The pre-processing techniques, for instance, those using the image generator with parameters like rotation, scaling,

shearing, sample-wise centering, and sample standardization, are important to the input data enhancement and the model's accuracy.

(Saul et al., 2019) argues that One of the major challenges is that rotation, scaling, shearing, and so on are essential to increasing and enhancing the dataset’s variety. The changes help the model integrate better with the variations common in such scenes: the angles, dimensions, and layout of the lungs besides the adjacent tissues. This is improved because it means the model can adopt the already learned patterns to new and unfamiliar data and minimize the possibility of overfitting. Matrix-wise centering and standardization are necessary to preserve the mentioned pixels' uniform dimension together with the entire data distribution. Centering is either normalization or the subtraction of the average pixel value of the set from each pixel, thus yielding a zero-mean dataset. This makes it possible to eliminate any potential bias the sample may have.

The basic operation of standardization is to rectify the pixel values to contain the same variance, to normalize it, and to make it more applicable for learning. An image generator can follow these principles to clean data, enrich the music with diversity, and boost the models' generalization ability. As a result, this contributes to ensuring better reliability and accuracy in figuring out pneumonia in X-ray pictures, which is a vital component in ensuring timely medical diagnosis and intervention.

3.5 Build model phase

3.5.1 CNN

In recent years, Deep Learning methods such as Convolutional Neural Networks (CNN or ConvNets) have been developed to process a substantial volume of data. Convolutional neural

networks (CNN) are deep learning networks offering a novel view of knowledge and are mainly utilized in image-based pattern detection.

As shown in the research by (Dutta ,2018) AI and Computer Vision evolved from the 1950s until 2012. This indicates that the scientists of that time, who were involved in AI creation, developed methods of recognizing visual data. Such was the genesis of the Computer Vision. Next, it highlights the key breakthrough in 2012, where researchers from the University of Toronto developed an AI algorithm that outperformed the best image recognition technology.

In 2012, Alex Krzyzewski created an AI system called Alexnet, which won the ImageNet Computer Vision competition with 85% accuracy. At the core of the AlexNet algorithm were convolutional neural networks (CNNs), which are neural networks inspired by human vision. CNNs are currently extensively used in computer vision which aligns with existing research (Krizhevsky et al., 2019).

Background of CNNs:

CNNs were first developed and used around the 1980s. The most that a CNN could do at that time was recognize handwritten digits. It was mostly used in the postal sector to read zip and pin codes. Remembering that any deep learning model requires a lot of data to train and computing resources is important. This was a major drawback for CNNs at that period; hence, CNNs were only limited to the postal sectors and failed to enter the world of machine learning.

In 2012, Alex Krizhevsky repurposed deep learning that uses multilayer neural networks by applying it to ImageNet datasets and abundant computing resources. This enabled researchers to know the importance of CNNs and bring it back to life.

A convolutional neural network (CNN/ConvNet) is a deep neural network mostly used for analyzing visual imagery in deep learning. When considering a neural network, the common association is with matrix multiplications. However, this is not the case with ConvNet. It employs a distinctive methodology known as Convolution. In mathematics, convolution is a mathematical process that combines two functions to create a third function that describes how one function affects the shape of the other.

In essence, the primary function of the ConvNet is to transform the images into a more manageable format while retaining essential properties necessary for accurate predictions.

Working of CNN:

Firstly, we must point out that the RGB images are pixel-valued matrices with three planes, while the grayscale images have a single plane. CNNs apply a filter/kernel (3×3 matrix) to an input image, creating convolved features that pass to the next layer, as shown in Figure 3.

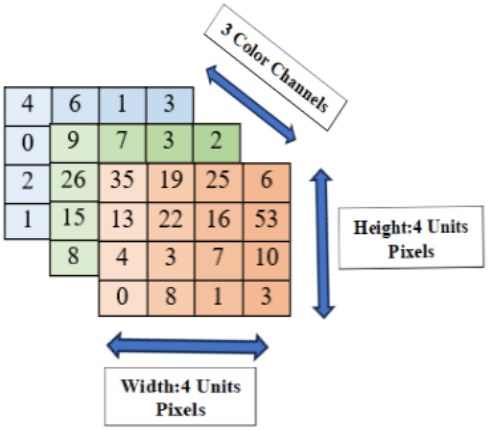


Figure 3 RGB images are pixel-valued matrices with three planes.

According to Figure 4, convolutional neural networks consist of artificial neurons that calculate weighted sums of inputs and output activation values, passing activation functions from one layer to the next.

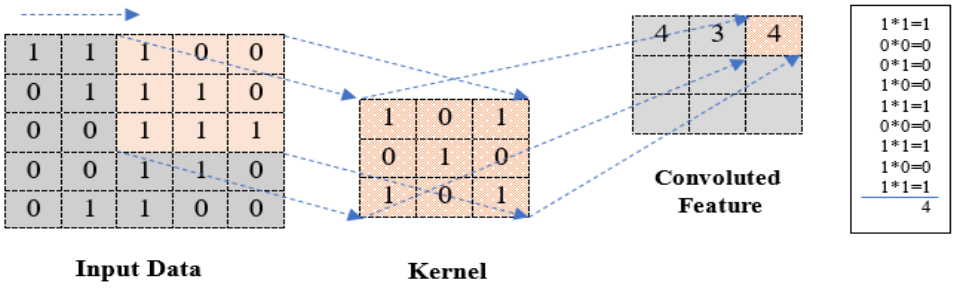


Figure 4. CNNs work by applying a filter/kernel (3×3 matrix)

In the first phase, as in Figure 5, the first layer extracts basic features like edges, followed by more complex ones like corners and combinational edges, and eventually identifies objects and faces.

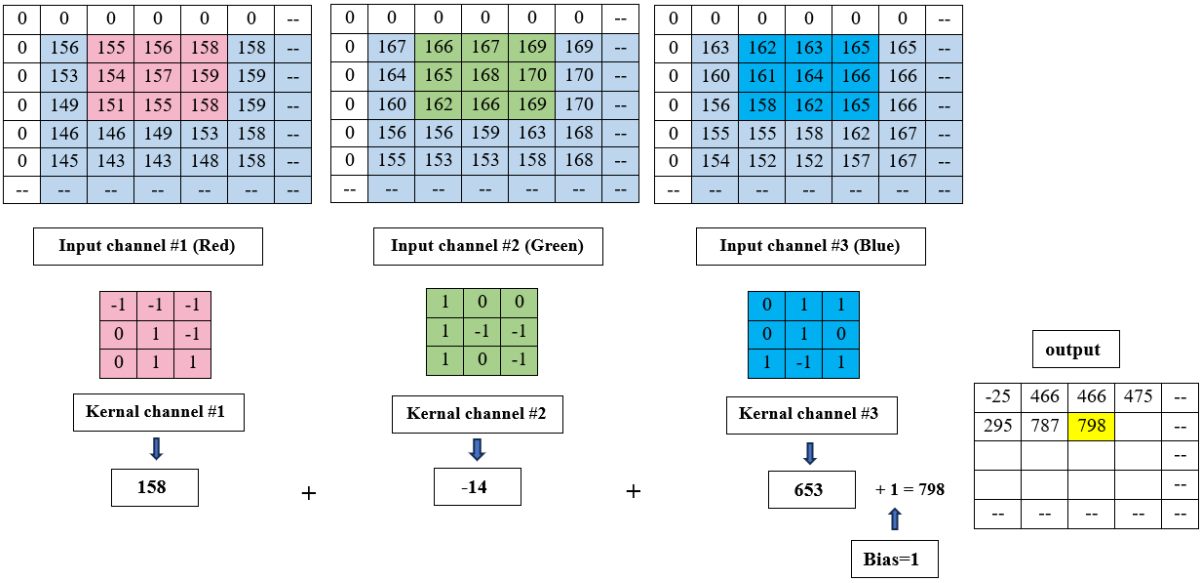


Figure 5, the first layer extracts basic features like edges.

After that, the final convolution layer outputs confidence scores based on the activation map, indicating the likelihood of an image belonging to a specific class, such as cats, dogs, or horses; this appears IN Figure 6.

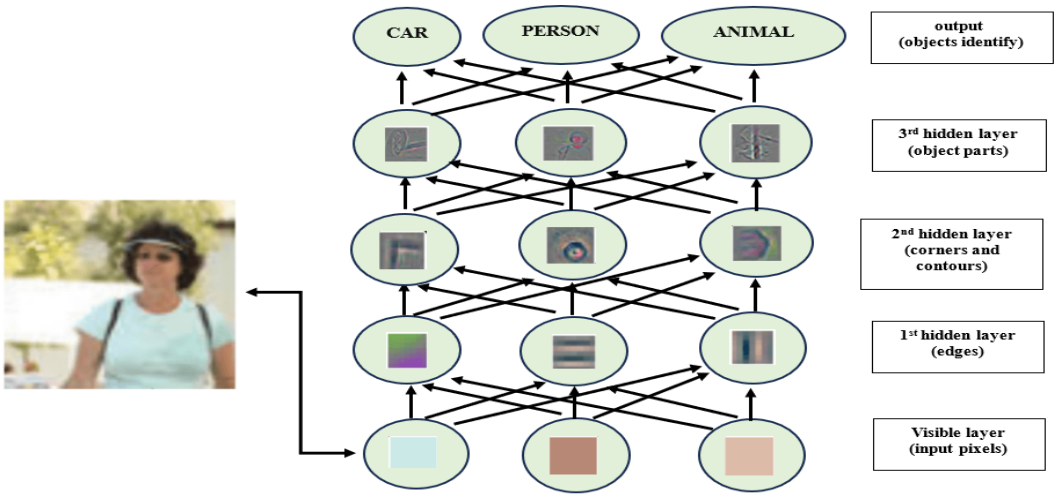


Figure 6 shows the final convolution layer outputs confidence scores based on the activation map.

On the other hand, the Pooling layer reduces the spatial size of the convoluted feature, reducing computational power, as shown in Figure 7. There are two types of pooling: the average and the max pooling, with Max Pooling having minimal issues. Max Pooling is a technique that finds the maximum pixel value from a kernel-covered area and acts as a Noise Suppressant, de-noising, and dimensionality reduction. At the same time, Average Pooling is a noise suppression mechanism that reduces dimensionality by aggregating all values from the kernel-covered portion of the image. Therefore, we can conclude that Max Pooling outperforms Average Pooling.

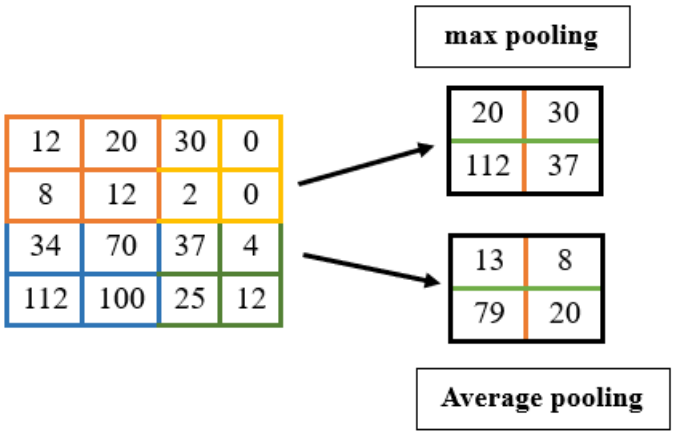


Figure 7: conclude that Max Pooling And Average Pooling

Limitations of CNNs:

CNNs, however, are known to provide in-depth results by detecting hidden patterns and details but, at the same time, are not suitable for understanding image content being vague.

Now, let us look into a concrete one. When a given image is fed into a CNN, the system successfully recognizes the presence of two individuals, a man in his mid-thirties and a girl, roughly ten years old. Nevertheless, we have different fantasies while thinking the same thing.

Maybe it refers to a day when a father and child are together for a picnic or camping. Maybe it is a schoolyard where the child successfully scored a goal, and his father's delight made him lift the child.

CNN's limitations in practical applications, such as being unable to block and remove inappropriate content completely, are evident in its use to moderate social media content.

Multiple studies have demonstrated that Convolutional Neural Networks (CNNs) that have been trained on ImageNet and other widely used datasets are deficient in detecting objects when they encounter them in varying lighting conditions and from novel perspectives.

Accordingly, the Convolutional neural networks revolutionized artificial intelligence by using it in facial recognition, image search, editing, and augmented reality. Despite their limitations, they're still far from replicating human intelligence's key components.

Multiple studies have demonstrated that Convolutional Neural Networks (CNNs) that have been trained on ImageNet and other widely used datasets are deficient in detecting objects when they encounter them in varying lighting conditions and from novel perspectives which has been demonstrated in previous research (Jaing et al., 2020).

Accordingly, Convolutional neural networks revolutionized artificial intelligence by using it in facial recognition, image search, editing, and augmented reality. Despite their limitations, they're still far from replicating human intelligence's key components.

3.5.2 VGG 16

(Jaing et al., 2020) demonstrate one such computer vision competition is the ImageNet Large Scale Visual Recognition Challenge, more commonly known as ILSVRC. There are two events where

teams compete annually. Object localization first involves detecting items inside a picture using 200 classifications. The second is image categorization, where each picture is labeled from a list of one thousand possible categories. VGG 16 was suggested in 2014 by Andrew Zisserman and Karen Simonyan of the Visual Geometry Group Lab at Oxford University for large-scale image recognition. This model won the 2014 ILSVRC competition in the first and second-place categories mentioned above. On the ImageNet dataset, which includes 14 million images from 1000 classes, this model attains a top-5 test accuracy of 92.7%. See Figure 8.

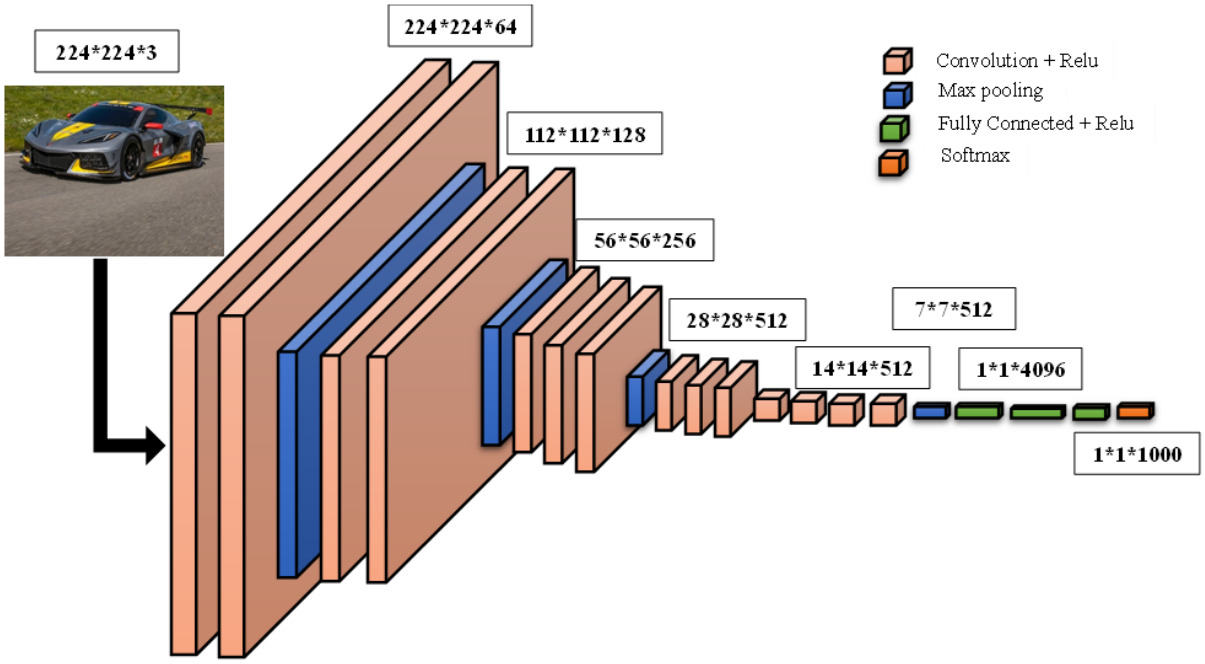


Figure 8 VGG16 Model

Working of VGG-16:

We must point out that all of the photos in the ImageNet collection are 224 x 224 pixels in size and feature RGB channels. As a result, the input tensor is (224, 224, 3). This model takes an image as input and returns a vector with a thousand values; see Equation 1.

$$\hat{y} = [\widehat{y_0}\ \widehat{y_1}\ \widehat{y_2}\ \widehat{y_3} \dots \widehat{y_{999}}\ ] \qquad (1)$$

The above vector shows the likelihood of classifying the related class. Consider a model that assigns a probability of 1, a probability of 0.05, a probability of 0,03, a probability of 0.72, a probability of 0.05, a probability of 999, and a probability of 0 to each of the previous classes. Thus, this will have the following classification vector, see Equation 2.

$$\hat{y} = [\widehat{y_0} = 0.1\ 0.05\ 0.05\ 0.03 \dots \widehat{y_{780}} = 0.72 \dots \widehat{y_{999}} = 0.05\ ] \qquad (2)$$

Our usage of the softmax function ensures that these probabilities sum to 1.

Then, we add the five candidates with the highest probability to vector C, as in equation 3.

$$C = [780\ 0\ 1\ 2\ 999\ ] \qquad (3)$$

Hence, our ground truth vector, denoted as G and it, is defined as the matrix as follows in equation 4:

$$G = [G_0\ G_1\ G_2\ ] = [780\ 2\ 999\ ] \qquad (4)$$

Furthermore, the following is the definition of our Error function using the following equation 5:

$$E = \frac{1}{2}\sum_{K} d(c_i, G_k)[Tex] \qquad (5)$$

Where:

[Tex] if c_{i} = G_{k} and d = 0 else, then d = 1 [/Tex] consequently.

In this case, the loss function is calculated using Equation 6 and 7:

$$E = \frac{1}{3} d(c_i, G_1) + d(c_i, G_2) + d(c_i, G_3) \quad (6)$$

So:

$$E = \frac{1}{3}(0 + 0 + 0)[Tex] \quad (7)$$

Where, kernel 6pc E \, = \, 0 \\ [/Tex]. The loss equals zero since the predicted top-5 matrix contains all ground truth categories.

VGG Architecture:

The network accepts an image of 224 by 224 pixels with three channels of colors as the input. The subsequent two layers follow with a 64 deepness and a filter size 3*3 with the same padding. Sixteen at the end of a max pooling layer with stride (2, 2), they have got two convolutional layers with a filter size of (3, 3) and a filter size of 128. The following stage presents a layer with max-pooling (2, 2) that is identical to the layer before. Further, two more convolutional layers are of 3x3 level with 256 filters. Secondly, there are two groups of 3 convolution layers and one max pool layer for data down sampling as revealed in (Sharma & Guleria, 2023).

Every filter (3, 3) is represented by 512 filters. The filters in the convolutional layers are uniformly padded. After that, the illustrated image is fed as input into the multi-layer stack of convolutional layers. The filters utilized in the convolution and max-pooling layers are 3*3 in size, and this procedure is opposite to the 11*11 filters used in AlexNet and the 7*7 filters used in ZF-Net. Within some levels, a 1*1 pixel is employed to modify the quantity of input channels. A 1-pixel

padding, known as the same padding, is applied after each convolution layer to preserve the spatial features of the image.

This procedure is continued with the convolution and max-pooling layers series, we obtained a feature map with dimensions of (7, 7, 512). We transform this output into a feature vector of shape (1, 25088) by flattening it. Following this, there are three fully connected layers. As in Figure 9, The first and second layers receive input from the previous feature vector and produce a vector. The third layer generates 1000 channels for 1000 classes in the ILSVRC . In other words, the third fully connected layer utilizes the softmax function to classify the 1000 classes. The activation

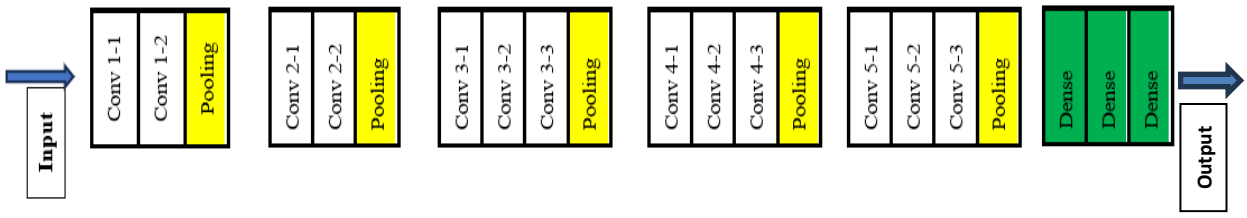


Figure 9 VGG 16 Architecture

function used in all the hidden layers is the Rectified Linear Unit. It is very powerful regarding computational efficiency.

Configuration:

The table 3 provided below enumerates various VGG models. There are two variations of VGG-16, namely VGG-16C and VGG-16D. The only notable distinction between them is that, except for a few convolution layers, they employ a (3, 3) filter size convolution instead of (1, 1). The first model contains 134 million parameters, while the second model contains 138 million parameters.

Table 3 VGG 16 Configuration

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
Input (224*224 RGB image)					
Conv3-64	Conv3-64 **LRN**	Conv3-64 **Conv3-64**	Conv3-64 Conv3-64	Conv3-64 Conv3-64	Conv3-64 Conv3-64
Maxpool					
Conv3-128	Conv3-128	Conv3-128 **Conv3-128**	Conv3-128 Conv3-128	Conv3-128 Conv3-128	Conv3-128 Conv3-128
Maxpool					
Conv3-256 Conv3-256	Conv3-256 Conv3-256	Conv3-256 Conv3-256	Conv3-256 Conv3-256 **Conv1-256**	Conv3-256 Conv3-256 **Conv3-256**	Conv3-256 Conv3-256 Conv3-256 **Conv3-256**
Maxpool					
Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512 **Conv1-512**	Conv3-512 Conv3-512 **Conv3-512**	Conv3-512 Conv3-512 Conv3-512 **Conv3-512**
Maxpool					
Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512 **Conv1-512**	Conv3-512 Conv3-512 **Conv3-512**	Conv3-512 Conv3-512 Conv3-512 **Conv3-512**
Maxpool					
FC-4096					
FC-4096					
FC-1000					
SOFT-MAX					

Object Localization in Image:

Rectifying object localization requires the output replacement of passing objects' class scores with the coordinates of the bounding box to determine the exact positioning. A bounding box location is expressed by a 4-dimensional vector whose elements are (x and y coordinated of center point and height and breadth). Compartmentalization of localization can be conducted in two ways. The first one is an object identification problem, and its solution is to use a shared bounding box between different bounding boxes and a 4-parameter vector as the output. The second variation is a class box. It is specifically related to the class. This arrangement results in a 4000-parameter

vector as the output. The scientists used both approaches (A), making VGG-16 (D) the base model (B). Besides that, we have to move from [classification loss to regression loss], such as Mean Squared Error(MSE), which downsizes the difference perspective of the predicted target and the real value.

Outcomes:

VGG demonstrated one of the top configurations in the 2014 ILSVRC competition, along with other architectures. Apart from its classification function, the algorithm created a subclassification model with an accuracy 0.07. While Inception-V3, a close second with a 63.6% accuracy rate, got Google's attention, GoogLeNet, a 6.7% classification error model, stole the spotlight. 66%, bringing to the forefront that it was not only the first to recreate the animal's movements, but it also ranked top on the localization task with a localization error rate of only 25%. 32%.

Drawbacks of VGG 16:

1. Inadequate training speed is the main criticism, considering that the original VGG model required 2-3 weeks of training on GPU Titan Nvidia.

2. The ImageNet parameters in VGG-16 are in megabytes; thus, the size is 528 MB. Therefore, it is characterized by enormous table usage and throughput that hamper efficiency.

3. With almost 138 million parameters as a boon, burgeoning gradients also surface.

Additional progress: ResNet (Restnet) was added to decrease the exploding gradient problems of VGG-16 and reuse the convolutional layers as many researchers have found (Sitaula & Hossain , 2021).

3.5.3 DenseNets

DenseNets, an abbreviation for Densely Connected Convolutional Networks, represents a progressive advancement in these techniques that can improve the depth of deep convolutional networks.

(Wang et al., 2020) points out that we have observed the progression of 5 layers to VGG, which had 19 layers, and ResNets, which have exceeded 100 and even 1000 layers.

Motivation:

The challenges emerge when Convolutional Neural Networks (CNNs) are designed with more layers. This occurs due to the elongated pathway that information must traverse from the input layer to the output layer and vice versa for the gradient. The length of this pathway is significant enough that the information can dissipate or disappear before reaching its intended destination.

DenseNets streamline the connection structure between layers in other architectures, such as Highway Networks, Residual Networks, and Fractal Networks.

The problem that DenseNets solves:

Surprisingly, DenseNets uses fewer parameters than a typical CNN of the same size because they don't have to learn redundant feature maps.

In addition, certain iterations of ResNets have demonstrated that numerous layers have minimal impact and can be omitted. Indeed, the ResNets have many parameters since each layer has its own set of weights to be learned.

Another issue encountered with deep networks was the challenge of training them due to the flow of information and gradients. DenseNets addresses this problem by allowing each layer to directly

access the gradients from the loss function and the original input image which has been demonstrated in previous research (Sanghvi et al., 2023) . This appears in Figure 10.

Structure:

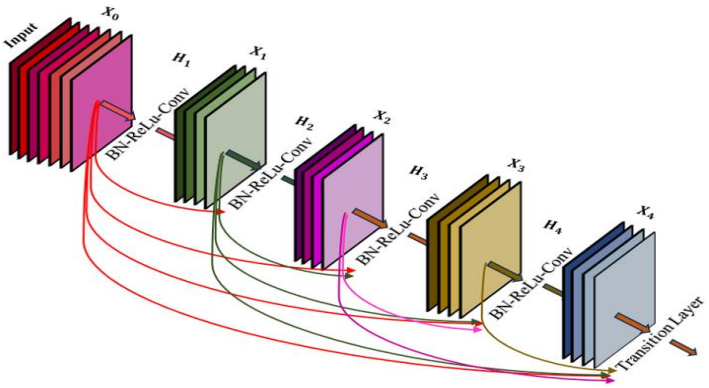


Figure 10 Five layers DenseNet

Conventional feed-forward neural networks establish a connection between the output of one layer and the next layer by applying a combination of operations. It has been observed that this composite typically consists of a convolution operation or pooling layers, batch normalization, and an activation function.

The mathematical expression for this would be as the following equation 8:

$$X_l = H_l\,(X_{l-1}) \qquad (8)$$

ResNets expanded upon this behavior by including skip connections, resulting in the reformulation of equation 9 as follows:

$$X_l = H_l\,(X_{l-1}) + X_{l-1} \qquad (9)$$

DenseNets differ from ResNets at this point. DenseNets employ concatenation rather than summation to combine the output feature maps of a layer with the incoming feature maps.

Thus, the equation 10 undergoes another transformation and takes on a new form:

$$X_l = H_l\left([X_0, X_1, \ldots, X_{l-1}]\right) \qquad (10)$$

The same problem appeared when working on ResNets, where grouping feature maps is not feasible when their sizes vary. Whether the grouping involves addition or concatenation. Similarly to ResNets, DenseNets are organized into DenseBlocks, where the feature map dimensions remain consistent inside each block while the number of filters varies.

The process by which intermediary layers interpolate data is called Transition Layers. These output layers are formed by batch normalization followed by 1 x 1 convolution and 2 x 2 pooling.

Channel-wise, natural language processing systems are growing step by step. So when we set H_l to generate k feature maps on each iteration, then we can extend this to the l-th layer, and when we set H_l to generate k feature maps on each iteration, then we can extend this to the l-th layer according to equation 11::

$$k_l = k_0 + k * (l - 1) \qquad (11)$$

The hyperparameter k represents the growth rate, determining the information incorporated into the network at each layer. Each layer accesses the previous feature maps, contributing to the common knowledge.

DenseNets-B:

DenseNets-B is a variant of DenseNets that utilizes 1x1 convolution to decrease the size of feature maps before the 3x3 convolution, enhancing computational efficiency. The letter B is positioned after the Bottleneck layer, which you are already acquainted with from previous work on ResNets.

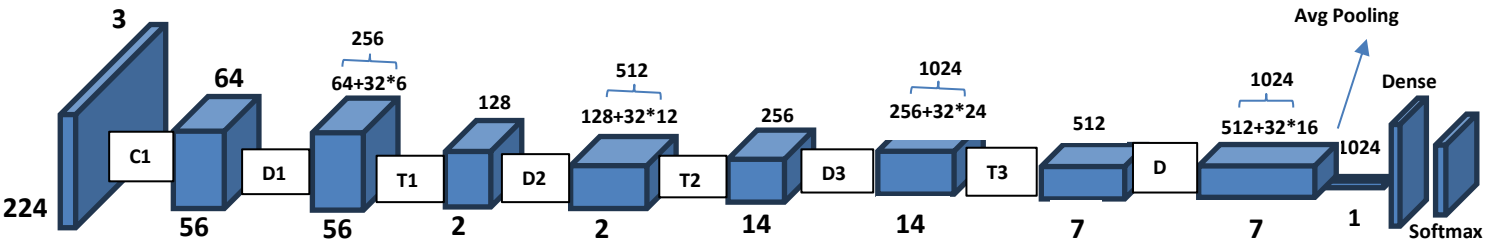
DenseNets-BC:

DenseNets-BC is a slight advancement over DenseNets-B, specifically to decrease the number of output feature maps. From this, we can see that the compression factor (theta) determines the reduction. This is very useful instead of having a specific number of feature maps at a certain layer since the researchers will have theta times the number of feature maps. Indeed, it falls within the interval of [0–1]. When theta is equal to 1, DenseNets will remain unchanged, but if theta is different from 1, it will be

Researchers shall delve into the specifics of the ImageNet dataset to be consistent with the research on ResNets. Nevertheless, the architecture is less complex for CIFAR-10 or SVHN due to the smaller input volumes, see Table 4.

Table 4 DenseNet Configurations

Layers	**Output size**	**DenseNets-121**	**DenseNets-169**	**DenseNets-201**	**DenseNets-264**
Convolution	**112 X 112**	7 **X 7 conv, stride 2**			
Pooling	**56 X 56**	3 **X 3 max pool, stride 2**			
Dense Block (1)	**56 X56**	[1 *X* 1 *conv* 3 *X* 3 *con*	[1 *X* 1 *conv* 3 *X* 3 *con*	[1 *X* 1 *conv* 3 *X* 3 *cor* **6**	[1 *X* 1 *conv* 3 *X* 3 *cor* **6**
Transition Layer (1)	**56 X 56**	1 **X 1 conv**			
	28 X 28	2 **X 2 average pool, stride 2**			
Dense Block (2)	**28 X 28**	[1 *X* 1 *conv* 3 *X* 3 *cor* **12**	[1 *X* 1 *conv* 3 *X* 3 *con* **12**	[1 *X* 1 *conv* 3 *X* 3 *cor* **12**	[1 *X* 1 *conv* 3 *X* 3 *cor* **12**
Transition Layer (2)	**28 X 28**	1 **X 1 conv**			
	14 X 14	2 **X 2 average pool, stride 2**			



Dense Block (3)	**14 X 14**	[1 *X* 1 *conv* 3 *X* 3 *cor*	[1 *X* 1 *conv* 3 *X* 3 *con*	[1 *X* 1 *conv* 3 *X* 3 *cor*	[1 *X* 1 *conv* 3 *X* 3 *cor*
Transition Layer (3)	**14 X 14**	1 **X 1 conv**			
	7 X 7	2 **X 2 average pool, stride 2**			
Dense Block (4)	**7 X 7**	[1 *X* 1 *conv* 3 *X* 3 *cor* **16**	[1 *X* 1 *conv* 3 *X* 3 *con*	[1 *X* 1 *conv* 3 *X* 3 *cor*	[1 *X* 1 *conv* 3 *X* 3 *cor*
Classification layer	**1 X 1**	**7 X 7 global average pool**			
		1000D fully-connected, softmax			

Figure 11 Dense and Transition Blocks

Dense and Transition Blocks:

However, due to the significantly higher number of connections in DenseNets, the visualization becomes more intricate than VGG and ResNets, which is useful in some cases and gives good accuracy. Figure 11 is a straightforward representation of the architectural design of DenseNet-121, the specific variant of DenseNet that will be the primary focus of this study. This is because it is the most basic version of DenseNet among those specifically developed for the ImageNet dataset.

Subsequently, Figure 11 can be compared with Table 4 on DenseNet-121. The measurements within each volume correspond to the width and depth, while the numbers above indicate the dimensions of the feature maps.

The bypass connections in DenseNet are explained by adding feature maps to each new volume, which matches the dimension of the previous volume. This concatenates new information to the previous volume, which is being reused. Note that the growth rate mentioned at the beginning of this section is 32.

After this, multiply this indicator by the number of layers of dense type inside the dense block, and you will get the volume of each dense layer.

Dense Layers:

Adding feature maps involves amplifying the immediate volume with an additional 32 maps, yielding an increase in size from 64 to 256 as of Article 6.

Transition blocks are performed as 1x1 convolutions with 128 filters and a 2x2 average pooling, which downsamples the volume by a factor of two and summarizes the information from feature maps. Each Dense or Normal Convolution block with a 3x3 Kernel size and a Stride of 1 would keep the volume the same. Each Transition Block would consist of a 1x1 Kernel and a Stride of 2 to reduce the volume by half and the feature maps by a quarter, see Figure 12.

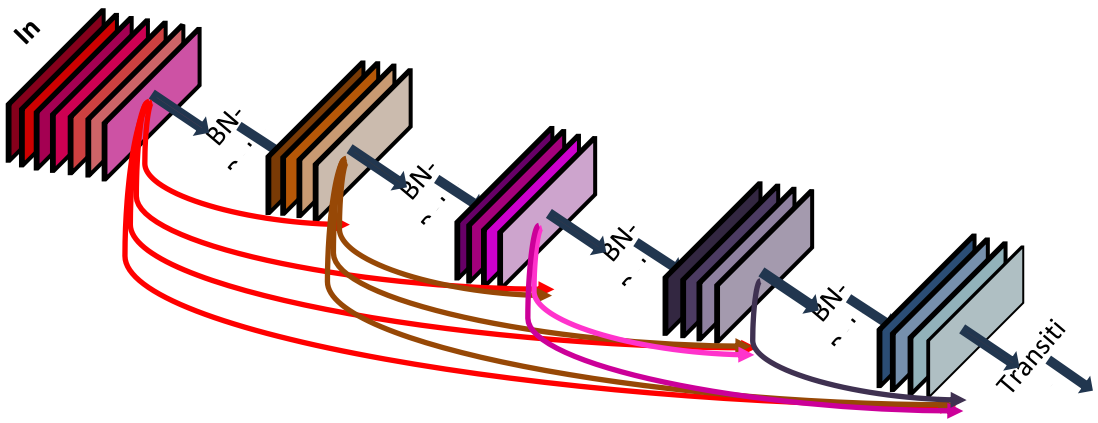


Figure 12 Dense Layers

Nevertheless, the accurate and sound grasp of the entire Dense Layer contained in each Dense Block gives one true and thorough knowledge of the topic. Hence, at the first Dense Block, the first Dense Level is obtained by increasing the number of loop connections to produce 32 feature maps of 128 filters with a 3x3 convolution. This operation is more costly than the previous one. Consequently, the degree of connectivity and amount of data flowing from each Dense Layer inside a Dense Block are fused; hence, the network's collective knowledge is extended by adding new information as noted in the literature (Mahajan et al., 2019).

3.5.4 ResNet

ResNet (Residual Network) is a deep learning model introduced as a special tool for image processing to solve computer vision tasks. The architecture is a CNN, and this special kind of CNN can take advantage of hundreds to thousands of convolutional layers. Earlier CNN designs had performance limitations due to the capacitation problem, which caused the models to be almost overwhelmed by the sheer number of layers. However, researchers had to solve the vanishing gradient issue when adding more layers to the network.

In the study by (Pant et al., 2020), Neural networks get trained using a backpropagation technique that depends on gradient descent, which involves iteratively adjusting the weights to minimize the loss function. Excessive layering can gradually diminish the gradient through repeated multiplications, resulting in its disappearance. This, in turn, causes performance to reach a saturation point or even decline with the addition of each layer.

ResNet offers a novel approach to address the vanishing gradient, called "skip connections." ResNet employs stacking several identity mappings, convolutional layers initially designed to have no effect. It then bypasses these layers and uses the activations from the preceding layer. The network is compressed into fewer layers by skipping, accelerating the initial training process.

Subsequently, while retraining the network, all layers expand, enabling the network's residual sections to explore a wider range of features within the input image.

ResNet models often employ a skip connection strategy where two or three layers are bypassed simultaneously, including nonlinearity and batch normalization in between. Highway Nets, more sophisticated versions of ResNet topologies, can learn "skip weights" that decide the appropriate amount of layers to bypass.

What Is a Residual Block?

Residual blocks constitute the key building blocks of ResNet and are, therefore, essential to this network type. As for the example of VGG16 (previous architecture), convolutional layers continue forming a chain with alternating batch normalization and nonlinear activation layers. ReLu befits in better. This method will be used where only a few convolutional layers are available, and for VGG, about 19 layers is the maximum limit. Nevertheless, things moved one way or another when subsequent research found that adding more network layers could significantly enhance the performance of CNNs.

ResNet architecture considerations that include a side input from the output to a pool of convolution blocks to be fed again into the network are shown in Figure 12 below.

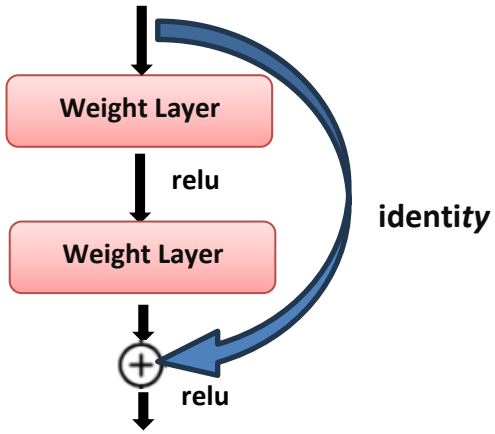


Figure 13 a series of convolution blocks

The provided visual represents a standard residual block. This may be implemented in Python using the expression output = F(x) + x, where x represents an input to the residual block and output from the preceding layer. F(x) corresponds to a Convolutional Neural Network (CNN) component comprising multiple convolutional blocks.

According to that, this method enhances the smoothness of the gradient flow during the backpropagation process, allowing the network to effectively handle a larger number of layers, ranging from 50 to 150. Bypassing a link does not impose any extra processing burden on the network.

Conversely, the practice of integrating the input from a previous layer into the output of a succeeding layer has become very common. This method has been effectively applied in several neural network topologies, including UNet and Recurrent Neural Networks (RNN).

Variants of ResNet:

1. ResNeXt: ResNeXt is a modified model derived from the original ResNet architecture. It utilizes the building block shown below.

The ResNeXt model integrates the ResNet approach of layer repetition while introducing a flexible and straightforward method for implementing the split, transform, and merge techniques. The building block resembles the Inception network, facilitating a range of transformations, including 1×1 Conv, 3×3 Conv, 5×5 Conv, and MaxPooling operations. Contrary to the Inception model, the ResNeXt model incorporates the addition and merging of layered transformations, which appears in Figure 14.

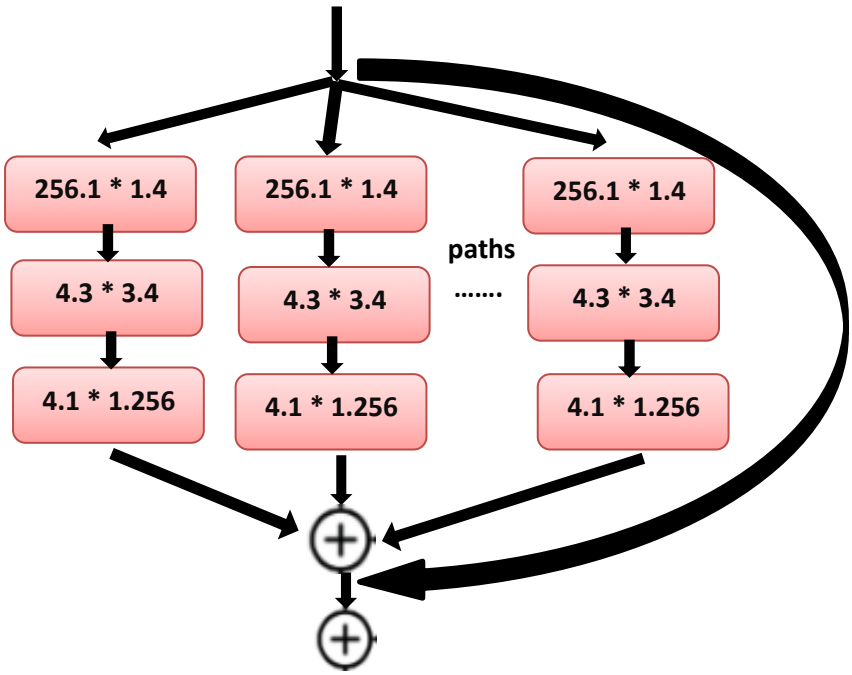


Figure 14 the building block

This model introduces an additional measure of cardinality known as the independent path number. Additionally, it encompasses the current measurements of depth and height. The authors empirically illustrated the significance of incorporating an additional dimension to enhance precision. Enhanced cardinality facilitates the expansion of the network and its depth, particularly when the width and depth dimensions yield diminishing benefits for conventional models.

ResNeXt, a model compared to the Inception network, is easier to train due to its ability to train over multiple datasets and its single adjustable hyperparameter, see Figure 15.

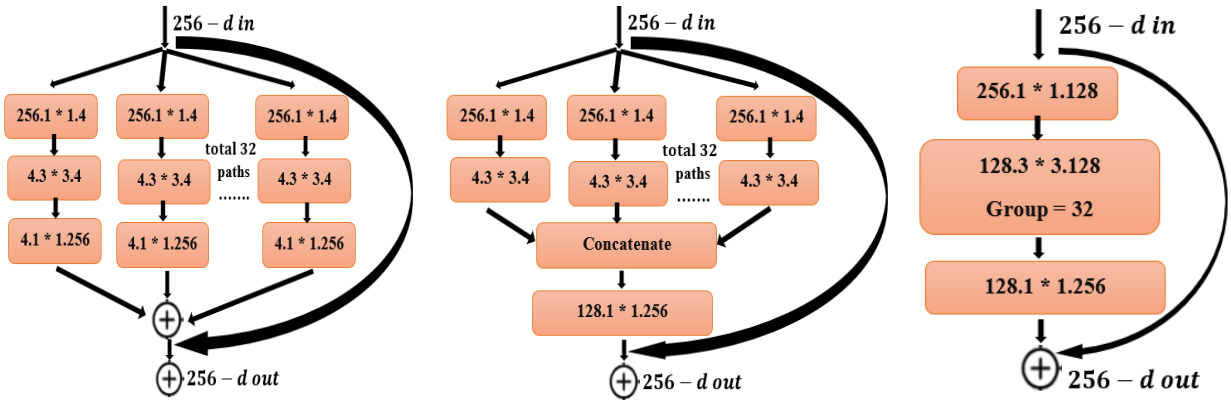


Figure 15 The proposed ResNeXt building block consists of three versions.

Wide ResNet:

Wide ResNet is an enhanced iteration of the original Deep ResNet model. Users can expand or reduce the network's breadth or depth without affecting performance. Network accuracy is not improved by increasing depth. The model, initially presented in a 2016 publication titled "Wide Residual Networks," has subsequently undergone revisions, such as in 2017.

Deep residual networks provide exceptional precision and can facilitate the efficient execution of a machine's diverse tasks, including image recognition. Nevertheless, deep networks persistently encounter challenges from vanishing or expanding gradients and overall degradation. A deep ResNet does not ensure that all residual blocks are included in operations.

Research by (Mustapha et al., 2020) indicates that the model may omit several blocks, or a limited number of residual blocks may make it to the larger contributing block. To address this problem, the researcher of Wide ResNet considered turning off random blocks. Research has concluded that a wider network can be more efficient than a narrower one for these tests.

The Wide ResNet architecture comprises a sequence of ResNet blocks, each adhering to the Batch Normalization-ReLU-Conv structure. The visual representation of this configuration is provided in the table 5 below:

Table 5 The Wide ResNet

Group name	Output size	Block type = B(3.3)
Conv1	32×32	$[3 \times 3, 16]$
Conv2	32×32	$[3 \times 3, 16 \times k\ 3 \times 3, 16 \times k] \times N$
Conv3	16×16	$[3 \times 3, 32 \times k\ 3 \times 3, 32 \times k] \times N$
Conv4	8×8	$[3 \times 3, 64 \times k\ 3 \times 3, 64 \times k] \times N$
Avg-pool	1×1	$[8 \times 8]$

According to (Ikechukwu et al., 2020), A wide ResNet is composed of five groups or layers, like Figure 16. The contested residual block is of type B(3,3). In all networks, the conv1 group remains unchanged, whereas the other convolutional groups differ according to the k value that specifies the breadth of the network. An average-pool layer and a classification layer are positioned after

the convolutional group. These variable measures can be adjusted to provide various representations of the residual blocks:

1. The convolution type used in the previous example is B(3, 3), which is composed of 3×3 convolutional layers. One such alternative is to combine. 3×3 convolutions with a 1X1 convolutional layer.
2. The number of convolutions per block is a statistic that assesses the depth of the block based on its impact on the model's performance.
3. Regularly monitoring the depth and width of each residual block is crucial for assessing performance and computational complexity.
4. Dropout layer—To mitigate overfitting, including a dropout layer after each convolution in every residual block is advisable.

5.

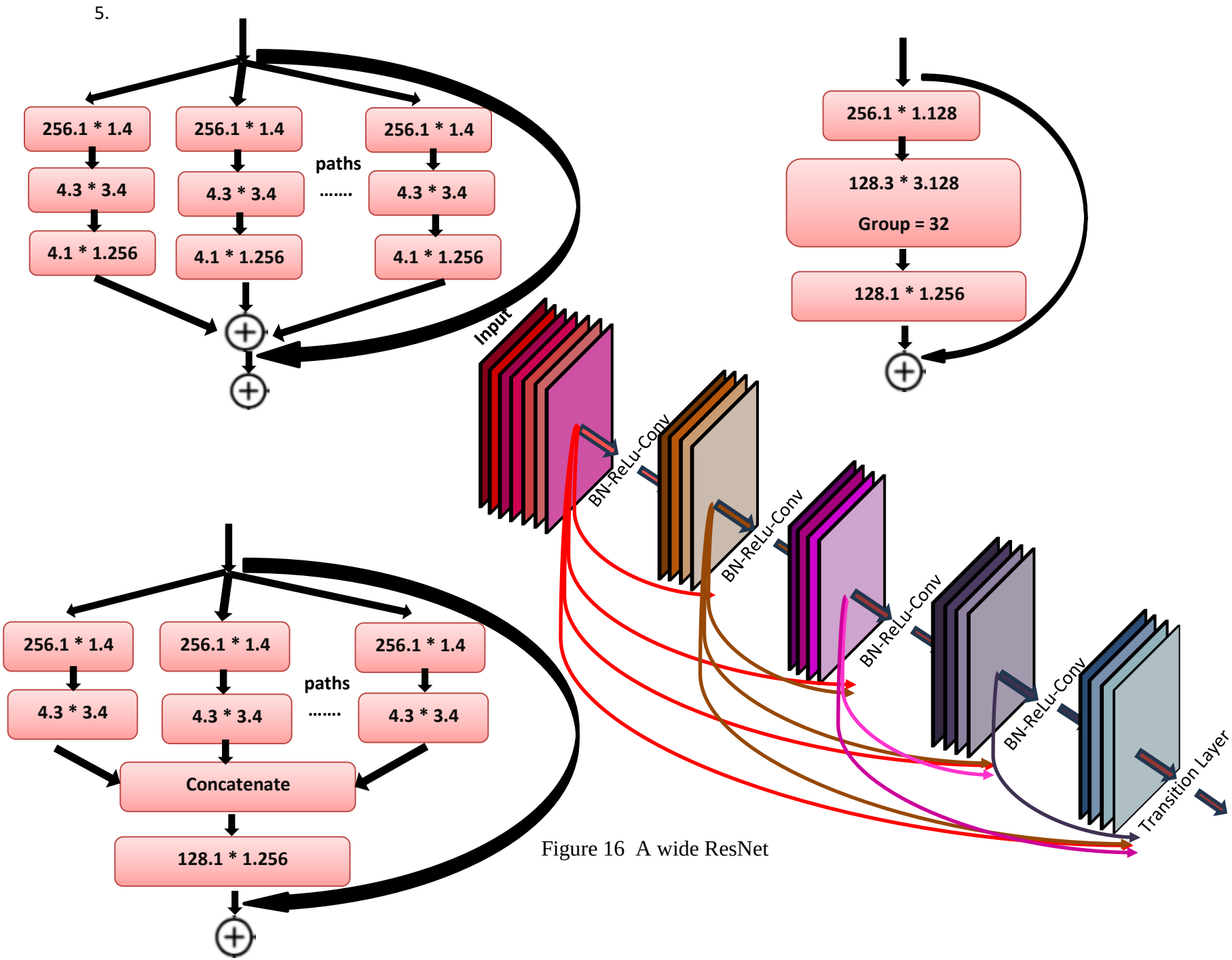


Figure 16 A wide ResNet

3.5.5 GoogLeNet

In 2014, the research paper "Going Deeper with Convolutions" introduced Google Net, also known as Inception V1. Researchers at Google developed this groundbreaking neural network architecture in partnership with several universities. This design emerged as the victor in the ILSVRC 2014 image categorization contest. It has resulted in a substantial reduction in error rate compared to previous winners AlexNet (the winner of ILSVRC 2012) and ZF-Net (the winner of ILSVRC 2013), and a much lower error rate than VGG (the runner-up in 2014). This architecture employs 1×1 convolutions within the central part of the framework and global average pooling as detailed by (Militante et al., 2020)

Features of GoogLeNet:

The GoogLeNet architecture differs greatly from previous state-of-the-art architectures, such as AlexNet and ZF-Net. The model utilizes many approaches, such as 1×1 convolution and global average pooling, to develop a more intricate architecture. In the realm of architecture, we will examine many techniques.

A 1×1 convolution refers to a convolutional operation using a filter size of 1×1. This operation is commonly used in convolutional neural networks to perform dimensionality reduction. .notably, The purpose of these convolutions is to decrease the number of parameters (weights and biases) in the structure. Simultaneously, reducing the parameters increases the depth of the structure. Now, let's analyze an example of a 1×1 convolution.

For example, if we want to perform a 5×5 convolution with 48 filters without using a 1×1 convolution as an intermediate step, see Figures 17 and 18:

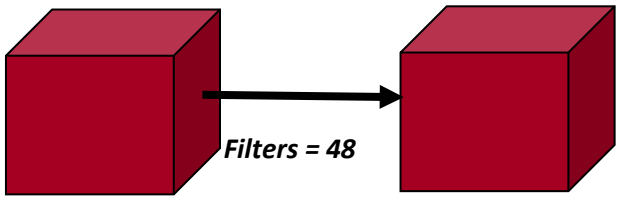


Figure 17 a 5×5 convolution with 48 filters

The total number of operations is calculated as (14 x 14 x 48) x (5 x 5 x 480) = 112.9 M.

Using a 1×1 convolution:

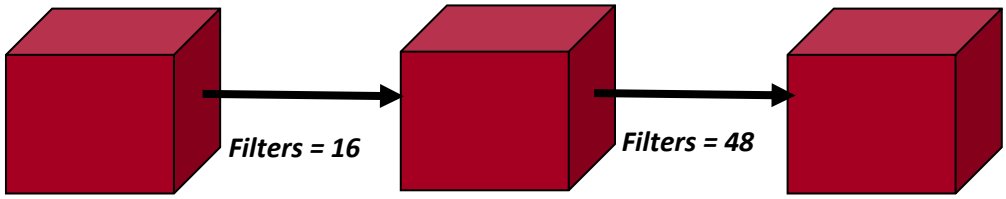


Figure 18 Google operations

The total number of operations is calculated as: (14 x 14 x 16) x (1 x 1 x 480) + (14 x 14 x 48) x (5 x 5 x 16) = 1.5M + 3.8M = 5.3M which is much smaller than 112.9M.

Global Average Pooling:

Previously implemented architectures, including AlexNet, utilized entirely connected layers at the network's end. In numerous architectures, most parameters are located in these entirely connected layers, which raises the computation cost.

The GoogLeNet architecture incorporates a technique known as global average pooling, which is applied after the network. This layer performs spatial average pooling on a feature map with dimensions of 7×7, resulting in a downscaled feature map with dimensions of 1×1. Additionally,

this reduces the count of trainable parameters to zero and results in a 0.6% enhancement in the top-1 accuracy.

Inception Module:

The inception module differs from previous architectures like AlexNet and ZF-Net by implementing a fixed convolution size for each layer.

Based on the work of (Sharma & Guleria, 2024), it can be observed that In this module, parallel operations are done on the input using 1×1, 3×3, and 5×5 convolutions, as well as 3×3 max pooling. Figure 19 shows that these operations' outputs are piled together to produce the final result. The concept is that convolution filters of varying sizes will effectively address objects at various scales.

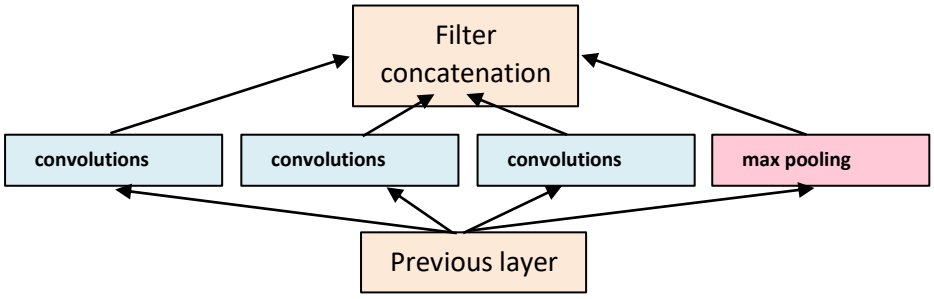


Figure 19 Inception module, naïve version

Auxiliary Classifier for Training:

The inception architecture incorporates intermediate classifier branches inside its structure, exclusively used for training. The branches comprise a 5×5 average pooling layer with a stride of 3, a 1×1 convolution layer with 128 filters, two fully connected layers with 1024 and 1000 outputs, respectively, and a softmax classification layer. The generated loss of these layers contributed to overall loss with a weight of 0.3. These layers effectively address the gradient vanishing issue and contribute to regularization.

Model Architecture:

The architectural features of GoogleNet are presented below and organized in a Layer-by-Layer manner. See Figure 20 and Table 6.

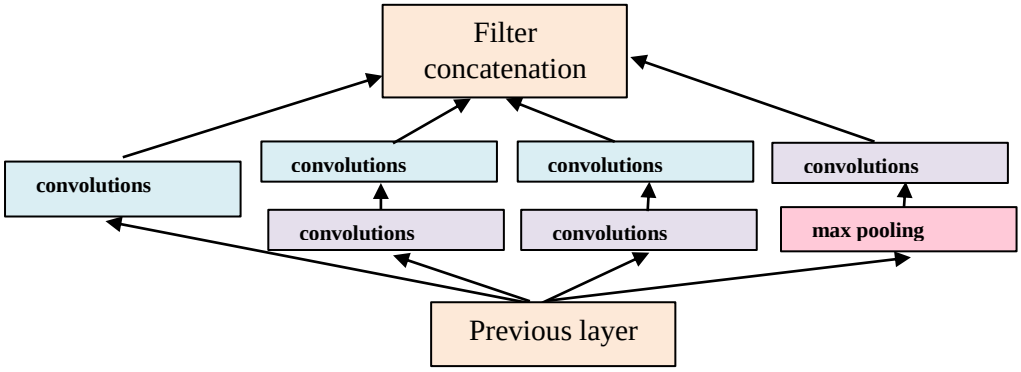


Figure 20 Inception module with dimension reductions

Table 6 shows the architectural features of GoogleNet

type	Patch size /stride	Output size	Depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	Pool proj	params	ops
convolution	7×7 / 2	112×112×64	1							2.7k	34M
Max pool	3×3 / 2	56×56×64	0								
convolution	3×3 / 1	56×56×192	2		64	192				112k	360M
Max pool	3×3 / 2	28×28×192	0								
Inception(3a)		28×28×256	2	64	96	128	16	32	32	159k	128M
Inception(3b)		28×28×480	2	128	128	192	32	96	64	380k	304M
Max pool	3×3 / 2	14×14×480	0								
Inception(4a)		14×14×512	2	192	96	208	16	48	64	364k	73M
Inception(4b)		14×14×512	2	160	112	224	24	64	64	437k	88M
Inception(4c)		14×14×512	2	128	128	256	24	64	64	463k	100M
Inception(4d)		14×14×528	2	112	144	288	32	64	64	580k	119M
Inception(4e)		14×14×832	2	256	160	320	32	128	128	840k	170M
Max pool	3×3 / 2	7×7×832	0								
Inception(5a)		7×7×832	2	256	160	320	32	128	128	1072k	54M
Inception(5b)		7×7×1024	2	384	192	384	48	128	128	1388k	71M
Avg pool	7×7 / 1	1×1×1024	0								
Dropout(40%)		1×1×1024	0								
Linear		1×1×1000	1							1000k	1M
softmax		1×1×1000	0								

The architecture has a total depth of 22 layers. This architecture was created with a focus on optimizing computational effectiveness. The basic idea is that the architecture can run on individual devices, even those with limited processing capacity. It also includes two auxiliary classifier layers coupled to the output of the Inception (4a) and Inception (4d) layers.

The architectural specifics of auxiliary classifiers are as follows:

1. The pooling layer has an average filter size of 5x5 and a stride of 3.

2. The Rectified Linear Unit (ReLU) function reduces the dimension through a 1x1 convolution with 128 filters.

3. The layer is fully connected, with 1025 outputs and ReLU activation.

4. The dropout regularization process.

5. A softmax classifier with 1000 classes.

This architectural framework acquires photos of 224x224 pixels magnitudes by RGB color channels of three dimensions. All the convolutions used within this design are the activation functions ReLu referred to as Rectified Linear Units.

Results:

From the detailed investigation presented by (Umar et al., 2022), it is apparent that GoogLeNet emerged as the champion at ILSRVRC 2014, securing the top position in both the classification and detection tasks. The categorization task has an error rate of 6.67%, which ranks it among the top 5. A collection of six GoogLeNets achieves a mean average precision (mAP) of 43.9% on the ImageNet test set.

3.6 Matrices

3.6.1 What Is a Confusion Matrix?

A confusion matrix is a deep learning technique to evaluate a classification model's performance by showing its accuracy. The output includes the counts of true positives, true negatives, false positives, and false negatives. This matrix assists in evaluating model performance, detecting misclassifications, and enhancing forecast precision.

A confusion matrix is a square matrix of size N x N used to evaluate the performance of a classification model, where N represents the total number of classes being predicted. The matrix compares the actual target values with the predicted values generated by the deep learning model. As highlighted by (Lui et al., 2018) in their comprehensive analysis, This offers a thorough outlook on the effectiveness of our categorization model and the specific types of mistakes it produces. In a binary classification task, a 2 x 2 matrix is utilized, comprising four values as depicted table 7below:

Table 7 In a binary classification

		ACTUAL VALUES	
		POSITIVE	NEGATIVE
PREDICTED VALUES	POSITIVE	**TP**	**FP**
	NEGATIVE	**FN**	**TN**

The matrix consists of columns representing actual values and rows representing predicted values of the target variable.

3.6.2 Important Terms in a Confusion Matrix:

- True Positive (TP): where the positive predicted value is the same as the actual sample.
- True Negative (TN): where the negative predicted value is the same as the actual sample.
- False Positive (FP): where a bad issue has been found, but the prediction is that the end value is good.
- False Negative (FN) created a positive but negative prediction error.

For example, A classification dataset with 1000 data points is fitted with a classifier like logistic regression or decision tree, resulting in the below Table 8:

Table 8 is a binary classification example

		ACTUAL VALUES	
		POSITIVE	NEGATIVE
PREDICTED VALUES	POSITIVE	**560**	**60**
	NEGATIVE	**50**	**330**

The Confusion matrix will consist of several values:

- The True Positive (TP) value is 560, indicating that the model accurately categorized 560 data points from the positive class.

- The True Negative (TN) value is 330, indicating that the model accurately identified 330 data points in the negative class.
- The False Positive (FP) value is 60, indicating that the model erroneously identified 60 negative class data points as the positive class.
- The False Negative (FN) value is 50, indicating that the model misclassified 50 positive class data points as negative class.

According to the detailed research conducted by (Luque et al., 2019), his classifier performed well on our dataset because of the higher number of true positive and negative values.

3.6.3 How to Calculate Confusion Matrix for a 2-class classification problem?

To compute the confusion matrix for a binary classification task, you must be aware of the following:

- True positives (TP) refer to the samples accurately identified as positive.
- True negatives (TN) refer to the samples accurately predicted as negative.
- False positives (FP) refer to the samples that were inaccurately classified as positive.
- False negatives (FN) refer to the number of samples that were inaccurately classified as negative.

In addition, the confusion matrix is utilized to compute various measures like accuracy, precision, recall, and F1 score as reported by (Luque et al., 2019).

Precision vs. Recall:

Precision measures the proportion of true positive cases among all anticipated positive cases. We can determine the Precision by equation 13:

$$Precision = \frac{TP}{TP + FP} \qquad (13)$$

This will ascertain the reliability of our model.

Recall measures the proportion of true positive cases that our model successfully identified. Here is the method to calculate Recall using the following equation 14:

$$Recall = \frac{TP}{TP + FN} \qquad (14)$$

Precision is valuable when minimizing false positives, which is more important than minimizing false negatives. Moreover, it is crucial in music or video recommendation systems, e-commerce websites, etc. Incorrect outcomes may result in customer attrition and adversely affect the company

in line with the conclusions of (Chicco et al., 2021)

Recall is a valuable metric where minimizing False Negatives is more important than minimizing False Positives. In medical instances, avoiding missing true positive cases is crucial, even if it means raising a false alarm.

There will be situations where it is unclear whether Precision or Recall is more crucial.

3.6.4 What Is F1-Score:

According to equation 15, the F1-score effectively captures both trends in a single value, indicating that increasing model precision leads to decreased recall and vice versa.

$$F1 - score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} \qquad (15)$$

The F1-score is a harmonic means of Precision and Recall, providing a comprehensive assessment of both measurements. The greatest value occurs when Precision equals Recall.

(Grandini et al., 2020) notes that there is a caveat. The F1-score's interpretability is low. Therefore, whether our classifier is optimized for precision or recall is unclear. We use it with other assessment criteria to comprehensively understand the outcome.

Chapter 4

Experiments and Results

4.1 Introduction

This chapter reviews the practical steps used to process the data. Following the preprocessing step, many machine and deep learning techniques designed for image processing and recognition tasks are utilized for classification. The classification learner tool in Jupyter Notebook Anaconda is utilized for classification, where the selection of techniques is contingent upon the specific category. The most appropriate approaches are then chosen for the classification process.

Jupyter in Anaconda is used on the Apple MacBook Pro "M2 Pro." The Apple MacBook Pro "M2 Pro" has 8GB RAM and 256G SSD m2. The device gives results in a time of more than 6 hours. Five techniques have been applied: CNN, DenseNet, VGG16, ResNet, and InseptionNet - GoogleNet. The classification results of the five algorithms will also be presented and compared to identify which algorithm has the best performance in classification with 10 and 50 training epochs. These particular epoch values were chosen to compare the performance of the algorithms with small and long training tasks. 10 epochs is a decision based on an early evaluation to check a system's initial learning behavior and avoid overfitting, 50 epochs provide a comprehensive learning of the patterns in chest X-ray.medical images. The aim is to select the best performing deep learning model, which may be the most accurate with the least computational complexity. This comparison will indicate which one of these approaches is more appropriate for accurate classification and prediction in our context..

4.1 Dataset Description

The dataset consists of chest X-rays collected from healthy people compared to those with chest ailments, whether viruses or bacteria cause them. Below are examples of X-rays that illustrate both possibilities.

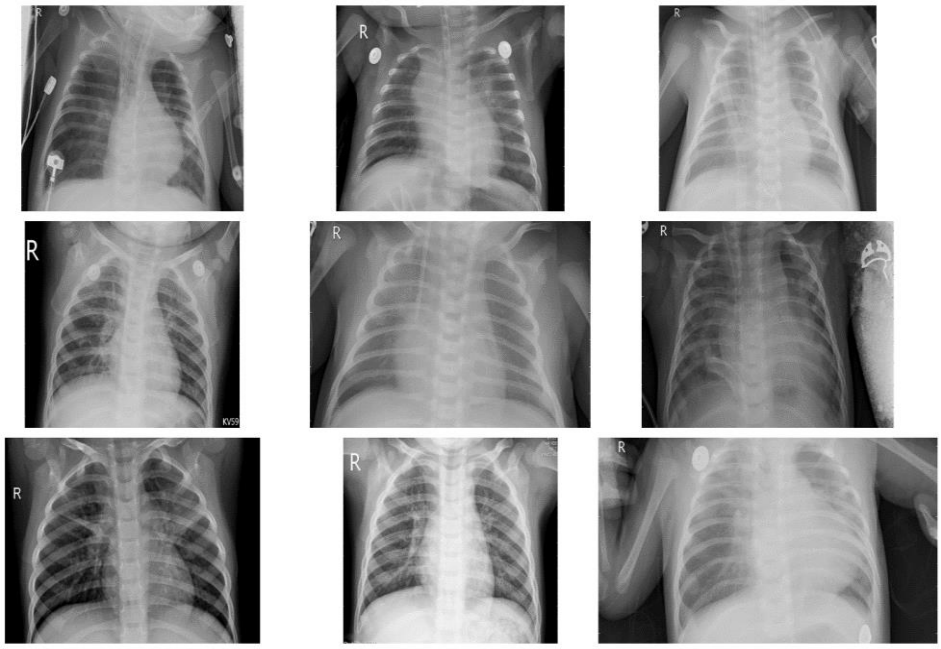


Figure 21 Unhealthy Lung

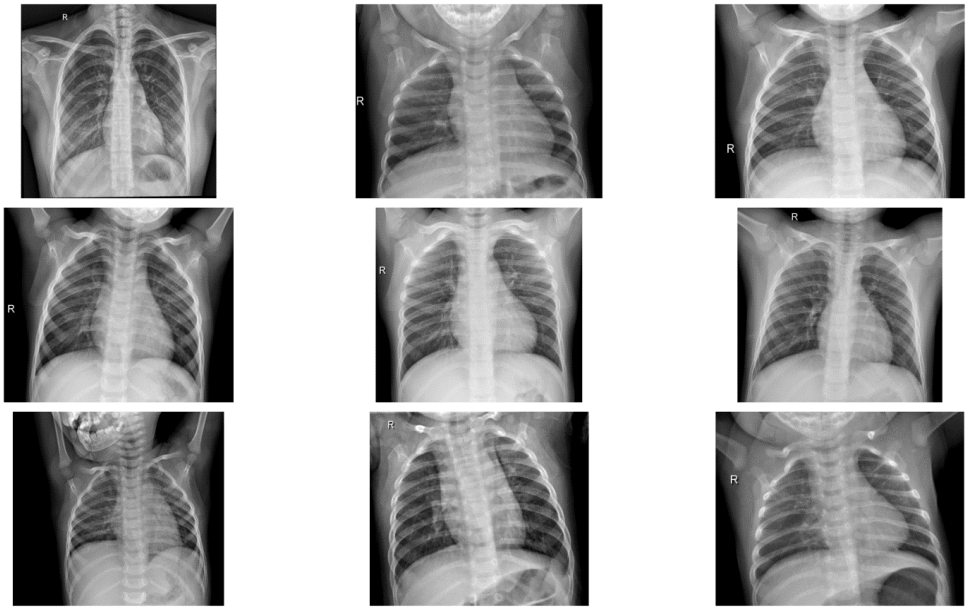


Figure 22 Healthy Lung

This classification offers valuable insights to physicians when they examine or assess chest radiographs containing abnormal findings. Thus, radiologists can improve the accuracy of diagnosis and the quality of treatment. When put together, the data becomes vital for aiding systems in training and evaluating AI algorithms in image analysis and standalone diagnosis. By deftly scrutinizing the pixel intensity distribution, it became obvious that a great variation was shown in Figure 22, unlike the dataset. Through this discovery, it is emphasized why efficient preprocessing is required to detect and correct distortion before holding homogeneity of data. Each of these hurdles calls for outstanding problem consideration and resolution in the processing phase training convolutional neural networks (CNNs) for vision analysis applications. The preprocessing would eliminate the inconsistency problem, allowing for more precise pattern recognition and detail detection to achieve accurate classification and interpretation.

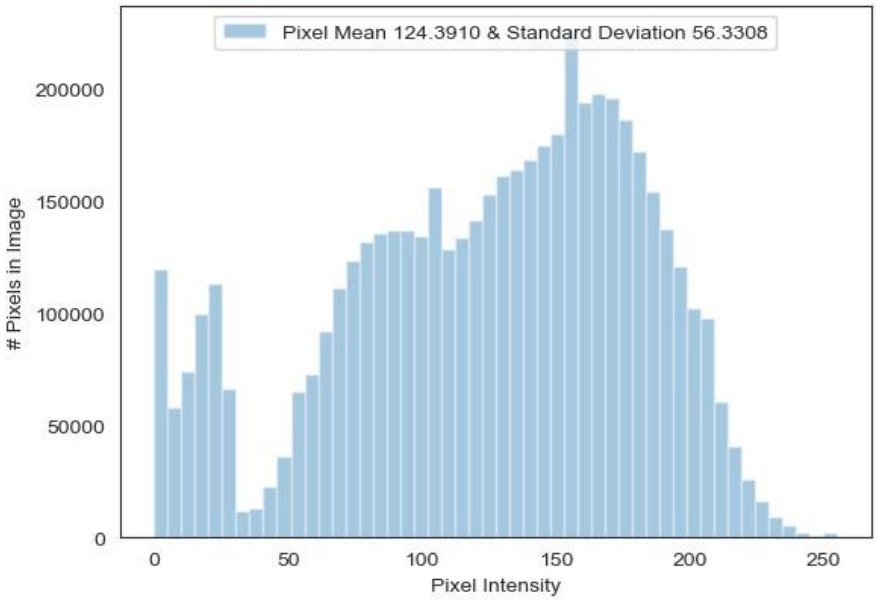


Figure 23 Distribution of Pixel Intensities in the Image before preprocessing.

Figures 25 and 26 below are the charts that reflect the images taken before and after processing. These visual representations help the reader better understand the major effects of one's daily life processing treatments. The photographs' superiority and grade will become obvious when you see the originals and processed images. Through this comparison, we have established the efficiency of the processing strategies in ensuring that the images are fully improved and useful for convolutional neural networks (CNNs). Visualization reflects the role of image processing in refinement, which aids in how efficiently the data is read and analyzed in AI.

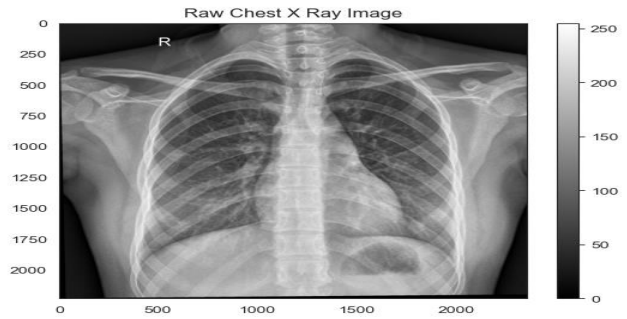


Figure 24 The Image Before Preprocessing Process.

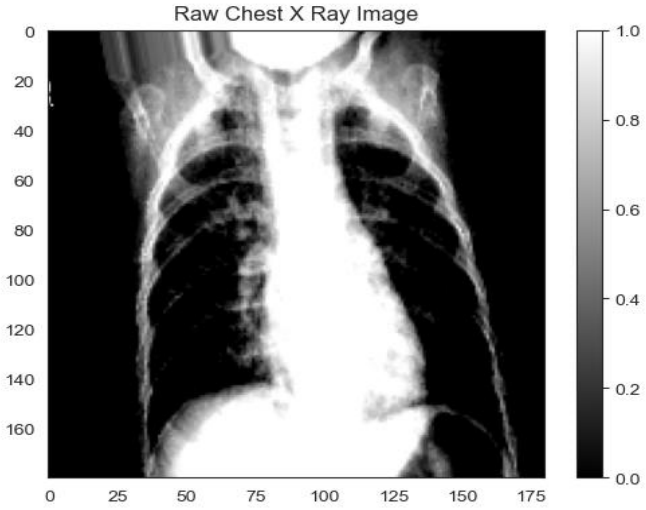


Figure 25 The Image After Preprocessing Process.

As one looks at the histogram of the pixel density, where a horizontal graph demonstrates a decline in the distribution spread, it becomes clear that there is a disparity in data collection. Hence, there is a higher degree of clustering information when cutting across the pixel value of the concentrated data distribution than in the previous sample. Progressive down sampling indicates the network becomes concentrated and performs better in image analysis tasks to improve accuracy. Reducing the data dispersion remains highly significant in considering the efficiency of DL techniques and in making it possible to accurately characterize healthy traits in both academic and real applications.

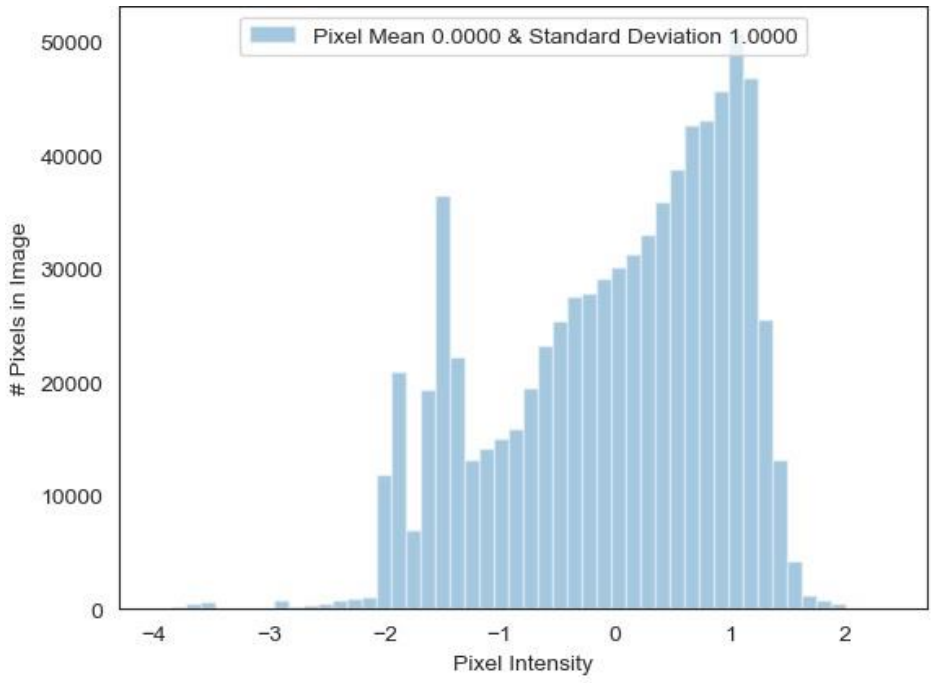


Figure 26 Distribution of Pixel Intensities in the Image After Preprocessing.

4.2 Deep Learning Techniques Results

The results in this section are evaluated using deep learning algorithms to identify pneumonia picture datasets and categorize them into two classes: Pneumonia and Normal. The measurement criteria discussed in the previous chapter are employed to evaluate the rating approaches. Scoring scales provide a way to assess the quality of the administered method and its effect. Classification matrices are based on the quality of deep learning classes applied across every scenario. After processing all the images in the data set, they were used in three stages to train the different neural networks: the roles of training, the validation message, and the testing uses.

Five famous types of neural networks were chosen for the image data set: Conventional Neural Network (CNN), DenseNet, VGG-16, ResNet, and Googlenet DL neural networks. After that, the number of training times was changed (10, 50) respectively to compare the results of the algorithms. In addition to observing what happens to the results when the number of training times increases. Test samples were used to compare the ability of these algorithms to predict after the network was built using training and validation data. This section reviews the different results when applying deep learning techniques specialized in detection and classification, providing valuable insights for the practical application of these algorithms in the medical field.

4.2.1 CNN Experiment Result

CNN technology is used to detect pneumonia images and classify them into a normal or pneumonia image. The input variables are 624, and the number of training iterations is 10.

The traditional neural network was applied according to the network architecture attached below:

To know the network's classification ability, we can use the following flow table 9 :

Table 9: CNN classification matrices for ten training iterations.

TRUE LABEL	normal	150	43
	pneumonia	84	347
		normal	Pneumonia
		PREDICTED LABEL	

From the previous table, we perform accuracy calculations and other parameters and extract the results as the following Table10 shows:

Table 10: CNN classification matrices for ten training iterations.

	Class 0	Class 1
TP	150	347
TN	347	150
FP	84	43
FN	43	84
(Recall)	0.78	0.81
False Negative Rate	0.22	0.19
(Precision)	0.64	0.89
False Discovery Rate	0.36	0.11
Specificity	0.81	0.78
FPR = (1-Specificity)	0.19	0.22
Accuracy	0.80	
F1 score	0.70	0.85
Macro-F1	0.77	
Macro-(Recall)	0.79	
Macro-(Precision)	0.77	
weighted-f1	0.80	
weighted-Recall	0.80	
weighted-Precision	0.81	

The algorithm attained a classification accuracy of 80%. Upon further analysis of supplementary classification measures, it is found that the Macro-F1 value is 0.77, and the weighted F1-score is 0.80. On the other hand, the F-score with weighted measurements is higher than the macro-F1 value, indicating that the classifier algorithm successfully and correctly sorts class 1, which is more important than class 2. The sensitivity and specificity numbers for each class are also notable. In the first class, the recorded values for sensitivity and 1-specificity are 0.78 and 0.19, respectively. In the second class, these values are 0.81 and 0.22, respectively.

Also, figure 27 below shows the accuracy and loss tracked over ten epochs in the CNN algorithm.

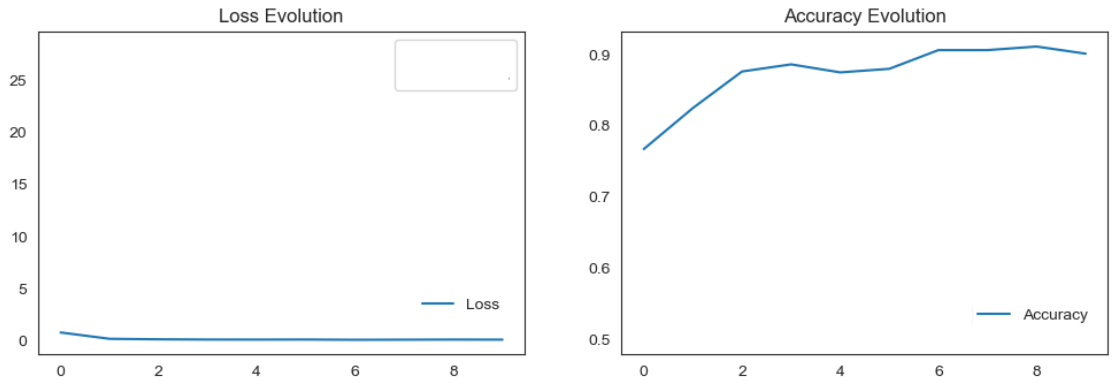


Figure 27 CNN Accuracy and Loss Evolution Carves for ten iterations.

On the other hand, we will review Table 11, the algorithm's results after increasing the number of training iterations to 50 iterates. Table 12 shows the accuracy result and other parameters as below:

Table 11 CNN classification results for 50 training iterates.

TRUE LABEL	normal	126	12
	pneumonia	108	378
		normal	Pneumonia
		PREDICTED LABEL	

Table 12 CNN classification matrices for 50 training iterates.

	Class 0	Class 1
TP	126	378
TN	378	126
FP	108	12
FN	12	108
(Recall)	0.91	0.78
False Negative Rate	0.09	0.22
(Precision)	0.54	0.97
False Discovery Rate	0.46	0.03
Specificity	0.78	0.91
FPR = (1-Specificity)	0.22	0.09
Accuracy	0.81	
F1 score	0.68	0.86
Macro-F1	0.77	
Macro-(Recall)	0.85	
Macro-(Precision)	0.75	
weighted-f1	0.82	
weighted-Recall	0.81	
weighted-Precision	0.87	

It is important to mention that the classification accuracy is 81 %. Additionally, when examining additional classification matrices, we observe that the Macro-F1 value is 0.0.77, and the weighted-f1 result is 0.82. It is worth mentioning that the weighted-f1 classification values increased compared to the Macro-F1 classification value. This suggests that the system successfully classified the first class, which carries more importance than the others. The values of sensitivity and 1-specificity were as follows: The sensitivity and 1-specificity values for the first class are 0.

91 and 0.22, respectively. For the second class, the values are 0.78 and 0.09. In addition, Figure 28 below displays the tracking path in each iteration, allowing for accuracy and loss observation.

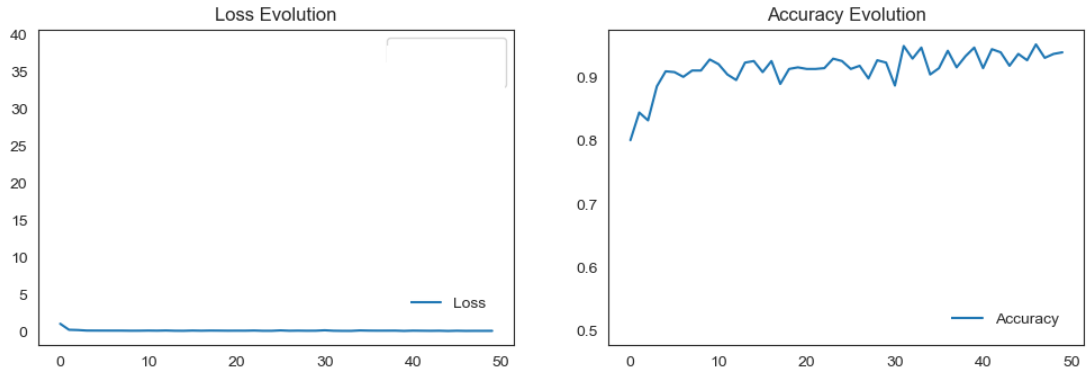


Figure 28 CNN Accuracy and Loss Evolution Carves for 50 training iterates.

Figure 29 below makes it abundantly clear that the results obtained after fifty epochs were only slightly superior to those obtained after ten epochs that came before them. Based on this, it appears that engaging in more repetitions of epochs is not useful.

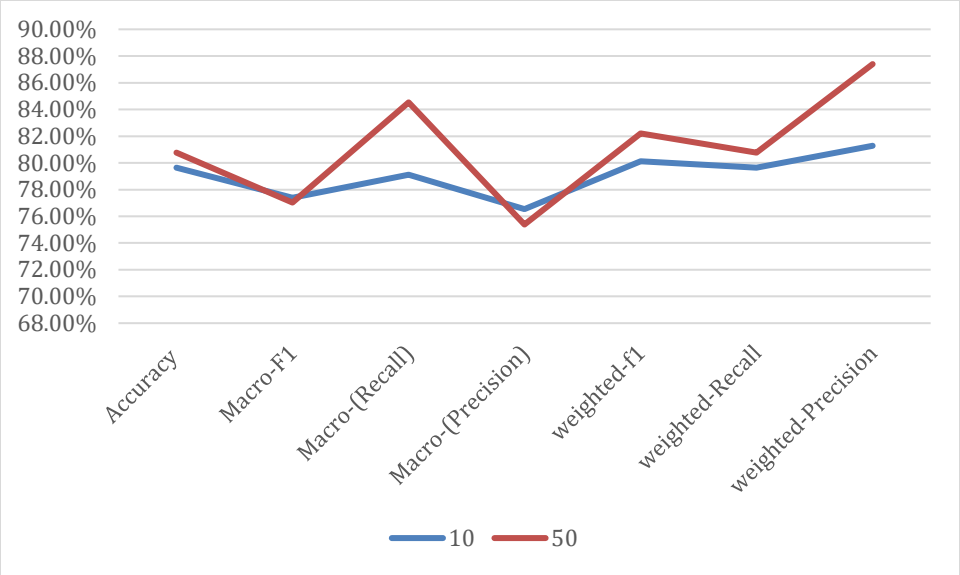


Figure 29 Comparison between 10 epoch and 50 epoch using Classification Matrices

4.2.2 DenseNet Experiment Result

To classify images as either normal or pneumonia, the DenseNet technique is also used to classify the image dataset. Ten training iterations are used, and there are 624 input variables. The traditional neural network was applied according to the network architecture of this technique.

The results from running this technique are shown in the following table 13:

Table 13 DenseNet classification results for ten training iterations.

TRUE LABEL	normal	160	9
	pneumonia	74	381
		normal	Pneumonia
		PREDICTED LABEL	

From the previous table, we perform accuracy calculations and other parameters and extract the results as Table 14 shows:

Table 14 DenseNet classification matrices for ten training iterates.

	Class 0	Class 1
TP	217	320
TN	320	217
FP	17	70
FN	70	17
(Recall)	0.76	0.95
False Negative Rate	0.24	0.05
(Precision)	0.93	0.82
False Discovery Rate	0.07	0.18
Specificity	0.95	0.76
FPR = (1-Specificity)	0.05	0.24
Accuracy	0.86	
F1 score	0.83	0.88
Macro-F1	0.86	
Macro-(Recall)	0.85	
Macro-(Precision)	0.87	
weighted-f1	0.86	
weighted-Recall	0.86	
weighted-Precision	0.87	

The classification's accuracy is 86%, which is worth noting. Furthermore, upon analyzing supplementary classification matrices, we note that the Macro-F1 score is 0.86, whereas the weighted-f1 outcome is 0.86. Notably, the weighted-f1 classification values increased compared to the Macro-F1 classification value. This indicates that the system effectively categorized the first class, which holds greater significance than the other two classes. The values of sensitivity and 1-specificity were as stated: The sensitivity and false positive rate values for the first class are 0.947 and 0.163, respectively. The values for the second class are 0.837 and 0.05. Furthermore, the figures below exhibit the trajectory of the tracking process in each iteration, enabling monitoring accuracy and loss.

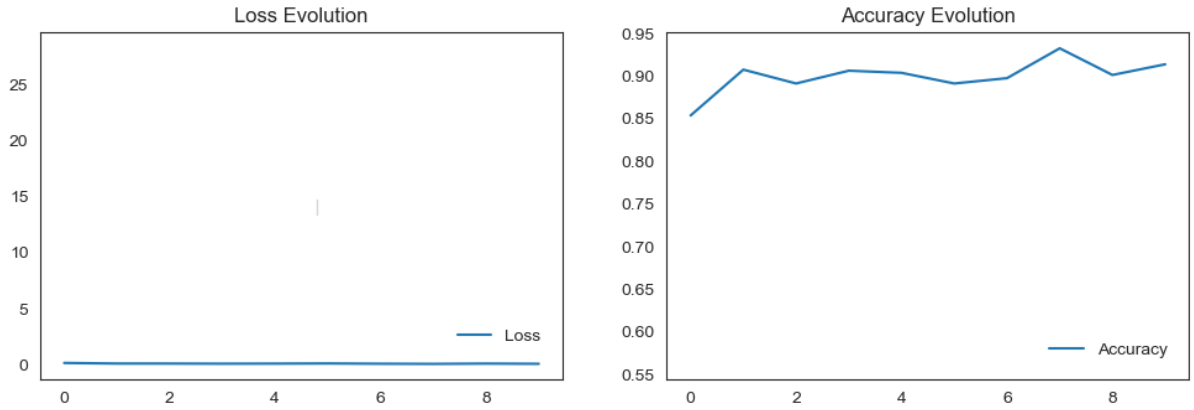


Figure 30 DenseNet Loss and Accuracy Evolution curves for ten iterations.

On the other hand, we will review in Table 4.7 the algorithm's results after increasing the number of training iterations to 50 iterates. Table 4.8 shows the accuracy result and other parameters as below:

Table 15 DenseNet classification results for 50 training iterates.

TRUE LABEL	normal	217	70
	pneumonia	17	320
		normal	Pneumonia
		PREDICTED LABEL	

Table 16 DenseNet classification matrices for 50 training iterates.

	Class 0	Class 1
TP	160	381
TN	381	160
FP	74	9
FN	9	74
(Recall)	0.95	0.84
False Negative Rate	0.05	0.16
(Precision)	0.68	0.98
False Discovery Rate	0.32	0.02
Specificity	0.84	0.95
FPR = (1-Specificity)	0.16	0.05
Accuracy	0.87	
F1 score	0.79	0.90
Macro-F1	0.85	
Macro-(Recall)	0.89	
Macro-(Precision)	0.83	
weighted-f1	0.87	
weighted-Recall	0.87	
weighted-Precision	0.90	

It is crucial to specify that classification accuracy stands at 80.06%. Furthermore, upon scrutinizing further classification matrices, it is evident that the weighted-f1 outcome is 0.858, and the Macro-F1 value is 0.428. It is noteworthy to remark that the weighted-f1 classification values exhibited an increase in comparison to the macro-f1 value. This implies that the system effectively categorized the initial class, which is of greater significance in comparison to the remaining two classes. The sensitivity and one-specificity values were as follows: The sensitivity and specificity values of 0.756 and 0.050 pertain to the first class. The values for the second class are 0.244 and 0.949. In addition, the tracking path for each iteration is illustrated in the figures below, enabling loss and accuracy monitoring.

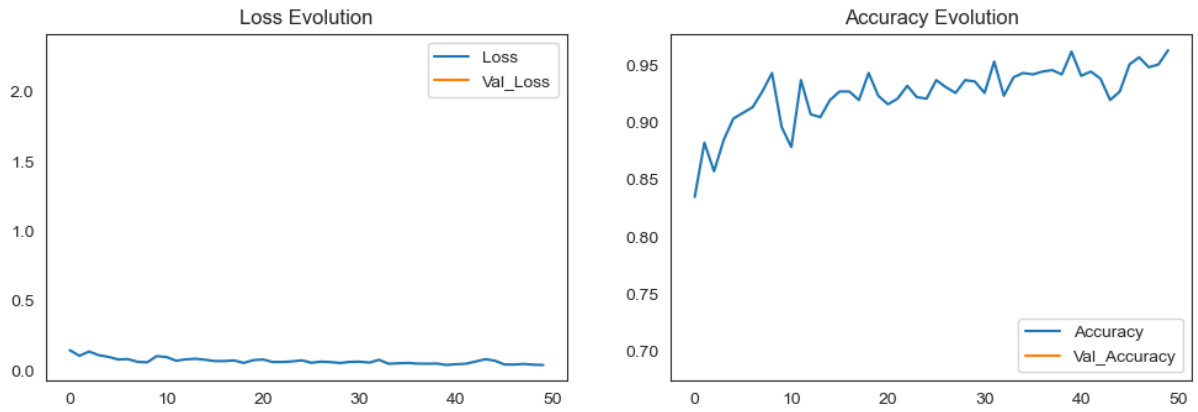


Figure 31 DenseNet Accuracy and Loss Evolution Carves for 50 iterates.

In light of the above graph, it has been established that the model after ten epochs has better results than the model after 50 epochs. The scores of 50 epochs are higher than those obtained after ten epochs, considering accuracy and weighted F1 scores. The Macro F1 score is even greater after fifty epochs. Even though the sample size differences in the classes affect the weighted F1, the average across all classes influences the Micro F1 score. This may occur since the weighted F1 score reflects how many samples are in each class.

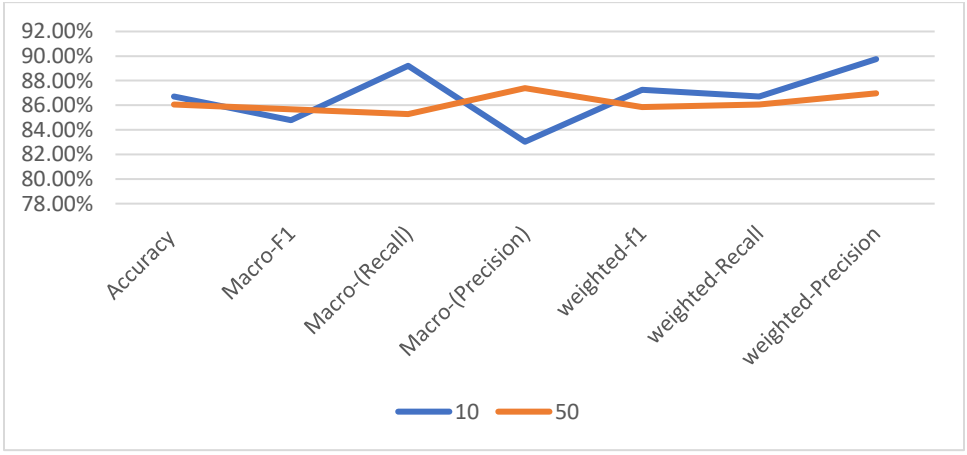


Figure 32 Comparison between 10 epoch and 50 epoch using Classification Matrices

4.2.3 VGG-16 Experiment Result

The VGG-16 method categorizes the picture collection and detects Pneumonia images, determining whether they are Pneumonia or Normal. The input variables are 624 values, and the number of training iterations is set to 10.

The results from running this technique are shown in the following Table 17:

Table 17 VGG-16 classification results for ten training iterations

TRUE LABEL	normal	56	9
	pneumonia	178	381
		normal	Pneumonia
		PREDICTED LABEL	

From the previous table, we perform accuracy calculations and other parameters and extract the results as the following Table18 shows:

Table 18 VGG-16 classification matrices for ten training iterates.

	Class 0	Class 1
TP	56	381
TN	381	56
FP	178	9
FN	9	178
(Recall)	0.86	0.68
False Negative Rate	0.14	0.32
(Precision)	0.24	0.98
False Discovery Rate	0.76	0.02
Specificity	0.68	0.86
FPR = (1-Specificity)	0.32	0.14
Accuracy	0.70	
F1 score	0.37	0.80

Macro-F1	0.59	
Macro-(Recall)	0.77	
Macro-(Precision)	0.61	
weighted-f1	0.76	
weighted-Recall	0.70	
weighted-Precision	0.90	

It is vital to note that the categorization accuracy is 70.%. Furthermore, when we examine additional classification matrices, we find that the Macro-F1 value is 0,59 and the weighted-F1 result is 0.76. It is worth noting that the weighted-f1 classification values rose when compared to the Macro-F1 classification value. This indicates that the system correctly classified the first class, which is more important than the others. The values for sensitivity and 1-specificity were as follows: The sensitivity and 1-specificity values for the first class are 0.86 and 0.31, respectively. The second-class values are 0.68 and 0.14. In addition, the figures below show the tracking path in each iteration, allowing for accuracy and loss observation.

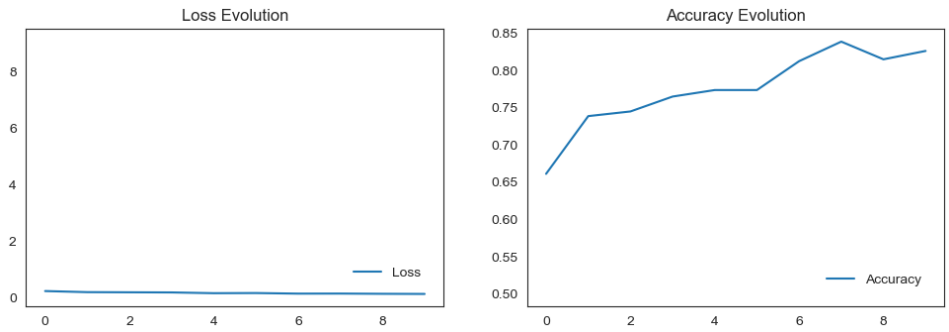


Figure 33 VGG-16 Accuracy and Loss Evolution Carves for ten iterates.

On the other hand, we will review in Table 4.11 the algorithm's results after increasing the number of training iterations to 50 iterates. Table 4.12 shows the accuracy result and other parameters as below:

Table 19 VGG-16 classification results for 50 training iterates.

TRUE LABEL	normal	150	25
	pneumonia	84	365
		normal	Pneumonia
		PREDICTED LABEL	

Table 20 VGG-16 classification matrices for 50 training iterates.

	Class 0	Class 1
TP	150	365
TN	365	150
FP	84	25
FN	25	84
(Recall)	0.86	0.81
False Negative Rate	0.14	0.19
(Precision)	0.64	0.94
False Discovery Rate	0.36	0.06
Specificity	0.81	0.86
FPR = (1-Specificity)	0.19	0.14
Accuracy	0.83	
F1 score	0.73	0.87
Macro-F1	0.80	
Macro-(Recall)	0.84	
Macro-(Precision)	0.79	
weighted-f1	0.83	
weighted-Recall	0.83	
weighted-Precision	0.85	

Not to be overlooked is the 82.53% categorization accuracy. Furthermore, we find that the weighted-f1 result is 0.83, and the Macro-F1 value is 0.80 when looking at various classification matrices. The weighted-F1 classification values increased compared to the Macro-F1 classification value, which is noteworthy. This implies that the system correctly classified the first class, which is more significant than the others. The following were the values of 1-specificity and sensitivity: For the first class, the sensitivity and 1-specificity values are 0. 86 and 0,19 respectively. The results for the second class are 0.81 and 0.14. Furthermore, the tracking path in each iteration is shown in the figures below, enabling accuracy and loss observation. More than that, figure 28 depicts a tracking way by which accuracy and loss are observed.

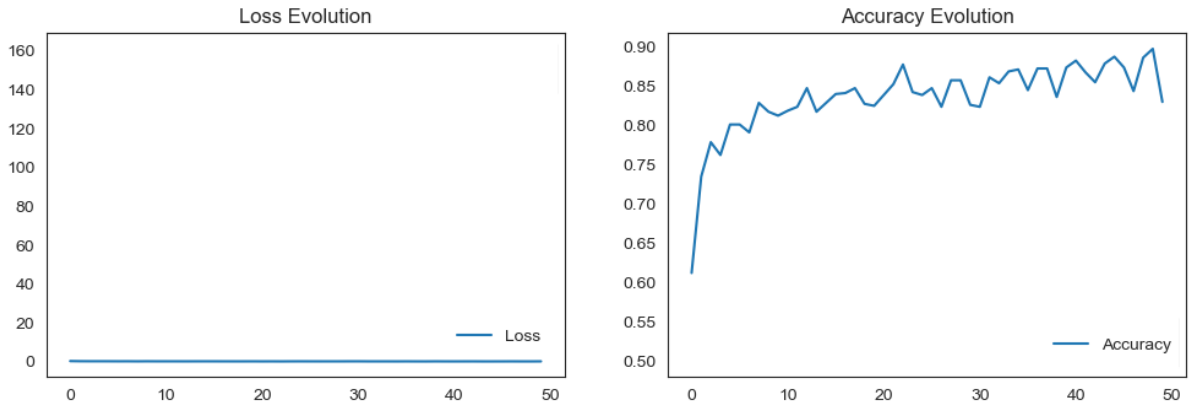


Figure 34 VGG-16 Accuracy and Loss Evolution Carves for 50 iterates.

Based on Figure 35, it is clear that the results acquired after 50 epochs are markedly better than those achieved after only 50 epochs across all criteria. Based on this finding, we may infer that the technique's performance can be improved by increasing the number of training iterations.

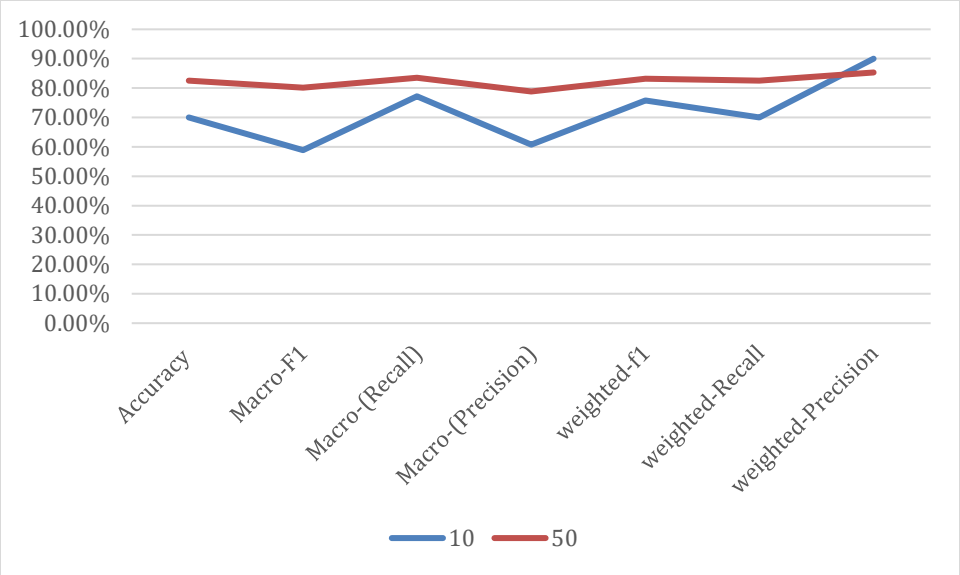


Figure 35 Comparison between 10 epoch and 50 epoch using Classification Matrices

4.2.4 ResNet Experiment Result

The ResNet method, which is also used to classify images of X-ray chests into normal and pneumonia classics, was used to classify the dataset. The model went through a total of ten training iterations and used 624 input variables. What follows is a description of the typical neural network's architecture:

The X-ray chest dataset was categorized using the ResNet technique, a practical tool we trained to detect pneumonia in photos and distinguish between normal and pneumonia images. The algorithm was subjected to 10 training iterations, each with 624 input variables.

The result from running this technique is shown in the following table 21:

Table 21 ResNet classification results for ten training iterations

TRUE LABEL	normal	64	14
	pneumonia	170	376
		normal	Pneumonia
		PREDICTED LABEL	

From the previous table, we perform accuracy calculations and other parameters and extract the results as the following Table 22 shows:

Table 22 ResNet classification matrices for ten training iterates.

	Class 0	Class 1
TP	64	376
TN	376	64
FP	170	14
FN	14	170
(Recall)	0.82	0.69
False Negative Rate	0.18	0.31
(Precision)	0.27	0.96
False Discovery Rate	0.73	0.04
Specificity	0.69	0.82
FPR = (1-Specificity)	0.31	0.18
Accuracy	0.71	
F1 score	0.41	0.80
Macro-F1	0.61	
Macro-(Recall)	0.75	
Macro-(Precision)	0.62	
weighted-f1	0.75	
weighted-Recall	0.71	
weighted-Precision	0.88	

Notably, the categorization accuracy stands at 70 %. We further find that the weighted-f1 result is 0.75, and the Macro-F1 value is 0.61 when looking at several classification matrices. Noteworthy is the rise in the weighted-f1 classification values over the Macro-F1 classification value. This implies that, out of the two classes, the first one is more important and was classified by the algorithm successfully. The following were the values of 1-specificity and sensitivity: First-class and 1-specificity values are 0.31 and 0.82, respectively. In the second class, the result is 0.18 and 0.69. The figures below show the tracking path in every iteration to facilitate further accuracy and loss observation.

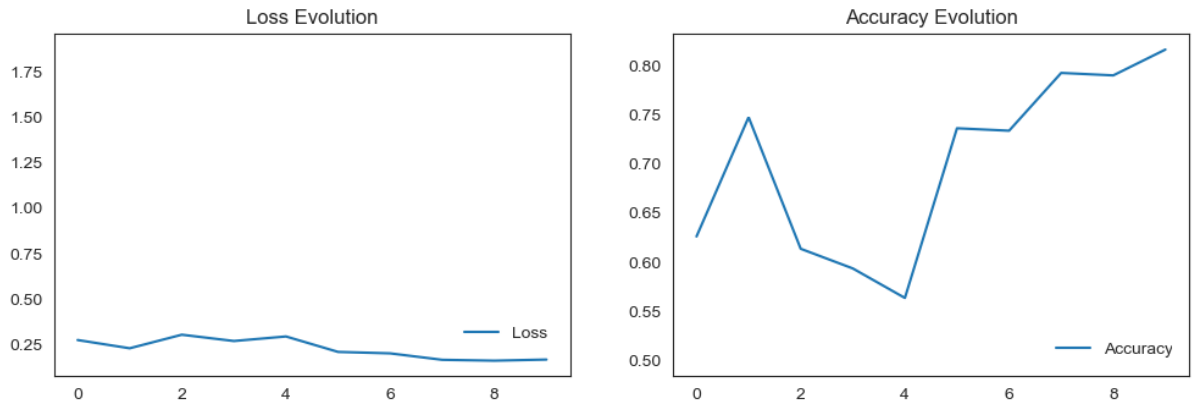


Figure 36 ResNet Accuracy and Loss Evolution Carves for ten iterates.

On the other hand, we will review in Table 23, the algorithm's results after increasing the number of training iterations to 50 iterates. Table 24 shows the accuracy result and other parameters as below:

Table 23 ResNet classification results for 50 training iterations.

TRUE LABEL	normal	48	0
	pneumonia	186	390
		normal	Pneumonia
		PREDICTED LABEL	

Table 24 ResNet classification matrices for 50 training iterates.

	Class 0	Class 1
TP	48.00	390.00
TN	390.00	48.00
FP	186.00	0.00
FN	0.00	186.00
(Recall)	1.00	0.68
False Negative Rate	0.00	0.32
(Precision)	0.21	1.00
False Discovery Rate	0.79	0.00
Specificity	0.68	1.00
FPR = (1-Specificity)	0.32	0.00
Accuracy	0.70	
F1 score	0.34	0.81
Macro-F1	0.57	
Macro-(Recall)	0.84	
Macro-(Precision)	0.60	
weighted-f1	0.77	
weighted-Recall	0.70	
weighted-Precision	0.94	

It is important to mention that the classification accuracy is 70%. Additionally, when examining additional classification matrices, we observe that the Macro-F1 value is 0.57, and the weighted-f1 result is 0.77. It is worth mentioning that the weighted-f1 classification values increased

compared to the Macro-F1 classification value. This suggests that the system successfully classified the first class, which carries more importance than the other two classes. The values of sensitivity and 1-specificity were as follows: The sensitivity and 1-specificity values for the first class are 1 and 0.32, respectively. For the second class, the values are 0.68 and 0. In addition, the figures below display the tracking path in each iteration, allowing for accuracy and loss observation.

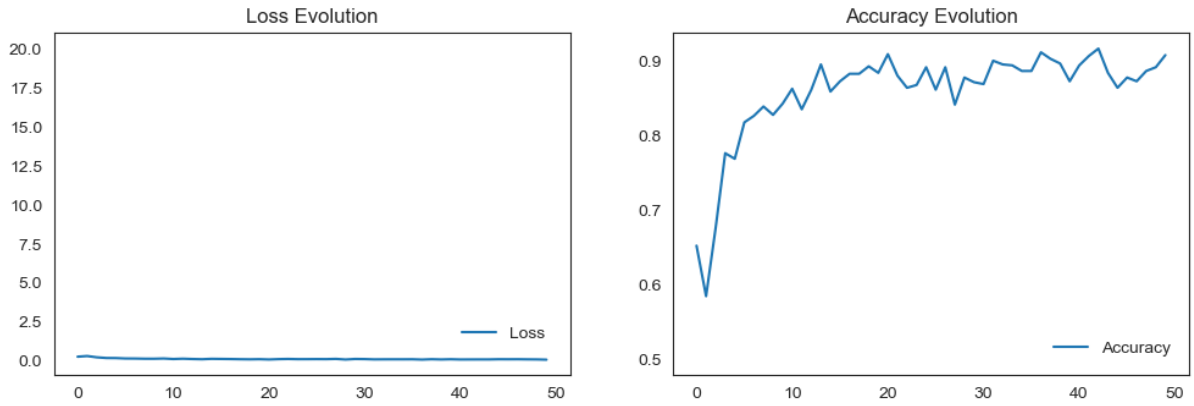


Figure 37 ResNet Accuracy and Loss Evolution Carves

The below figure 38 shows that performance after ten epochs is better than after 50. The accuracy values and weighted F1 score following ten epochs are much greater than those following 50 epochs. That said, the Macro F1 score is higher after 50 epochs.

The macro F1 score was averaged across all classes, while the weighted F1 score was quantified by the number of instances in each class, which led to the difference in scores. Trends suggest that the result after 50 epochs is better than after 10.

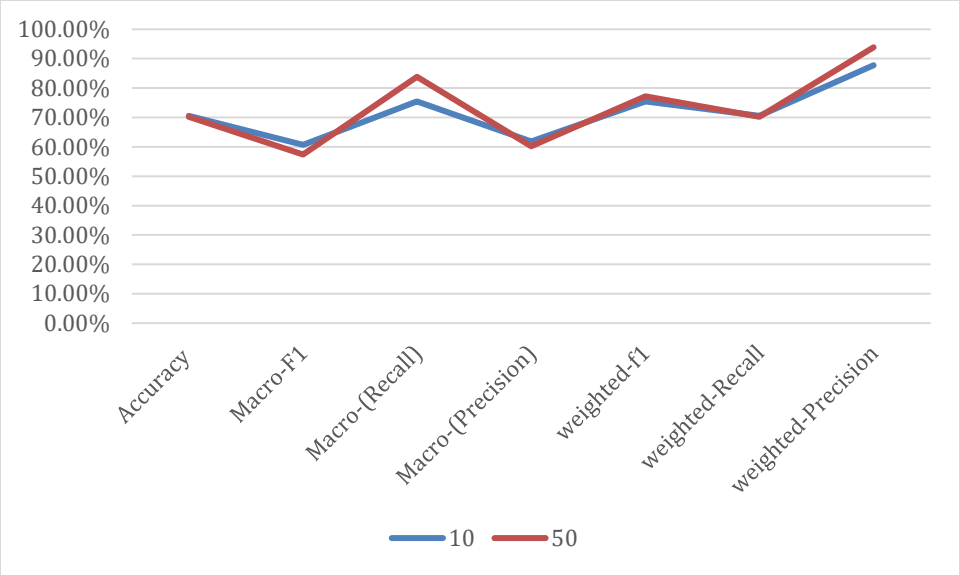


Figure 38 Comparison between 10 epoch and 50 epoch using Classification Matrices

4.2.5 GoogLeNet Experiment Result

Pneumonia images can be detected and classified as normal or pneumonia using the dataset that is classified using the GoogLeNet algorithm. Ten training iterations are used, and there are 624 input variables.

The results from running this technique are shown in the following table 25:

Table 25 GoogleNet classification results for ten training iterations

TRUE LABEL	normal	219	177
	pneumonia	15	213
		normal	Pneumonia
		PREDICTED LABEL	

From the previous table, we perform accuracy calculations and other parameters and extract the results as the following Table 26 shows:

Table 26 GoogLeNet classification matrices for ten training iterates.

	Class 0	Class 1
TP	219	213
TN	213	219
FP	15	177
FN	177	15
(Recall)	0.55	0.93
False Negative Rate	0.45	0.07
(Precision)	0.94	0.55
False Discovery Rate	0.06	0.45
Specificity	0.93	0.55
FPR = (1-Specificity)	0.07	0.45
Accuracy	0.69	
F1 score	0.70	0.69
Macro-F1	0.69	
Macro-(Recall)	0.74	
Macro-(Precision)	0.74	
weighted-f1	0.69	
weighted-Recall	0.69	
weighted-Precision	0.79	

It is worth noting that the classification accuracy stands at 70 %. In addition, when analyzing other classification matrices, we can see that the Macro-F1 value is 0.69, and the weighted-f1 result is 0.69. It's worth noting that the weighted-f1 classification values increased compared to the Macro-F1 classification value. The system appears to have effectively identified the first class, which holds greater significance than the other two. The sensitivity and 1-specificity values were as follows: The sensitivity and 1-specificity values for the first class are 0.55 and 0.0.7, respectively. The values for the second class are 0.93 and 0.45. Furthermore, the figures below provide a visual

representation of the tracking path in each iteration, enabling precise observation of accuracy and loss.

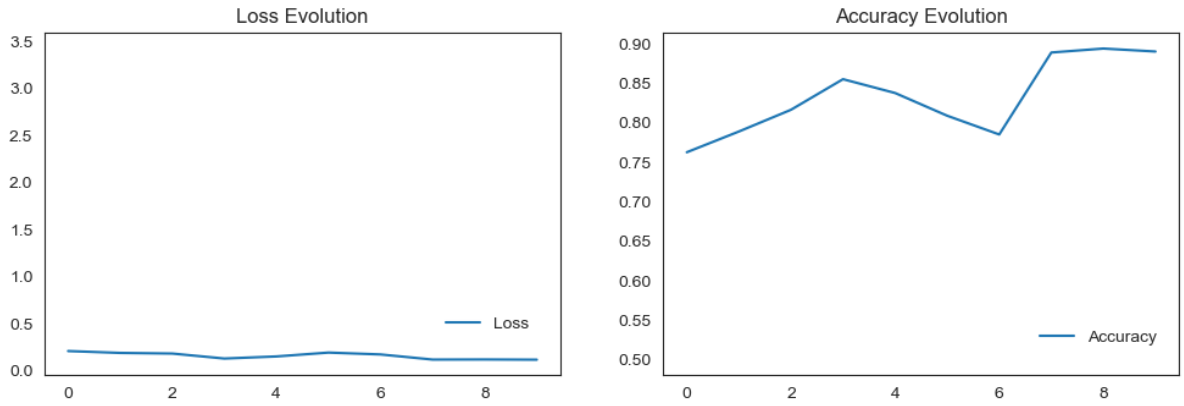


Figure 39 GoogleNet Accuracy and Loss Evolution Carves for ten iterates.

On the other hand, we will review in Table 27, the algorithm's results after increasing the number of training iterations to 50 iterates. Table 28 shows the accuracy result and other parameters as below:

Table 27 GoogleNet classification results for 50 training iterations.

TRUE LABEL	normal	**35**	**4**
	pneumonia	**199**	**366**
		normal	Pneumonia
		PREDICTED LABEL	

Table 28 GoogleNet classification matrices for 50 training iterates.

	Class 0	Class 1
TP	35	366
TN	366	35
FP	199	4
FN	4	199
(Recall)	0.90	0.65
False Negative Rate	0.10	0.35
(Precision)	0.15	0.99
False Discovery Rate	0.85	0.01
Specificity	0.65	0.90
FPR = (1-Specificity)	0.35	0.10
Accuracy	0.66	
F1 score	0.26	0.78
Macro-F1	0.52	
Macro-(Recall)	0.77	
Macro-(Precision)	0.57	
weighted-f1	0.75	
weighted-Recall	0.66	
weighted-Precision	0.93	

Notably, the categorization accuracy stands at %66 percent. Furthermore, when we look at other classification matrices, we find that weighted-f1 is 0.75 and Macro-F1 is 0.52. The weighted-f1 classification values were higher than the Macro-F1 classification value, which is important to note. This indicates that the system correctly identified the most important of the three classifications. The following were the findings for sensitivity and 1-specificity: 0.90 is the sensitivity, and 0.35is the 1-specificity for the first classes. The second group has values of 0.65

and 0.1. Figure 40 below also shows each iteration's tracking path, which helps observe accuracy and loss.

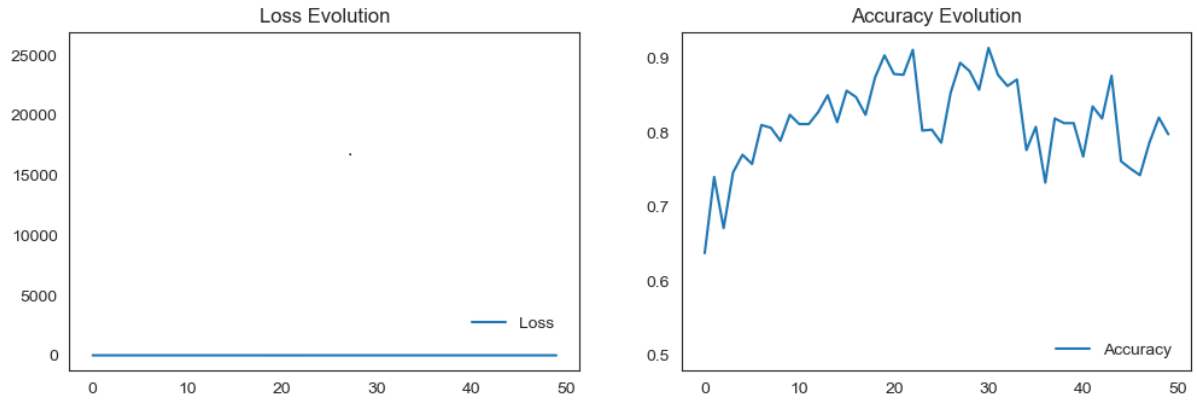


Figure 40 GoogLeNet Accuracy and Loss Evolution Carves

Figure 41 shows that performance after ten epochs is better than after 50. The accuracy values and weighted F1 score following ten epochs are much greater than those following 50 epochs. That said, after 50 epochs, the Macro-F1 value is higher. The macro F1 score, on the other hand, takes averages over all classes, while the weighted F1 score dives in proportion with the number of samples in each class. This is why the weighted score may be different for the classes.

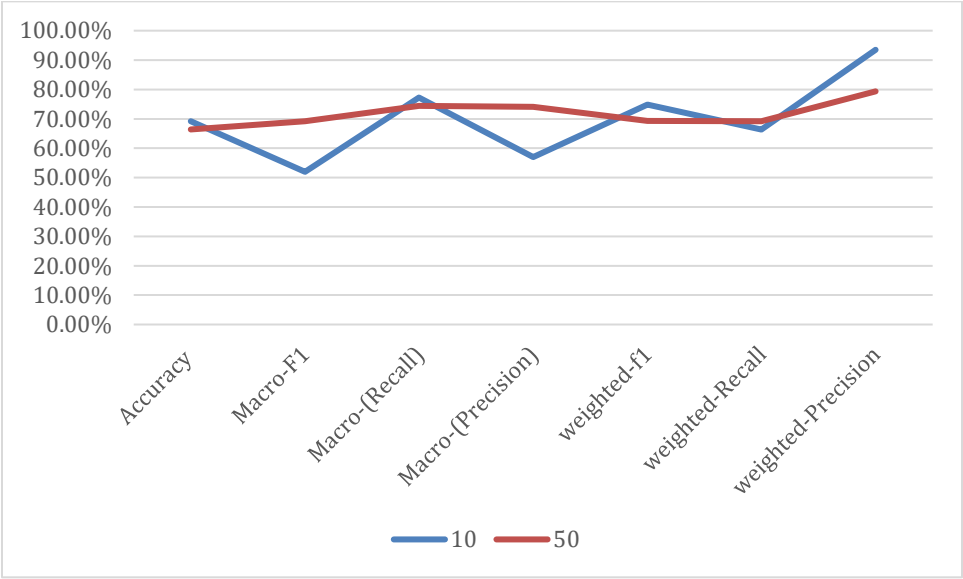


Figure 41 Comparison between 10 epoch and 50 epoch using Classification Matrices

4.3 Results Discussion

To evaluate the efficiency of several deep learning algorithms, namely Convolutional Neural Networks (CNN), DenseNet, VGG-16, ResNet, and GoogLeNet, their advanced mechanisms over the other algorithms will be considered for the given task. The accuracy of such algorithms is a set of several aspects, such as structures of neural networks, performance requirements, and analysis dataset characteristics. Every algorithm possesses features that enable its success, namely, parametrical efficiency, capacity of feature extraction, and a spectrum of layer architectures. Consequently, what algorithm you should be using should be based on a good understanding of these requirements, and you will see that this is in line with what the issue domain can tolerate because you have evaluated these factors meticulously.

Figure 42 below shows a concise comparison of various algorithms:

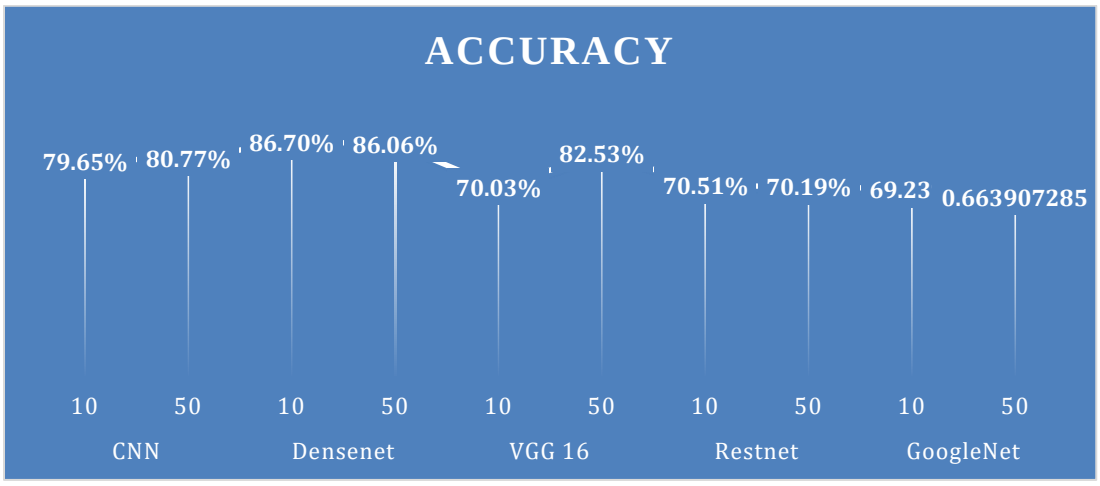


Figure 42 Comparison between algorithms using Accuracy Classification Matrices

The effectiveness of the following algorithms, DenseNet, VGG-16, CNN, ResNet, and GoogLeNet, were tested on the chest dataset. The Densenet model showed exceptional

performance when looping ten and 50 epochs, highlighting its efficiency in capturing complex patterns within the dataset. The observed rapid convergence indicates that DenseNet is compatible with this type of dataset, which shares properties similar to the chest dataset, where the features are densely correlated.

On the other hand, the VGG-16 algorithm shows satisfactory results after 50 epochs and decreases when 10 epochs are repeated. Its clear and consistent structure can be explained by the fact that it will encounter difficulties when faced with datasets containing complex patterns outside the model design, such as a population of chest datasets, and needs more cost for training its network.

The CNN model also confirmed its ability to extract useful features. After completing ten epochs, its performance was at par with the expected accuracy. The precision (correctness of the model) gradually reduced when the model ran 50 epochs. CNNs are, in most cases, highly successful for applications related to image processing, so they also prove very efficient for cases like this, see Figure 43.

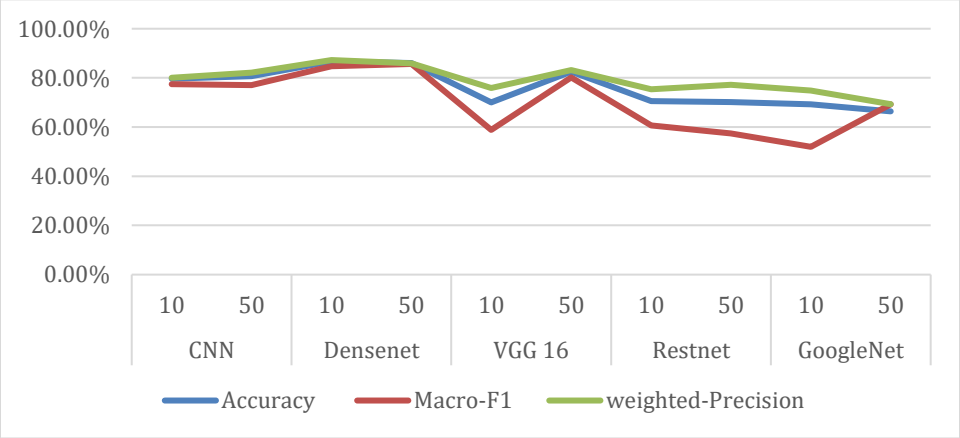


Figure 43 Comparison between 10 epoch and 50 epoch using Classification Matrices

The ResNet and GoogLeNet models did not perform better than the rest. Therefore, based on the given dataset, the study determined that using higher expansion architectures such as GoogLeNet and ResNet, the other models in the general evaluation, was not of value.

In short, the findings suggest that DenseNet, ResNet, and CNN have the potential to cope with the chest data, but VGG-16 and GoogLeNet may involve reduced effectiveness. This conclusion demonstrates why appropriate deep learning architecture should be chosen after paying attention to the specific features of the dataset in the first place because the latter is key to achieving a high level of accuracy.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

It goes without saying that the system for diagnosing illnesses in chest X-ray photos requires several highly credible steps. First, we preload the X-ray data, which brings several cases into a couple of pictures. These pictures are essential because they are used as a reference, which is critical during the model's training and validation process. To address the fourth research question, this phase helps the model distinguish between normal and abnormal pulmonary conditions by learning from labeled examples.

The following phase of data annotation involves labeling all dataset images with normal and pneumonia categories using supervised learning. This annotation process supports the model's ability to differentiate between pulmonary conditions, directly addressing our fourth research question.

Following the data annotation phase, we embark on data augmentation, a crucial step that helps generate additional dataset diversity. This process involves creating thousands of new examples, enhancing the model's versatility, and enabling it to recognize new data. We achieve this by modifying the pictures with scaling, rotating, and flipping techniques. Among these, flipping and small-angle rotations proved most effective, supporting our second research question on augmentation impact.

The subsequent phase commences the preprocessing stage, for which making the images uniform is a prerequisite. The next stage includes brightening the images and increasing the required contrast when pictures are aligned uniformly, and pixels are trimmed to the standard level. These preprocessing steps significantly improved model accuracy, as investigated in our third research question.

Our model then includes deep-learning techniques capable of processing chest X-ray pictures. This classifier will be run multiple times on the labeled chest X-ray image dataset to differentiate various chest conditions and validate uniformly. Models such as DenseNet, CNN, and VGG-16 were evaluated to identify the most effective algorithm, in line with our first research question.

The evaluation of the algorithms' performance of methods DenseNet, VGG-16, CNN, ResNet, and GoogLeNet has been carried out on the chest dataset. The DenseNet model, given ten and 50 epochs, demonstrated great success, capturing complex data patterns and thus highlighting its efficiency. The high degree of convergence observed in DenseNet proves that it is an ideal model for the given dataset. It shows signs of a dense structure with highly correlated patches. This highlights DenseNet's superiority in handling chest X-ray images for disease detection.

On the other hand, the VGG-16 algorithm works well after 50 epochs and shows a decrease in quality when another ten epochs are repeated, which is a surprising result. This is in contrast to the CNN model. The CNN model has shown the ability to extract vital aspects from the dataset. At the end of ten epochs, its performance was similar to the expected accuracy. However, there was a continuous drop in the model's accuracy after the 50th epoch ended, which was an unexpected finding. This performance variation supports our effort to compare algorithm effectiveness, as posed in the first research question.

There was also variance in performance, as other approaches did better than the ResNet and GoogLeNet ones. Based on our dataset, this study found no merit in using complex architectures like GoogLeNet and ResNet compared to the other models in general for this type of dataset. This suggests that simpler or moderately deep models may generalize better on chest X-ray datasets.

Lastly, the model is tested on different cases to pinpoint the range of applicability. Validating the model and applying it to clinical hospitals will allow doctors and healthcare providers, in general, to develop a practical tool for examining chest diseases such as pneumonia, among others. This final testing phase reinforces the clinical relevance of our findings and the applicability of selected models in real diagnostic workflows.

5.2 Future works

The next step of the work in the paper is the thorough research of the efficiency of the deep learning algorithms, including the CNN, DenseNet, VGG-16, ResNet, and GoogLeNet, taking attention to the main pluses and minuses, which are related to the given task. The analysis will examine several dimensions, including system exquisiteness, hardware resource demand, and database specifications used. These algorithms will be evaluated based on the excellent parameter efficiency, the algorithm feature extraction capability, and the type of neural network architecture. One can determine a suitable algorithm for a task's needs by assessing these variables. This, in turn, gives confidence that the algorithm is fitted for the particular domain situation.

Moreover, the experiments will entail the comparison of the algorithms on the test dataset to outline the areas that need to be addressed and those that have the potential. The training and performance testing algorithms using various parameters will provide an understanding of the capability of algorithms to discriminate X-rays containing pneumonia-related information. The findings of the experiments generated by these algorithms will be an important consideration in classifying and selecting the most appropriate algorithm for enhancing the accuracy in diagnosing thyroid cases.

References

Pham, L., McLoughlin, I., Phan, H., Tran, M., Nguyen, T., & Palaniappan, R. (2020, July). Robust deep learning framework for predicting respiratory anomalies and diseases. In 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC) (pp. 164-167). IEEE.

Khan, W., Zaki, N., & Ali, L. (2021). Intelligent pneumonia identification from chest x-rays: A systematic literature review. *IEEE Access*, *9*, 51747-51771.

Ervural, S., & Ceylan, M. (2021). Convolutional neural networks-based approach to detect neonatal respiratory system anomalies with limited thermal image. *Traitement du Signal*.

Bhatt, H., & Shah, M. (2023). A Convolutional Neural Network ensemble model for Pneumonia Detection using chest X-ray images. *Healthcare Analytics*, *3*, 100176.

Pal, J., & Das, S. (2023). A Convolutional Neural Network (CNN)-Based Pneumonia Detection Using Chest X-ray Images. In *Using Multimedia Systems, Tools, and Technologies for Smart Healthcare Services* (pp. 63-82). IGI Global.

Ignat, A., & Găină, R. A. (2023). Improving Diagnostics of Pneumonia by Combining Individual Hypotheses on Chest X-Ray Images. In *Advances in Smart Healthcare Paradigms and Applications: Outstanding Women in Healthcare—Volume 1* (pp. 43-64). Cham: Springer Nature Switzerland.

Gabruseva, T., Poplavskiy, D., & Kalinin, A. (2020). Deep learning for automatic pneumonia detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 350-351).

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.

Barhoom, A. M., & Abu-Naser, S. S. (2022). Diagnosis of pneumonia using deep learning.

Perna, D., & Tagarelli, A. (2019, June). Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 50-55). IEEE.

Surden, H. (2019). Artificial intelligence and law: An overview. *Georgia State University Law Review*, *35*, 19-22.

Mustapha, F. Z., Haruna, A. A., & Muhammad, U. T. (2020). An Overview of Artificial Intelligence. *Journal of Applied Sciences & Environmental Sustainability*, *6*(12), 60-74.

Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. *Computer Science Review*, *40*, 100379.

Tian, C., Fei, L., Zheng, W., Xu, Y., Zuo, W., & Lin, C. W. (2020). Deep learning on image denoising: An overview. *Neural Networks*, *131*, 251-275.

Ibrahim, A. U., Ozsoz, M., Serte, S., Al-Turjman, F., & Yakoi, P. S. (2021). Pneumonia classification using deep learning from chest X-ray images during COVID-19. *Cognitive Computation*, 1-13.

Ieracitano, C., Mammone, N., Versaci, M., Varone, G., Ali, A. R., Armentano, A., ... & Morabito, F. C. (2022). A fuzzy-enhanced deep learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images. *Neurocomputing*, *481*, 202-215.

Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. J. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement*, *145*, 511-518.

Singh, S., Kumar, M., Kumar, A., Verma, B. K., & Shitharth, S. (2023). Pneumonia detection with QCSA network on chest X-ray. *Scientific Reports*, *13*(1), 9025.

Račić, L., Popović, T., & Šandi, S. (2021, February). Pneumonia detection using deep learning based on convolutional neural network. In *2021 25th International Conference on Information Technology (IT)* (pp. 1-4). IEEE.

Chandra, T. B., & Verma, K. (2020). Pneumonia detection on chest x-ray using machine learning paradigm. In *Proceedings of 3rd International Conference on Computer Vision and Image Processing: CVIP 2018, Volume 1* (pp. 21-33). Springer Singapore.

Darici, M. B., Dokur, Z., & Olmez, T. (2020). Pneumonia detection and classification using deep learning on chest x-ray images. *International Journal of Intelligent Systems and Applications in Engineering*, *8*(4), 177-183.

Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R., & Ibragimov, B. (2019). Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. *Computers & electrical engineering*, *78*, 388-399.

Varshni, D., Thakral, K., Agarwal, L., Nijhawan, R., & Mittal, A. (2019, February). Pneumonia detection using CNN based feature extraction. In *2019 IEEE international conference on electrical, computer and communication technologies (ICECCT)* (pp. 1-7). IEEE.

Yi, R., Tang, L., Tian, Y., Liu, J., & Wu, Z. (2023). Identification and classification of pneumonia disease using a deep learning-based intelligent computational framework. *Neural Computing and Applications*, *35*(20), 14473-14486.

Bhatt, H., & Shah, M. (2023). A Convolutional Neural Network ensemble model for Pneumonia Detection using chest X-ray images. *Healthcare Analytics*, *3*, 100176.

Yang, Y., Mei, G., & Piccialli, F. (2022). A deep learning approach considering image background for Pneumonia identification using explainable AI (XAI). *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Yang, Y., & Mei, G. (2022). Pneumonia recognition by deep learning: A comparative investigation. *Applied Sciences*, *12*(9), 4334.

Manickam, A., Jiang, J., Zhou, Y., Sagar, A., Soundrapandiyan, R., & Samuel, R. D. J. (2021). Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures. *Measurement*, *184*, 109953.

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.

Jain, R., Nagrath, P., Kataria, G., Kaushik, V. S., & Hemanth, D. J. (2020). Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. *Measurement*, *165*, 108046.

Ayan, E., & Ünver, H. M. (2019, April). Diagnosis of pneumonia from chest X-ray images using deep learning. In *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)* (pp. 1-5). Ieee.

Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. J. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement*, *145*, 511-518.

Garstka, J., & Strzelecki, M. (2020, September). Pneumonia detection in X-ray chest images based on convolutional neural networks and data augmentation methods. In *2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)* (pp. 18-23). IEEE.

Saul, C. J., Urey, D. Y., & Taktakoglu, C. D. (2019). Early diagnosis of pneumonia with deep learning. *arXiv preprint arXiv:1904.00937*.

Dutta, S. (2018). An overview on the evolution and adoption of deep learning applications used in the industry. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1257.

Krizhevsky, S. A., Sutskever, I., & Hinton, G. E. (2019). From Photographic Image to Computer Vision.

Sharma, N., Jain, V., & Mishra, A. (2018). An analysis of convolutional neural networks for image classification. *Procedia computer science*, *132*, 377-384.

Jiang, Z. P., Liu, Y. Y., Shao, Z. E., & Huang, K. W. (2021). An improved VGG16 model for pneumonia image classification. *Applied Sciences*, *11*(23), 11185.

Sharma, S., & Guleria, K. (2023). A deep learning based model for the detection of pneumonia from chest X-ray images using VGG-16 and neural networks. *Procedia Computer Science*, *218*, 357-366.

Sitaula, C., & Hossain, M. B. (2021). Attention-based VGG-16 model for COVID-19 chest X-ray image classification. *Applied Intelligence*, *51*(5), 2850-2863.

Wang, K., Jiang, P., Kong, D., Sun, B., & Shen, T. (2023). Improving Accuracy of Pneumonia Classification Using Modified DenseNet. *Journal of Digital Imaging*, *36*(4), 1507-1514.

Sanghvi, H. A., Patel, R. H., Agarwal, A., Gupta, S., Sawhney, V., & Pandya, A. S. (2023). A deep learning approach for classification of COVID and pneumonia using DenseNet-201. *International Journal of Imaging Systems and Technology*, *33*(1), 18-38.

Zhang, K., Guo, Y., Wang, X., Yuan, J., & Ding, Q. (2019). Multiple feature reweight densenet for image classification. *IEEE access*, *7*, 9872-9880.

Mahajan, S., Shah, U., Tambe, R., Agrawal, M., & Garware, B. (2019, March). Towards evaluating performance of domain specific transfer learning for pneumonia detection from x-ray images. In *2019 IEEE 5th international conference for convergence in technology (I2CT)* (pp. 1-6). IEEE.

Pant, A., Jain, A., Nayak, K. C., Gandhi, D., & Prasad, B. G. (2020, July). Pneumonia detection: An efficient approach using deep learning. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.

Ikechukwu, A. V., Murali, S., Deepu, R., & Shivamurthy, R. C. (2021). ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. *Global Transitions Proceedings*, *2*(2), 375-381.

Ikechukwu, A. V., Murali, S., Deepu, R., & Shivamurthy, R. C. (2021). ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. *Global Transitions Proceedings*, *2*(2), 375-381.

Militante, S. V., Dionisio, N. V., & Sibbaluca, B. G. (2020, August). Pneumonia detection through adaptive deep learning models of convolutional neural networks. In *2020 11th IEEE control and system graduate research colloquium (ICSGRC)* (pp. 88-93). IEEE.

Sharma, S., & Guleria, K. (2024). A systematic literature review on deep learning approaches for pneumonia detection using chest X-ray images. *Multimedia Tools and Applications, 83*(8), 24101-24151.

Umar Ibrahim, A., Ozsoz, M., Serte, S., Al-Turjman, F., & Habeeb Kolapo, S. (2022). Convolutional neural network for diagnosis of viral pneumonia and COVID-19 alike diseases. *Expert Systems, 39*(10), e12705.

J. Liu et al., "Prediction of rupture risk in anterior communicating artery aneurysms with a feed-forward artificial neural network," Eur. Radiol., vol. 28, no. 8, pp. 3268–3275, 2018.

A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," Pattern Recognit., vol. 91, pp. 216–231, 2019.

D. Chicco, N. Tötsch, and G. Jurman, "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation," BioData Min., vol. 14, no. 1, pp. 1–22, 2021.

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756.

الملخص

يُعَدّ أربع مرات من أكثر الأمراض للوفيات العالمية، خصوصًا بين الأطفال دون سن الخامسة، مما يجعل التشخيص مبكرًا ودقيقًا للوصول إلى تحسين فرص العلاج والنجاة. في هذا البحث، تم تطوير نظام يعتمد على تقنيات التعلم الأساسية لتحليل صور الأشعة السينية للصدر، بغرض الاستثمار بين الحالات الطبيعية والإصابة بالالتهاب. وقد استُخدمت لذلك مجموعة بيانات عالمية مكونة من (5856) صورة مطلوبة من منصة Kaggle.

بدأت العمل بخطوات الكاميرا التمهيدية، والتي تشارك في مشاركة الصور، مخصصة للإضاءة والتباين، والتأكد من إزالة التشويش والاختلافات البصرية غير المرئية. كما تم تطبيق تقنيات زيادة البيانات (Data Augmentation) مثل الانعكاسات والتدوير البسيط، بهدف تعزيز تنوع البيانات وتقنية قوية للتطورات في مجال التفاعل عند مواجهة صور جديدة. وهذه الخطوات الفاعلة لها في رفع الارتباطات التفصيلية والتحليلات التوزيعية البكسلية للسيدات.

تم اختبار خمسة نماذج من الشبكات العصبية المتخصصة: CNN، DenseNet، VGG-16، ResNet، GoogLeNet عبر مرحلتي تدريب مكونتين من (10) و(50) حقبة، وذلك لتدريب الأداء بين التدريب قصير المدى وطويل المدى. لا تزال النتائج ما يلي:

- حقق نموذج DenseNet أداءً أفضل، حيث تم تحقيق دقة تصل إلى حوالي (87%) بقيم مرتفعة لمؤشرات F1 والاسترداد، ما يعكس التدقيق في التنسيق الدقيق والتداخل السريع مع طبيعة البيانات الطبية.
- أظهر نموذج VGG-16 تحسنًا ملحوظًا عند (50) حقة تقريبًا بوتت (83%)، ولكنه كان أقل استقرارًا مع زيادة عدد التكرارات.
- حقق نموذج CNN نتائج جيدة بعد (10) حقبات بنسبة (81٪-80)، ولكنه عانى من انخفاض الأداء في التدريب الأطول.
- أما ResNet وGoogLeNet فقد سجلا أداءً أضعف (حوالي 66–71%)، مما يشير إلى أن البنى التقنية الأكثر ابتكاراً ليست بالضرورة اختيار الأفضل مع هذا النوع من البيانات.

تشير هذه النتائج إلى أن الدراسات متوسطة العمق مثل DenseNet توفر مطبخًا ضروريًا للتشخيص والدقة، في حين أن الدراسات الأكثر مخاطرًا قد لا تحقق مزايا إضافية مع بيانات محدودة أو خصائص مستوردة مثل الصور الرئيسية. وتبرز أهمية اختيار البنية اللازمة لخصائص البيانات لأداء أفضل ممكن.

ختامًا، يمكن القول إن النموذج المقترح يُظهر إمكانية واعدة المشاهير العملي في المستشفيات والمراكز الصحية، حيث يسهم في دعم الباحثين عبر توفير أداة مساعدة دقيقة وغير موثوقة لتشخيص حاجتهم، مع إمكانية الوصول إلى نطاق البحث المستقبلي، وتعزيزات إضافية في الخوارزمية المتكاملة ودمجها في أنظمة الدعم الطبية.