
PNEUMONIA DETECTION USING DEEP LEARNING

Pneumonia Detection Using Deep Learning

SALIM HABIB UNIVERSITY
(FORMERLY BARRETT HODGSON UNIVERSITY)



Wardah Ali

Eesha Qureshi

Omama Ahmed Farooqi

A project report submitted in partial fulfillment of the
Requirements for the award of the degree of Bachelor of
Computer Science

Faculty of Information Technology
Department of Computer Science
Salim Habib University, Karachi, Pakistan

Year 2022

Pneumonia Detection Using Deep Learning

Submitted by

Wardah Ali

(F18CSC05)

Eesha Qureshi

(F18CSC24)

Omama Ahmed Farooqi

(F18CSC26)

Faculty of Information Technology

Department of Computer Science

Final Year Project Thesis Presented to

Dr. Rizwan Ahmed Khan

Faculty of Information Technology

Department of Computer Science

Salim Habib University, Karachi, Pakistan

Year 2022

DECLARATION

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Salim Habib University or other institutions.

Signature : _____

Name : _____

Reg No. : _____

Signature : _____

Name : _____

Reg No. : _____

Signature : _____

Name : _____

Reg No. : _____

Date : _____

APPROVAL FOR SUBMISSION

We certify that this project report entitled **PNEUMONIA DETECTION USING DEEP LEARNING** was prepared by Wardah Ali, Eesha Qureshi and Omama Ahmed Farooqi, has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Computer Science at Salim Habib University. Approved by,

Signature : _____

Supervisor : Dr. Rizwan Ahmed Khan

Date : _____

The copyright of this report belongs to Salim Habib University according to the Intellectual Property Policy of Salim Habib University amended on July 2nd 2021. Due acknowledgement shall always be made of the use of any material contained in, or derived from, this report.

© 2022 Salim Habib University. All right reserved.

ACKNOWLEDGEMENTS

We would like to thank everyone who had contributed to the successful completion of this project. Moreover we would also like to express our gratitude to our research supervisor, Dr. Rizwan Ahmed Khan for his invaluable advice, guidance and his enormous patience throughout the development of this research.

In addition, we would also like to express our gratitude to our loving parents and friends who helped and gave us encouragement.

PNEUMONIA DETECTION USING DEEP LEARNING

ABSTRACT

People all over the globe are affected by pneumonia but deaths due to it are highest in Sub-Saharan Asia and South Asia. In recent years, the overall incidence and mortality rate of pneumonia regardless the utilization of effective vaccines and compelling anti-biotics, has escalated. Thus, pneumonia remains a disease that needs spry prevention and treatment. The widespread prevalence of pneumonia is of great importance to the research community to come up with a framework that helps detect, diagnose and analyze diseases accurately and promptly. But the major difficulties faced by the Artificial intelligence research community is that only a few authentic datasets for chest diseases are publically available and even these available datasets are highly imbalanced. For this matter, we present a deep learning framework based on pre-trained deep learning models to aid medical practitioners for efficient and early pneumonia diagnosis and to cater the existing class imbalance problems. The proposed framework utilizes state-of-the-art deep learning models i.e. ResNet, Xception-50 and VGG-16 to compare their robust features and results while The Generative Adversarial Network GAN, specifically a combination of Deep Convolutional Generative Adversarial Network (DCGAN) and Wasserstein GAN gradient penalty (WGAN-gp) was applied on the minority class "Pneumonia" for augmentation, whereas Random Under-Sampling (RUS) was done on the majority class "No Findings" to deal with the imbalance problem.

Contents

1. Introduction	xiv
1.1 Identified Research Gap	xv
1.2 Scope of the Proposed Work	xvi
1.3 Organization of the Thesis	xvi
2. Background (Literature Review)	xviii
2.1 Convolutional Neural Network	xxi
2.2 Transfer Learning	xxii
2.3 Class Imbalance	xxvi
2.4 Generative Adversarial Network GANxxvii
3. Dataset Analysis	xxxi
3.1 Data Descriptionxxxii
3.2 Prior Work on the Chest X-Ray8 Dataset	xxxiii
4. Methodology	xxxv
4.1 Data Pre-processingxxxv
4.1.1 Exploratory Data Analysis (EDA)	xxxvi
4.1.2 Data Imbalance Techniques	xxxviii
4.2 Data Augmentation	xxxix
4.2.1 Generative Adversarial Network (GAN)	xxxix
4.2.2 Deep Convolutional Generative Adversarial Network (DCGAN)	xlii
4.3 Wasserstein GAN gradient penalty (WGAN-gp)	xlv
4.4 Explanation about the Transfer Learning models	xlix

4.4.1	AlexNet	xlix
4.4.2	ResNet	li
4.4.3	Xception	liii
4.4.4	The Visual Geometry Group (VGG-16)	liii
5.	Experimental Result	lvi
5.1	Performance analysis of the models	lvi
5.2	Deliverables	lxiii
6.	Conclusion and Future Work	lxiv

List of Figures

Figure 2.1 – Standard Steps that are being followed in Conventional Machine Learning Algorithms	xix
Figure 2.2 – Artificial Neural Network Architecture	xxi
Figure 2.3 – Convolutional Neural Network Architecture	xxii
Figure 2.4 – Original pneumonia images from CheXnext and their CycleGAN generated pairs. Last row shows the difference between Original and Generated images; by [1]	xxviii
Figure 2.5 – Comparison of sample quality between different GAN architectures and the extended version of the ChestX-ray8 dataset i.e "ChestX-ray14" by [2] . . .	xxix
Figure 3.1 – Dataset details	xxxii
Figure 3.2 – No findings X-ray (left) and Pneumonia X-ray (right)	xxxiii
Figure 3.3 – Important attributes of Chest X-ray8, According to this table, males are more likely to develop nosocomial bacterial and community acquired pneumonia than females and it's more likely to happen to people between the age group of 40-60.	xxxiii
Figure 4.1 – Overview of the Statistics of Chest X-Ray8 dataset	xxxvi
Figure 4.2 – Correlation of the statistical variables	xxxvi
Figure 4.3 – Principle Component Analysis (PCA) projection of the ChestX-Ray8 dataset stratified by class (Pneumonia and No Finding) of 60k data-points	xxxvii
Figure 4.4 – t-SNE (TDistributed Stochastic Neighbor Embedding) of the ChestX-Ray8 dataset stratified by class (Pneumonia and No Finding) of 60k data-points	xxxvii

Figure 4.5 – The two models in GAN (Generator and Discriminator) that are trained simultaneously via an adversarial method	xl
Figure 4.6 – Generative Adversarial Network (GAN) applied on images of five girls from our batch with 64x64 image resolution.	xli
Figure 4.7 – Images generated through conventional augmentation techniques	xlii
Figure 4.8 – DCGAN Generator and Discriminator Architecture	xliv
Figure 4.9 – DCGAN distorted results of 128x128 image resolution with Binary Cross-Entropy Loss function at 80K iteration	xlv
Figure 4.10 –Generator and Discriminator architecture employed for the presented Pneumonia classification task	xlvii
Figure 4.11 –Generator's cost during GAN training (left). Generator's cost during DCGAN training with Wasserstein's loss (right) by [3]	xlviii
Figure 4.12 –WGAN-gp generated results applied on the minority class "Pneumonia" with 64x64 image resolution, these results were further used in our classification task.	xlix
Figure 4.13 –Architecture of AlexNet: Convolution, max-pooling, LRN and fully connected (FC) layer (Alex Krizhevsky et al. [4])	1
Figure 4.14 –Training error (left) and Test error (right) on CIFAR-10, The deeper network has higher training error and thus test error [5]	li
Figure 4.15 –Residual Learning: a building block (Kaiming He et al. [5])	li
Figure 4.16 –Architecture of VGG-16: Convolution, max-pooling, and dense layers by [49]	liv
Figure 4.17 –Complete methodology of the proposed Pneumonia classification model . .	lv
Figure 5.1 – ResNet training accuracy and loss on Chest X-Ray8 extracted dataset . . .	lvii
Figure 5.2 – VGG-16 training accuracy and loss on Chest X-Ray8 extracted dataset . .	lviii
Figure 5.3 – Alexnet training accuracy and loss on Chest X-Ray8 extracted dataset . .	lviii
Figure 5.4 – Xception training accuracy and loss on Chest X-Ray8 extracted dataset . .	lviii
Figure 5.5 – ROC curves of the three models used for classification without applying WGAN loss function for augmentation along with their respective AUC score through which we determine our error metric	lix

Figure 5.6 – ROC curves of the three models used for classification after applying WGAN loss function with DCGAN architecture, along with their respective AUC score through which we determine our error metric	lix
Figure 5.7 – The Generative Adversarial Network (GAN) loss details after applying the Wasserstein Loss with gradient penalty (WGAN-gp)	lx
Figure 5.8 – Confusion matrix of Xception model on the two classes "Pneumonia" and "No Findings"	lxii
Figure 5.9 – Confusion matrix of ResNet model on the two classes "Pneumonia" and "No Findings"	lxii
Figure 5.10 –Confusion matrix of VGG-16 model on the two classes "Pneumonia" and "No Findings"	lxii
Figure 5.11 –The Graphical User Interface (GUI) of the presented Pneumonia classification model based-on pre-trained architectures	lxiii

List of Tables

Table 1.1 – The most prevailing procedural techniques and weaknesses in the writings and literature of X-Ray imaging technique	xvii
Table 5.1 – The ChestX-Ray8 dataset after catering the imbalance problem	lv
Table 5.2 – Classification Report of state-of-the-art on the "ChestX-Ray8" dataset for pneumonia classification on the test set	lvii
Table 5.3 – Model Results on the train set for comparison	lxii

1. INTRODUCTION

Pneumonia is a principal infectious source of death in children worldwide. In accordance with the statistics issued by the World Health Organization (WHO), a number of 740,180 children that were below the age of 5 died because of pneumonia in 2019 [6]. People all over the globe are affected by pneumonia but deaths due to it are highest in Sub-Saharan Africa and South Asia. In 2017, more than half of the deaths in five countries i.e. the Republic of Congo, Pakistan, India, Nigeria, and Ethiopia was from childhood pneumonia. This clearly tells how a country's income has a strong correlation with the child mortality rate from pneumonia [7]. In 2018, Watkins et al. [8] called pneumonia "the ultimate disease of poverty" in a statement of the journal named, 'The Lancet'. Pneumonia is the third leading cause of death in elderly deaths (people greater than or equal to 80 years of age) in Japan [9]. Pneumonia is an inflammation of the bronchi, alveoli, bronchioles, and interstitial lungs. The most common types of pneumonia are viral and bacterial pneumonia which possess significant health threats. Pneumonia substantially occurs due to pathogenic microbial infections, immune function damage, allergies as well as drug factors. In recent years, the overall incidence and mortality rate of pneumonia regardless the utilization of effective vaccines and compelling anti-biotics, has escalated. Thus, pneumonia remains a disease that needs spry prevention and treatment. Computer-Aided Diagnosis (CAD) is a very popular batch of techniques that assist doctors to detect and interpret various types of abnormalities in medical imaging, to diagnose and analyze diseases accurately and promptly [10]. Familiar examples related to CAD projects that aid in the examination and diagnoses of lung infection and diseases involves Chest X-rays, Magnetic Resonance Imaging (MRI), chest CT, bronchoscopy, etc. Chest X-rays are mainly used to detect lung-related diseases which include pneumonia. Conventionally, manual inspection of chest X-rays is done by the radiologist in order to detect and diagnose pneumonia and other lung diseases, however, it

can lead to a prolonged diagnosis process. For example,

1. About 2/3 of persons around the world still do not have the means to get their disease diagnosed by a radiologist in accordance to a report by the World Health Organization (WHO).
2. Fatigue and the concentration of medical experts/radiologists can affect diagnosis.
3. Inspection of a large number of X-rays on a daily basis can be exhausting.
4. Availability of a medical expert at all times having the knowledge to precisely analyze multiclass images as there exists a large number of chest-related diseases.

One course of action in-order to overcome the issues that are mentioned above, is to propose an intelligent framework that detects pneumonia in an effective and efficient manner. The performance of the proposed framework to diagnose whether a person does or does not have pneumonia will be measured by accuracy, specificity, sensitivity, AUC-score, and ROC curve. In order to detect pneumonia, we will be using a transfer learning based-approach to extricate meaningful features on the Chest X-Ray8 dataset [11].

1.1 *Identified Research Gap*

Pneumonia causes inflammation of the lungs which usually results in loss of human life all around the world thus premature detection of pneumonia is crucial so as to escalate the survival ratio and effective treatment. Despite of the fact that AI is helping to develop effective models for medical image analysis and early diagnosis, there remains quite a number of difficulties to be resolved until AI can have a substantial impact on clinical practice. One of the major difficulties is that only a few authentic datasets for chest diseases are publically available and even these available datasets are highly imbalanced. Data imbalance occurs whenever the targeted class has an uneven data distribution, such as one class label has a small number of observations while the other has a large number.

For this matter we present a deep learning framework based on pre-trained transfer-learning

models to aid medical practitioners for efficient and early pneumonia diagnosis and to cater the existing class imbalance problems.

1.2 Scope of the Proposed Work

The deep learning models ensure better results if they are trained on a large amount of data but typically medical studies offer a limited amount. Even if the medical imaging datasets offer a large amount of data, a lot of them suffer from imbalance problems [12]. Another challenge of deep learning algorithms especially in medical imaging is selecting an appropriate classifier/model and extracting the most relevant features or predicting more accurately. Our project is in response to the above-mentioned problems. Data analysis is the first and foremost objective of this project in order to explore and see how many diseases(classes) exist in the 'CXR8' dataset and if there are any statistical links between the classes. It is further followed by the extraction of relevant data from the dataset needed for this project. The second and most important objective is to deal with the data imbalance between the extracted classes. For this problem, data imbalance techniques were applied to increase the samples of minority class and decrease the data of majority class. Lastly, use of existing pre-trained transfer learning models and comparison between them to see which model accurately detects if pneumonia is present or not.

1.3 Organization of the Thesis

The ensemble of this thesis is organized into six chapters. Chapter 1 bestows the introduction to the thesis. Chapter 2 confers the literature review on CNN and Transfer Learning. Chapter 3 accords with the dataset analysis. The methodology of the project is the main focus of Chapter 4. Chapter 5 summarizes the results of the experiments and sets the stage for Chapter 6, which finishes the document with Conclusion and Future Work.

Tab. 1.1: The most prevailing procedural techniques and weaknesses in the writings and literature of X-Ray imaging technique

Data	Deep Learning Approach
<p>Most Prevailing</p> <ul style="list-style-type: none"> Conventional data augmentation techniques such as cropping, flipping, translation, noise injection are common in medical image analysis. <p>Weaknesses</p> <ul style="list-style-type: none"> For the medical image analysis task, determine the best dataset size thresholds. Explore data imbalance techniques, under-sampling is done for the majority class and over-sampling is done for the minority class (Smote). Evolutionarily/genetics algorithm are also being used for sampling of imbalanced data. 	<p>Most Prevailing</p> <ul style="list-style-type: none"> For smaller datasets, feature-extracting Transfer Learning (TL) models are used, and for bigger datasets, TL models are fine-tuned. <p>Weaknesses</p> <ul style="list-style-type: none"> For making each TL technique's accuracy, performance, and memory, optimal, decide whether a bigger dataset or a better CNN model selection is most significant. In the medical setting, whether transfer learning is relatively more accurate than training the model from scratch. To determine the best models (optimal) for medical imaging task, a greater immersion on rigorous bench-marking is required.

2. BACKGROUND (LITERATURE REVIEW)

Using machine learning (ML) and artificial intelligence (AI) methodologies, a large amount of research work has already been carried out for the diagnosis of chest diseases. The diagnosis of pneumonia and chronic obstructive pulmonary diseases has been applied using artificial immune systems and neural networks [13]. AI's successful contribution includes better quantitative assessment in identifying intricate image patterns from data in a robust and automated manner. AI is also being used as a tool to easily and accurately assist physicians and radiologists for various tasks of medical imaging i.e. disease detection, prognosis or diagnosis etc. Moreover, the chances of uncertainty and error in decision making for medical imaging analysis can be decreased using AI. The two categories of AI Algorithms are conventional AI (Machine learning) and Contemporary AI (Deep Learning). In this chapter, we will first discuss the background of the two branches of AI i.e. Machine Learning and Deep Learning Algorithms. Then Convolutional Neural Networks (CNNs) followed by transfer learning and pre-trained models of transfer learning for medical imaging.

Machine Learning Models

According to Arthur Samuel machine learning is a subfield of Artificial intelligence (AI) [14] that includes algorithms which allows a machine to parse data, perform a function in order to learn from that data, and then apply what it has learned to make informed decisions. The aim of machine learning algorithms is to learn and get progressively better over time so it can make better decisions without programming explicitly. As shown in Fig 2.1, initially the input to ML algorithms in terms of medical imaging can be manually handcrafted, features that basically distinguish image characteristics like area, shape, texture, image pixels etc. [15]. The extracted input features are then further processed to select most relevant features

through feature selection algorithms. The goal of machine learning algorithms is to combine the chosen relevant features into one value. Supervised and Unsupervised learning are the two main subtypes of machine learning. The supervised learning algorithms use labeled data for training i.e. algorithm learns from prior existing knowledge (classification). On the contrast, the algorithms in unsupervised learning learn patterns from data, associating them with data points of discovered clusters (clustering) [16]. Numerous cutting-edge ML algorithms have been applied to the evaluation and diagnosis of pneumonia. In 2020, AlMamlook et al. [17] proposed a model to increase accuracy and efficiency for the classification of normal(healthy) from abnormal (sick) Chest X-rays. They used 7 pre-existing state-of-the-art Machine learning models and techniques along with well-known Convolution Neural Network (CNN) models achieving an overall accuracy of 98.46%. These include Random Forest (RF), Decision Tree (DT), Naive Bayes (NB), Support Vector Machine (SVM), Linear Dicriminant Analysis (LDA), K-Nearest Neighbour (KNN), and Logistic Regression (LR) [4]. Machine Learning (ML) models perform better if they are trained on a large amount of data but typically medical studies offer a limited amount. Due to the small number of images per study, machine learning becomes difficult and insipid. Another challenge of ML Algorithm especially in medical imaging is selecting an appropriate classifier and extracting most relevant features.

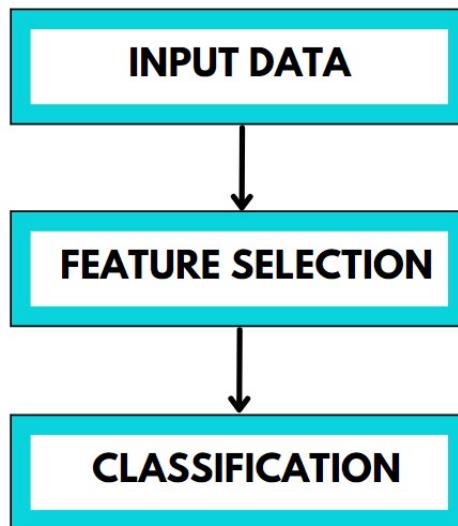


Fig. 2.1: Standard Steps that are being followed in Conventional Machine Learning Algorithms

Deep Learning Models

Unlike manually handcrafted features in ML algorithms, deep learning (DL) algorithms do not require explicit and manual process of extracting features from the original data in order to predict (output). The advancement in the hardware for parallel computing and the development of learning techniques[18], [19] made the learning of deep models like deep neural networks(DNN) possible. The breakthrough in deep learning [20] made the way for unprecedented progress in Computer Vision (CV) aspects i.e detection, classification and semantic segmentation [21]. There is a huge momentum of transforming from traditional to deep learning techniques in the pathology image computing community. Compared to conventional ML models, DL model provides robust results in classification and segmentation problems as they are able to incorporate every feature from the input data.

Deep Learning is basically a sub-field of ML that creates an “Artificial Neural Network (ANN)” by structuring algorithms in layers that learns and makes intelligent decisions just like a human brain. The motivation behind ANN is the biological pattern of neurons in the human brain and leads to a learning system that is substantially more capable than the traditional ML models. The fundamental building block of an ANN is an artificial neuron. The neuron is provided with weighted inputs to produce a prediction (output). As shown in Fig 2.2, prior to applying a nonlinear activation function to produce an output, the neuron first adds up the inputs it has received. [22]. The nonlinear activation function applied can be Tanh, Sigmoid, Rectified Linear Unit (ReLu) etc. The neurons/nodes in ANN can send and receive signals and are linked to one another by a singular or multi-hierarchical layers. The nonlinear activation function’s output determines how the node responds to send and receive stimuli. Each neuron in the network has a weight attached to its input, which is helpful for transmitting data to the output layer and might as well affect the given input. The hidden layer receives a feature vector with a weighted score of the input data. Moreover, the hidden layer, which is composed of activation units, uses the features vector from the upper layer to execute calculations before producing an output. The hidden layer’s weighted output is carried by the output layer, which also predicts the relevant class, and is composed of activation units, each of which matches to a label or class in the dataset. [22] [23].

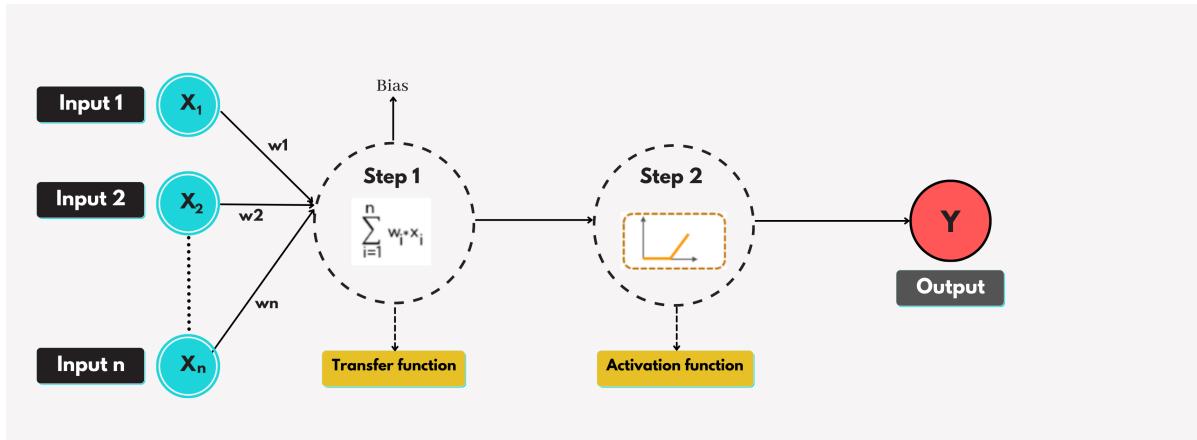


Fig. 2.2: Artificial Neural Network Architecture

2.1 Convolutional Neural Network

Convolutional Neural Networks, explored in the early '90s [24] have grown in popularity as a machine learning technique for a variety of applications, including medical image analysis. A very important milestone of CNN was archived by Yann Lecun, Leon Battou, and Yoshua Bengio [25] in 1998 that is nowadays well-known as LeeNet-5. Convolutional neural networks are designed to automatically acquire knowledge of spatial hierarchies of features through back-propagation algorithms. There are two fundamental requirements for fully training a CNN from scratch: a large labelled dataset, and a lot of computing and memory power. Such huge labelled datasets are often not available in clinical settings. It takes a lot of time and effort to create a large labelled dataset, especially when the number of patients with a certain medical condition is not enough [26]. For example, Stephen et al. [27] created their own convolutional neural network (CNN) model that classified pneumonia on chest X-ray images with high accuracy. Its most significant flaw is its high development cost, which makes the deployment of the model difficult. The authors of [28] introduce the Darknet model as a new method for automatic COVID-19 detection using unprocessed chest X-ray images. Their proposed methodology is made to provide accurate diagnostics for binary classification tasks (COVID vs. No-Findings) and multi-class classification tasks (COVID vs. No-Findings vs. Pneumonia). They reported a multi-class classification accuracy of 87 percent and a binary classification accuracy of 98.08

percent on 25 COVID-19, 100 normal, and 100 pneumonia images.

Multiple building components, including convolution, pooling, and fully connected layers, make up CNN [29]. As shown in Fig 2.3, the convolutional layer's task is to apply convolution operation on an input image with learnable filters that are in a moving window fashion. Thus, creating input volume for the next convolutional layer.

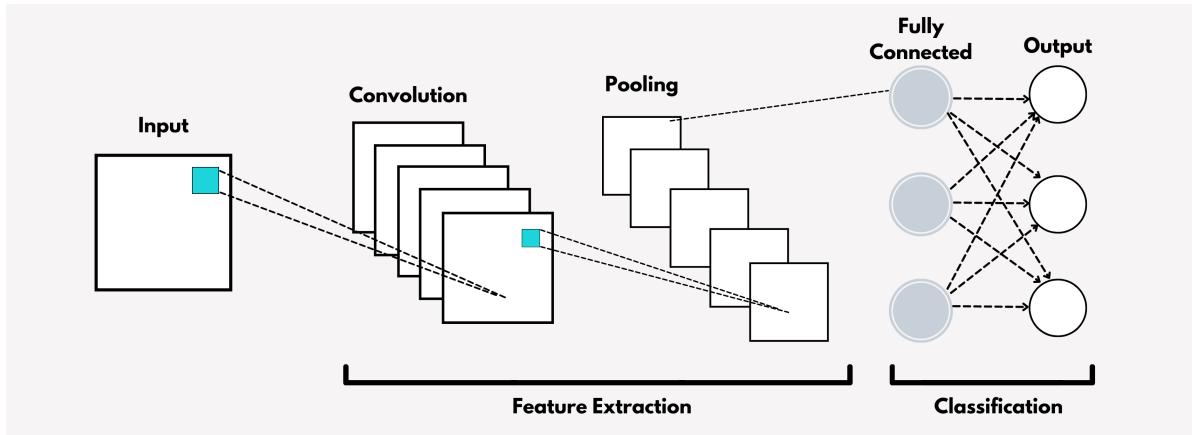


Fig. 2.3: Convolutional Neural Network Architecture

2.2 Transfer Learning

An alternative to training the CNN model from scratch, a Transfer Learning-based approach is used. As this methodology requires minimal CPU power and few to none training samples, it [30] became the most preferred method for implementing deep learning for medical image analysis. By leveraging transfer learning, the insights obtained from huge non-medical data can be transferred to address a specific medical problem (for e.g, AlexNet [4], VGGNet [31], and ResNet [5]) can be transmitted to a specific convolutional neural network (CNN) model to address a diagnostic imaging problem).

Transfer learning pre-trained model: AlexNet

After LeNet, Alex Krizhevsky [4] along with others proposed a wider and deeper CNN model and won the Imagenet large-scale visual recognition challenge (ILSVRC) in 2012. Against all the traditional computer vision and machine learning approaches, AlexNet achieved remarkable recognition accuracy. Sundaram and Lakhana [32] used AlexNet and Neural Network technology to partition and additionally amplify the data, without doing pre-training in order to acquire an AUC score of 94-95 percent. The best-performing classifier's accuracy of the area under the curve was 0.99, which was basically an ensemble of the GoogLeNet and AlexNet deep convolutional neural networks. The area under the curve of the pre-trained models was greater than that of the untrained models ($P < .001$). Further Augmentation of the dataset increased accuracy (P values for GoogLeNet and AlexNet were 0.02 and 0.03, respectively). Togaçar et al. [33] used chest X-rays for pneumonia detection in the lungs. They utilized convolutional neural networks as a means for extracting features by using the already present models of CNN which are AlexNet and VGG16. These models gather a huge number of parameters from the images, and they employed feature selection methods to cut down on the number of features. Additionally, they used traditional machine learning classifiers including LDA, DT, and linear regression for the pneumonia diagnosis and obtained great results, demonstrating the significance of DL and classification methods. Liang, G. et al. [34] proposed an approach to reduce calibration error while maintaining classification accuracy. The approach of [34] was evaluated across various datasets and architectures that included AlexNet [4], ResNet-50 [5], DenseNet-121 [35], and SqueezeNet1-1 [36]. Muhammed Talo et al. [37] proposed a method using the following pre-trained models: AlexNet, Vgg-16, ResNet-34, ResNet-18, and ResNet-50, to automatically categorise MR scans into normal, neoplastic, degenerative, cerebrovascular, and inflammatory disorders classes. The outcome of the proposed model of [37] was to help physicians to corroborate their manual findings of MRI images. They compared the classification performance of these state-of-the-art architectures in which they discovered that ResNet-50 outclassed the other four pre-trained models in their classification task, they attained an accuracy of 95.23% on their model. A. Titoriya and S. Sachdeva [38] made an effort to use deep convolution state-of-the-art architecture AlexNet to

learn and predict breast cancer from histopathological images. They fine-tuned the architecture by changing the final fully connected and classification layer.

Transfer learning pre-trained model: ResNet

Optimizing deeper networks is usually hard to perform, thus the authors of the ResNet architecture [5] employed a shallower model that has already been trained with extra layers to accomplish identity mapping, making the shallower network perform similarly to the deeper network, hence solving the degradation problem by including residual mapping. A classification approach that can evaluate chest X-rays and aid in a precise COVID-19 diagnosis was proposed by Shelke et al. [39], they used four classes for the classification of chest X-rays i.e. normal, pneumonia, tuberculosis (TB), and COVID-19. Additionally, the X-rays that show COVID-19 are divided into three severity categories: moderate, medium, and severe. VGG-16, a deep learning model with which they achieved test accuracy of 95.9 %, was used to classify pneumonia, tuberculosis, and normal, then they applied ResNet-18 for severity classification, reaching test accuracy up to 76 %, while the DenseNet-161 was employed to separate normal pneumonia from COVID-19 with a test accuracy of 98.9 %. Ansari et al. [40] used a pre-trained ResNet-50 model for pneumonia diagnostics, It was retrained on two distinct datasets of chest x-ray images using transfer learning and attained an accuracy of 96.76%. Liang et al. [41] used a novel network utilizing residual structure to get around the depth model's over-fitting and degradation issues, the input of the network was processed through 49 convolutional layers for child pneumonia image classification. Chouhan et al. [42] used standard network architectures (ResNet18, AlexNet, DenseNet121, InceptionV3 and GoogLeNet) and They presented an ensemble model that integrates the results from all the previously trained models. This strategy beat individual models, and they achieved state-of-the-art efficiency in pneumonia recognition with their proposed ensemble model, reaching an accuracy of 96.4% [43] used images of colon glands to train ResNet-18 and ResNet-50 in their study. The models were trained to discern between benign and malignant colorectal cancer. They further evaluated the prototypes using three different forms of testing data (20%, 25%, and 40% of the whole datasets). In three types of testing data, their empirical results showed that ResNet-50 performs more consistently than ResNet-18 in terms of accuracy, sensitivity, and specificity values. Narin et al. [44] proposed

an efficient alternative diagnosis option i.e. an automatic detection system for the prevention of COVID-19 virus from spreading among individuals. In this research, five models (ResNet50, ResNet101, ResNet152, InceptionV3 and Inception-ResNetV2) for the detection of patients with coronavirus pneumonia, pre-trained convolutional neural networks have been proposed, by utilizing chest X-ray radiographs. They employed three different binary classifiers with four classes (COVID-19, normal (healthy), bacterial pneumonia, and viral pneumonia) with five-fold cross-validation. The pre-trained ResNet50 model outperformed the other four models in terms of classification when taking into consideration the performance results obtained.

Transfer learning pre-trained model: Xception

Xception [45] or extreme inception is a deep convolutional neural network that introduces new layers to the already existing inception architecture. On the ImageNet dataset [46], Xception achieved the third best results. Ayan & Ünver [47] in 2019, developed a computer-aided diagnosis system which detected the presence of pneumonia in chest-X Rays. The transfer learning models used to pre-train were Xception and VGG16 on the ImageNet [46] dataset. According to the results, the accuracy of both models were 0.82% and 0.87% respectively. After comparing the results, it was concluded that Xception performed well in diagnosing Pneumonia in chest-XRays while VGG16 performed well in diagnosing normal chest-X Rays. Hashmi et al. [48] devised a weighted classifier that combined predictions of the five state-of-the-art transfer learning architectures namely Xception, ResNet18, DenseNet121, InceptionV3, and MobileNetV3. Each of the five pre-trained models were fine-tuned individually for pneumonia classification. The devised weighted classifier out-performed the individual pre-trained models by achieving an accuracy of 98.43%.

Transfer learning pre-trained model: VGG-16

The Visual Geometry Group (VGG) [49] architecture is a standard Convolutional Neural Network (CNN) architecture consisting multiple layers. It was based on the study of how to make these networks deeper. By using the Visual Geometry Group (VGG16) idea from Convolutional neural network (CNN) Naveen et al. [50] performed Pneumonia classification using Chest X-ray images for the purpose of aiding doctors in their decision-making process. In-order to gain better precision in training and validation, deep learning models were calibrated

using the VGG-16 model. Furthermore, their model was able to achieve a 95.67 % test precision and 96 % AUC score on the dataset. According to the study of [51], the already existing models for the classification of COVID-19 through Chest X-Ray images perform poorly because they fail to take into account the spatial relationships between the region of interests (ROIs) in CXR images, which could be used to determine the likely areas where COVID-19 has an impact on human lungs. Thus, the authors of [51] presented a novel attention-based VGG-16 model which basically uses an attention module on top of the VGG-16 model for the classification of COVID-19 Chest X-Ray images, they used three COVID-19 CXR datasets to assess their approach. Furthermore, their evaluation findings show that their strategy was effective for both training parameters and classification accuracy, thus they inferred from this outcome that their novel approach is more suited for the classification of COVID-19 Chest X-Ray images. The research of Habib et al. [52] proposed using Chest X-Ray images to diagnose pneumonia using an ensemble method. To extract features from provided X-ray images, two fine-tuned CNN models, CheXNet and VGG-19, were combined and trained. Following that, these features were then combined for classification. Moreover, their ensembled feature vector is subjected to the Random Under Sampling (RUS), Random Over Sampling (ROS), and Synthetic Minority Oversampling Technique (SMOTE) in order to address the issue of data irregularity in their dataset. They [52] further discovered that on the provided standard dataset, Random Forest outperformed the other approaches in terms of performance measures. They additionally compared their proposed method to current ones and it revealed that their method achieves increased classification accuracy, AUC values, and prediction accuracy of 98.93%.

2.3 Class Imbalance

Remarkable progress has been achieved in the field of medical image classification with the advancement in deep learning. Generally the deep models require a large number of labeled samples for training. But in many clinical and medical cases, it can be extremely hard to find/collect a balanced dataset in order to train because of high and low prevalence of some diseases which results in the imbalance between classes. Imbalance between classes means that the number of samples between different classes of the same dataset are not balanced. The

performance of models can be significantly affected in a negative way due to data imbalance as many models only perform well on balanced data [53]. In a neural network, if two classes (binary classification) of a dataset are highly imbalanced, the cost function will automatically be biased towards the majority class. Class imbalance also results as a major problem for algorithms that perform intelligent classification. Because of this, it is extremely important to select different evaluation metrics other than accuracy to address the imbalance problem as accuracy does not offer a fair measure of the algorithm's performance [54]. In 2004, Batista et al. [55] proposed a non-heuristic method to help combat the class imbalance in the dataset which was Random Under Sampling. The samples were randomly eliminated from the majority class which resulted in obtaining a balanced dataset. Lujan Garcia et al. [56] also proposed Random Under Sampling (RUS) to deal with highly imbalanced Pneumonia training and validation sets. They randomly eliminated 2534 Chest X-ray images in order to obtain an equal number of instances in each class. In 2017, Antin et al. [57] focused on binary classification of Chest X-rays with and without pneumonia, on one of the most extensive datasets available i.e. ChestX-ray8 compiled by NIH [11], to deal with imbalance in the dataset, data augmentation was performed by randomly flipping the images in horizontal direction. Moreover, weighted binary cross entropy was used as a loss function to account for the class imbalance. Srivastav et al. [58] applied the Deep Convolutional Generative Adversarial Network (DCGAN) augmentation technique and produced synthetic data in order to cater the imbalance problem of the dataset they were working on. Their model was able to achieve an accuracy of 94.5%.

2.4 Generative Adversarial Network GAN

The Generative Adversarial Networks (GANs) [59] have attained remarkable success in computer vision and natural language processing as being amongst one of the most innovative deep learning models in recent years. The concept of game theory is used by adversarial networks or GANs in particular, as they are trained to play a minimax game with a discriminator and a generator network, which aims to maximize a given objective function whereas a discriminator attempts to minimize the very same objective function, thus the term "adversarial".

GANs (Generative Adversarial Networks) [59] have increased in power, generating astonishing

realistic images that closely resemble the content of datasets they were trained to copy. The recurring question that arises in medical imaging is that whether GANs can generate acceptable medical data in the same way they make realistic RGB images.

Across multiple data regimens, GAN data-augmented models and standard augmented models trained on Chest X-Ray images (CheXpert dataset) were compared by Sundaram et al. [60], their findings demonstrate that GAN-based augmentation proved to be a useful method for addressing medical datasets with class imbalance problems. They further indicated there comparison results through AUC performance gains.

Another GAN-based framework was proposed by Malygina et al. [1] in order to cater the class imbalance problem which occurred in their model. Their classification task included 3 modalities namely pneumonia, fibrosis and pleural-thickening from one of the most extensive datasets available i.e. "ChestX-ray14". They applied CycleGAN and trained it on unpaired images such that it generates images from the opposite class for each input image. Furthermore, their results show that the classifier performance greatly improved for the pneumonia class however they couldn't achieve considerable changes for pleural-thickening, and also observed degradation of classifier quality on fibrosis, thus they concluded that the proposed GAN architecture is insufficient to handle such complex instances as fibrosis.

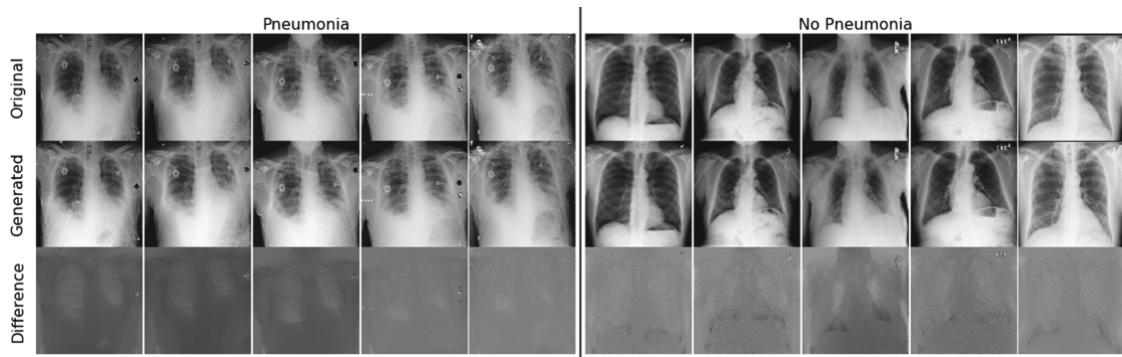


Fig. 2.4: Original pneumonia images from CheXnext and their CycleGAN generated pairs. Last row shows the difference between Original and Generated images; by [1]

Additionally on the "ChestX-ray14" dataset which is an extended version of the ChestX-ray8 dataset by the National Institute of Health (NIH), a Progressive Growing GAN (PGGAN)

architecture was used by Segal et al. [2] to perform unsupervised X-ray synthesis and to have radiologists assess the samples' clinical realism. They further applied the Fréchet Inception Distance (FID) which is basically a metric for evaluating the quality of the X-ray synthesis and discovered that the quality of the generated images is comparable to that of usual high-resolution tasks where PGGAN is used. FID works by integrating a set of real and generated images in the final average pooling layer of an ImageNet-trained Inception Net architecture. Furthermore, to evaluate the true realism of the generated images, they introduced a classification model trained on synthetic data and evaluated its performance on real scans. They concluded that it is possible to create realistic, medically credible images that capture most of the variability seen in ordinary X-rays; nevertheless, significant improvements in the replication of minor details in the X-ray abnormalities are still needed to get fully indistinguishable images.

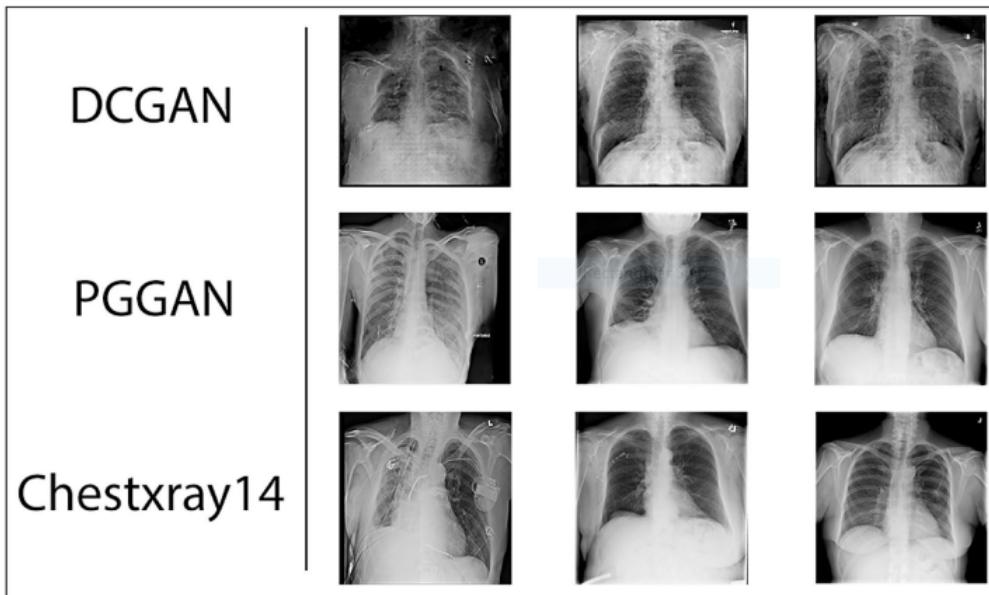


Fig. 2.5: Comparison of sample quality between different GAN architectures and the extended version of the ChestX-ray8 dataset i.e "ChestX-ray14" by [2]

GANs are used as a mean for creating more training data such that it improves classifier accuracy. Thus by using generative models, Motamed et al. [61] present a new GAN architecture namely Inception Augmentation GAN (IAGAN) for augmentation of chest X-rays to

perform semi-supervised detection of COVID-19 and pneumonia. They used two publicly available chest X-ray datasets to train their proposed IAGAN: one with normal and pneumonia images, and the other with normal, pneumonia, and COVID-19 images. They further demonstrated that a trained IAGAN model can produce new X-ray images independent of image labeling while also improving generative model accuracy and tested the IAGAN model's performance by training a DCGAN for anomaly detection (AnoGAN) through which they achieved improved results in terms of area under the receiver operating characteristic (ROC) curve (AUC) (0.90 on dataset 1 and 0.76 on dataset 2), sensitivity (0.82), and specificity (0.84). [61] concluded by contrasting their suggested augmentation GAN model with Deep Convolutional GAN (DCGAN) and conventional augmentation techniques (rotate, zoom, etc.), they demonstrate that their GAN-based augmentation method performs better than other augmentation techniques in terms of having trained a GAN to detect anomalies in X-ray images.

The authors of [62] used a transfer learning based approach by utilizing the VGG-16 model for pneumonia detection. The dataset used for their model was the Mendeley data of chest X-rays and performed the binary classification task (Pneumonia vs. No Findings), however they discovered that there was a class imbalance problem in their dataset as there were fewer images of a healthy person (No Findings). Thus in-order to cater the class imbalance problem Srivastav et al. [62] applied the Deep Convolutional Generative Adversarial Network (DCGAN) augmentation technique and produced synthetic data in order to oversample the dataset. Their model was able to achieve an accuracy of 94.5%.

Furthermore, Shin et al. [63] used a GAN-based technique to produce synthetic brain tumor MRI data and then tested the effectiveness of segmentation networks trained on the data. The segmentation networks trained purely with synthetic data don't really come close to those who were trained on real data, performance wise.

3. DATASET ANALYSIS

Both anatomically and functionally, imaging plays a critical role in examining lungs [64]. Chest imaging is used for a variety of purposes, including diagnosis, monitoring severity of disease, and screening. Computer-aided diagnosis (CAD) is a widely used collection of methods for detecting various types of anomalies in medical imaging. CAD systems that aid in the diagnosis of lung diseases involve chest X-ray images (X-Ray), Computed Tomography (CT) scans, and Magnetic Resonance Imaging (MRI). However, many cities, particularly in poor nations, lack access to diagnostic imaging such as CT and MRI. Although X-Ray scans may not be as accurate as CT and MRI, they are the least expensive amidst the radiological tests, allowing people with low-income to have access to these kind of tests. In many parts of the world, chest X-Ray, also referred to as radiography, has been the most routinely used diagnostic imaging technique for the diagnosis of lung disease [65]. This technology has numerous advantages, including simplicity, low radiation, low cost, huge amount of data, and wide availability. Machine learning models using X-Ray images show better accuracy and less training time [66] since Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) consist of multiple images to form a 3D image which makes the system complex. Also X-Ray images have larger public datasets available comparatively to MRI and CT.

For this Pneumonia classification model, the dataset that was opted is among the biggest datasets for chest X-rays available i.e. the Chest X-Ray8 dataset (CXR8). The Chest X-Ray8 dataset is published by the National Institute Of Health (NIH), they also published an extended version of this dataset i.e. Chest X-Ray14 (CXR14), introduced by Wang et al. [11] which contains 14 thoracic pulmonary diseases. The CXR8 dataset was used in this study instead of CXR14, as we only had to extract the pneumonia class from the dataset, thus it was irrelevant to use the dataset with more number of diseases.

3.1 Data Description

The ChestX-Ray8 dataset [11] was used in the development of this deep learning algorithm. This dataset is one of the most extensive chest X-Ray datasets available to the experimental research-based association, which is published on the National Institute Of Health (NIH). It contains 32,717 distinctive patients' frontal view X-ray images, averaging 112,120 in total, including many people with advanced lung disease. The X-Ray images from the CXR8 dataset were promptly extricated from the Digital Imaging and Communications in Medicine (DICOM) file, further the authors describe the image preprocessing strategy as restricting to resize the images to 1024x1024 resolution and constructing bounding boxes for the unique manifestation of the disease on each image. Each image in the ChestX-Ray8 dataset was annotated with the text-mined eight common thoracic labels/diseases (where every image has multi-labels or normal otherwise), generated with the use of natural language processing from the corresponding radiological records. Data preparation is often stated to be 70% of the effort in the data mining field thus Wang et al. [11] mined the radiology reports via two concepts, namely DNorm and MetaMap, DNorm is a disease recognition and normalization machine learning approach which assigns a unique ID to every mention of keywords in the report whereas MetaMap uses an ontology-based technique for the detection of biomedical text-corpus (keywords). The CXR8 dataset is freely available for research and performance evaluation of various computer-assisted detection techniques. Out of the 8 diseases in the dataset, we selected two classes of the dataset to work on i.e. Pneumonia and No findings Fig 3.2.

TYPE	NO. OF X-RAY IMAGES
NO FINDINGS	63,115
PNEUMONIA	322
TOTAL	63,437

Fig. 3.1: Dataset details

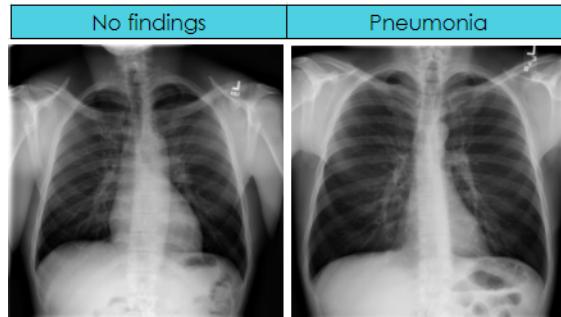


Fig. 3.2: No findings X-ray (left) and Pneumonia X-ray (right)

	Attributes	CX8
	Patients	30,805
	View	Frontal
Sex	Female	43.51%
	Male	56.49%
Age	0-20	6.09%
	20-40	25.96%
	40-60	43.83%
	60-80	23.11%
	80+	1.01%

Fig. 3.3: Important attributes of Chest X-ray8, According to this table, males are more likely to develop nosocomial bacterial and community acquired pneumonia than females and it's more likely to happen to people between the age group of 40-60.

3.2 Prior Work on the Chest X-Ray8 Dataset

Seewan et al. [67] presented an automatic classification of chest X-ray images linked to tuberculosis, they used two datasets from the National Library of Medicine (NLM) namely, Shenzhen No.3 Hospital X-ray set, and the NIH ChestX-ray8 dataset, they found out that by utilizing the Shenzhen hospital dataset, the model displayed an AUC score of 0.8502 and 0.9845, for detecting Tuberculosis while the AUC was dropped to 0.7054 for the ChestX-Ray8 dataset. It was further discovered that the final model predicted about 36.51% of aberrant pathology radio-graphs in the ChestX-ray8 dataset that had TB connections. Karargyris et al. [68]

developed an age prediction model by using the NIH ChestX-ray8 dataset, their results also focused on the stats in the dataset apart from images. The Chest X-Ray8 dataset is used in a recent study where [69] depicts that images having blur, low contrast, or other flaws have been shown to reduce the accuracy of classification models in studies. These difficult-to-identify images necessitate the usage of image metadata/ textual-information in order to fully comprehend and classify the image accuracy within a model. Thus, Caseneuve et al. [69] proposed a Sobel/Scharr operator edge detector followed by a CNN classifier for automatically identifying images as clear or deficient for further classification to avoid the necessity of textual input for classification. Another recent study where Bouslama et al. [70] presented a technique to automatically identify Cardiomegaly pathology from X-Ray images by using a Deep Convolutional Neural Network U-Net based architecture. As a result they were able to detect Cardiomegaly with an accuracy of 93 to 94 %, which was the greatest percentage recorded in the literature at the time of writing. Also according to the authors of this paper, this is the first study to use the U-Net technique to assess deep learning methods for detecting thoracic diseases. Pranav et al. [71] also utilized the chest X-Ray dataset to diagnose pneumonia with 85 percent accuracy. For the pneumonia patients, additional down-scaling and metadata are added to the images. The focus is on disease identification, with the image preparations being glossed over. Dunnmon et al. [72] aims at escalating the dataset capacity to improve the CNN classifier efficiency. The final images were preprocessed for data augmentation, histogram equalisation, mean standard deviation standardization, and downscaling, according to the paper. The published findings were found to be remarkable, which came out to be approximately 96%.

4. METHODOLOGY

4.1 Data Pre-processing

The original data set was created with the goal of identifying widespread thoracic pathology diseases [11] and consists of approx. 112,120 frontal view X-Ray images. Every X-ray image is categorized with one or more labels that demonstrate a link between diseases. Each image in the ChestX-Ray8 dataset was annotated with the text mined eight common thoracic diseases/labels, however out of the 8 diseases we extracted out the relevant classes that were to be used in our model i.e. "Pneumonia" and "No Finding". We observed that there was a class imbalance problem in the extracted dataset, as the count of images in the "Pneumonia" class was approximately 322 whereas in "No Finding" it was 63,115, creating a huge divergence in the model. This imbalance problem was catered by using imbalance techniques (Under-Sampling and Over-Sampling) and by using a new & innovative data augmentation technique known as Generative Adversarial Network (GAN) [59]. Furthermore, the Chest X-Ray8 dataset consisted of statistical data apart from images which included 11 variables out of which 4 variables are categorical and 7 are numeric, more statistics of CXR8 dataset are demonstrated in Fig 4.1, however as displayed in Fig 4.2, these variables didn't had an impact on the model as they weren't contributing in the classification of our model i.e. Pneumonia diagnosis; thus, they were ruled out for this classification problem.

4.1.1 Exploratory Data Analysis (EDA)

Overview Alerts 23 Reproduction

Dataset statistics

Number of variables	11
Number of observations	112120
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	33.8 MiB
Average record size in memory	315.9 B

Fig. 4.1: Overview of the Statistics of Chest X-Ray8 dataset

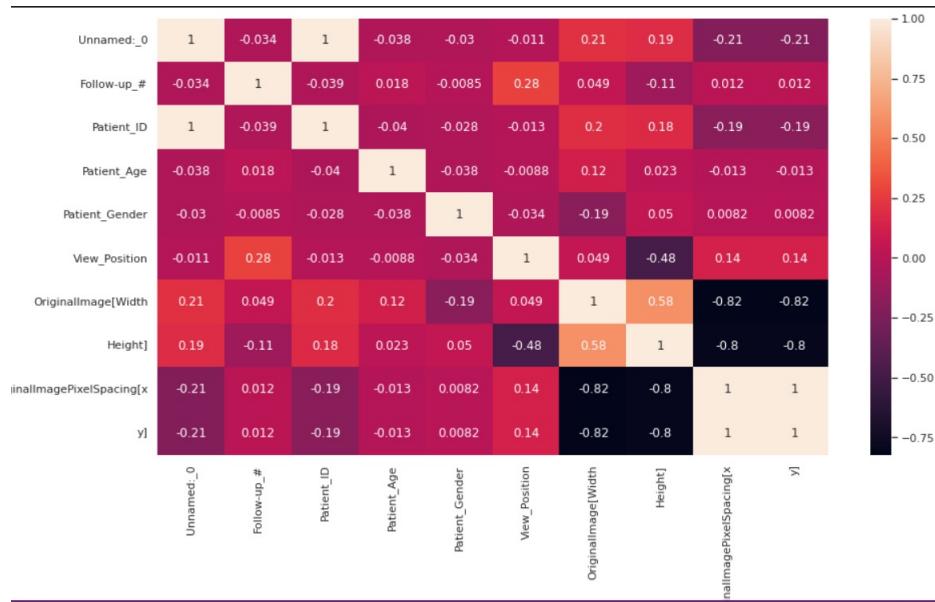


Fig. 4.2: Correlation of the statistical variables

A brief exploration of the data was also performed using unsupervised techniques. We start by plotting the principle components and categorizing them by class. It was observed that

the principal components do not seem to be clustered according to class as shown in Fig 4.3. The t-SNE (TDistributed Stochastic Neighbor Embedding) was also applied as it has been successful for visualizing high dimensional data [73]. Fig 4.4. illustrates the two-dimensional visualization however it doesn't seem to be clustered and cannot be distinguished by class.

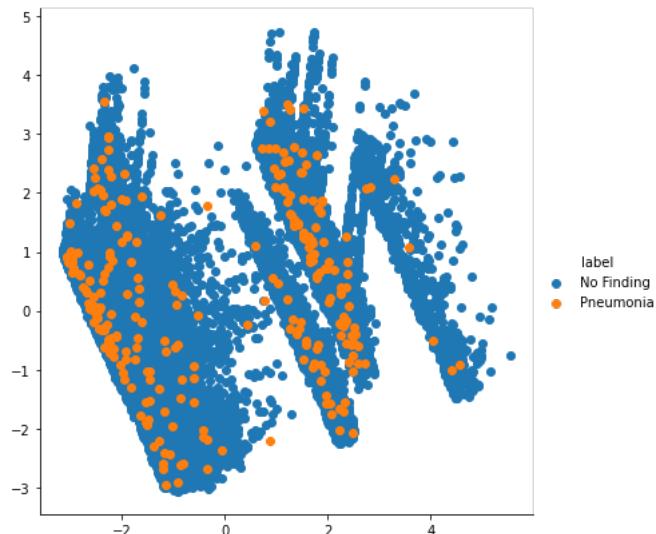


Fig. 4.3: Principle Component Analysis (PCA) projection of the ChestX-Ray8 dataset stratified by class (Pneumonia and No Finding) of 60k data-points

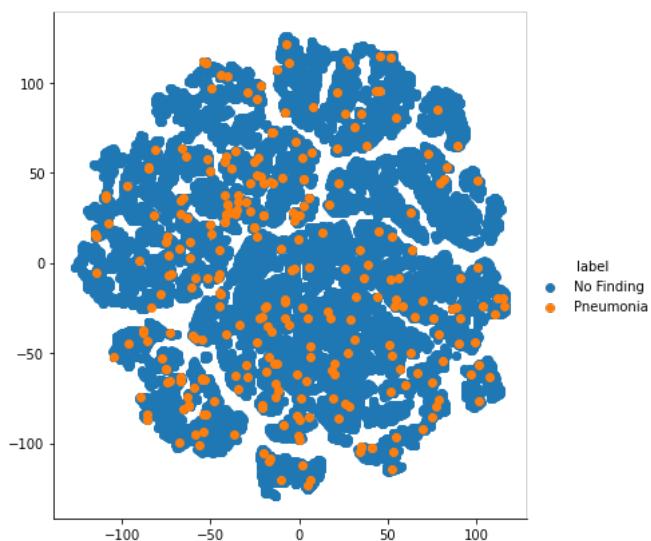


Fig. 4.4: t-SNE (TDistributed Stochastic Neighbor Embedding) of the ChestX-Ray8 dataset stratified by class (Pneumonia and No Finding) of 60k data-points

4.1.2 Data Imbalance Techniques

In real-world applications, class imbalance is one of the most everlasting challenges. that may challenge the typical supervised learning task. Multiple techniques for handling data imbalance have also been explored [74]. However, for the binary classification task, this issue occurs when quantity of samples in one class greatly surpasses the quantity of samples in the other class. Our dataset was highly imbalanced with "No finding" covering 99.46% of the total dataset while having only 322 images i.e less than 1% of pneumonia. Such highly imbalanced data leads to biasness in the model for the majority class and thus leads to poor testing accuracy. In order to tackle this problem, we employed a combination of data sampling techniques.

Data Sampling: Sampling procedures are used to overcome problems with a dataset's distribution. It involves artificially resampling the data either by under-sampling the majority class or by over-sampling the minority class.

Under-sampling Technique:

The most common approach is Random Under-Sampling (RUS) [75] which equalizes class by randomly removing the majority class samples (which is 'No Finding' in our case). We reduced 60,361 images of the 'No Finding' class to 30,000. The major drawback of this technique is the loss of samples that might have been useful for classifiers. Therefore, we explored the clustering-based under-sampling [76] and **Evolutionary Under-Sampling (EUS)** [77] method which uses an evolutionary/genetic algorithm to find the best subset of the majority class. The chromosomes and the fitness function are critical components in EUS. The fitness function analyzes the fitness of each alternative solution, whereas a chromosome encodes a possible solution to the problem. We look forward to applying these techniques and seeing their impact on the accuracies of the models in the coming months.

Oversampling Technique:

There are several oversampling methods like random oversampling [78] that randomly makes duplicates of the minority class, Synthetic Minority Oversampling Technique (SMOTE) [79] that creates new samples along the lines of a randomly chosen point and its nearest neighbor, and data augmentation which generates images by slightly modifying the existing images by rotating, flipping or changing the contrast.

4.2 Data Augmentation

4.2.1 Generative Adversarial Network (GAN)

Instead of using conventional data augmentation techniques (cropping, translation, rotation etc.), we enforced an innovative and state-of-the-art technique for data augmentation i.e. The Generative Adversarial Network (GAN) [59] for catering the imbalance problem in our dataset. GANs create one-of-a-kind images that closely resemble the original dataset's feature distribution. A generator and a discriminator are the two models in the GAN, simultaneously trained via an adversarial process as shown in Fig 4.5. The discriminator learns to distinguish between actual and fake images, while the generator learns to produce images that resemble real images. Until the discriminator is unable to distinguish between actual and fake images, we keep training both of these models. Using GAN can lead to the production of data keeping the criticality of the case in mind. It has shown promising results in image generation tasks while considering how difficult data assembling can be in the healthcare sector due to privacy concerns. GAN has been quickly embraced in a variety of healthcare and biomedical applications. With GAN, a neural network can be capable of generating hyper-realistic human faces [80], [81]. Moreover, GAN has also proved to be very beneficial in medical image analysis [82].

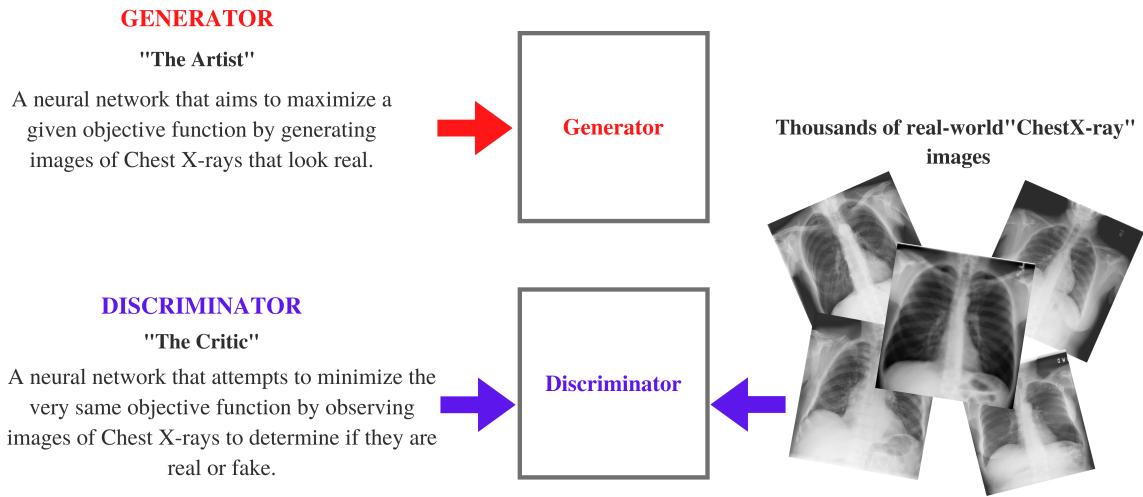


Fig. 4.5: The two models in GAN (Generator and Discriminator) that are trained simultaneously via an adversarial method

Furthermore, we applied GAN on the facial recognition dataset of our university's CS batch-2018. As shown in Fig 4.6, we didn't acquire quite clear and accurate results. Therefore, improved and optimized codes for GAN that'll give state-of-the-art results are being explored as in order for the model to give adequate results, the quality of the images should be clear and accurate, thus the intention is to investigate and apply this technique in the upcoming months.



Fig. 4.6: Generative Adversarial Network (GAN) applied on images of five girls from our batch with 64x64 image resolution.

We also implemented a few of the conventional techniques in the beginning: rotation that rotates the image to a certain degree, zoom modifies the image by zooming in or out, shear transformation which fixes all the points in the given line while shifting other points parallel to the distance proportional to their perpendicular distance from the line and horizontal flip to mirror the image in a horizontal direction.

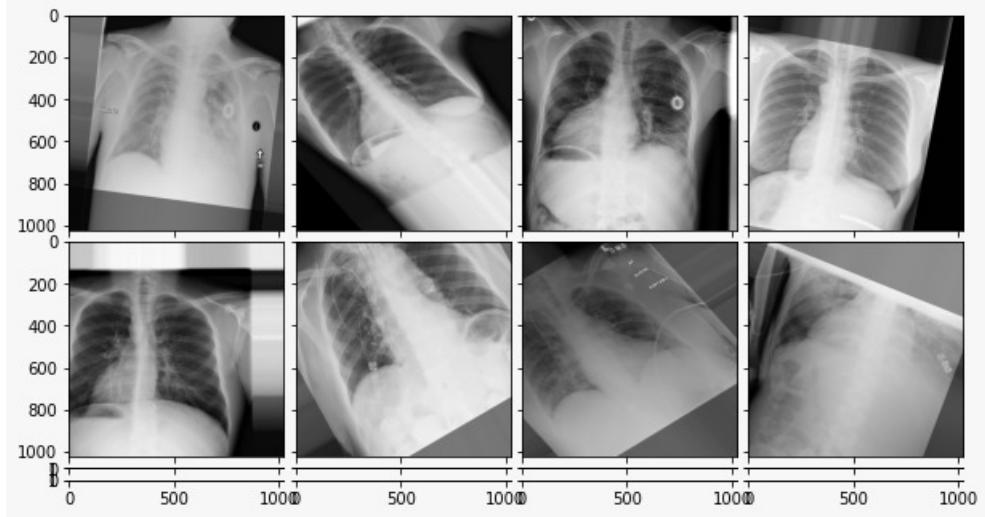


Fig. 4.7: Images generated through conventional augmentation techniques

4.2.2 Deep Convolutional Generative Adversarial Network (DCGAN)

An extension of the Generative Adversarial Network (GAN) that came out in 2016 is the Deep Convolutional Generative Adversarial Network (DCGAN) [83]. DCGAN is one of the most well-known and effective GAN implementations. DCGAN primarily consists of convolution layers in place of multi-layer perceptron, the convolution layers are implemented without any fully connected or max pooling layers that are used in vanilla GAN. For downsampling and the upsampling, DCGAN employs transposed convolution and convolutional stride. In general, the convolution network looks for spatial correlations to identify areas of correlation within an image. As a result, a DCGAN architecture would probably be more suited for image/video data, whereas a simple GAN's fundamental concept can be used to a larger range of domains because the model's specifics are left up to the individual model architectures. For our classification model DCGAN was trained on the minority class in our dataset i.e. Pneumonia (322 images) through which up-to 30,000 synthetic images were generated of the pneumonia class.

The DCGAN architecture guidelines are summarized as follows:

- Max pooling is entirely replaced with convolutional stride.
- For upsampling, transposed convolution is employed.

- Remove layers that are fully connected.
- Use batch normalization on all layers except the input layer of the generator and output layer of the discriminator.
- ReLU activation function is used in the generator excluding the output which uses the tanh function.
- Employ Leaky ReLU in the discriminator.

Initially, after applying the Deep Convolutional Generative Adversarial Network (DCGAN) model for oversampling the minority class in our dataset i.e. "Pneumonia", it was observed that the DCGAN model wasn't generating competent and reliable results as shown in Fig 4.9. Subsequently, the reason for these inadequate results were explored, to which it was discovered that the Binary Cross-Entropy Loss (loss function used in the DCGAN model) depicted below, wasn't converging well. The reasons might be: as the generator enhances throughout the training process, the discriminator's performance deteriorates because the discriminator cannot distinguish between real and fake. Hence the discriminator essentially tosses a coin to generate its prediction. This evolution is one possible issue for the GAN's overall convergence because the discriminator's feedback might have become less important through time. Moreover if the GAN proceeds to train even after the discriminator has provided fully randomized feedback, the generator begins to train on garbage responses, and its own quality might deteriorate.

Additionally, because the discriminator loss can occasionally grow to such a size that it prevents the GAN model from updating, the loss function may have reached zero for this reason as well. It may also occur if a huge number of weights are added and the model becomes overwhelmed, thus failing to update. Moreover, since the discriminator's confidence-values are a single value that can only be between 0 and 1, and the target is to get as close to 1 as possible, the calculated gradients decrease significantly and eventually become zero, thus as a result, the generator cannot obtain much information and cannot learn. Consequently, this could result in an excellent discriminator but it will lead to a weak generator.

$$BCELoss = \frac{-1}{N} \sum_{i=1}^N y_i.log(p(y_i)) + (1 - y_i).log(1 - p(y_i)) \quad (4.1)$$

As earlier GAN research exploited momentum to speed up training, Adam variant of stochastic gradient descent with a learning rate of 0.001, was used for optimization for both the generator and discriminator in our model. However, it was further determined that the recommended learning rate of 0.001 was too high and thus 0.0002 was used instead. Nonetheless the results were still not adequate.

Layer (type)	Output shape	Param #
<hr/>		
dense_1 (Dense)	(None, 8192)	827392
leaky_re_lu_5 (LeakyReLU)	(None, 8192)	0
reshape (Reshape)	(None, 8, 8, 128)	0
conv2d_transpose (Conv2DTran)	(None, 16, 16, 128)	262272
leaky_re_lu_6 (LeakyReLU)	(None, 16, 16, 128)	0
conv2d_transpose_1 (Conv2DTr)	(None, 32, 32, 128)	262272
leaky_re_lu_7 (LeakyReLU)	(None, 32, 32, 128)	0
conv2d_transpose_2 (Conv2DTr)	(None, 64, 64, 128)	262272
leaky_re_lu_8 (LeakyReLU)	(None, 64, 64, 128)	0
conv2d_transpose_3 (Conv2DTr)	(None, 128, 128, 128)	262272
leaky_re_lu_9 (LeakyReLU)	(None, 128, 128, 128)	0
conv2d_5 (Conv2D)	(None, 128, 128, 3)	24579
<hr/>		
Total params:	1,901,059	
Trainable params:	1,901,059	
Non-trainable params:	0	
<hr/>		

Layer (type)	Output shape	Param #
<hr/>		
conv2d (Conv2D)	(None, 128, 128, 128)	3584
leaky_re_lu (LeakyReLU)	(None, 128, 128, 128)	0
conv2d_1 (Conv2D)	(None, 64, 64, 128)	147584
leaky_re_lu_1 (LeakyReLU)	(None, 64, 64, 128)	0
conv2d_2 (Conv2D)	(None, 32, 32, 128)	147584
leaky_re_lu_2 (LeakyReLU)	(None, 32, 32, 128)	0
conv2d_3 (Conv2D)	(None, 16, 16, 128)	147584
leaky_re_lu_3 (LeakyReLU)	(None, 16, 16, 128)	0
conv2d_4 (Conv2D)	(None, 8, 8, 128)	147584
leaky_re_lu_4 (LeakyReLU)	(None, 8, 8, 128)	0
flatten (Flatten)	(None, 8192)	0
dropout (Dropout)	(None, 8192)	0
dense (Dense)	(None, 1)	8193
<hr/>		
Total params:	602,113	
Trainable params:	602,113	
Non-trainable params:	0	

Fig. 4.8: DCGAN Generator and Discriminator Architecture



Fig. 4.9: DCGAN distorted results of 128x128 image resolution with Binary Cross-Entropy Loss function at 80K iteration

To advance, we basically adopted a fresh strategy for a new cost function and made some required changes in our model by using the same parameters and network structure provided by DCGAN, however updating the generator and discriminator parameters suggested by a distinct GAN architecture i.e. Wasserstein GAN gradient penalty (WGAN-gp).

4.3 Wasserstein GAN gradient penalty (WGAN-gp)

Up-till now we have seen that Generative Adversarial Networks GANs are excellent generative models, however they may suffer from training instabilities. Thus after some thorough research, it was discovered that one way to solve this issue and the issues mentioned in Sec 4.2.2 was to use Wasserstein GAN [3] as it makes progress towards stable training of GANs. The Wasserstein GAN gradient penalty (WGAN-gp) [84] is a generative adversarial network that uses the gradient norm penalty and the Wasserstein loss formulation to achieve Lipschitz continuity [84].

The main benefit of The Wasserstein GAN gradient penalty (WGAN-gp) is its convergence. It improves training stability, hence making it easier to train. The details of the Wasserstein loss with gradient penalty are as follows:

- The difference between the desired value of the discriminator's output for actual images and the discriminator's output expected value for artificially generated fake images

is what constitutes Wasserstein's loss.

- The discriminator wishes to increase the above gap, whereas the generator aims to decrease it.
- WGAN-gp employs gradient penalty rather than weight clipping to impose the Lipschitz constraint.
- Batch normalization is not anymore employed in the critic (discriminator) since it connects a set of inputs with a set of outputs. What we require is to be able to determine the gradients of each output relative to its corresponding inputs.
- So, the Generator's ultimate goal is to raise the mean of the fake output produced by the discriminator. Whereas, the discriminator's objective is Wasserstein loss along with weighted penalty.

$$W - Loss = E_{x \sim P_g}[D(x)] - E_{x \sim P_r}[D(x)] + \lambda E_{x \sim P_x}[(\|\nabla_x D(x)\|_2 - 1)^2] \quad (4.2)$$

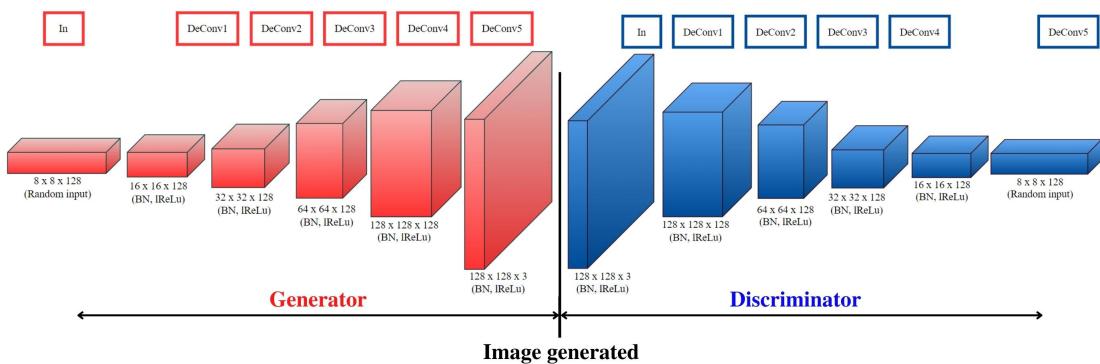


Fig. 4.10: Generator and Discriminator architecture employed for the presented Pneumonia classification task

In vanilla GAN, the loss evaluates how effectively it deceives the discriminator instead of measuring the image quality. The generator loss in GAN is not reduced even as the image quality rises as shown in Fig 4.11 and as a result, we are unable to determine progress from its value.

On the other hand, the more desirable image quality is reflected by the WGAN loss function as the loss drops significantly, and sample quality also improves as shown in Fig 4.12. For WGAN, two key contributions are:

- In experiments, it shows no signs of mode collapse.
- The generator can continue to learn once the critic (discriminator) performs effectively.

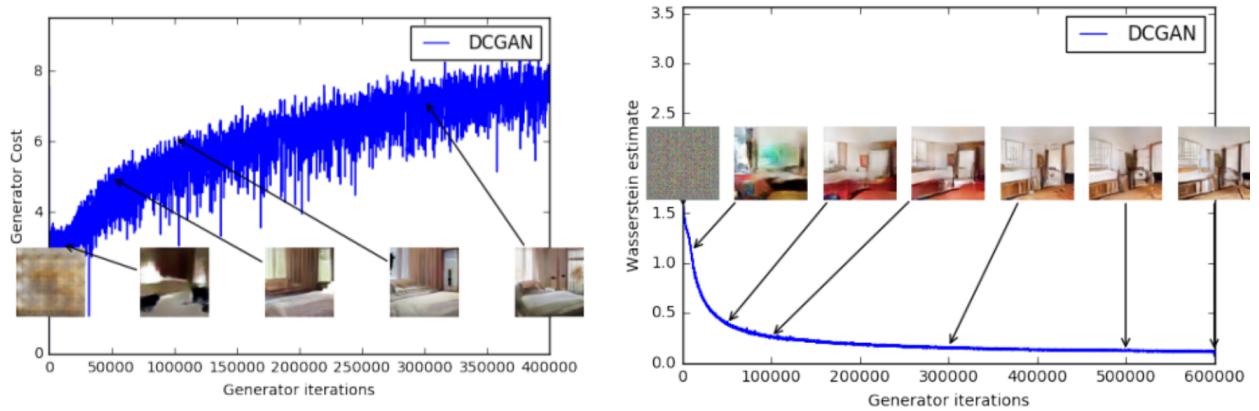


Fig. 4.11: Generator's cost during GAN training (left). Generator's cost during DCGAN training with Wasserstein's loss (right) by [3]

We utilized the Wasserstein's loss which is basically used by the WGAN-gp model instead of the Binary Cross Entropy Loss because of the issues discussed previously in Sec 4.2.2. During WGAN training, the discriminator is trained several times for each step, whereas the generator is trained once per step. As a result, we trained the discriminator more than the generator, therefore the generator was updated after every 5 epochs, while the discriminator was updated after each epoch. Following this, we combined the modifications from DCGAN and WGAN into a single model because the areas under their control are distinct. We applied the convolution layers used in DCGAN in both Generator and Discriminator while WGAN-gp was utilized to do the following:

- Adjust the loss function (Wasserstein's loss was used instead of BCE loss)
- The Rmsprop optimizer was used for discriminator while Adam optimizer was used for generator.

- The activation function used in the discriminator was removed as WGAN-gp takes in the difference between the discriminator's output's expected value for real images and its expected value for the generated fake images.



Fig. 4.12: WGAN-gp generated results applied on the minority class "Pneumonia" with 64x64 image resolution, these results were further used in our classification task.

4.4 Explanation about the Transfer Learning models

Following the successful GAN implementation for augmentation of the minority class "Pneumonia" and doing Random Undersampling (RUS) for the majority class "No Finding", the imbalance problem in the dataset was finally catered. Consequently, now for the classification task, we employed a Transfer learning approach where we utilized four transfer learning models namely, ResNet, Xception, VGG-16 and Alexnet to extract features and then these features were fed into the classifiers of the various models.

4.4.1 AlexNet

AlexNet [4] has been an important milestone in the field of Computer Vision after it won the Imagenet large-scale visual recognition challenge (ILSVRC) with significantly improved accuracy. ILSVRC evaluates algorithms for image classification and object detection at large scale. AlexNet's network has eight layers with learnable parameters. Its layers consist of a combination of 5 convolutional layers(three pooling layers and two fully connected layers). The activation function used in each function is ReLU [85] except for the output layer(it uses Softmax).

Main Features of AlexNet architecture:

- It uses ReLU non-linear activation function that efficiently solves the vanishing gradient problem as compared to sigmoid and tanh used in other neural networks.
- To prevent overfitting, it adds data augmentation and dropout in the network layer.
- To reduce the size of the network, overlapping pooling was also used.
- It uses multiple parallel GPUs for accelerating computational throughput during training.

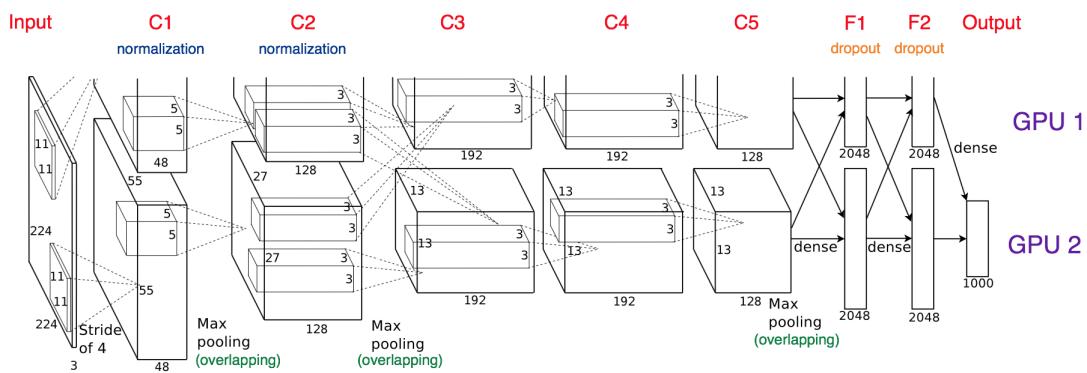


Fig. 4.13: Architecture of AlexNet: Convolution, max-pooling, LRN and fully connected (FC) layer (Alex Krizhevsky et al. [4])

The dimension of input image for the AlexNet has to be $227 \times 227 \times 3$ and the first convolution layer converts input image with 96 kernels sized at $11 \times 11 \times 3$ having a stride of 4 pixels, which is the input to the second layer.

AlexNet Experimental details:

- Comprises of eight layers incorporating learning parameters.
- RGB images as input to the model.
- Combination of max pooling layers with 5 convolution layers.
- Incorporates 3 fully connected layers.
- Relu is used the act function in each layer.
- Moreover, has 2 dropout layers.

- Output layer comprises of softmax as the activation function.
- The architecture, hence, comprises of 62.3m parameters.

4.4.2 ResNet

In theory, the deeper the neural network the better the performance of the network as the intuition is that by adding additional layers they can progressively learn more complex features.

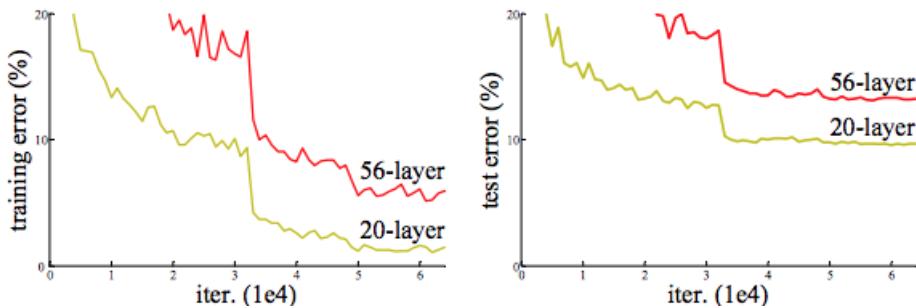


Fig. 4.14: Training error (left) and Test error (right) on CIFAR-10, The deeper network has higher training error and thus test error [5]

This plot by Kaiming He et al. [5] denies the above theoretical fact and tell us that adding more layers to a neural network can actually do the opposite i.e. more layers = lower accuracy. Practically learning the identity function (mapping input directly to the output) is very difficult in a finite amount of time and data as it converges very slowly (degradation). Kaiming He et al. [5] diminished the problem of training very deep networks with the introduction of a new neural network layer: The Residual Network.

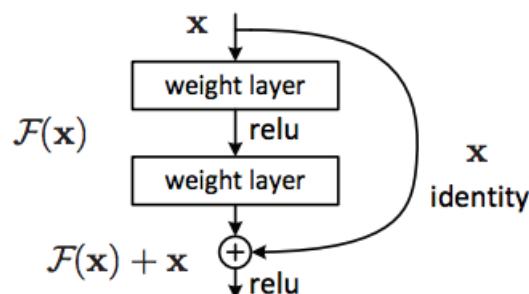


Fig. 4.15: Residual Learning: a building block (Kaiming He et al. [5])

The goal of the residual network is to skip connections and perform identity mapping. The identity mapping has no parameters and serves only to add the results (output) from the previous layer to the next layer. The identity shortcut connections don't add any extra parameters or complexity to the computation. Thus through these skip connections, ResNet solves the degradation problem, which allows it to take advantage of adding hundreds of new layers without diminishing results. However, when we're skipping connections we can have the problem of x and $F(x)$ having different dimensions, such as when we apply the convolution operation, the spatial resolution of the image shrinks thus linear projection is used to expand the channels of shortcut to match the residual.

$$y = F(x, W_i) + W_s x \quad (4.3)$$

ResNet Experimental Details:

- After each convolution and before activation, batch normalization is utilized.
- He Initializer [86] (named after his name Kaiming He, which is used to cater the vanishing/exploding gradient problem during back-propagation, it takes into account the non-linearity of activation functions, such as ReLU activations.)
- A maximum batch size of 256 was used.
- When the error rate reaches a plateau, the learning rate (i.e. 0.1) is divided by ten.
- Trained for 60 x 104 iterations.
- Weight decay of 0.0001, momentum of 0.9.
- Did not use Dropout.
- Test-time augmentation (10-crop testing): They didn't just predict on the test image, what they did is that they took 10 crops from the test image and then the model predicts on each of the crop images and further they average the prediction to form the final prediction.

4.4.3 Xception

Xception or Extreme inception [87] is based on stronger hypothesis of the inception module i.e cross-channel correlations and spatial correlations can be mapped entirely separately [87]. The Inception architecture [88] is based upon the idea how a convolutional network can approximate and cover the easily available dense components through its optimal local sparse structure. It [88] computes 1×1 , 3×3 , and 5×5 convolutions within the same layer of the network and then the output of these filters are stacked along the channel dimension and before being passed on to the next layer in the network. Xception is an extension of inception module, following depthwise separable convolution. It has 36 convolutional layers with residual connections. All Convolution and Separable Convolution layers are followed by batch normalization. The residual connections proposed by He et al. [5] helps the architecture in convergence of both speed and classification accuracy.

Xception Experimental Details:

- Optimizer: SGD is preferred for imagenet dataset.
- Momentum: 0.9.
- Initial learning rate: 0.045.
- Learning rate decay: decay of rate 0.94 every 2 epochs.
- Uses dropout with the rate of 0.5 but if the dataset is very large dropout is not needed.
- uses L2 regularization of weight decay at the rate of 1e-5.

4.4.4 The Visual Geometry Group (VGG-16)

In 2014, Karen Simonyan and Andrew Zisserman of Oxford University's Visual Geometry Group Lab proposed VGG-16 [49]. Their proposed model won 1st and 2nd place in the object localization and image classification category. The model achieved a top-5 test accuracy of 92.7% on the ImageNet dataset that contains 1000 classes and a total of 14 million images. VGG-16 has sixteen layers with weights (learnable parameters). There are thirteen

convolutional layers, three dense layers, and five max pooling layers making a total of 21 layers. The reason it is named VGG-16 is that only 16 layer comes with learnable parameters.

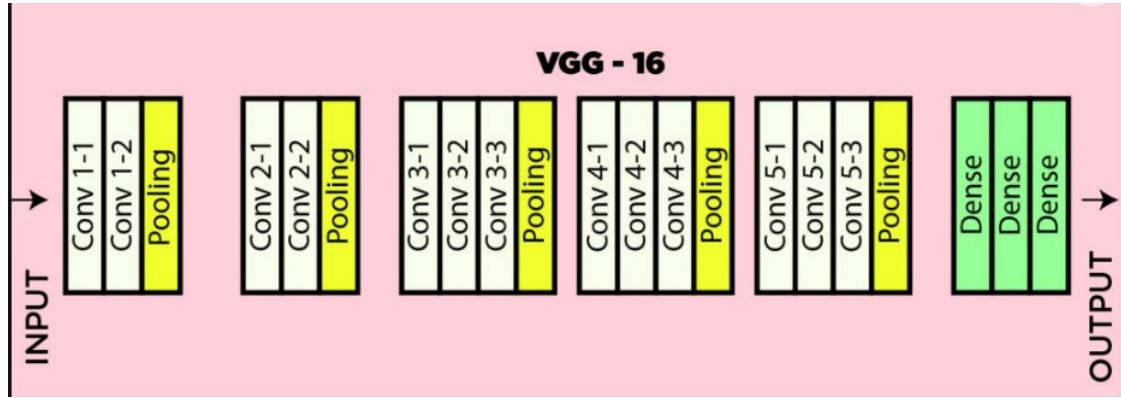


Fig. 4.16: Architecture of VGG-16: Convolution, max-pooling, and dense layers by [49]

VGG-16 Experimental Details:

- VGG-16 takes input tensor size as 244×244 with three RGB channels,
- Instead of having a vast number of hyper-parameters, VGG-16 is focused on having a convolution layer of 3×3 with a stride of 1.
- The number of filters in the first convolution layer or Conv-1 are 64. The filters in Conv-2 are 128, Conv-3 are 256, and Conv-4 and Conv-5 have 512 filters.
- The first two layers of the fully connected layers have 4096 channels and the third has 1000 channels.
- The final layer has softmax as the activation function.

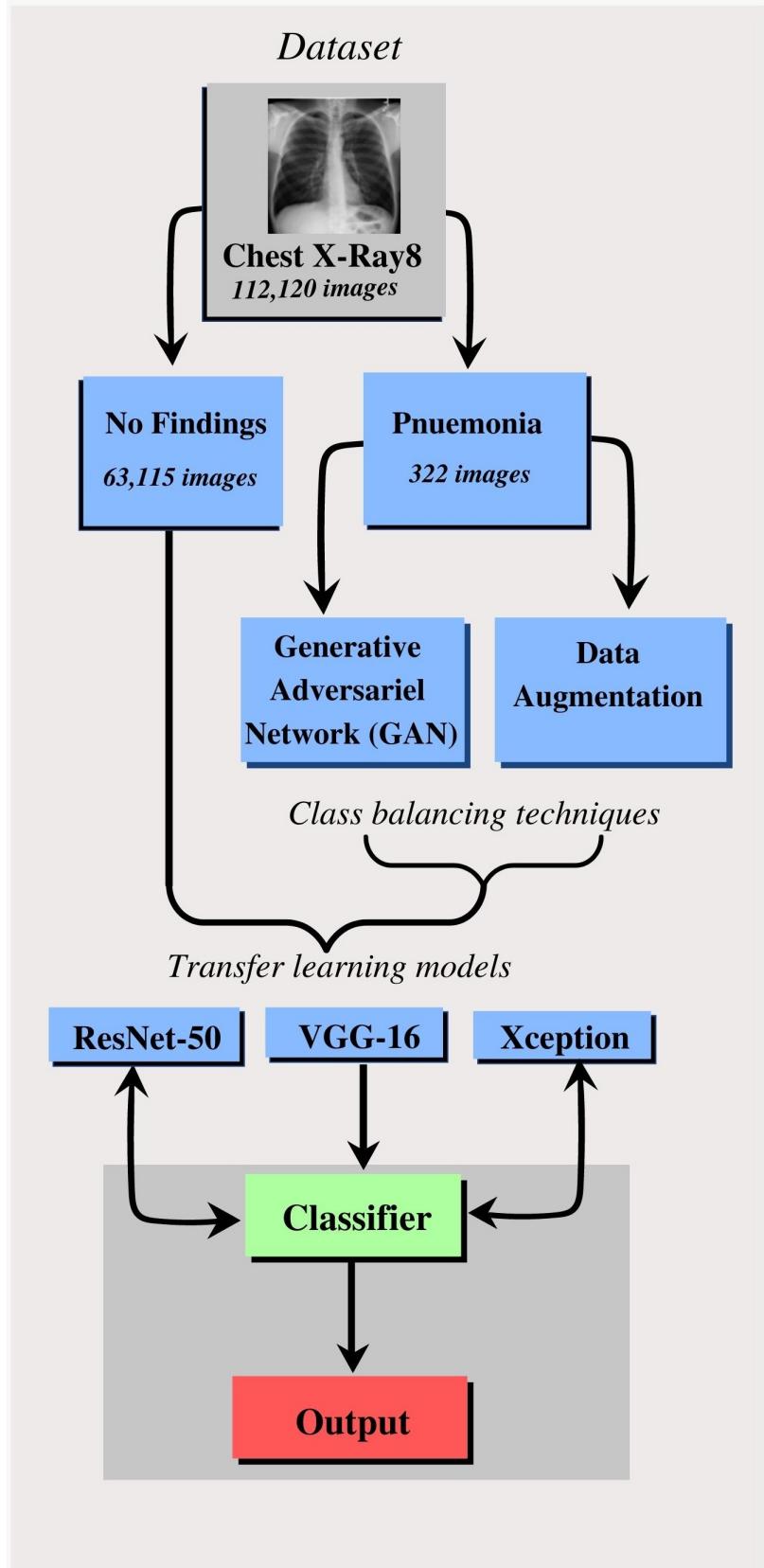


Fig. 4.17: Complete methodology of the proposed Pneumonia classification model

5. EXPERIMENTAL RESULT

In this chapter, we demonstrate the results of the experiments conducted for the pneumonia classification model, as well as the training and test findings acquired.

5.1 *Performance analysis of the models*

We performed certain experiments on the training set, by training the baseline models, and then testing the model with the test set.

The ChestX-Ray8 dataset was split into 70% train, 15% test and 15% validation sets after which the imbalance problem that was found in the dataset was catered by performing Random under-sampling on the majority class (No Finding) and applying an innovative and state-of-the-art technique for data augmentation i.e. The Generative Adversarial Network (GAN) for the minority class (Pneumonia). We then compared the state-of-the-art four baseline models (ResNet-50, VGG-16, Xception, Alexnet) used, on the basis of sensitivity (recall), precision and F1-Score as depicted in Table 5.2, through which we observed that the model still isn't giving state of the art results on the dataset used for this classification task, therefore further research and investigation is being performed to generate much better results in future.

Tab. 5.1: The ChestX-Ray8 dataset after catering the imbalance problem

Classes	Before Augmentation	After Augmentation
No Findings	63,115	30,000
Pneumonia	322	30,000

Tab. 5.2: Classification Report of state-of-the-art on the "ChestX-Ray8" dataset for pneumonia classification on the test set

Model	Epochs	Precision (%)	Recall (%)	F1-Score(%)
ResNet-50 (No Finding)	10	50%	54%	52%
ResNet-50 (Pneumonia)	10	51%	47%	49%
VGG-16 (No Finding)	10	49%	50%	50%
VGG-16 (Pneumonia)	10	50%	49%	50%
Xception (No Finding)	10	51%	51%	51%
Xception (Pneumonia)	10	51%	51%	51%

In order to depict a more precise analysis of the baseline models, we demonstrate the training and validation accuracy along with their losses on the augmented dataset.

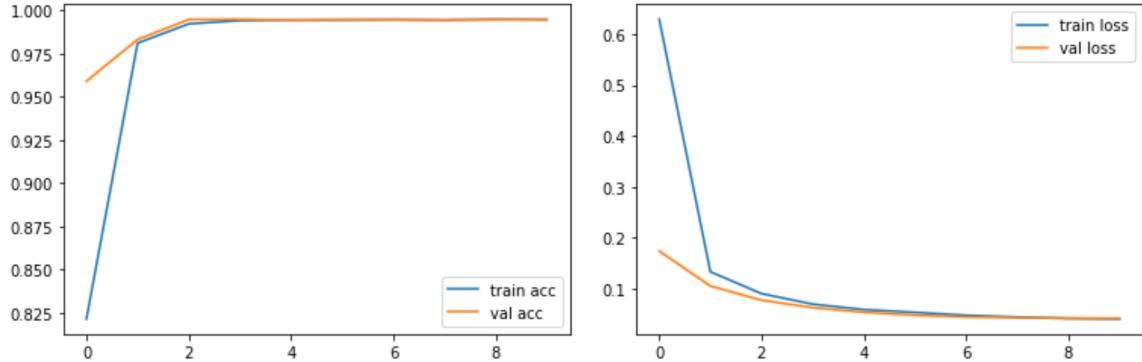


Fig. 5.1: ResNet training accuracy and loss on Chest X-Ray8 extracted dataset

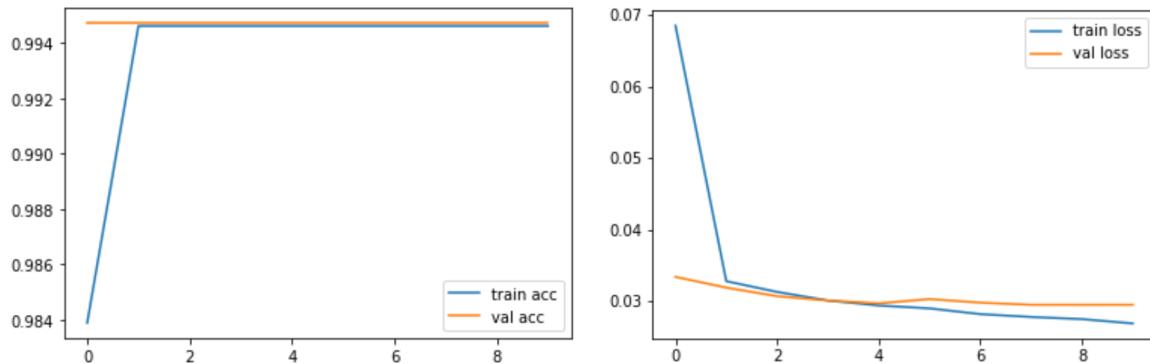


Fig. 5.2: VGG-16 training accuracy and loss on Chest X-Ray8 extracted dataset

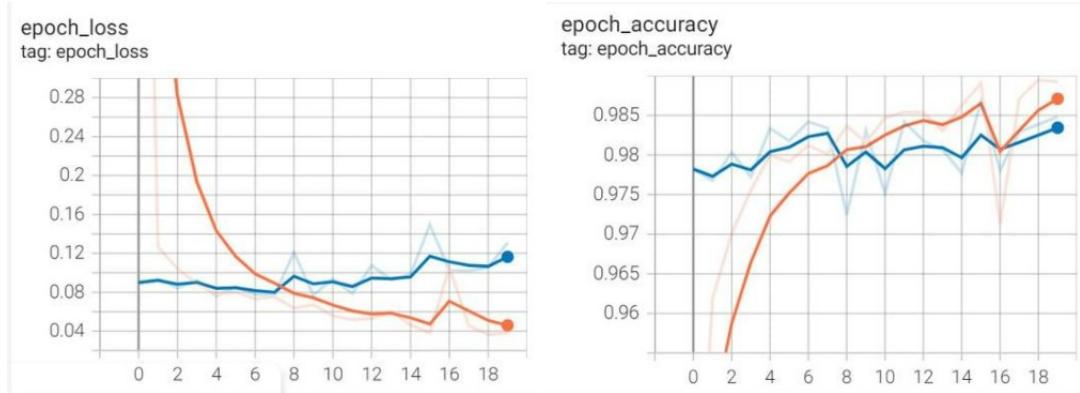


Fig. 5.3: Alexnet training accuracy and loss on Chest X-Ray8 extracted dataset

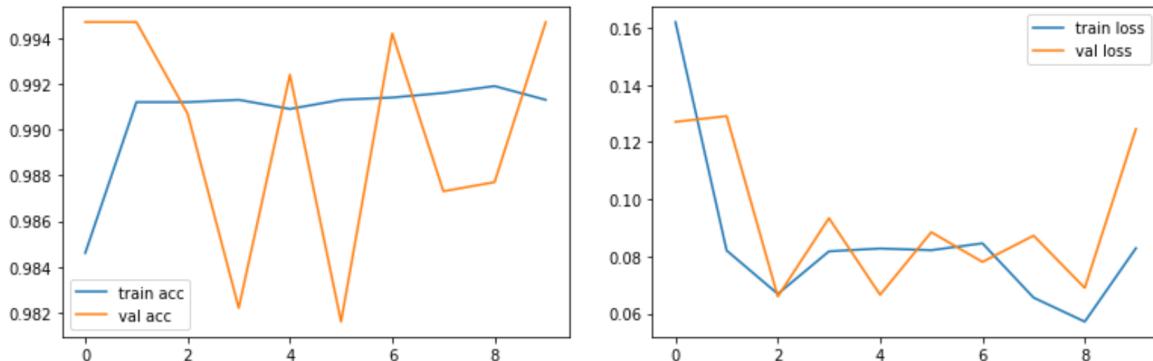


Fig. 5.4: Xception training accuracy and loss on Chest X-Ray8 extracted dataset

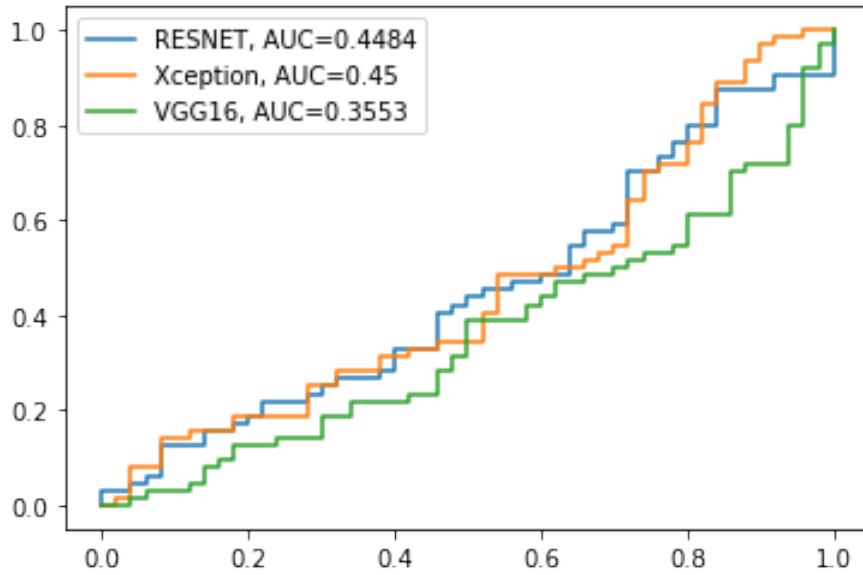


Fig. 5.5: ROC curves of the three models used for classification without applying WGAN loss function for augmentation along with their respective AUC score through which we determine our error metric

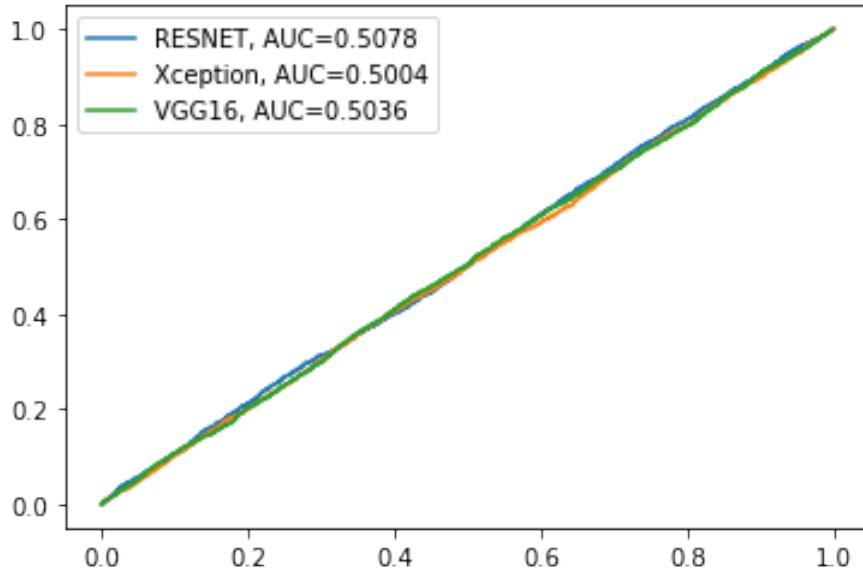


Fig. 5.6: ROC curves of the three models used for classification after applying WGAN loss function with DCGAN architecture, along with their respective AUC score through which we determine our error metric

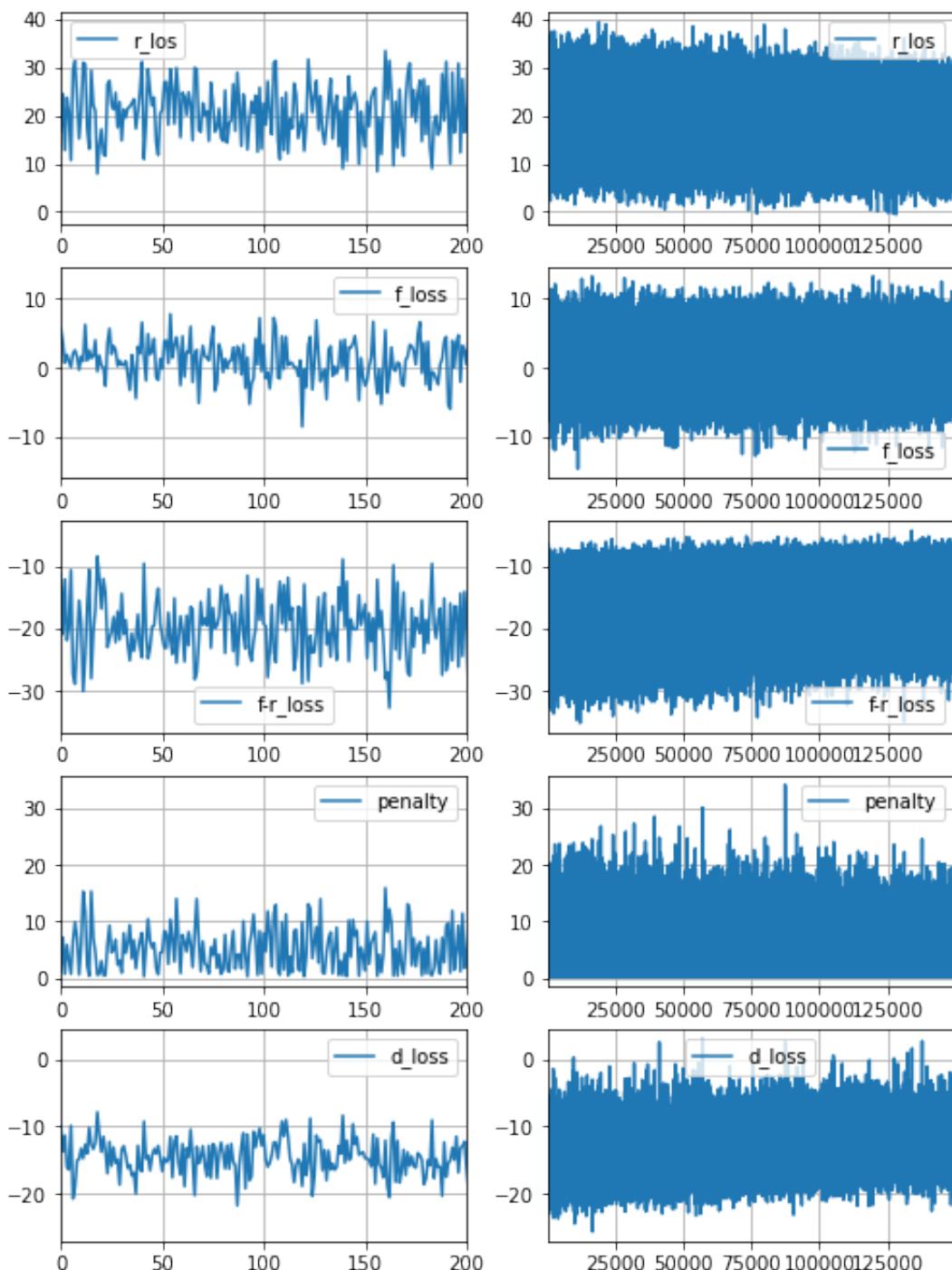


Fig. 5.7: The Generative Adversarial Network (GAN) loss details after applying the Wasserstein Loss with gradient penalty (WGAN-gp)

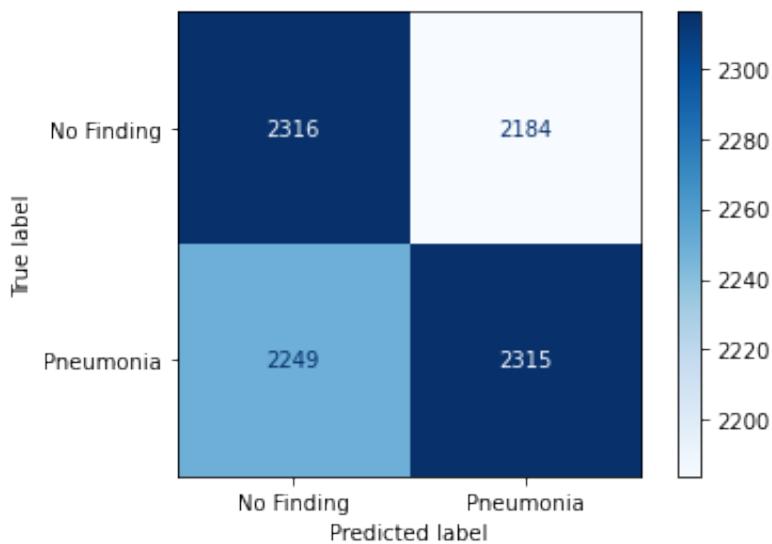


Fig. 5.8: Confusion matrix of Xception model on the two classes "Pneumonia" and "No Findings"

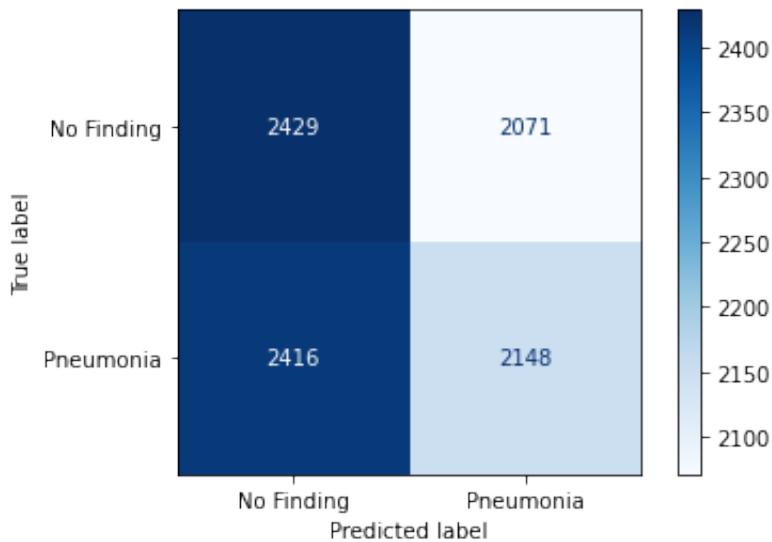


Fig. 5.9: Confusion matrix of ResNet model on the two classes "Pneumonia" and "No Findings"

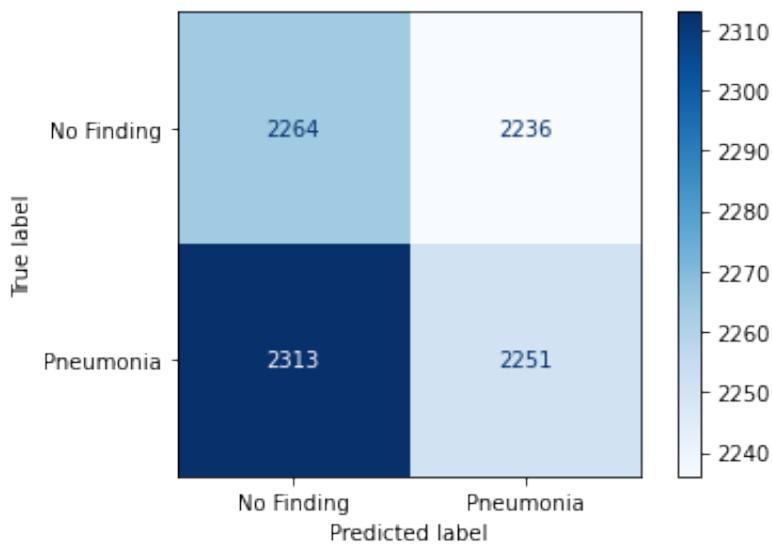


Fig. 5.10: Confusion matrix of VGG-16 model on the two classes "Pneumonia" and "No Findings"

The results of each pre-trained model in terms of the training set used for this classification task are depicted in Table 5.3.

Tab. 5.3: Model Results on the train set for comparison

Model	Epochs	Loss	Accuracy	Validation Loss	Validation Accuracy
ResNet-50	10	47.90%	77.96%	36.25%	89.14%
VGG-16	10	9.06%	97.64%	8.52%	97.86%
Xception	10	6.48%	99.18%	8.79%	99.47%

5.2 Deliverables

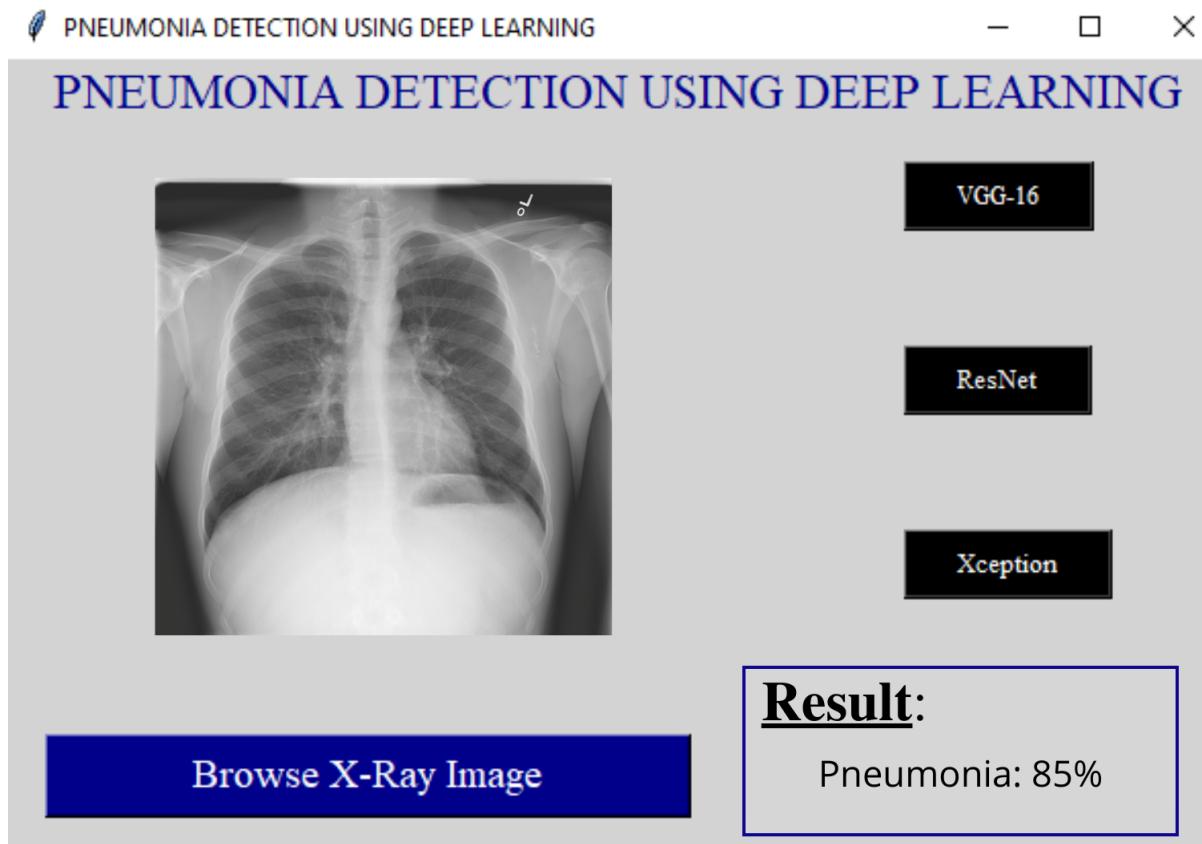


Fig. 5.11: The Graphical User Interface (GUI) of the presented Pneumonia classification model based-on pre-trained architectures

6. CONCLUSION AND FUTURE WORK

Pneumonia is one of the biggest threat to human life all over the world. Early diagnosis of pneumonia is essential to choose the best treatment method and further prevent against the infection that is endangering the patient's life. An X-Ray scan is frequently used to aid in the diagnosis of pneumonia. The abscess and fluid produced by pneumonia creates radiopaque portions (white patches) in the lungs. Regardless of the presence of pneumonia on the X-Ray images, the diagnosis is always dependent on the doctor's expertise and experience. However, due to a scarcity of competent medical radiologists in developing nations, examining chest X-Rays is a difficult undertaking as it is vulnerable to subjective uncertainty. As a result, an automated system is constantly needed to quicken the image analysis process and assist radiologists in early diagnosis of the deadly Pneumonia disease. These automated solutions allow radiologists to get another opinion and thus make better and precise decisions in regard to pneumonia diagnosis.

Thus, if the gap in diagnosing various types of pneumonia is not bridged with robust automated techniques of chest disease identification, the healthcare sector may be forced to face undesirable situations. An automated Computer Aided Diagnosis (CAD) system was proposed in this thesis study to aid medical practitioners for diagnosis. The system uses a transfer learning approach for the classification of X-ray images into 2 classes "Pneumonia" and "No Finding". Three different pre-trained CNN models were applied namely, ResNet-50, VGG-16, AlexNet and Xception to extricate features and then these features were fed into the classifiers of the various models, and the results were gathered of distinct architectures.

As previously mentioned, we've applied augmentation techniques for oversampling and undersampling the data, as the extracted dataset used for this classification model was highly imbalanced. Thus, we enforced an innovative and state-of-the-art technique for data augmen-

tation i.e. The Generative Adversarial Network (GAN) which was improvised by using a combination of Deep Convolutional Generative Adversarial Network (DCGAN) and Wasserstein GAN gradient penalty (WGAN-gp), applied on the minority class "Pneumonia" for augmentation. Whereas Random Under-Sampling (RUS) was done on the majority class "No Findings" through which both the classes were balanced as depicted in Table 5.1

Furthermore, we look forward to explore additional classification techniques as well which may lead to develop our own CNN model instead of using transfer learning approach. Additionally we intend to explore the ensemble framework as instead of training a single classifier, we can train a variety of classifiers and clusters altogether for producing better results.

BIBLIOGRAPHY

- [1] T. Malygina, E. Ericheva, and I. Drokin, “Data augmentation with gan: Improving chest x-ray pathologies prediction on class-imbalanced cases,” in *Analysis of Images, Social Networks and Texts*, W. M. P. van der Aalst, V. Batagelj, D. I. Ignatov, M. Khachay, V. Kuskova, A. Kutuzov, S. O. Kuznetsov, I. A. Lomazova, N. Loukachevitch, A. Napoli, P. M. Pardalos, M. Pelillo, A. V. Savchenko, and E. Tutubalina, Eds. Cham: Springer International Publishing, 2019, pp. 321–334.
- [2] B. Segal, D. M. Rubin, G. Rubin, and A. Pantanowitz, “Evaluating the clinical realism of synthetic chest x-rays generated using progressively growing gans,” *SN Computer Science*, vol. 2, no. 4, p. 321, 2021. [Online]. Available: <https://doi.org/10.1007/s42979-021-00720-7>
- [3] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in *International conference on machine learning*. PMLR, 2017, pp. 214–223.
- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- [5] K. He, X. Zhalng, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Dec. 2015.
- [6] “Roa fact sheets, 2021,” 2021.

- [7] B. Dadonaite and M. Roser, “Pneumonia,” *Our World in Data*, 2018, <https://ourworldindata.org/pneumonia>.
- [8] D. Sridhar, “Pneumonia: a global cause without champions,” *The Lancet*, vol. 392, pp. 718–719, Sep. 2018.
- [9] K. Kondo, K. Suzuki, M. Washio, S. Ohfuchi, W. Fukushima, A. Maeda, and Y. Hirota, “Effectiveness of 23-valent pneumococcal polysaccharide vaccine and seasonal influenza vaccine for pneumonia among the elderly – selection of controls in a case-control study,” *Vaccine*, vol. 35, pp. 4806–4810, 08 2017.
- [10] K. Doi, “Computer-aided diagnosis in medical imaging: Historical review, current status and future potential,” *Computerized Medical Imaging and Graphics*, vol. 31, no. 4, pp. 198–211, 2007, computer-aided Diagnosis (CAD) and Image-guided Decision Support. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0895611107000262>
- [11] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” *CoRR*, vol. abs/1705.02315, 2017. [Online]. Available: <http://arxiv.org/abs/1705.02315>
- [12] L. Gao, L. Zhang, C. Liu, and S. Wu, “Handling imbalanced medical image data: A deep-learning-based one-class classification approach.” *Artificial intelligence in medicine*, vol. 108, p. 101935, Aug 2020.
- [13] O. Er, C. SERTKAYA, F. Temurtas, and A. Tanrikulu, “A comparative study on chronic obstructive pulmonary and pneumonia diseases diagnosis using neural networks and artificial immune system,” *Journal of medical systems*, vol. 33, pp. 485–92, 12 2009.
- [14] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.

- [15] X. Tang, “The role of artificial intelligence in medical imaging research,” *BJR open*, vol. 2, no. 33178962, pp. 20190031–20190031, Nov. 2019. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7594889/>
- [16] A. Nazir and R. A. Khan, “Network intrusion detection: Taxonomy and machine learning applications,” in *Machine Intelligence and Big Data Analytics for Cybersecurity Applications*. Springer, 2021, pp. 3–28.
- [17] R. Al Mamlook, S. Chen, and H. Bzizi, “Investigation of the performance of machine learning classifiers for pneumonia detection in chest x-ray images,” 07 2020.
- [18] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks.” *Science (New York, N.Y.)*, vol. 313, pp. 504–7, Jul 2006.
- [19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 06 2014.
- [20] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–44, 05 2015.
- [21] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014.
- [22] Y. B. Ian Goodfellow and A. Courville, “Deep learning.”
- [23] R. Hecht-Nielsen, “Theory of the backpropagation neural network,” *International 1989 Joint Conference on Neural Networks*, pp. 593–605 vol.1, 1989.
- [24] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980. [Online]. Available: <https://doi.org/10.1007/BF00344251>

- [25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [26] A. Borjali, A. F. Chen, O. K. Muratoglu, M. A. Morid, and K. M. Varadarajan, “Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network.” *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*, vol. 38, pp. 1465–1471, Jul 2020.
- [27] O. Stephen, M. Sain, U. Maduh, and D. Jeong, “An efficient deep learning approach to pneumonia classification in healthcare,” *Journal of Healthcare Engineering*, vol. 2019, pp. 1–7, 03 2019.
- [28] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, “Automated detection of covid-19 cases using deep neural networks with x-ray images,” *Computers in Biology and Medicine*, vol. 121, p. 103792, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482520301621>
- [29] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology.” *Insights into imaging*, vol. 9, pp. 611–629, Aug 2018.
- [30] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, *A Survey on Deep Transfer Learning: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4–7, 2018, Proceedings, Part III*, 10 2018, pp. 270–279.
- [31] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [32] P. Lakhani and B. Sundaram, “Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks.” *Radiology*, vol. 284, pp. 574–582, Aug 2017.
- [33] M. Toğaçar, B. Ergen, Z. Cömert, and F. Özyurt, “A deep feature learning model for pneumonia detection applying a combination of mrmr feature selection and machine

- learning models,” *IRBM*, vol. 41, no. 4, pp. 212–222, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1959031819301174>
- [34] G. Liang, Y. Zhang, X. Wang, and N. Jacobs, “Improved trainable calibration method for neural networks on medical imaging classification,” *ArXiv*, vol. abs/2009.04057, 2020.
- [35] G. Huang, Z. Liu, L. van der Maaten, and K. Weinberger, “Densely connected convolutional networks,” 07 2017.
- [36] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “Squeezezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size,” *arXiv preprint arXiv:1602.07360*, 2016.
- [37] M. Talo, O. Yildirim, U. B. Baloglu, G. Aydin, and U. R. Acharya, “Convolutional neural networks for multi-class brain disease detection using mri images,” *Computerized Medical Imaging and Graphics*, vol. 78, p. 101673, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0895611119300886>
- [38] A. Titoriya and S. Sachdeva, “Breast cancer histopathology image classification using alexnet,” in *2019 4th International conference on information systems and computer networks (ISCON)*. IEEE, 2019, pp. 708–712.
- [39] A. Shelke, M. Inamdar, V. Shah, A. Tiwari, A. Hussain, T. Chafekar, and N. Mehendale, “Chest x-ray classification using deep learning for automated covid-19 screening,” *SN Computer Science*, vol. 2, no. 4, p. 300, 2021. [Online]. Available: <https://doi.org/10.1007/s42979-021-00695-5>
- [40] N. Ansari, A. Faizabadi, S. Motakabber, and M. Ibrahimy, “Effective pneumonia detection using resnet based transfer learning,” *Test Engineering and Management*, vol. 82, pp. 15 146 – 15 153, 02 2020.
- [41] G. Liang and L. Zheng, “A transfer learning method with deep residual network for pediatric pneumonia diagnosis,” *Computer Methods and Programs in Biomedicine*,

- vol. 187, p. 104964, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169260719306017>
- [42] V. Chouhan, S. K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damaševičius, and V. H. C. de Albuquerque, “A novel transfer learning based approach for pneumonia detection in chest x-ray images,” *Applied Sciences*, vol. 10, no. 2, 2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/2/559>
- [43] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, “Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer,” *Procedia Computer Science*, vol. 179, pp. 423–431, 2021, 5th International Conference on Computer Science and Computational Intelligence 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050921000284>
- [44] A. Narin, C. Kaya, and Z. Pamuk, “Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks,” *Pattern Analysis and Applications*, vol. 24, no. 3, pp. 1207–1220, 2021. [Online]. Available: <https://doi.org/10.1007/s10044-021-00984-y>
- [45] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807.
- [46] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [47] E. Ayan and H. M. Ünver, “Diagnosis of pneumonia from chest x-ray images using deep learning,” in *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science (EBBT)*, 2019, pp. 1–5.
- [48] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, “Efficient pneumonia detection in chest xray images using deep transfer learning,” *Diagnostics*, vol. 10, no. 6, 2020. [Online]. Available: <https://www.mdpi.com/2075-4418/10/6/417>

- [49] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2015.
- [50] P. Naveen and B. Diwan, “Pre-trained vgg-16 with cnn architecture to classify x-rays images into normal or pneumonia,” in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2021, pp. 102–105.
- [51] C. Sitaula and M. B. Hossain, “Attention-based vgg-16 model for covid-19 chest x-ray image classification,” *Applied Intelligence*, vol. 51, no. 5, pp. 2850–2863, 2021. [Online]. Available: <https://doi.org/10.1007/s10489-020-02055-x>
- [52] N. Habib, M. M. Hasan, M. M. Reza, and M. M. Rahman, “Ensemble of chexnet and vgg-19 feature extractor with random forest classifier for pediatric pneumonia detection.” *SN computer science*, vol. 1, p. 359, 2020.
- [53] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, “A survey on addressing high-class imbalance in big data,” *Journal of Big Data*, vol. 5, no. 1, p. 42, 2018. [Online]. Available: <https://doi.org/10.1186/s40537-018-0151-6>
- [54] A. Fernández, S. del Río, N. V. Chawla, and F. Herrera, “An insight into imbalanced big data classification: outcomes and challenges,” *Complex and Intelligent Systems*, vol. 3, no. 2, pp. 105–120, 2017. [Online]. Available: <https://doi.org/10.1007/s40747-017-0037-9>
- [55] G. Batista, R. Prati, and M.-C. Monard, “A study of the behavior of several methods for balancing machine learning training data,” *SIGKDD Explorations*, vol. 6, pp. 20–29, 06 2004.
- [56] J. E. Luján-García, C. Yáñez-Márquez, Y. Villuendas-Rey, and O. Camacho-Nieto, “A transfer learning method for pneumonia classification and visualization,” *Applied Sciences*, vol. 10, no. 8, 2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/8/2908>

- [57] B. Antin, J. Kravitz, and E. Martayan, “Detecting pneumonia in chest x-rays with supervised learning,” 2017.
- [58] D. Srivastav, A. Bajpai, and P. Srivastava, “Improved classification for pneumonia detection using transfer learning with gan based synthetic image augmentation,” in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE, 2021, pp. 433–437.
- [59] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [60] S. Sundaram and N. Hulkund, “Gan-based data augmentation for chest x-ray classification,” *ArXiv*, vol. abs/2107.02970, 2021.
- [61] S. Motamed, P. Rogalla, and F. Khalvati, “Data augmentation using generative adversarial networks (gans) for gan-based detection of pneumonia and covid-19 in chest x-ray images,” *Informatics in Medicine Unlocked*, vol. 27, p. 100779, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352914821002501>
- [62] D. Srivastav, A. Bajpai, and P. Srivastava, “Improved classification for pneumonia detection using transfer learning with gan based synthetic image augmentation,” in *2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, 2021, pp. 433–437.
- [63] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. H. Michalski, “Medical image synthesis for data augmentation and anonymization using generative adversarial networks,” *ArXiv*, vol. abs/1807.10225, 2018.
- [64] C. Schaefer-Prokop and M. Prokop, “New imaging techniques in the treatment guidelines for lung cancer.” *The European respiratory journal. Supplement*, vol. 35, pp. 71s–83s, Feb 2002.

- [65] A. Gulati and R. Balasubramanya, “Lung imaging,” in *StatPearls [Internet]*. StatPearls Publishing, 2021.
- [66] F. Zhang, “Application of machine learning in ct images and x-rays of covid-19 pneumonia,” *Medicine*, vol. 100, no. 34516488, pp. e26 855–e26 855, Sep. 2021. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8428739/>
- [67] S. Sathiratanacheewin, P. Sunanta, and K. Pongpirul, “Deep learning for automated classification of tuberculosis-related chest x-ray: dataset distribution shift limits diagnostic performance generalizability,” *Heliyon*, vol. 6, no. 8, p. e04614, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2405844020314584>
- [68] A. Karargyris, S. Kashyap, J. T. Wu, A. Sharma, M. Moradi, and T. F. Syeda-Mahmood, “Age prediction using a large chest x-ray dataset,” *CoRR*, vol. abs/1903.06542, 2019. [Online]. Available: <http://arxiv.org/abs/1903.06542>
- [69] G. Caseneuve, I. Valova, N. LeBlanc, and M. Thibodeau, “Chest x-ray image preprocessing for disease classification,” *Procedia Computer Science*, vol. 192, pp. 658–665, 2021, knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 25th International Conference KES2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050921015556>
- [70] A. Bouslama, Y. Laaziz, and A. Tali, “Diagnosis and precise localization of cardiomegaly disease using u-net,” *Informatics in Medicine Unlocked*, vol. 19, p. 100306, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352914819304009>
- [71] P. Rajpurkar, J. A. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Y. Ding, A. Bagul, C. Langlotz, K. S. Shpanskaya, M. P. Lungren, and A. Ng, “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” *ArXiv*, vol. abs/1711.05225, 2017.
- [72] J. A. Dunnmon, D. Yi, C. P. Langlotz, C. Ré, D. L. Rubin, and M. P. Lungren, “Assess-

- ment of convolutional neural networks for automated classification of chest radiographs,” *Radiology*, vol. 290, no. 2, pp. 537–544, 2019.
- [73] L. van der Maaten and G. E. Hinton, “Visualizing data using t-sne,” *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [74] R. Longadge and S. Dongre, “Class imbalance problem in data mining review,” *arXiv preprint arXiv:1305.1707*, 2013.
- [75] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, “On the class imbalance problem,” *Fourth International Conference on Natural Computation, ICNC '08*, vol. Vol. 4, Oct. 2008.
- [76] M. Rahman and D. N. Davis, “Addressing the class imbalance problem in medical datasets,” *International Journal of Machine Learning and Computing*, vol. 3, p. 224, 04 2013.
- [77] I. Triguero, M. Galar, S. Vluymans, C. Cornelis, H. Bustince, F. Herrera, and Y. Saeys, “Evolutionary undersampling for imbalanced big data classification,” in *2015 IEEE Congress on Evolutionary Computation (CEC)*, 2015, pp. 715–722.
- [78] M. Hayati, S. Mutmainah, and S. Ghufran, “Random and synthetic over-sampling approach to resolve data imbalance in classification,” *International Journal of Artificial Intelligence Research*, vol. 4, p. 86, 01 2021.
- [79] Z. Zheng, Y. Cai, and Y. Li, “Oversampling method for imbalanced classification,” *Comput. Informatics*, vol. 34, pp. 1017–1037, 2015.
- [80] S. Hu, Y. Li, and S. Lyu, “Exposing gan-generated faces using inconsistent corneal specular highlights,” in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 2500–2504.
- [81] X. Yang, Y. Li, H. Qi, and S. Lyu, “Exposing gan-synthesized faces using landmark locations,” in *Proceedings of the ACM Workshop on Information Hiding and Multimedia*

- Security*, ser. IH&MMSec'19. New York, NY, USA: Association for Computing Machinery, 2019, p. 113–118. [Online]. Available: <https://doi.org/10.1145/3335203.3335724>
- [82] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, “Gans for medical image analysis,” *Artificial Intelligence in Medicine*, vol. 109, p. 101938, 2020.
- [83] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *arXiv preprint arXiv:1511.06434*, 2015.
- [84] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of wasserstein gans,” *Advances in neural information processing systems*, vol. 30, 2017.
- [85] A. F. Agarap, ‘Deep learning using rectified linear units (relu),’ Mar. 2018.
- [86] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034.
- [87] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” *CoRR*, vol. abs/1610.02357, 2016. [Online]. Available: <http://arxiv.org/abs/1610.02357>
- [88] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.