

HERIOT-WATT UNIVERSITY

FINAL YEAR DISSERTATION

**Automated Diagnosis of COVID-19 using Medical
Imagery**

Author:

Alister George Luiz

Supervisor:

Dr. Hani Ragab

A thesis submitted in fulfillment of the requirements

for the honours degree of Bachelor of Science

in the

School of Mathematical and Computer Sciences

April 2021



Declaration of Authorship

I, **Alister George Luiz**, confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g., ideas, equations, figures, text, tables, programs) are properly acknowledged at any point of their use. A list of the references employed is included.

Signed: **Alister George Luiz**

Date: **November 26, 2020**

“The only limit to our realization of tomorrow will be our doubts of today.”

Franklin D. Roosevelt

Abstract

The Coronavirus Disease (COVID-19) since its inception in late 2019 has spread all across the world and have led to an increased burden on healthcare professionals due to the urgent need for rapid disease diagnosis and effectuating quarantine protocols.

Currently, the Real-Time Reverse Transcription Polymerase Chain Reaction (RT-PCR) test recommended by the World Health Organization (WHO) remains the front runner in terms of COVID-19 diagnosis when compared to other testing mechanisms. But using this test as a primary diagnosis tool involves serious downsides, a few of them include the shortage in RT-PCR test kits and delays in receiving test results (up to 2 days).

The primary objective of this project is to develop an automated method for quick, efficient, and affordable COVID-19 diagnosis. We aim to utilize medical imagery such as chest X-rays and CT scans, apply deep learning classification techniques to identify lung patterns that are characteristic of COVID-19 infection.

We obtained 98.8% and 98.7% accuracy in the 10-fold cross validation with excellent precision, recall and F1-score for X-ray and CT scans respectively. We deployed an experimental web application (<http://40.76.124.61/>) to allow live testing of our model. We intend to publish our research paper on the experiments conducted for X-ray classification in the *Computers in Biology and Medicine* journal. After thorough clinical validation, such application could help reduce workload and limit virus exposure for healthcare professionals.

Keywords: COVID-19, SARS-CoV-2, Chest X-ray, Chest CT, Medical Image Classification, Deep Learning, Convolutional Neural Networks, Transfer Learning

Acknowledgements

I would like to thank Dr. Hani Ragab for his constant guidance and support throughout the tenure of this project. I also thank Dr Lynne Baillie, my second reader. Her comments, suggestions and feedback have helped shape this project into a sophisticated piece of work. I would also like to thank my parents whose prayers and encouragement made me who I am today ...

Contents

Declaration of Authorship	i
Abstract	iii
Acknowledgements	iv
Contents	v
List of Figures	viii
List of Tables	x
List of Abbreviations	xii
1 Introduction	1
1.1 Aim	1
1.2 Objectives	1
1.3 Extra Achievements	2
1.4 Manuscript Organization	2
2 Literature Review	3
2.1 The COVID-19 Pandemic Era	3
2.2 Diagnosing COVID-19	3
2.2.1 The RT-PCR Test	4
2.2.2 Medical Imagery	5
2.3 Deep Learning for Medical Imagery-based COVID-19 Detection	8
2.3.1 CT Based Diagnosis of COVID-19	8
2.3.2 X-ray Based Diagnosis of COVID-19	13
2.3.3 Heatmap Visualization of Deep Learning Results	14

2.4 Discussion	16
2.5 Conclusion and Research Questions	18
3 Project Implementation	19
3.1 Requirements Analysis	19
3.2 Development Environment	19
3.3 Source Code	20
3.4 X-ray Scans	20
3.4.1 Data Collection	20
3.4.2 Data Pre-processing	20
3.4.3 Methodology	22
3.5 CT Scans	29
3.5.1 Data Collection	29
3.5.2 Data Pre-processing	29
3.5.3 Methodology	32
3.6 Web Interface	38
4 Results and Evaluation	39
4.1 X-ray Scans	39
4.1.1 Performance Evaluation	39
4.1.2 Performance Comparison	42
4.2 CT Scans	46
4.2.1 Performance Evaluation	46
4.2.2 Performance Comparison	49
4.3 Web Interface	52
4.3.1 Hypothesis	52
4.3.2 Target Audience	52
4.3.3 Experiment Design	52
4.3.4 Activity 1: Tasks	52
4.3.5 Activity 2: System Usability Scale (SUS) Survey	54
4.3.6 Activity 3: Metrics Test	56
4.4 Requirements Validation	57

5 Conclusion and Future Work	59
5.1 Limitations	59
5.2 Future Work	60
5.3 Reflections	60
Bibliography	61
A Requirements Analysis and Evaluation Strategy	73
A.1 Requirements Analysis	73
A.1.1 Functional and Non-Functional Requirements	73
A.2 Implementation Workflow	75
A.3 Evaluation Strategy	76
A.4 Implementation Methodology	77
A.4.1 Data Collection	77
A.4.2 Data Pre-processing	78
A.4.3 Image Segmentation	78
A.4.4 Model Optimization	78
B Project Management	79
B.1 Risk Management	79
B.2 Project Plan	82
B.3 Development Methodology	82
B.4 Professional, Legal, Ethical and Social Issues	83
B.4.1 Professional and Legal Issues	83
B.4.2 Ethical and Social Issues	83
C The COVID-19 Pandemic Era	84
D COVID-19 Lung Characteristics	87
E Automated Diagnosis Workflow	88
F CT Image Segmentation Techniques	90
G SCRUM	92
H Paper: "A Deep Learning Method for COVID-19 Detection from X-ray Scans"	93

List of Figures

2.1 RT-PCR Test Negative Rate	5
2.2 CT Scan Images	7
2.3 U-Net Architecture	9
2.4 ResNet-50 Architecture	13
2.5 X-ray Heat Map	15
2.6 Attention Heat Map	16
3.1 Dataset Distribution for X-ray Scans	21
3.2 Sample of X-ray scans with labels for model training.	22
3.3 X-ray Model Architectures	23
3.4 A simple representation of Transfer Learning Technique [64]	25
3.5 Ensemble Learning Illustration for X-ray Scans	26
3.6 X-ray Workflow Illustration	28
3.7 Lung Parenchyma Extraction	30
3.8 Dataset Distribution	31
3.9 Sample of unprocessed CT scans with labels for model training.	31
3.10 CT Model Architectures	32
3.11 Ensemble Learning Illustration for CT Scans	35
3.12 CT Workflow Illustration	37
3.13 Diagnosis Portal Screenshots	38
4.1 DenseNet121 Confusion Matrix	40
4.2 DenseNet121 Model Training Performance Trends	40
4.3 ResNet50 Confusion Matrix	40
4.4 ResNet50 Model Training Performance Trends	41
4.5 VGG16 Confusion Matrix	41

4.6 VGG16 Model Training Performance Trends	41
4.7 X-ray Ensemble Model Confusion Matrix	42
4.8 X-ray Ensemble Model Training Performance Trends	42
4.9 UNet Model Confusion Matrix	46
4.10 UNet Model Training Performance Trends	46
4.11 Attention UNet Model Confusion Matrix	47
4.12 Attention UNet Model Training Performance Trends	47
4.13 UNet++ Model Confusion Matrix	48
4.14 UNet++ Model Training Performance Trends	48
4.15 CT Ensemble Model Confusion Matrix	48
4.16 CT Ensemble Model Training Performance Trends	49
4.17 Comparison of SUS Scores	56
A.1 Implementation Methodology	76
B.1 Project Plan	82
C.1 Dow Jones	85
C.2 COVID-19 Cases	85
C.3 Total COVID-19 Cases	86
C.4 Distribution of COVID-19 Cases	86
E.1 Mobile CT Platform	89
G.1 SCRUM	92

List of Tables

2.1 Findings at Initial Chest CT Examination in 21 Patients [8]	6
2.2 CT Image Segmentation Techniques in COVID-19 Quantification Applications [27] . .	12
2.3 COVID-19 Diagnosis Applications and their results from CT Image Segmentation [27]	12
2.4 X-ray Image Segmentation Techniques in COVID-19 Diagnosis Applications [27] . .	14
3.1 Training Dataset Post Augmentation	22
3.2 Dataset Distribution	26
3.3 Training Dataset Post Augmentation	31
3.4 Dataset Distribution for CT Scans	36
4.3 DenseNet121 Classification Report	40
4.6 ResNet50 Classification Report	40
4.9 VGG16 Classification Report	41
4.12 X-ray Ensemble Model Classification Report	42
4.13 X-ray Models Performance Comparison	43
4.14 Comparison with RT-PCR and CNN for X-ray scans	43
4.15 Comparison with Related Work	44
4.16 Comparison with Radiologist Finding	45
4.19 UNet Model Classification Report	46
4.22 Attention UNet Model Classification Report	47
4.25 UNet++ Model Classification Report	48
4.28 CT Ensemble Model Classification Report	48
4.29 CT Models Performance Comparison	49
4.30 Comparison with RT-PCR and CNN for X-ray scans	50
4.31 Comparison with Related Work	50
4.32 Comparison with Radiologist Finding	51

4.33 Activity 1 Results	53
4.34 Feedback on the Diagnosis Portal	54
4.35 System Usability Scale (SUS) Questions	55
4.36 System Usability Scale (SUS) Scores	55
4.37 System Usability Scale (SUS) Statistical Analysis	56
4.38 Activity 3 Likert Scale Results	57
4.39 Requirements Validation	58
A.1 Functional Requirements	74
A.2 Non-Functional Requirements	75
A.3 Evaluation Strategy	77
B.3 Risk Analysis	81
D.1 Chest X-ray imaging characteristics from multiple studies	87
E1 CT Image Segmentation Techniques in COVID-19 Applications [27]	91

List of Abbreviations

COVID-19	Coronavirus Disease 2019
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
RT-PCR	Reverse Transcription Polymerase Chain Reaction
CT	Computed Tomography
AI	Artificial Intelligence
CNN	Convolutional Neural Network
WHO	World Health Organization
RNA	Ribonucleic Acid
DNA	Deoxyribonucleic Acid
GGO	Ground Glass Opacity
ROI	Regions of Interest
CAP	Community Acquired Pneumonia

Chapter 1

Introduction

COVID-19, which was declared a global pandemic by the World Health Organization on March 11, 2020, has affected millions of lives worldwide in terms of both health and finances and have also had a severe global economic impact.

Over a billion tests have been carried across the world for diagnosing patients with COVID-19 [1], it is therefore the need of the hour to alleviate the burden on healthcare professionals who conducts these diagnoses on a day-to-day basis, and more importantly, minimize the exposure rate between patients and healthcare professionals.

1.1 Aim

This project aims to **automate the diagnoses of COVID-19 with medical imagery using Deep Learning**. The main focus being to achieve the highest diagnostic accuracy possible, minimizing the false-negative rate, which, if not accounted for may have adverse real-world implications.

The overall goal of this project is to develop an automated workflow that enables highly accurate rapid diagnoses of COVID-19 for both symptomatic and asymptomatic patients. This would lead to a safer environment for healthcare professionals by minimizing the rate of exposure therefore minimizing the spread of COVID-19.

1.2 Objectives

The objective of this project is to develop a deep learning classification model that rapidly diagnoses patients with COVID-19 using medical imagery and thereby provide assistance to healthcare professionals.

These are the primary objectives for this thesis:

1. Analyse X-ray imaging features of COVID-19.
2. Build a deep learning model that diagnoses patients with COVID-19 using chest X-rays.
3. Test the accuracy of the proposed COVID-19 diagnoses model by comparing with other similar models implemented previously.
4. Develop a deep learning API that can be used by healthcare professionals and medical facilities to diagnose COVID-19.
5. Optionally, build a deep learning model that diagnoses patients with COVID-19 using chest CT scans and compare the accuracy and results obtained from both models respectively.

1.3 Extra Achievements

In addition to the above objectives, we proposed an ensemble learning approach for COVID-19 diagnosis that achieved a higher accuracy than all the papers implemented for both X-ray and CT scans reported in this manuscript. We have also drafted a research paper on the experiments conducted for X-ray classification. We intend to publish it to the *Computers in Biology and Medicine* journal and is attached in Appendix H. Furthermore, we have also validated our results with the help of a senior Radiologist who has provided us his findings on a sample of X-ray and CT scans.

1.4 Manuscript Organization

This manuscript contains 5 chapters, starting with a comprehensive **Introduction** of the main objectives of this project. This is followed by the **Literature Review** chapter which aims to synthesize and sum up relevant research and implementations previously conducted in this same field. The **Project Implementation** chapter covers the dataset collection, pre-processing, and methodology used to implement this project. Following this, **Results and Evaluation** is where we showcase the results obtained and conduct a critical evaluation of our findings. The last chapter **Conclusion and Future Works** summarizes our experiments, provides an insight into the major challenges faced, and future works that we plan to carry out.

Chapter 2

Literature Review

We provide a comprehensive analysis of the existing research and methodologies pertaining to COVID-19 detection using medical imagery and deep learning in this chapter.

2.1 The COVID-19 Pandemic Era

Coronavirus disease (COVID-19) is a highly contagious respiratory disease caused by the newly discovered coronavirus. The virus mainly spreads through the discharge of saliva droplets when an infected person coughs or sneezes [2].

Most people affected by COVID-19 often show mild symptoms but those who have a compromised immune system such as older adults or those with underlying medical conditions are at a much higher risk of developing severe illness [3].

Therefore, one must follow the advice of medical professionals and adhere to social distancing protocols. This must be combined with other preventive measures such as maintaining personal hygiene and wearing masks to reduce the spread of the virus [4].

Further information about the novel coronavirus, the timeline demonstrating its spread across the world, and decisive events are discussed in Appendix C.

2.2 Diagnosing COVID-19

The following section discusses in detail the traditional procedure used in detecting COVID-19 using the real-time RT-PCR test but more importantly gives an insight on how medical imagery

could be used to achieve the same. The specific patterns and lesions observed from the lung scans of patients diagnosed with COVID-19 are also showcased in this section.

Besides these two approaches, Laguarta et al. utilized cough recordings to diagnose patients with COVID-19 [5]. Their CNN-based models achieved a sensitivity of 98.5% and specificity of 94.2% respectively for patients diagnosed with COVID-19 using an official test. For asymptomatic patients, the model achieves a sensitivity of 100% and specificity of 83.2% respectively.

2.2.1 The RT-PCR Test

The real-time RT-PCR test is a nuclear-derived method that detects the presence of specific genetic material in any pathogen which includes a virus. Scientists would be able to see the results even when the process is ongoing using the real-time RT-PCR test whereas the conventional RT-PCR provides the result only after the process is complete.

The real-time RT-PCR test has been widely used to detect viruses such as Ebola and Zika, and therefore is extensively used to detect COVID-19 in laboratories across the world [6]. The real-time RT-PCR test has been recommended by the WHO for COVID-19 diagnosis.

'Reverse Transcription' involves converting Ribonucleic Acid (RNA) to Deoxyribonucleic Acid (DNA), the reason for this being the ability to amplify specific parts of DNA which allows the scientists to spot strands of the virus among genetic information. Samples are collected from the patient's throat or nose, typically where the COVID-19 virus tend to gather.

Scientists add short DNA fragments, which complements the viral DNA. Therefore the virus, if present, leads to these added fragments to be attached to the target sections of the viral DNA. When marker labels attach to these DNA strands, a fluorescent dye is released, which is measured by the RT-PCR machine. When a certain threshold of fluorescence have passed the scientists could then diagnose the patient with COVID-19 [6].

Despite the high sensitivity and reliable diagnosis by the real-time RT-PCR test, there exist certain limitations which have led researchers to identify alternate methods of COVID-19 diagnoses, such as using medical imagery which includes X-Ray and CT scans.

During these trying times where the numbers keep rising rapidly, medical facilities are running short of RT-PCR test kits and therefore, are in dire need of alternate sources of diagnosis. Furthermore, the high false-negative rate of the real-time RT-PCR, which is as high as 100% before the

time of symptom onset and decreases only up to 64% on the day of symptom onset [7] could lead to severe consequences in the real world where the infected person could spread the virus as they are not under quarantine.

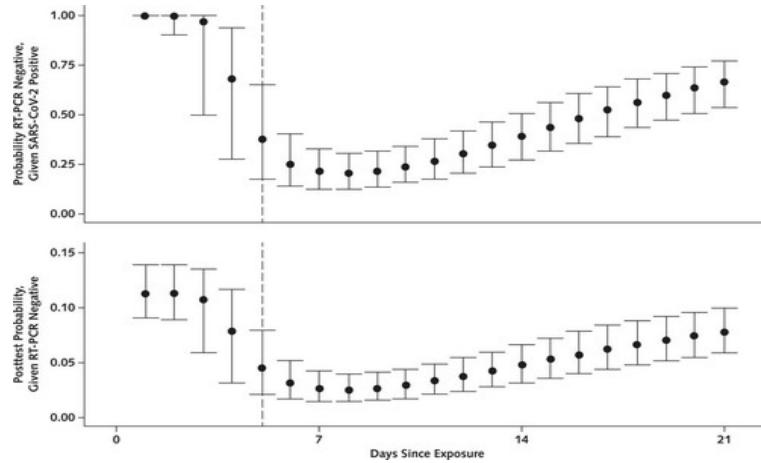


FIGURE 2.1: Probability of having a negative RT-PCR test result given SARS-CoV-2 infection (top) and of being infected with SARS-CoV-2 after a negative RT-PCR test result (bottom), by days since exposure [7]

Another important factor considering the daily rising numbers are the delays in receiving the results for the RT-PCR test. This places extra strain on the medical professionals in terms of workload and makes it difficult for them to apply safety protocols on suspected patients.

Therefore, due to these aforementioned limitations of the real-time RT-PCR test, finding a safer, more accurate, and faster diagnosis mechanism is essential. Thus, making COVID-19 diagnosis using medical imagery an ideal alternative candidate.

2.2.2 Medical Imagery

Medical Imagery such as X-rays and CT scans have proven to be a viable alternative to the RT-PCR test for COVID-19 detection due to the limitations mentioned above. The reduced exposure risks, and faster diagnosis time are also added benefits.

In this section, CT imaging features of the novel coronavirus shall be presented to lay a foundation for future sections where we illustrate how deep learning could learn these features and use it for automated real-time COVID-19 diagnosis.

Chung et al. conducted a study on 21 symptomatic patients infected with coronavirus admitted to three hospitals in provinces of Guangdong, Jiangxi, and Shandong respectively in China from

January 18th, 2020 to 27th, 2020. They aimed to identify potential imaging features of COVID-19 from CT scans with the help of two experienced fellowship-trained cardiothoracic radiologists with approximately 5 years of experience each.

The degree of lobe involvement was assessed and a "Total Severity Score" was assigned by summing up each of the individual lob scores. Patients were also re-evaluated in order to study the progression of features by the same two radiologists [8].

The observed common characteristics after evaluation are tabulated in Table 2.1. Other abnormalities such as Cavitation, Reticulation, Interlobular Septal Thickening, Calcification, and Bronchiectasis were also assessed.

Finding	No. of Patients
Ground-glass opacities and consolidation	
Absence of both ground-glass opacities and consolidation	3 (14%)
Presence of either ground-glass opacities or consolidation	18 (86%)
Presence of ground-glass opacities without consolidation	12 (57%)
Presence of ground-glass opacities with consolidation	6 (29%)
Presence of consolidation without ground-glass opacities	0 (0%)
Frequency of lobe involvement	
Right Upper Lobe	3 (14%)
Right Middle Lobe	1 (5%)
Right Lower Lobe	2 (10%)
Left Upper Lobe	3 (14%)
Left Lower Lobe	4 (19%)
Opacification distribution and pattern	
Rounded Morphology	7 (33%)
Linear Opacities	3 (14%)
Crazy-Paving Pattern	4 (19%)
Peripheral Distribution	7 (33%)
More than two lobes affected	15 (71%)
Bilateral lung disease	16 (76%)

TABLE 2.1: Findings at Initial Chest CT Examination in 21 Patients [8].

A follow-up chest CT scan was conducted on 8 of the initial 21 patients, within a range of 1 to 4 days. Only 1 patient out of the 8 had normal initial and follow-up CT scan results. 5 out of 8 experienced mild progression in the lung characteristics, the remaining 2 displayed moderate progression. Fortunately, none of the patients experienced severe progression.

The primary observations from this study on 21 patients include GGO's found in 12 patients and consolidation in 6 patients. There is also a high possibility the virus affects more than two lobes with bilateral involvement. Other observations include rounded morphology detected in 7 patients, reticulation in 3 patients, and crazy-paving in 4 patients [8].

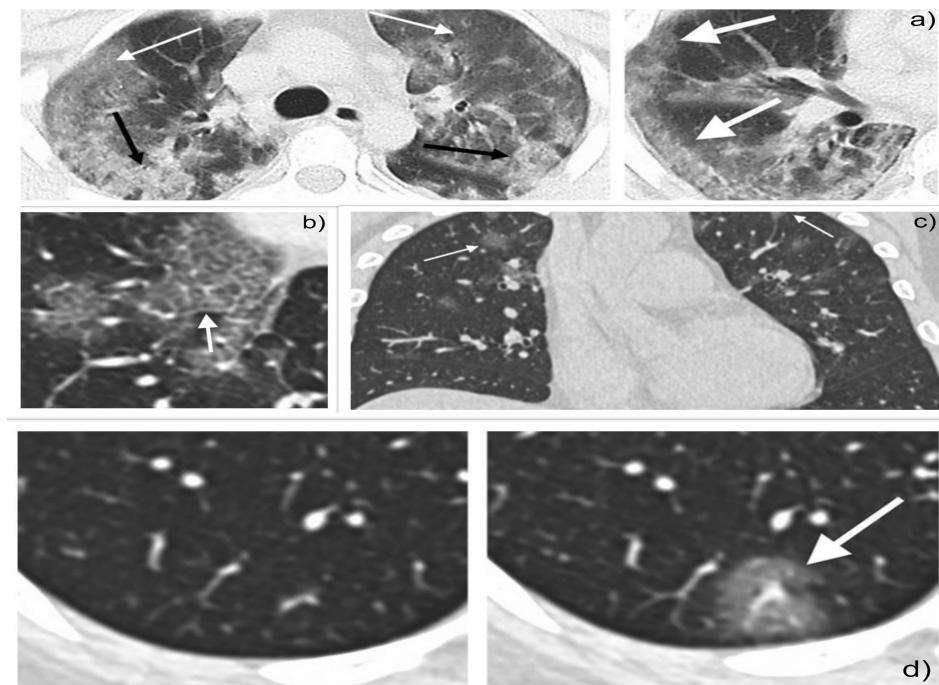


FIGURE 2.2: Observed lung CT Scan characteristics. a) White arrow indicates patchy GGO, and black arrow indicates consolidative pulmonary opacities. b) White arrow indicates GGO's with a rounded morphology. c) White arrow indicates crazy-paving pattern, GGO and interlobular septal thickening with intralobular lines. d) Follow-up CT scan progression. White arrow indicates new solitary, rounded, peripheral ground-glass lesion [8].

One obvious limitation of this study is the relatively low number of patients, with only 8 out of 21 carrying out a follow-up CT scan. As this study was conducted during the dawn of coronavirus, this number is certainly a very respective amount.

Another study by Morales et al. whose results are summarized in Appendix D, also observe similar lung characteristics. As we have now seen the prominent imaging characteristics of COVID-19, we shall now discuss how deep learning could learn these features and accurately diagnose patients with COVID-19 in real-time.

2.3 Deep Learning for Medical Imagery-based COVID-19 Detection

The coronavirus global pandemic spreading rapidly all across the world have forced scientists and researchers to identify alternative diagnosis mechanisms in addition to the RT-PCR test to overcome its limitations. As we have seen in previous sections, medical imagery such as X-rays and CT scans have played a vital role in combating the rising numbers by saving valuable time in diagnosis and reducing virus exposure.

Deep learning techniques have further enhanced COVID-19 diagnosis using medical imagery due to its rapid detection capabilities, fully automated and efficient diagnosis workflow, and assisting medical practitioners by highlighting observed COVID-19 lung characteristics similar to features discussed in Section 2.2.2. A detailed overview of the modern CT and X-ray systems enabling automated diagnosis workflow ensuring minimal virus exposure can be found in Appendix E.

To identify the regions of interest (ROIs), segmentation of the lung CT or X-ray scans are a vital pre-requisite. The former produces high-quality 3D images for detecting COVID-19 whereas the latter involves the ribs being projected onto soft-tissues in 2D.

As a result, segmentation in X-ray scans is more challenging as compared to CT scans. But on the other hand, X-ray scans are more widely accessible in medical facilities all across the world and are usually the first imaging modality used on patients suspected of COVID-19.

Keeping in mind these limitations, the next two subsections review the segmentation techniques using deep learning for both CT and X-rays respectively and discuss the results obtained.

2.3.1 CT Based Diagnosis of COVID-19

To identify the ROIs from a CT scan for diagnosis, deep learning techniques are extensively used. These techniques could be narrowed down to the three most prominent segmentation methods which are **U-Net** [9]–[14], **UNet++** [15], [16], and **VB-Net** [17] respectively.

From Section 2.2.2, we can infer that the main ROIs could be classified into two specific categories, which are lung-region and lung-lesion oriented methods. The latter is more of a challenge in terms of its detection as lesions could be in a variety of shapes and sizes, furthermore, locating its region also adds to the difficulty in identifying it.

The literature indicates U-Net architecture as most reliable in segmenting both lung regions and lesions in COVID-19 diagnosis applications. Designed by Ronneberger, U-Net as its name suggests has a U-shape architecture, such that it has a symmetric expansive and contracting path [18].

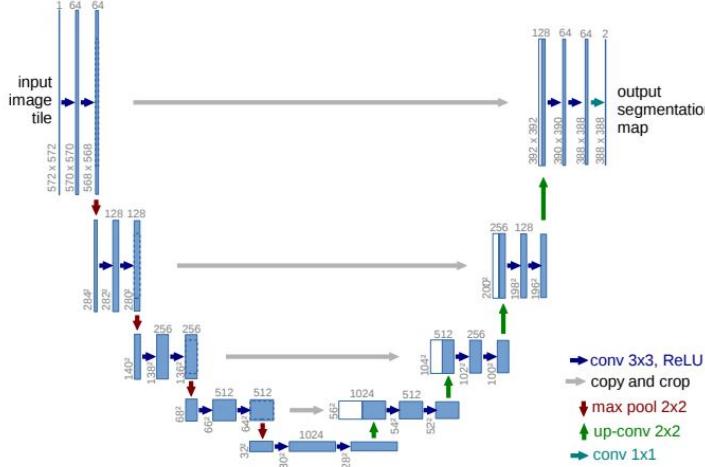


FIGURE 2.3: A representation of the U-Net architecture [18]

Various variants of the U-Net have been developed since its inception due to its high ability to learn visual semantics and therefore are suitable for many medical applications. They include the following:

- **3D U-Net** - Replaces the conventional U-Net layers with 3D layers. [19]
- **V-Net** - Utilizes residual blocks as the convolutional block, network optimization carried out through Dice Loss. [20]
- **VB-Net** - Combination of V-Net with a bottle-neck structure. [17]
- **UNet++** - A more complex version of U-Net, inserts a nested convolutional structure between the expansive and contracting path. [21]
- **Attention U-Net** - Integrates Attention Gates with U-Net architecture to focus on target structures of various shapes and sizes. [22]

Identifying and labeling the training data for the purpose of COVID-19 diagnosis is often time consuming and labour intensive, especially the manual detection of lesions. Shan et al. proposes a workaround which involves the "human-in-the-loop" strategy, where radiologists play an integral part in the training process [17]. Another similar suggestion from Yue et al. allowed the radiologists to provide initial seeds for the U-Net model [12].

Zheng et al. suggests an alternative approach, where unsupervised methods are used to generate pseudo segmentation masks for images to overcome the labeling process and thus avoid its limitations. In addition to reviewing results of lung segmentation from the U-Net model, Zheng et al. also utilizes a 3D CNN to predict the probability of COVID-19 using images obtained as an input [9]. With a dataset of 540 patients, 313 with COVID-19, and remaining without COVID-19 are used for training and testing purposes. The model attains a sensitivity of 90.7% and specificity 91.1%.

The lung segmentation's obtained could be used for COVID-19 diagnosis and for the quantification of data. The literature mentioned in this section includes both of the objectives.

Li et al. conducted a multi-center study by distinguishing COVID-19 from Community-acquired Pneumonia [14]. A combination of U-Net and ResNet-50 is proposed with the former used to extract lung regions using pre-processed 2D slices. The latter along with shared weights between the 2D slices combined with max-pooling is used for COVID-19 diagnosis. This study utilizes a large dataset with 1296 COVID-19 patients, 1735 with Community-acquired Pneumonia, and 1325 Non-Pneumonia patients. The model achieves a specificity of 96% and sensitivity of 90%.

An AI system developed by Jin et al. for rapid COVID-19 diagnosis, inputs segmented CT slices to the model [16]. Instead of 3D CNN, a ResNet-50 model is used for diagnosis and UNet++ for lung segmentation and lesion identification. The model trained on 1136 images, 723 COVID-19 positives, and 413 negatives achieves sensitivity and specificity of 97.4% and 92.2% respectively.

As for the quantification of data, both Cao et al. [10] and Huang et al. [11] monitor the longitudinal progression of COVID-19 using the CT segmentation of pulmonary opacities using the segmentation of the lung region and GGO. Therefore, the image segmentation obtained aids radiologists in infection identification, analysis, and diagnosis.

Most classification studies involve segregating COVID-19 patients from non-COVID-19 with most of the latter patients being segregated further into Pneumonia and Non-Pneumonia subjects.

Chen et al. developed a UNet++ model which segments lung lesions [15] using CT images of 51 COVID-19 patients and 55 patients with other diseases and diagnose patients with 95.2% accuracy thus reducing the reading time of radiologists by 65%. Given raw images to the model, it outputs prediction boxes displaying suspected regions, after further extraction and filtering, a logic linking of predictions is added that aims to aid radiologists in manual detection of the virus.

Jin et al. considers an alternative approach utilizing 2D Deeplab v1 and 2D ResNet-152 models for lung segmentation and lung-mask slice based classification of COVID-19 respectively [23]. The model achieves a respectable score of 94.1% sensitivity and 95.5% specificity using a dataset of 496 COVID-19 positive CT scan images.

The objective of the remaining studies besides COVID-19 diagnosis is its differentiation with common Pneumonia which primarily includes viral Pneumonia. The main reason for this objective being the very similar radiological appearances of both the diseases.

Wang et al. classifies between COVID-19 and viral Pneumonia using a 2D CNN model on delineated region patches [24]. Experiments conducted on chest CT scans from 44 COVID-19 and 55 Pneumonia patients with external testing resulting in 79.3% accuracy.

Experiments carried out by Song et al. employs OpenCV to segment 2D slices which include lung regions [25]. The 3D chest CT images resulted in 15 2D slices of complete lungs and each slice is fed into the deep learning based CT diagnosis system also called DeepPneumonia. A combination of a pre-trained ResNet-50 along with Feature Pyramid Network (FPN) which can extract specific ranked details from the images, coupled with an attention module to learn these extracted details are used to develop this model.

The dataset includes 88 COVID-19 patients, 101 bacterial Pneumonia and 86 healthy patients. The proposed model achieves a classification accuracy of 86%.

Table 2.2 and 2.3 summarizes all the COVID-19 image segmentation applications mentioned in this section, and the results of diagnosis experiments conducted across various medical facilities. An extended version of Table 2.2 can be found in Appendix F.

Study	Method	Target ROI
Zheng et al. [9]	U-Net	Lung
		Lesion
Cao et al. [10]	U-Net	Lung
		Lesion
Huang et al. [11]	U-Net	Lung
		Lung Lobes
		Lesion

Yue et al. [12]	U-Net	Lung Lobes
		Lesion
Gozel et al. [13]	U-Net	Lung
		Lesion
Tang et al. [26]	Commercial Software	Lung
		Lesion
		Trachea
		Bronchus

TABLE 2.2: CT Image Segmentation Techniques in COVID-19 Quantification Applications [27]

Study	Subjects	Method	Result
Zheng et al. [9]	313 COVID-19	U-Net	90.7% (Sens.)
	229 Others	CNN	91.1% (Spec.)
Li et al. [14]	468 COVID-19	ResNet-50	90.0% (Sens.)
	1551 CAP		96.0% (Spec.)
	1445 Non-pneu.		0.95 (AUC.)
Chen et al. [15]	51 COVID-19	UNet++	100% (Sens.)
	55 Others		93.6% (Spec.)
Jin et al. [16]	723 COVID-19	UNet++	97.4% (Sens.)
	413 Others		92.2% (Spec.)
Jin et al. [23]	496 COVID-19	CNN	94.1% (Sens.)
	1385 Others		95.5% (Spec.)
Song et al. [25]	88 COVID-19	ResNet-50	86.0% (Acc.)
	100 Bac. Pneu.		
	86 Normal		
Wang et al. [24]	44 COVID-19	CNN	79.3% (Acc.)
	55 Vir. Pneu.		

TABLE 2.3: COVID-19 Diagnosis Applications and their results from CT Image Segmentation [27]

* Bac. - Bacterial, Vir. - Viral, Pneu. - Pneumonia

As we have seen, the above studies result in promising diagnosis outcomes. Therefore, COVID-19 diagnosis with CT images could facilitate early detection of the coronavirus and also reduce the high exposure rates between patients and medical professionals.

2.3.2 X-ray Based Diagnosis of COVID-19

X-rays are most often the first imaging modality used on suspected patients, due to its wide availability in most clinics and medical facilities. As seen in Section 2.2.2, radiological signs include GGOs, consolidation, and opacification.

In order to detect these abnormalities in lung X-ray scans, three popular architecture's are used across various studies which are **ResNet** [28], **ResNet-50** [29], and **CNN** [30], [31].

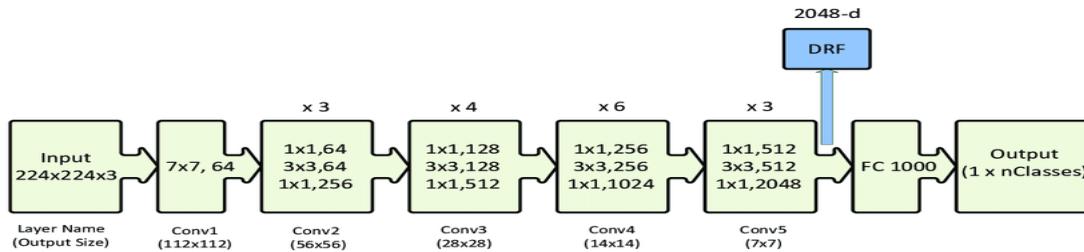


FIGURE 2.4: ResNet-50 architecture with residual units [32]

X-rays despite, as mentioned previously, being the first imaging modality for patients suspected with COVID-19 is less sensitive than 3D chest CT images. Anomalous chest radiographs are found in 69% of the patients initially during admission and this number increases to 80% after a certain period once hospitalized [33].

To estimate the uncertainty in COVID-19 prediction Ghoshal et al. [30] proposes a Bayesian CNN. X-rays of 70 COVID-19 patients are obtained from Cohen et al. [34] and others from Kaggle's chest X-ray images. Bayesian inference improves the detection accuracy of the model from 85.7% to 92.9% from the experiments conducted by the authors.

Narin et al. experiments with three deep learning models i.e., ResNet-50, InceptionV3 and Inception-ResNetV2 respectively, with the objective to detect COVID-19 from X-ray images [29]. The dataset includes X-rays from 50 COVID-19 patients and 50 normal scans. The results indicate that ResNet-50 achieves highest accuracy with 98% followed by InceptionV3 which attains 97%.

Zhang et al. also suggest a ResNet based model for COVID-19 detection [28]. But the model aims to achieve two objectives, one for COVID-19 classification and another for anomaly detection.

The experiment is conducted on a dataset containing X-rays from 70 COVID-19 patients and 1008 other X-rays. The anomaly detection score in-turn optimizes the COVID-19 classification score which reaches 96% from the experiments conducted by the authors.

Wang et al. proposes a deep CNN based model (COVID-Net) and achieves a testing accuracy of 83.5% [31]. The dataset used for the study include X-rays from patients diagnosed with both Bacterial and Viral Pneumonia. More specifically, 45 COVID-19 positive, 931 Bacterial Pneumonia, 660 Viral Pneumonia, and 1203 normal X-rays.

Table 2.4 summarizes the results obtained by the experiments discussed in this section.

Study	Subjects	Method	Result
Ghoshal et al. [30]	70 COVID-19	CNN	92.9% (Acc.)
	Others		
Zhang et al. [28]	70 COVID-19	ResNet	96.0% (Sens.)
	1008 Others		70.7% (Spec.)
Narin et al. [29]	50 COVID-19	ResNet-50	98.0% (Acc.)
	50 Normal		
Wang et al. [31]	45 COVID-19	CNN	83.5% (Acc.)
	931 Bac. Pneu.		
	660 Viral Pneu.		
	1203 Normal		

TABLE 2.4: X-ray Image Segmentation Techniques in COVID-19 Diagnosis Applications [27]

* Acc. - Accuracy, Sens. - Sensitivity, Spec. - Specificity

As seen in the above studies on X-ray images, the classification of COVID-19 from Pneumonia seems to be a repeating objective. The major limitation involves the lack of data available, therefore the generalizability and stability of the model are yet to be evaluated.

2.3.3 Heatmap Visualization of Deep Learning Results

Deep learning techniques are often regarded as "Black Boxes" concerning its training and classification mechanisms. In medical applications such as this topic where we aim to diagnose patients

with COVID-19 in real-time, understanding the reasoning behind the model's prediction is of utmost importance.

More than just an additional insight, visualization of distinct features from the lung segmentation images would allow radiologists to cross-check their findings with that of the deep learning model and thus allow for an even better and reliable diagnosis.

Saliency methods are a set of popular and powerful tools which allow researchers to analyze and understand deep learning decisions [35]. Various interpretation methods exist in deep learning with a few of them mentioned below:

- **Saliency Map** - Estimates specific parts of the image which contributes to highest layer activation [36].
- **Class Activation Mapping (CAM)** - Averages and adds the activation's of each feature map (Global Average Pooling) and uses this to highlight important regions [37].
- **Gradient-Weighted CAM (Grad-CAM)** - Calculates the gradient of the classification score with respect to the convolutional features [38].

The saliency maps shown in Figure 2.5 and 2.6 highlight regions in the lungs which as discussed in Section 2.2.2 exhibit the most common characteristics in patients diagnosed with COVID-19. GGO's, consolidations, lesions, and crazy-paving patterns are some of the more contributing features for diagnosis purposes by the deep learning model as indicated by the saliency maps.

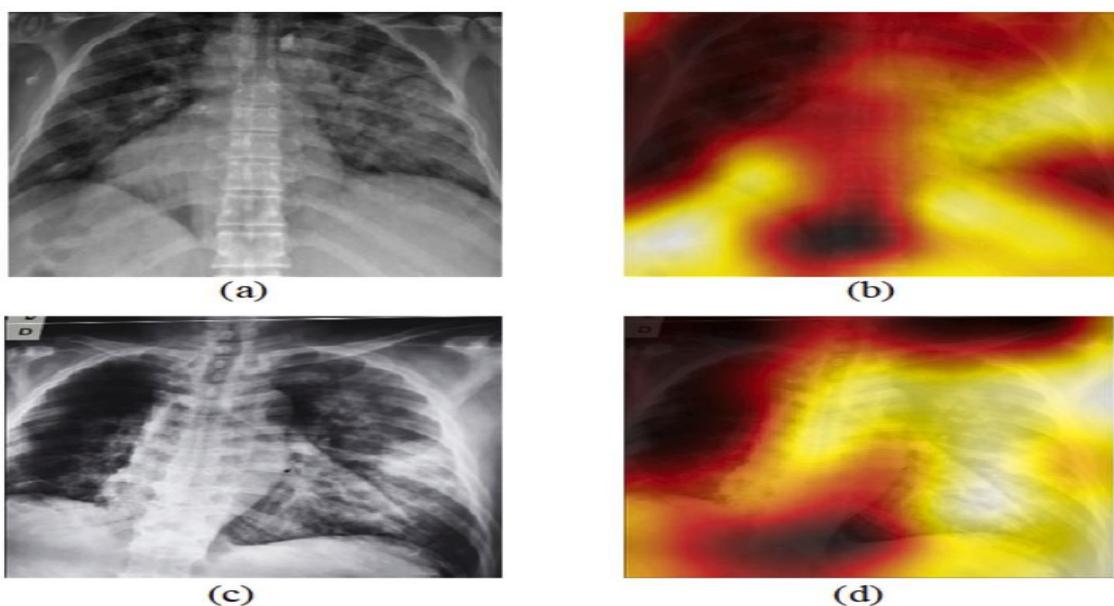


FIGURE 2.5: X-ray images and the corresponding heat maps [39]

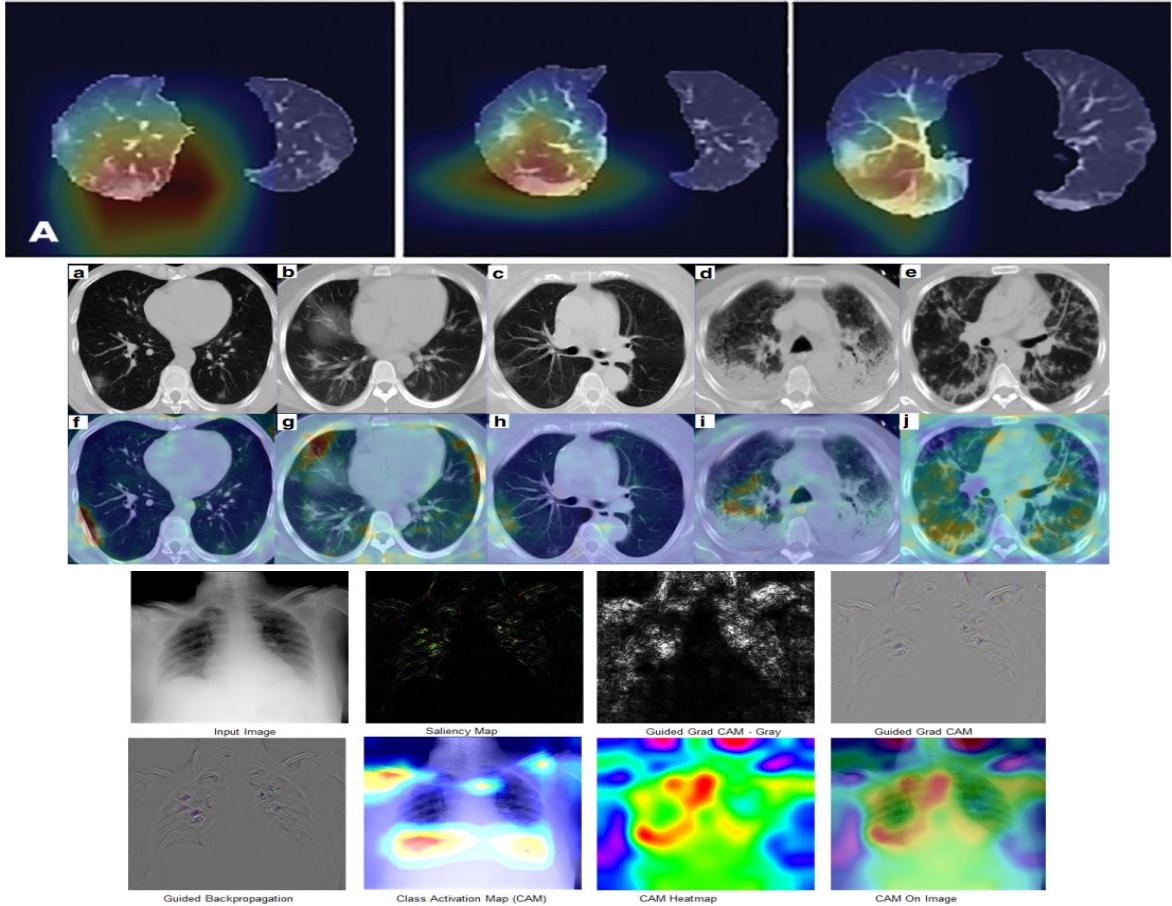


FIGURE 2.6: Saliency Maps displaying COVID-19 lung characteristics in CT and X-ray scans [14] [30] [40]

The results of saliency maps are therefore, in direct correlation with studies shown above which displays the most repeating COVID-19 lung characteristics. This provides additional assistance and assurance to radiologists who during diagnosis, look to identify the same lung characteristics.

2.4 Discussion

The integration of AI techniques in medical research has just scratched the surface. We have discussed many studies which have experimented on fully automated COVID-19 diagnosis employing deep learning methods and have yielded respectable results. But as mentioned, this is only the beginning and medical industry is due for a breakthrough soon.

Among many medical applications, where AI could potentially optimize and improvise standard procedures, automated imaging acquisition workflows as discussed previously, seem to be at the forefront. The overall efficiency of scanning procedure whether it be CT or X-rays could be enhanced and as a direct result, exposure to transferable viruses such as the coronavirus could be

decreased, thus protecting medical professionals.

Section 2.3.3 discusses Class Activation Mapping (CAM) which is a conventional technique to focus on prominent pixels or regions which lead to a resultant classification. Studies which have utilized this technique for COVID-19 diagnosis have seen a close correlation between the patterns identified by radiologists and the regions highlighted by the model therefore ensuring the reliability of the predictions given by the model.

Explainable Artificial Intelligence (XAI) methods [41], [42] are a recent addition to deep learning interpretability techniques which provides a finer localization map and exhibits more intricate details when compared to Class Activation Mapping (CAM) techniques.

Critical Review

Using medical imagery for disease diagnosis has variety of applications, for example, diagnosing COVID-19 as seen through experiments conducted by multiple studies. But one noticeable caveat is the observed negative radiological signs during preliminary stages of the disease. This could have critical consequences in real-life circumstances especially amid a global pandemic.

Many studies mentioned above use U-Net architecture for lung segmentation and variants of CNN's such as ResNet and ResNet-50 for COVID-19 diagnosis. One of the limiting factors of AI-based experiments is the difficulty in extracting the reasons as to why the deep learning model resulted in a certain classification.

COVID-19 applications employing deep learning often require accurate labeling of data, but this task often is very time-consuming and medical professionals due to the rising numbers often would not be able to carry out this procedure. This leads to incomplete and inaccurate labeling of data which proves to be an additional challenge to deep learning models who train on this dataset and therefore result in an incorrect diagnosis.

Exploiting unsupervised training techniques [43], [44] or exploring transfer learning methods [45] where filters applied on a completely different dataset could be reused for COVID-19 applications are viable options to mitigate the above labeling complication.

As the coronavirus was declared a global pandemic in early 2020, limited follow-up studies exist to identify treatment mechanisms and evaluating diagnosis tools. Few studies such as from Chung et al. [8], mentioned earlier did in fact carry out follow-up CT scans to recognize progression

patterns from segmented lung images, but due to the lack of participants when compared to the initial version, the results obtained could not be reliably applied in the real world.

Tracking of COVID-19 patients is essential for containing the spread of the virus. Therefore medical facilities should enforce strict patient follow-up protocols carrying out long-term monitoring and capturing progression data which could be used by studies to further develop their models and deploy them in the real world.

2.5 Conclusion and Research Questions

The coronavirus has affected millions of lives throughout the world. Front-line workers combat the virus tirelessly to counter the rapidly rising numbers daily. In this chapter, we have discussed how deep learning could provide a safe, efficient and accurate workflow for COVID-19 diagnosis and be a suitable alternative to the RT-PCR test.

It is important to couple the results obtained using the same imaging workflow discussed in this chapter, with clinical observations and laboratory results to provide reliable COVID-19 diagnosis. From the experiments and studies discussed, it is safe to conclude that AI if utilized effectively could play a vital role in accurate diagnosis and analysis of COVID-19, and therefore potentially save precious lives and ultimately lead to our victory against the coronavirus global pandemic.

Research Questions

Based on the gaps we identified in this chapter, we plan to answer the following research questions:

1. Is there a correlation between the ROIs detected by the deep learning model and the lung characteristics observed on COVID-19 patients?
2. Are the results obtained by the deep learning approach better when compared to the standard RT-PCR test?
3. Is it feasible to deploy and utilize the deep learning model in medical facilities and laboratories for rapid real-time COVID-19 diagnosis?

Chapter 3

Project Implementation

In this section we explore the implementation workflow followed throughout project development phase. This chapter is divided into two main sections, that is, X-rays and CT scans. Within each of these sections, we provide a comprehensive discussion on the approaches and concepts utilized. The implementation workflow ranges from Data Collection to Heatmap Visualization. Before we dive into these sections, we briefly describe our requirements and development environment.

3.1 Requirements Analysis

Our implementation is indeed focused around the Functional Requirements carefully curated during Stage 1 of this project. Most of our requirements revolved around the segmentation and diagnosis of chest X-rays and CT scans, with the former being a mandatory requirement.

Fortunately, we have been able to experiment with both X-rays and CT scans and successfully build a real-time diagnosis system. Other requirements focused on heatmap generation, multi-class diagnosis, and web interface, which have also been implemented. In addition, we have validated a subset of results obtained with findings observed by a senior Radiologist.

Our Functional and Non-Functional Requirements generated during Stage 1 along with the Evaluation Strategy used as the blueprint for Chapter 4, can be found in Appendix A.

3.2 Development Environment

We have utilized the free-to-use tier of Google Colaboratory to develop our deep learning models. Google provides access to Jupyter Notebooks powered by Tesla K80 GPU with 12 GB RAM and 32

GB disk space. It is also possible for free tier users to access limited number of Tensor Processing Units (TPUs). We have used the latter to enhance performance speed and save processing time.

3.3 Source Code

As mentioned in Section 3.2, we have used notebooks provided by Google Colab for developing and evaluating our models. All datasets and notebooks utilized in this project, Flutter code used for our Diagnosis Portal, Python scripts and additional resources are uploaded on our OneDrive repository¹. We have uploaded the Source Code on GitHub as per the requirements². We have also deployed our Diagnosis Portal that allows real-time COVID-19 Diagnosis³.

3.4 X-ray Scans

The following section illustrates the workflow utilized to collect data, develop and test our deep learning models. We also demonstrate various concepts used to improve model performance.

3.4.1 Data Collection

Our model is trained and tested on images from a Kaggle dataset⁴. The dataset is compiled from multiple sources [46]–[48]. It is an open-source database of COVID-19 cases which includes X-ray scans. It includes COVID-19 and Pneumonia cases as well as Healthy Chest Scans.

Furthermore, given the objective of our classification task to identify COVID-19 patterns, only the posterior-anterior view of the lungs was considered for training and testing. This view visualizes the bony thoracic cavity, mediastinum, and great vessels [49]. This choice allowed us to reduce the number of training instances and work with a less unbalanced dataset.

3.4.2 Data Pre-processing

As we successfully identified a suitable dataset and collected the required X-ray scans, we then pre-process the data and prepare it for model training purposes. This section highlights all the pre-processing procedures carried out.

¹The OneDrive Repository can be found here: https://heriotwatt-my.sharepoint.com/:f/g/personal/ag12_hw_ac_uk/EmzALshj8o1HrfeelAF8h7ABPoVhQBsnViMaiNgTp9Zc6g?e=FbzZhj

²GitHub Repository can be found here: <https://github.com/AlisterLuiz/Dissertation>

³Link to the Diagnosis Portal: <http://40.76.124.61/>

⁴Available at: <https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia>

3.4.2.1 Data Augmentation

To generate more training samples and reliable training results we have performed Data Augmentation. It was indeed necessary to balance the dataset before augmenting our scans.

In each fold of the 10-fold cross validation, we take 10% of dataset out to be used as validation dataset. The remaining 90% undergoes an augmentation step. The augmentation step uses three operators to produce slightly modified image, namely rotation, zooming and shearing. Image rotation rotates image by 5° . Zooming applies a +2% zoom. Shearing distorts an image along an axis (essentially converting rectangles into parallelograms), applies a 2° counter-clockwise distortion.

To perform augmentation with these three operators we have utilized Keras' Image Data Generator class which generates batches of tensor image data with real-time data augmentation [50]. Given memory limits of Google Colab, we have generated around 600 images per class using training data which was appended to our existing dataset. Therefore, augmented training data was fed into our deep learning model for training. A summary of our dataset is provided in Section 3.4.2.2.

3.4.2.2 Exploratory Data Analysis

As an initial insight into our dataset, we have plotted the number of X-ray scans per class before and after we have balanced the dataset, these figures are plotted in Figure 3.1. From our initial data collection, we received 575 COVID-19, 4273 Pneumonia, and 2583 Healthy Chest Scans. After balancing, our dataset comprised of 572 scans from each class respectively.

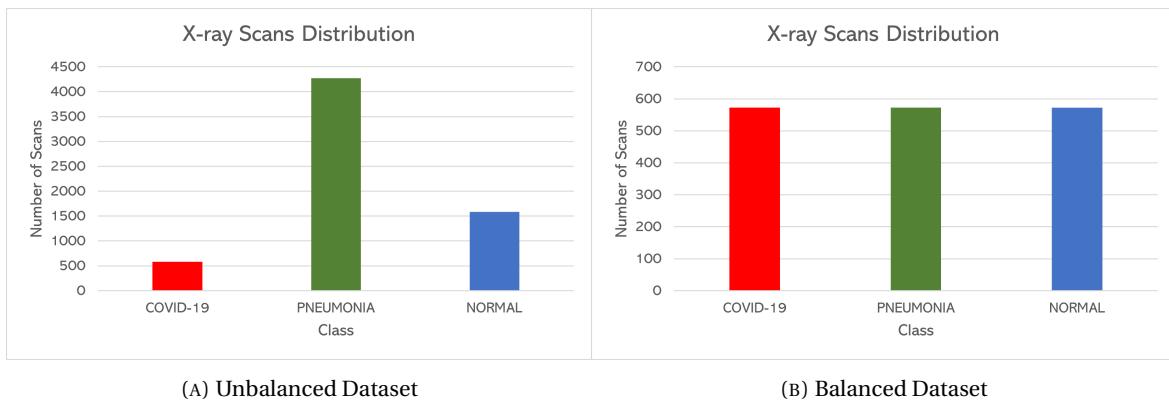


FIGURE 3.1: Dataset Distribution for X-ray Scans

Post augmentation we have been able to yield more scans per class to further enhance the generalizability of classification. Table 3.2 provides a detailed overview before and after augmentation. Figure 3.2 contains sample X-ray scans from our dataset along with their corresponding labels.

Disease	Total	Non-augmented Training	Augmented Training
COVID-19	572	468	1070
Pneumonia	572	468	1070
Normal	572	468	1070

TABLE 3.1: Training Dataset Post Augmentation

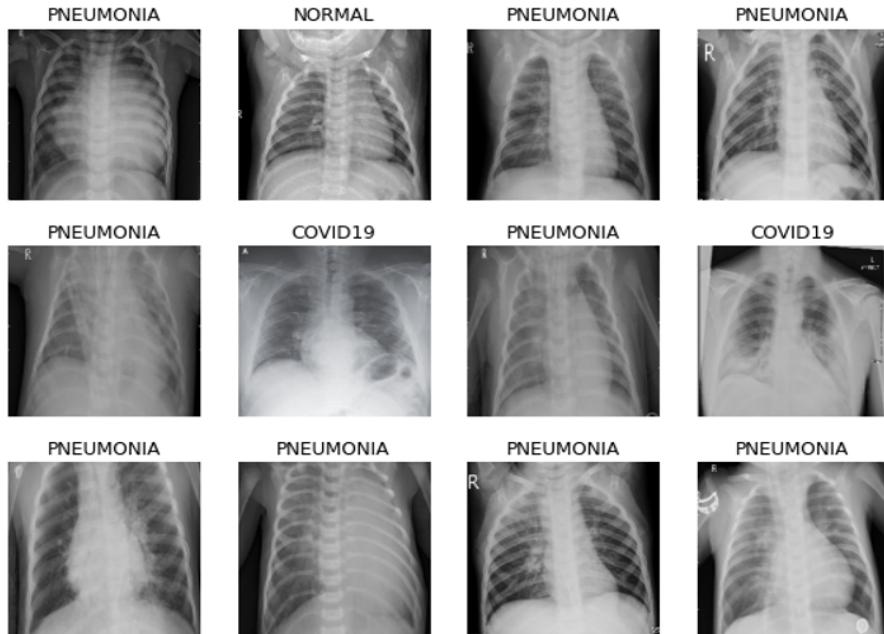


FIGURE 3.2: Sample of X-ray scans with labels for model training.

3.4.3 Methodology

As we have now explored our dataset, we present the methodology used to construct and develop our deep learning models.

3.4.3.1 Proposed Models

We have constructed, developed, and evaluated three pre-trained base models on the ImageNet dataset [51], that is, DenseNet121, ResNet50, and VGG16. Each of these base models were trained on the augmented dataset, and ensembled to further improve classification performance. The model architectures for each of these three models are provided in Figure 3.3.

Each of the pre-trained models we have experimented with has their own set of advantages. Dense-Nets are known to ease vanishing-gradient problem, reinforce feature propagation, promote feature reuse, and reduce the number of training parameters [52].

Residual Networks are known for their faster training speeds and powerful representational ability [53], [54]. Furthermore, from Chapter 2 we observe that ResNet variants seem to be a popular choice among studies performing image classification.

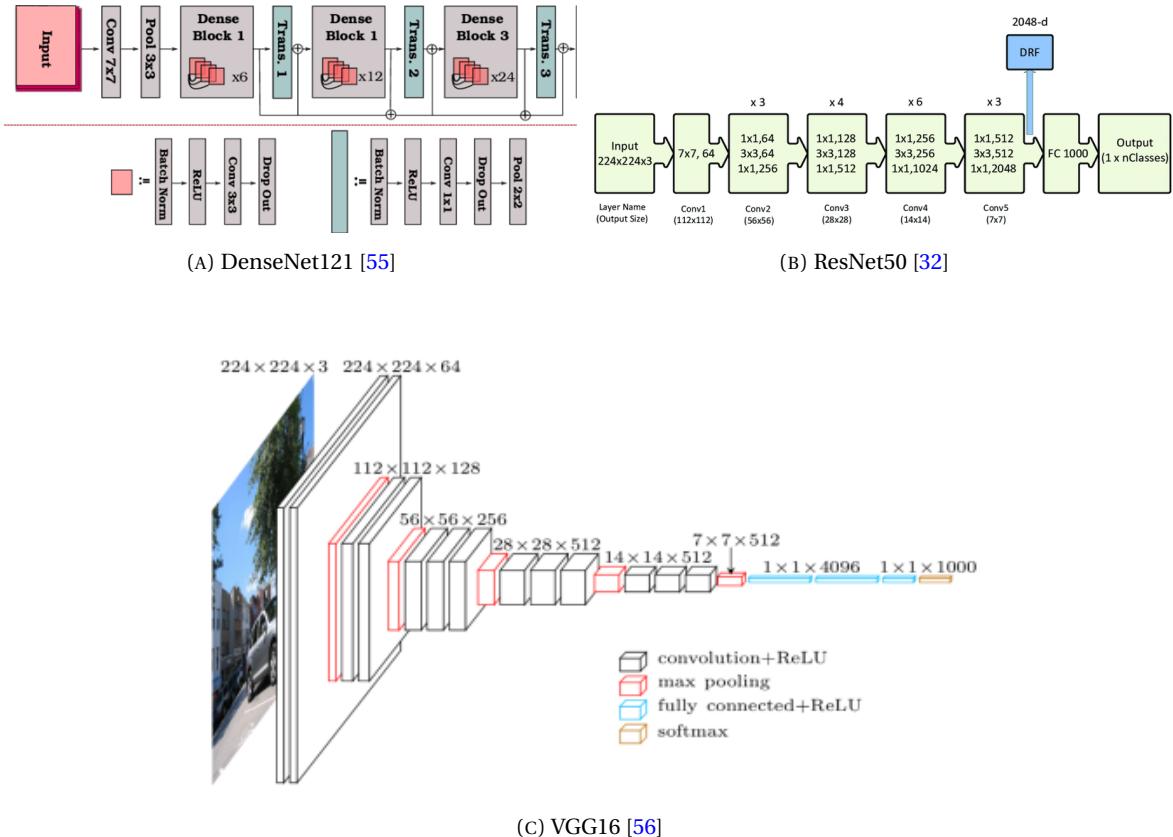


FIGURE 3.3: X-ray Model Architectures

VGG is the simplest model among the three and is usually used for benchmarking [57]. Built as a deep CNN, VGG outperforms baselines on many tasks and datasets outside of ImageNet [58].

These pre-trained models were readily available on the Keras applications library [59]. We have used these pre-trained models as a starting point to build our classification models.

```
pretrainedDenseNet = tf.keras.applications.DenseNet121(input_shape = (imageSize,
    imageSize, 3), weights = 'imagenet', include_top = False)
```

```
pretrainedResNet = tf.keras.applications.ResNet50(input_shape = (imageSize,
    imageSize, 3), weights = 'imagenet', include_top = False)
```

```
pretrainedVGG = tf.keras.applications.VGG16(input_shape = (imageSize, imageSize,
3), weights = 'imagenet', include_top = False)
```

3.4.3.2 Keras Callbacks

The Keras Callbacks API is an effective utility which aids in enhancing model training. A Callback is an object that can perform actions at various stages of training, usually in between epochs or when training begins or ends [60]. For training our deep learning models we utilized three Callback functions, that is, EarlyStopping, ReduceLROnPlateau, and ModelCheckpoint.

EarlyStopping (ES) ends training when a monitored metric has stopped improving [61]. For each of our deep learning models we have used the ES Callback with the following parameters.

```
esCallback = tf.keras.callbacks.EarlyStopping(monitor='val_loss', patience=15,
verbose=1)
```

We monitor validation loss and terminate our training as soon as model converges. An added benefit of using this Callback is reduced processing time required to train our deep learning models.

The next Callback, ReduceLROnPlateau reduces the learning rate when a certain metric has stopped improving, in our case, the validation loss. This Callback proved to be fruitful while training our deep learning models. Reducing the Learning Rate by a factor of 0.5, helped the models to converge faster and overcome learning plateaus [62].

```
lrReduce = tf.keras.callbacks.ReduceLROnPlateau(monitor='val_loss', factor=0.5,
min_delta=0.0001, patience=8, verbose=1)
```

ModelCheckpoint, simply saves the best performing model during training depending on the given parameter [63].

```
mcpSave = ModelCheckpoint('Path to Save Model', save_best_only=True,
monitor='val_loss', mode='min')
```

Utilizing each of our these three callbacks allowed us to increase the efficiency and performance of our models.

3.4.3.3 Transfer Learning

Transfer learning is a machine-learning technique where a model trained for an application with some dataset is reused on another dataset. The pre-trained model approach to transfer learning consists of using the source model as a starting point for training the target model.

In order to reduce training time and improve accuracy, we adopted the pre-trained transfer learning approach. We have used a model trained with large volume of images from the ImageNet dataset for classification tasks [51]. Using the pre-trained model, we trained only the final layer so as to alter its weights to suit our dataset. All other layers used fixed weights obtained from the prior training using the ImageNet dataset. A very simple illustration of the Transfer Learning approach is shown in Figure 3.4.

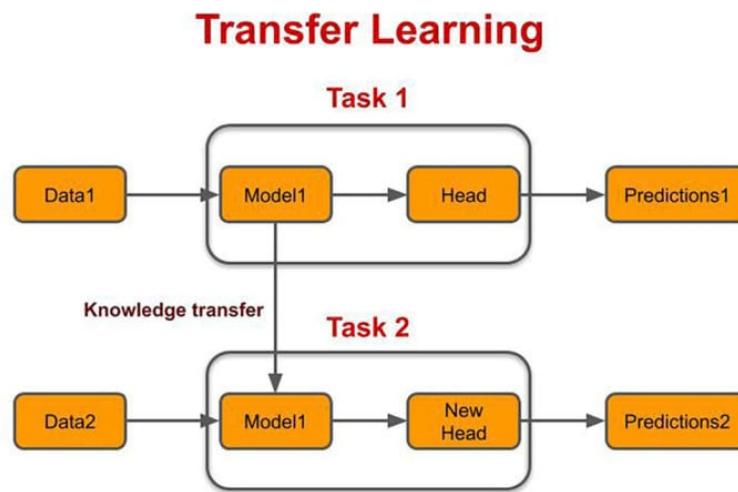


FIGURE 3.4: A simple representation of Transfer Learning Technique [64]

3.4.3.4 Ensemble Learning

Given our base models, an ideal approach to improve upon the classification performance was to implement Ensemble Learning. Ensemble Learning involves combining several classifiers to solve a particular problem. The combined model generally achieves optimum predictive performance when compared to any of the constituent learning algorithms alone [65], [66].

Our approach involves utilizing the best performing model after cross-validation from each of the three base models. We take the last dense layer of each model, that is, without the heads, and

provide the network an opportunity to learn from these dense layers before applying the softmax function. Indeed, we have frozen the layer weights of our base models, before ensembling.

Keras' Concatenate layer [67] allows us to easily merge or stack these dense layers and yield the final prediction after applying the softmax activation. As expected, we have observed an increase in classification performance through ensembling our base models. An illustration of the same is provided in Figure 3.5.

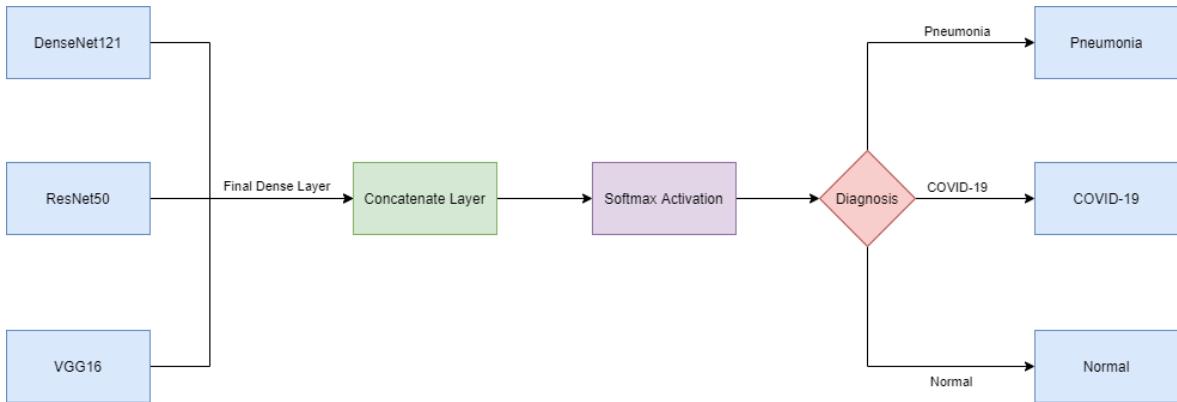


FIGURE 3.5: Ensemble Learning Illustration for X-ray Scans

3.4.3.5 Model Training

We have used 10-fold cross validation and leave an eleventh fold to test the 10 models on it as unseen data. Indeed, 10-fold cross validation produces 10 models that, collectively, saw all the training dataset. We have seen our dataset distribution is Section 3.4.2.2. We built a balanced dataset of 572 samples of each of the classes (572 is a multiple of 11). The size of the eleventh fold, to which we refer to from now on as *unseen data*, is 52.

Disease	Total	Non-augmented Training	Augmented Training	Non-augmented Testing	Unseen Data
COVID-19	572	468	1070	52	52
Pneumonia	572	468	1070	52	52
Normal	572	468	1070	52	52

TABLE 3.2: Dataset Distribution

At each of the 10 iterations of 10-fold cross validation, we take 10% of the dataset out to be used for testing. The remaining 90% undergoes the augmentation steps as mentioned in Section 3.4.2.1.

The augmentation of each of the 468 samples of each class (90% of 520) resulted in creating 1070 samples per class. Since we have three classes, COVID-19, Pneumonia and Normal, the total number of samples in the final training dataset is of 3,210. Only the training dataset was augmented; no augmentation was applied to the test dataset. Table 3.2 shows the size of the data we used.

We trained our models for 50 epochs with a batch size of 210 per forward/backward pass. We have also initialized all the callbacks mentioned in Section 3.4.3.2. Post training, the results were evaluated using scikit-learns classification report function [68]. A confusion matrix was also displayed after every epoch indicating the model performance [69], which also allows us to extract useful metrics and determine the best performing model. A sample snippet of code which calls the fit function is provided below.

```
history = model.fit(augmentedDataX, YCatAug, validation_data = (XTrain[test],
                                                               YCatVal), callbacks=[lrReduce, esCallback, mcpSaveDenseNet], epochs=50,
                                                               batch_size=210)
```

3.4.3.6 Heatmap Visualization

For the purpose of building a more interpretable model, we generated heatmaps for the provided X-ray scans. The heatmaps highlight the regions that led to the classification by our model. We calculated the Grad-CAM heatmaps with the help of Keras' Computer Vision examples [70]. The same Grad-CAM class activation visualization example provided by François Chollet is utilized in our implementation [71]. We have generated Grad-CAM heatmaps from each of our three base models. We have utilized Numpy [72] to average obtained heatmaps and produce the final consolidated heatmap.

Grad-CAM heatmaps highlight regions in the lungs which, as discussed in Section 2.3.3, exhibit the most common characteristics in patients diagnosed with COVID-19. These include features such as GGO's, consolidations, lesions, and crazy-paving patterns, which are some of the most contributing features to diagnosis. This technique provides interpretation means that would assist radiologists in identifying the same lung characteristics as with traditional segmentation-based methods. Regions in red are the most influential in making the decision, whereas blue regions are the least influential.

3.4.3.7 Workflow Summary Illustration

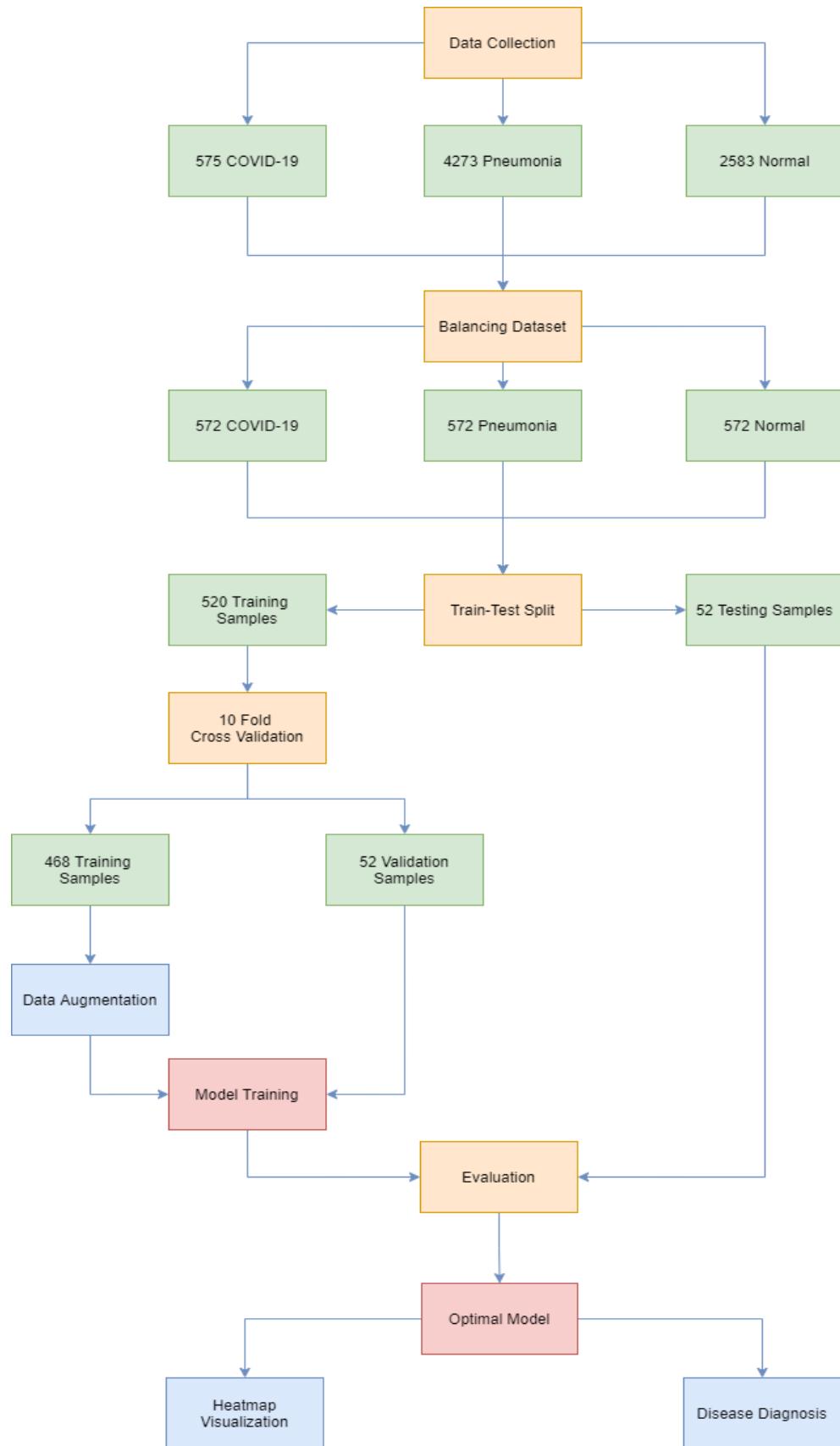


FIGURE 3.6: X-ray Workflow Illustration

3.5 CT Scans

The following section explores the second part of our project, that is, CT scans. A similar workflow has been utilized for CT classification purposes with minor changes in between such as Lung Parenchyma extraction as an additional pre-processing step.

3.5.1 Data Collection

We have used an open-source Kaggle CT dataset⁵ for training and testing our models. All the images present in this dataset was collected by the authors from the Frontier Exploration Program of Huazhong University of Science and Technology, the program for HUST Academic Frontier Youth Team, and the Fundamental Research Funds for the Central Universities [73], [74].

This dataset is comprised of two parts, the original and pre-processed CT scans of COVID-19 and Normal patients readily available for model training purposes. The code to pre-process the original CT scans was given by the author of this dataset. Applying them we were able to extract the lung parenchyma, which proved to be a vital pre-requisite to train CT data.

3.5.2 Data Pre-processing

In the following subsections, we provide a brief overview on the various Data Pre-processing techniques we have applied.

3.5.2.1 Lung Parenchyma Extraction

Even though the Kaggle dataset already contains the pre-processed dataset, it was necessary for us to understand the process involved in extracting the lung parenchyma, so as to do the same for incoming unseen samples in a real-time application.

This process for extracting the lung parenchyma is as per the suggestions by Ning et al. [74]. The source code involved a fill water function that applied certain filter masks to further fine tune the extracted lung parenchymas. An example of a CT scan image before and after extracting lung parenchyma is displayed in Figure 3.7. Indeed, the processed CT images were used for model training and evaluation purposes.

⁵Available at <https://www.kaggle.com/azaemon/preprocessed-ct-scans-for-covid19>

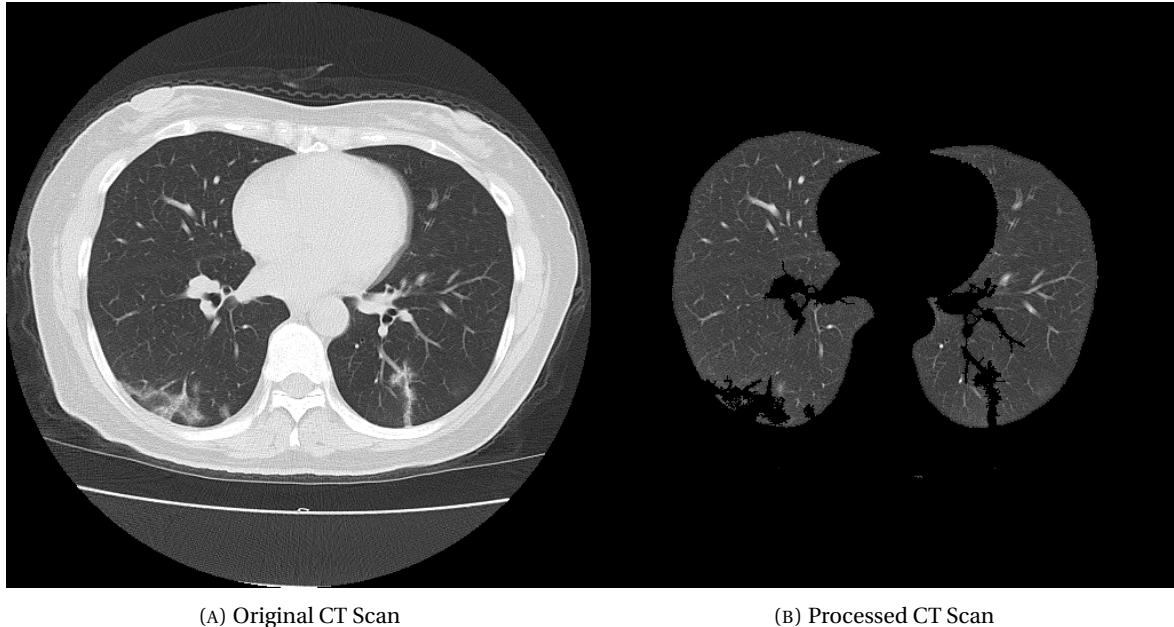


FIGURE 3.7: Lung Parenchyma Extraction

3.5.2.2 Data Augmentation

Similar to our X-ray pre-processing workflow we have augmented the CT data so as to improve the reliability of our model prediction and handle unpredictabilities of real-world data. The same augmentation parameters were applied at each of the 10 iterations of 10-fold cross validation, that is, 5° rotation, +2% zooming, and 2° shearing. Keras' Image Data Generator class was once again utilized for performing real-time data augmentation [50]. Once again prioritizing the limits of Google Colab, we have generated around 2000 augmented images per class and is appended to the existing dataset.

3.5.2.3 Exploratory Data Analysis

To familiarize ourselves with the data, the first step was to plot the number of CT scans per class. Similar to X-ray scans, the plots include the counts before and after balancing the dataset. The plots are displayed in Figure 3.8. The number of training scans post augmentation is tabulated in Table 3.3. After collecting data from the Kaggle Dataset we have ended up 4001 COVID-19 and 9979 Normal scans. After balancing our dataset we have ended up with 3993 scans per class.

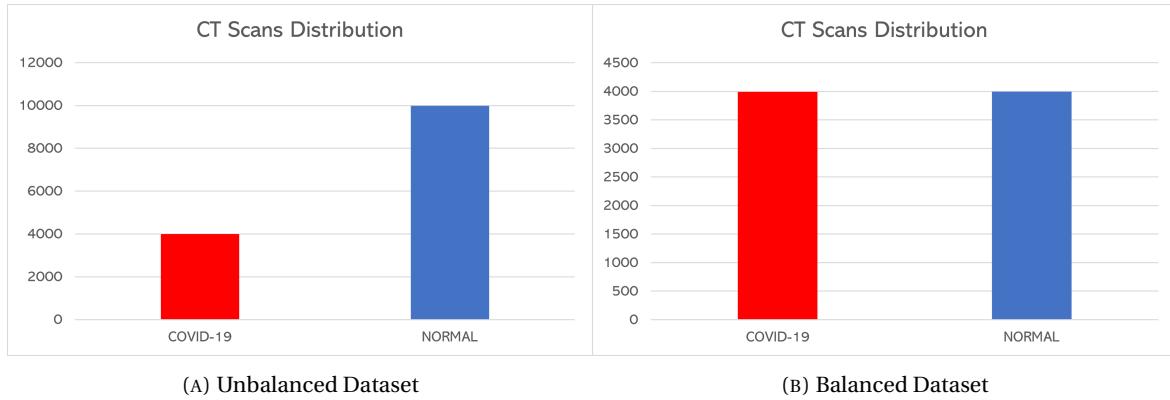


FIGURE 3.8: Dataset Distribution

Disease	Total	Non-augmented Training	Augmented Training
COVID-19	3993	3267	5270
Normal	3993	3267	5270

TABLE 3.3: Training Dataset Post Augmentation

We have displayed a subset of our original CT scans dataset along with their labels in Figure 3.9.

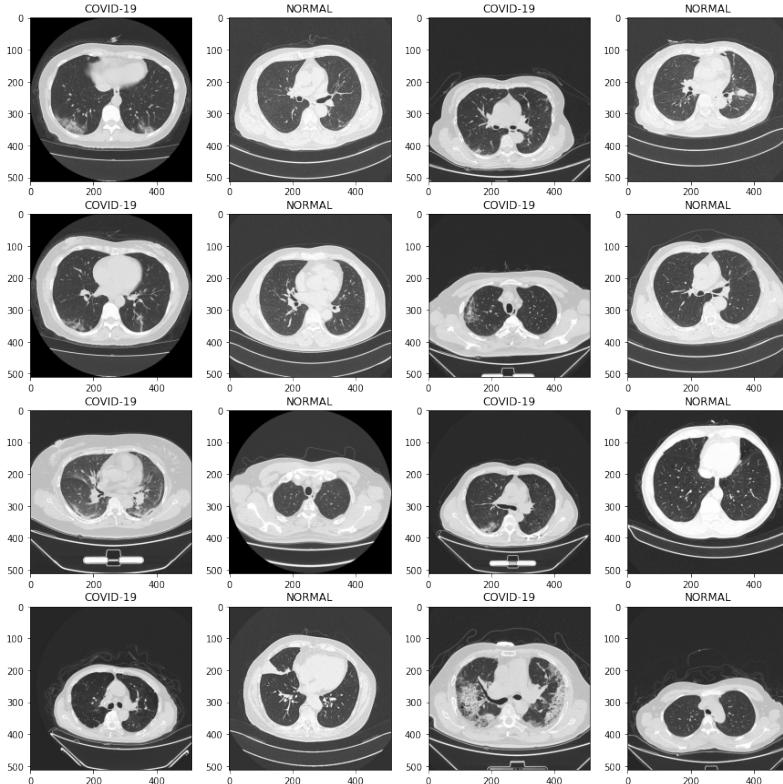


FIGURE 3.9: Sample of unprocessed CT scans with labels for model training.

3.5.3 Methodology

Our proposed methodology for CT scans is similar to that of X-ray's to keep our workflow consistent. There are indeed changes in our proposed models and certain parameter values. The following sections elaborates on these differences.

3.5.3.1 Proposed Models

For CT scans we have experimented with three models, that is, UNet, UNet++, and Attention UNet. These variants of the base UNet model seems to be popular among literature. The UNet++ and Attention UNet models were ensembled to improve the classification performance, similar to our X-ray diagnosis workflow. Each of these three model architectures are provided in Figure 3.10.

The UNet model was originally created for Bio-medical Image Segmentation purposes [18]. As its name indicates, the architecture consists of a 'U' structure, as shown in Figure 3.10a. The basic idea revolves around the notion that the same feature maps used for image contraction is utilized to expand a vector to a segmented image. This would help in preserving the structural integrity of the image and greatly reduce image distortion [75]. The UNet architecture is comprised of three sections, which are, contraction, bottleneck, and expansion respectively.

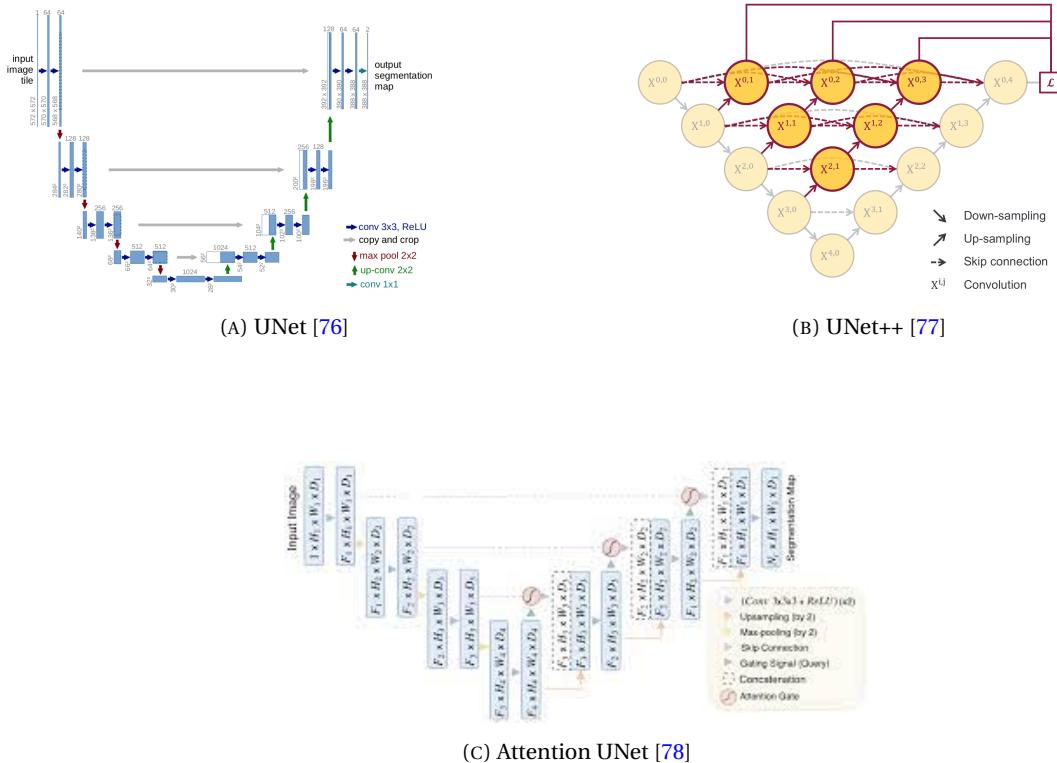


FIGURE 3.10: CT Model Architectures

UNet++ is another powerful architecture for medical image segmentation [21]. UNet++ has three major additions when compared to the original UNet model proposed by Ronneberger, that is, redesigned skip pathways, dense skip connections, and deep supervision [79]. One caveat of using the UNet++ due to these additions is that the training time is almost doubled when compared to a simple UNet.

Attention UNet aims to only highlight the relevant activations during training, thus significantly reducing computation time. This is made possible using attention gates which uses additive soft attention. As its name suggests, the name pays attention or focuses on certain parts of the provided image [22], [80]. When compared to a UNet model, the computation time is only slightly longer. The code snippets initializing each of the three models have been provided below.

```
pretrainedUNet = kerasModels.unet_2d((imageSize, imageSize, 3), filter_num=[64,
    128, 256, 512, 1024], n_labels=2, stack_num_down=2, stack_num_up=2,
    activation='ReLU', output_activation='Sigmoid', batch_norm=True, pool=False,
    unpool=False, backbone='ResNet50V2', weights='imagenet', freeze_backbone=True,
    freeze_batch_norm=True, name='unet')

pretrainedUNetPlus = kerasModels.unet_plus_2d((imageSize, imageSize, 3),
    filter_num=[64, 128, 256, 512, 1024], n_labels=2, stack_num_down=2,
    stack_num_up=2, activation='ReLU', output_activation='Sigmoid',
    batch_norm=True, pool=False, unpool=False, backbone='ResNet50V2',
    weights='imagenet', freeze_backbone=True, freeze_batch_norm=True,
    name='unetplus')

pretrainedAttUNet = kerasModels.att_unet_2d((imageSize, imageSize, 3),
    filter_num=[64, 128, 256, 512, 1024], n_labels=2, stack_num_down=2,
    stack_num_up=2, activation='ReLU', atten_activation='ReLU', attention='add',
    output_activation='Sigmoid', batch_norm=True, pool=False, unpool=False,
    backbone='ResNet50V2', weights='imagenet', freeze_backbone=True,
    freeze_batch_norm=True, name='attunet')
```

3.5.3.2 Keras Callbacks

We have utilized the same set of Callbacks used for our X-ray models, that is, EarlyStopping, ReduceLROnPlateau, and ModelCheckpoint. However, there have been differences in the parameters especially for EarlyStopping as the UNet models tend to stop training before reaching convergence. Therefore, it was necessary for us to increase the patience parameter depending the variant of UNet model. The following code snippets showcase the parameters used for EarlyStopping Keras Callback.

```
UNetESCallback = tf.keras.callbacks.EarlyStopping(monitor='val_loss', patience=80,
                                                    verbose=1)

AttUNetESCallback = tf.keras.callbacks.EarlyStopping(monitor='val_loss',
                                                      patience=80, verbose=1)

UNetPlusESCallback = tf.keras.callbacks.EarlyStopping(monitor='val_loss',
                                                       patience=140, verbose=1)
```

Furthermore, for the UNet++ model we have provided custom parameters for the ReduceLROnPlateau Callback to achieve optimal convergence and thereby increase classification accuracy.

```
UNetPluslrReduce = tf.keras.callbacks.ReduceLROnPlateau(monitor='val_loss',
                                                        factor=0.5, min_delta=0.0001, patience=16, verbose=1)
```

3.5.3.3 Transfer Learning

We have applied the Transfer Learning technique for CT scan classification as well to avail its benefits such has reduced training time and improvement in accuracy as discussed in Section 3.4.3.3. The only difference from our X-ray scan implementation is that we have used a different Python package, Keras UNet Collection [81], which contains pre-trained UNet model and its variants on the large ImageNet dataset [51].

This package allowed us to easily access different UNet variants and contains various tweakable parameters allowing us to customize or fine-tune our models for training and evaluation purposes. From the code snippet we can clearly see the weights parameter which initializes the model with ImageNet weights. Similar to our X-ray implementation, we have only set the final layer to be trainable.

```
pretrainedUNet = kerasModels.unet_2d((imageSize, imageSize, 3), filter_num=[64,
128, 256, 512, 1024], n_labels=2, stack_num_down=2, stack_num_up=2,
activation='ReLU', output_activation='Sigmoid', batch_norm=True, pool=False,
unpool=False, backbone='ResNet50V2', weights='imagenet', freeze_backbone=True,
freeze_batch_norm=True, name='unet')
```

3.5.3.4 Ensemble Learning

We have ensembled the UNet, Attention UNet and UNet++ models to further boost our classification performance on the CT dataset. We have utilized a similar ensembling workflow but with different base models as discussed in Section 3.4.3.4. We have extracted the final dense layer, froze all the trained layers from each of our base models to prevent re-training, and simply merged these dense layers before applying the softmax function. Once again, through ensembling we have observed an increase in model classification performance. Figure 3.11 displays an illustration of our ensemble learning workflow for CT scans.

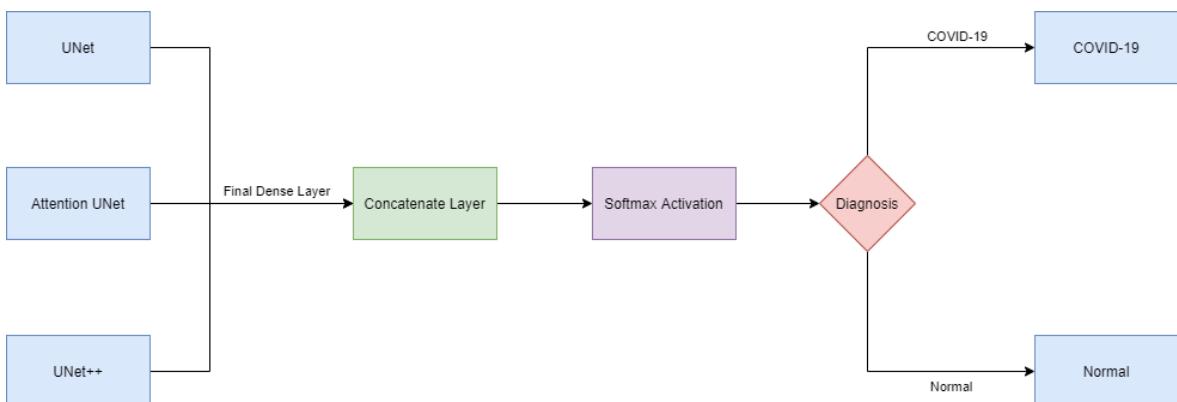


FIGURE 3.11: Ensemble Learning Illustration for CT Scans

3.5.3.5 Model Training

We have once again utilized 10-fold cross validation to train our CT models. We have also left an eleventh fold to test our model on unseen data. As seen from Section 3.5.2.3, we have built a balanced dataset comprised of 3993 samples from each class (3993 is a multiple of 11). The size of the eleventh fold also referred to as *unseen data*, is 363.

At each iteration of the 10-fold cross validation, 90% of the training data undergoes the augmentation steps and the remaining 10% is used for validation purposes.

Disease	Total	Non-augmented Training	Augmented Training	Non-augmented Testing	Unseen Data
COVID-19	3993	3267	5270	363	363
Normal	3993	3267	5270	363	363

TABLE 3.4: Dataset Distribution for CT Scans

Augmentation of 3267 samples from each class resulted in 5270 samples as our augmented training dataset. Our final training dataset is therefore comprised of 10,540 samples as we have two classes, COVID-19 and Normal. Indeed, only the training dataset was augmented, our testing dataset was left aside only for validation purposes. Table 3.4 displays the size of our data.

We have trained our UNet and Attention UNet models for 200 epochs each, whereas for UNet++ it took around 300 epochs to achieve convergence. The batch size was 1024 samples per forward/backward pass. Our ensemble model was trained for 50 epochs to give it an opportunity to learn from the final dense layers from each of our base models before returning a prediction. Similar to X-ray scans, we have initialized the same Callbacks, and returned the classification report [68] and confusion matrix [69] after every fold which allowed us to determine the best performing model.

3.5.3.6 Heatmap Visualization

For interpreting our model results we have generated heatmaps in addition to the diagnosis result. The same Grad-CAM technique utilizing Keras library [70] was applied in the case of CT scans to highlight common characteristics in patients diagnosed with COVID-19. We have once again averaged the heatmaps produced by each of our three base models with the help of Numpy [72] to obtain the final heatmap. We believe these heatmaps would aid radiologists in identifying the same lung characteristics more efficiently.

There is only one difference in our implementation when compared to X-rays. Given a test image, it was necessary for us to extract the lung parenchyma before proceeding with the diagnosis and generating the heatmaps. The heatmap filter generated was superimposed to our unprocessed CT input in order to highlight the most critical regions.

3.5.3.7 Workflow Summary Illustration

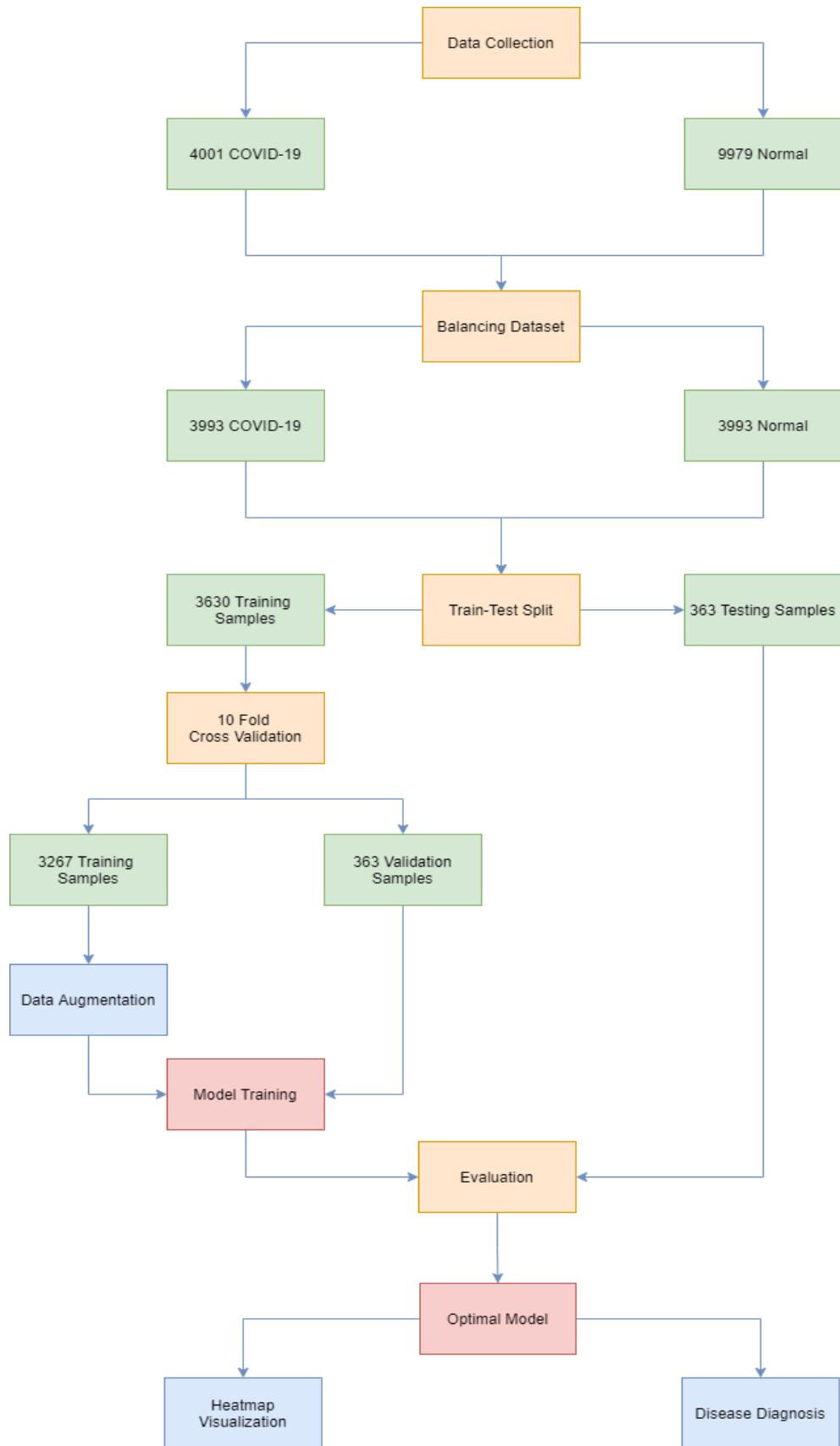


FIGURE 3.12: CT Workflow Illustration

3.6 Web Interface

Our Web Interface is a simple Diagnosis Portal allowing users to input X-ray or CT scans and retrieve real-time diagnosis and a heatmap highlighting critical regions. The front-end of the web application is built using Google's Flutter UI Toolkit. A Flask application is also deployed serving as backend containing Python scripts that retrieve diagnosis results. We have used an Ubuntu Virtual Machine offered by Microsoft Azure to deploy our Python Scripts and Web Application⁶.

The Flask API provides an interface for the Flutter app, which takes the scan as input and returns diagnosis and heatmap. To ensure responsiveness of our Portal, we have utilized a custom responsive folder architecture [82]. We have uploaded a video demonstration of the Web Application, along with screenshots on OneDrive⁷. A subset of screenshots is displayed in Figure 3.13.



FIGURE 3.13: Diagnosis Portal Screenshots

⁶Link to the Diagnosis Portal: <http://40.76.124.61/>

⁷Web Application Screenshots and Video Presentation can be found here: https://heriotwatt-my.sharepoint.com/:f/g/personal/agl2_hw_ac_uk/EgtAqrerqXZIhD3EPQJ0JBsBJ3VTXhCZv_9pm_R01Bf8pA?e=z3mHHj

Chapter 4

Results and Evaluation

In this chapter we analyze the performance of each of our trained models. We conduct a thorough performance evaluation and compare the results obtained with the RT-PCR test and a basic CNN. We also compare our results with related work highlighted in Section 2. The heatmaps obtained have also been compared to the findings observed by a professional Radiologist.

4.1 X-ray Scans

The following sections evaluates results obtained by each of our X-ray models, followed by comparing them across various parameters.

4.1.1 Performance Evaluation

We evaluate the performance of our three models, DenseNet121, ResNet50, and VGG16, across various metrics followed by showcasing the performance increase obtained by our Ensemble model. These sections are supplemented with tables, plots highlighting trends in loss and accuracy, and confusion matrices respectively. As we have performed 10-fold cross validation, we have obtained 10 confusion matrices whose values we have summed up to calculate average metric scores [83].

4.1.1.1 Pre-trained Models

The first model implemented is DenseNet121. Our 10-fold cross validation resulted in 94.6% average accuracy on the balanced dataset. The average and best accuracy obtained on unseen data are 95.1% and 96.8% respectively. We have displayed classification report results indicating Precision, Recall, and F1-scores per class in Table 4.3 and confusion matrix in Figure 4.1, followed by trends in accuracy and loss in Figure 4.2, of the best performing model.

	Precision	Recall	F1-Score
COVID-19	97.7%	97.7%	97.7%
Normal	93.6%	92.7%	93.1%
Pneumonia	92.6%	93.5%	93.0%
Accuracy		94.6%	

TABLE 4.3: DenseNet121 Classification Report

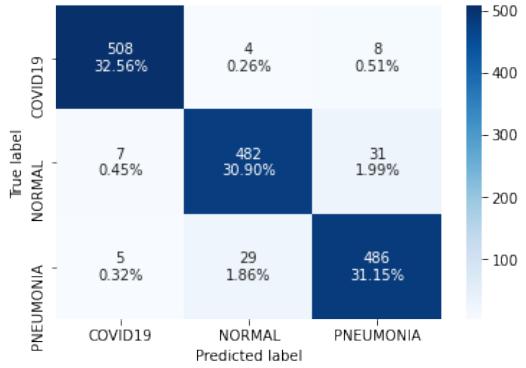


FIGURE 4.1: DenseNet121 Confusion Matrix

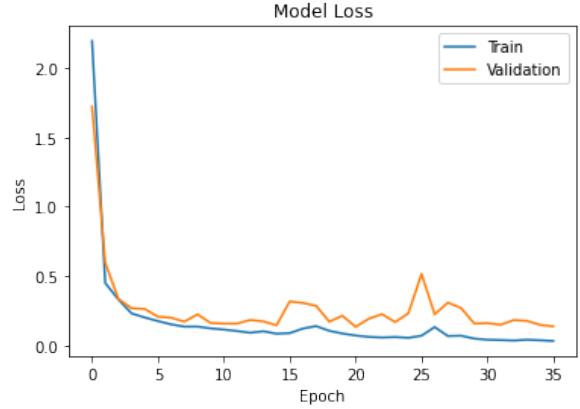


FIGURE 4.2: DenseNet121 Model Training Performance Trends

Next we have ResNet50 which is popular among literature as seen in Section 2. We have achieved an average accuracy of 95.8% after 10-fold cross validation on the balanced dataset. The average and best accuracy obtained on unseen data are 97.6% and 98.1% respectively. We have plotted the classification report, confusion matrix, and trends in accuracy and loss in Table 4.6, Figure 4.3, and Figure 4.4 respectively.

	Precision	Recall	F1-Score
COVID-19	98.9%	99.0%	98.9%
Normal	94.9%	93.3%	94.1%
Pneumonia	93.6%	95.0%	94.3%
Accuracy		95.8%	

TABLE 4.6: ResNet50 Classification Report

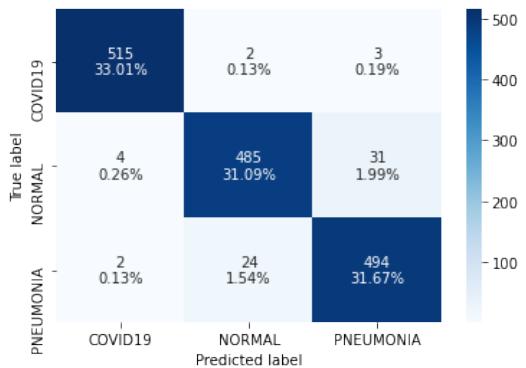


FIGURE 4.3: ResNet50 Confusion Matrix

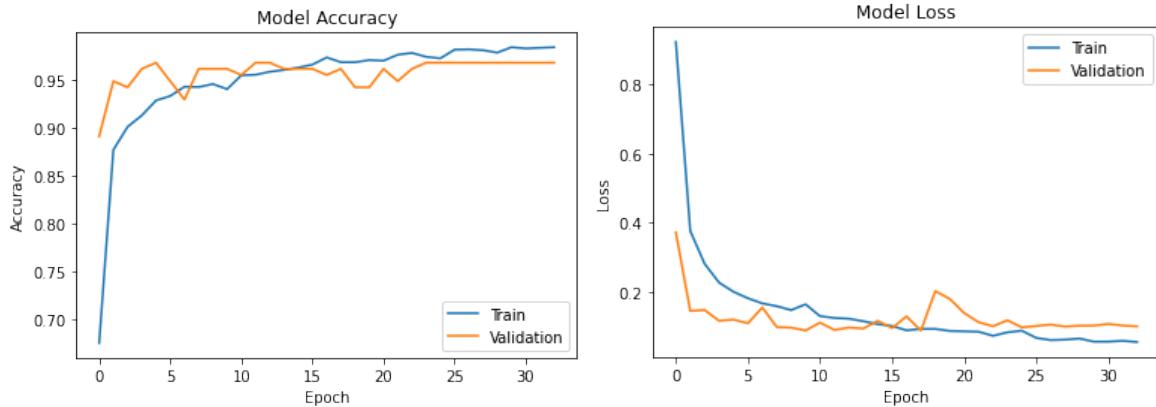


FIGURE 4.4: ResNet50 Model Training Performance Trends

The final base model is VGG16. We have achieved an average accuracy of 95.6% after 10-fold cross validation on the balanced dataset. The average and best accuracy obtained on unseen data are 95.3% and 96.8% respectively. The classification report, confusion matrix, and trends in accuracy and loss is displayed in Table 4.9, Figure 4.5, Figure 4.6 respectively.

	Precision	Recall	F1-Score
COVID-19	97.5%	98.9%	98.2%
Normal	94.1%	94.6%	94.3%
Pneumonia	95.3%	93.5%	94.4%
Accuracy		95.6%	

TABLE 4.9: VGG16 Classification Report

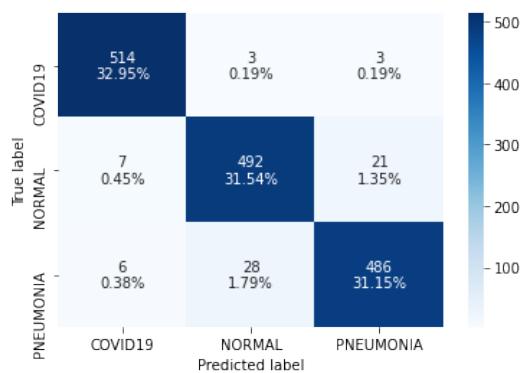


FIGURE 4.5: VGG16 Confusion Matrix

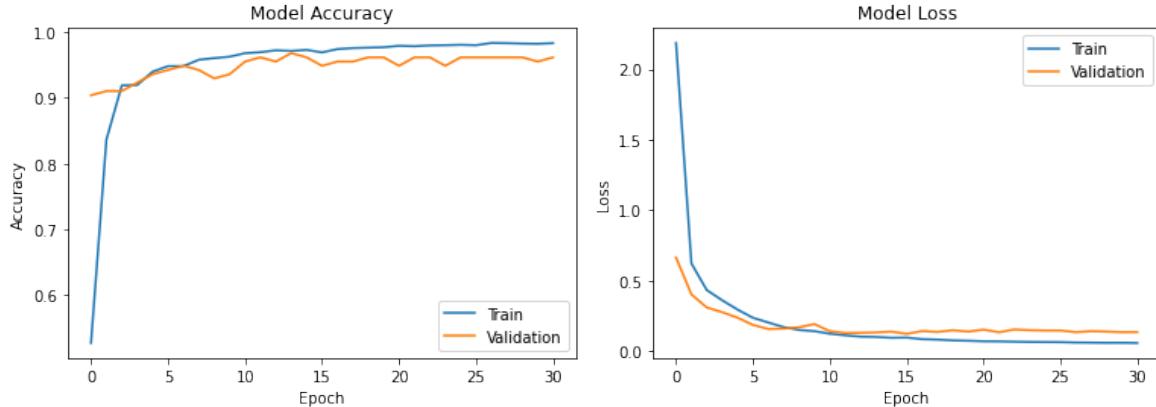


FIGURE 4.6: VGG16 Model Training Performance Trends

4.1.1.2 Ensemble Model

After ensembling each of the three base models discussed in Section 4.1.1.1, we have evaluated the performance utilizing the classification report function, confusion matrix and trends in accuracy and loss. These are displayed in Table 4.12, Figure 4.7, and Figure 4.8 respectively. We have achieved an average accuracy of 98.8% after 10-fold cross validation on the balanced dataset. The average and best accuracy obtained on unseen data are 97.8% and 98.1% respectively.

	Precision	Recall	F1-Score
COVID-19	99.8%	99.8%	99.8%
Normal	98.3%	98.5%	98.4%
Pneumonia	98.3%	98.1%	98.2%
Accuracy		98.8%	

TABLE 4.12: X-ray Ensemble Model Classification Report

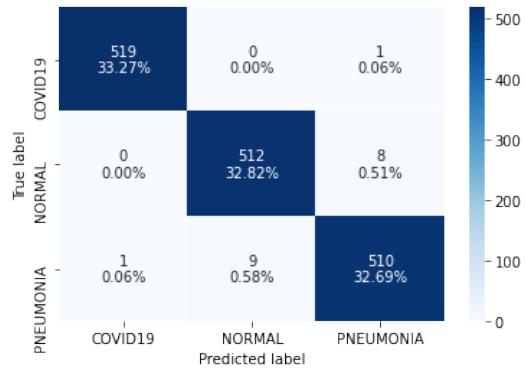


FIGURE 4.7: X-ray Ensemble Model Confusion Matrix

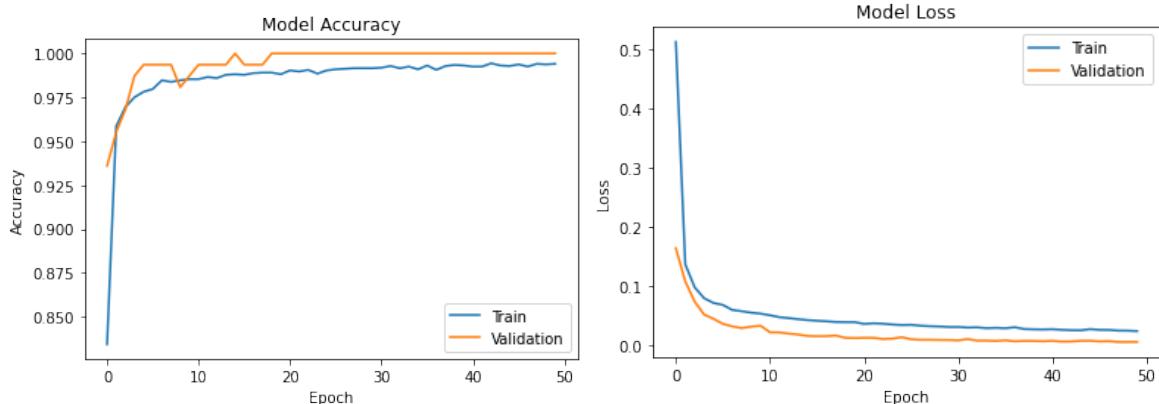


FIGURE 4.8: X-ray Ensemble Model Training Performance Trends

4.1.2 Performance Comparison

In the following sections we compare the performance of our models with respect to each other and determine the best performing model, followed by comparing the best results with that of the RT-PCR test and CNN. We also compare our results with related studies observed in the literature. We conclude this section by comparing our heatmap results with that of the findings obtained by a professional Radiologist who has volunteered to help us in this study.

4.1.2.1 Comparison of Trained Models

We compare each of our base X-ray models along with the ensemble model across various parameters such as Accuracy, Precision, Recall, F1-Score, Specificity, and Sensitivity respectively. The average metric scores are tabulated in Table 4.13.

	Accuracy	Precision	Recall	F1-Score	Specificity	Sensitivity
DenseNet121	94.6%	94.6%	94.6%	94.6%	97.3%	94.6%
ResNet50	95.8%	95.8%	95.8%	95.8%	97.9%	95.8%
VGG16	95.6%	95.6%	95.6%	95.6%	99.0%	97.8%
Ensemble Model	98.8%	98.8%	98.8%	98.8%	99.4%	98.8%

TABLE 4.13: X-ray Models Performance Comparison

We observe that out of the base models, ResNet50 and VGG16 have very comparable results with the former having a slight performance increase when compared to the latter. However, the Ensemble Model clearly outperforms each of our base models as expected.

4.1.2.2 Comparison with RT-PCR and CNN

We compare the results given by our Ensemble model, only for **COVID-19 class**, with the RT-PCR test according to the study conducted by the Infectious Diseases Society of America (IDSA) [84]. We also did implement a CNN-based model to compare our model with deep learning models that use state of the art techniques. Table 4.14 summarizes the comparison and uses three metrics, namely accuracy, sensitivity, and specificity.

	Accuracy	Sensitivity	Specificity
RT-PCR	94.0%	84.2%	98.9%
CNN	93.9%	85.2%	98.2%
Our Approach	99.9%	99.9%	99.8%

TABLE 4.14: Comparison with RT-PCR and CNN for X-ray scans

As expected, our Ensemble model performs significantly better when compared to a basic CNN model. More importantly, our approach also exhibits better performance when compared to the traditional RT-PCR test across all metrics.

4.1.2.3 Comparison with Related Work

In this section, we compare the performance of the Ensemble model with similar studies found in literature. We shall compare the results obtained by our best performing model with the scores reported by studies discussed in Section 2. Table 4.15 tabulates these results.

	Accuracy	Precision	Recall	F1-Score	Specificity	Sensitivity
Wang et al. [31]	83.5%	98.9%	91.0%	94.7%	99.5%	91.0%
Ghoshal et al. [30]	92.9%	66.7%	85.7%	75.0%	99.4%	85.7%
Zhang et al. [28]	96.0%	59.6%	96.0%	73.5%	70.7%	96.0%
Narin et al. [29]	98.0%	76.5%	91.8%	83.5%	96.6%	91.8%
Ensemble Model	98.8%	98.8%	98.8%	98.8%	99.4%	98.8%

TABLE 4.15: Comparison with Related Work

Our X-ray Ensemble model outperforms the results obtained by other studies across most metrics. We believe that various approaches considered such as Data Augmentation, Transfer Learning, and Ensemble Learning have helped our models achieve optimal classification performance.

4.1.2.4 Comparison with Radiologist Findings

To further verify that the heatmaps produced by the model are in conformance with the observed COVID-19 characteristics, we have provided a subset of X-rays scans to a senior Radiologist ¹ and have asked to highlight the critical regions and provide their diagnosis results.

Table 4.16 compares the Radiologist's annotations and diagnosis with those produced by our model along with the heatmap result. We can observe that the heatmaps are in direct compliance with the Radiologists' findings. Precisely, visually, the heatmaps show that our predictive model highlights at least one or more of the critical regions identified by the Radiologist. Furthermore, all areas highlighted by the Radiologist are shades of either red or purple, but never blue.

We have also verified that the heatmaps we obtained correspond to the regions highlighted from studies conducted by Harmon et al. [40] and Li et al. [14]. The lower lung lobes seem to be the most affected in patients diagnosed with COVID-19, this correlates with the observations from Chung et al. [8] and therefore, provides additional validation that lung characteristics such as GGO's, consolidations, and lesions are major contributing factors to COVID-19 detection.

¹The authors thank Dr. G.R. Mahadevan (MD, IDCCM,D.Diab, Specialist in COVID-19) from Maya Nursing Home,Tamil Nadu for annotating the X-ray and CT images.

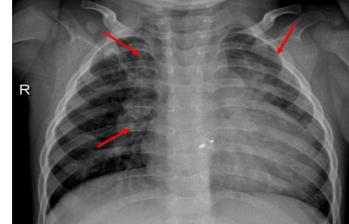
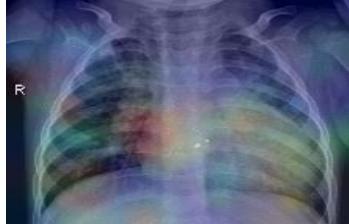
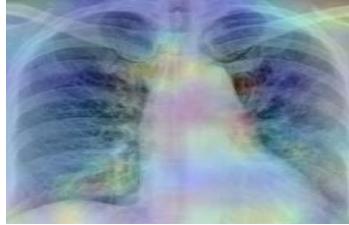
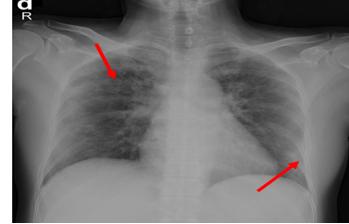
Original	Radiologist Finding	Heatmap
		
	Pneumonia	Pneumonia
		
	Pneumonia	Pneumonia
		
	Pneumonia	Pneumonia
		
	COVID-19	COVID-19
		
	COVID-19	COVID-19

TABLE 4.16: Comparison with Radiologist Finding

4.2 CT Scans

In this section we conduct a comprehensive evaluation and comparison of the performance across each of our CT models.

4.2.1 Performance Evaluation

The following sections evaluates the performance of each of our base models, UNet, Attention UNet, and UNet++. We have supplemented these sections with tables, plots indicating the trends in loss and accuracy, and confusion matrices respectively.

4.2.1.1 Pre-trained Models

The first model we have experimented with is UNet. We have achieved an average accuracy of 94.9% after performing 10-fold cross validation on the balanced dataset. The average and best accuracy obtained on unseen data are 94.6% and 95.7% respectively. The classification report, confusion matrix, and trends in accuracy and loss have been displayed in Table 4.19, Figure 4.9, and Figure 4.10 respectively.

	Precision	Recall	F1-Score
COVID-19	95.2%	94.7%	94.9%
Normal	94.7%	95.2%	94.9%

Accuracy **94.9%**

TABLE 4.19: UNet Model Classification Report

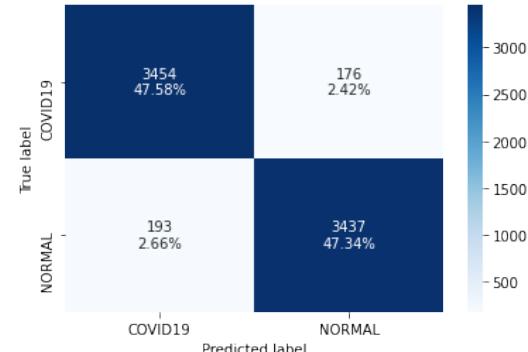


FIGURE 4.9: UNet Model Confusion Matrix

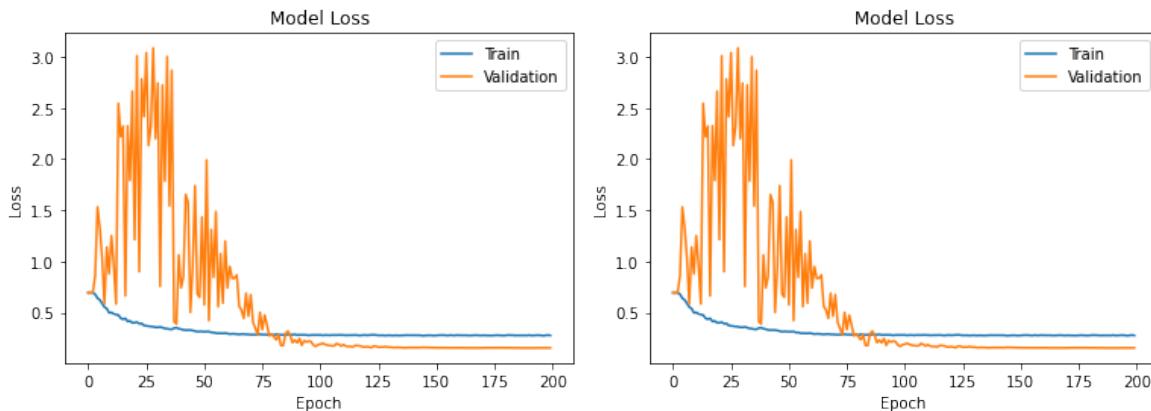


FIGURE 4.10: UNet Model Training Performance Trends

Next, we trained our Attention UNet model. We achieved an average accuracy of 98.1% after 10-fold cross validation on the balanced dataset. The average and best accuracy obtained on the unseen data are 98.3% and 99.2%. We have displayed the classification report, confusion matrix, and trends in accuracy and loss in Table 4.22, Figure 4.11, and Figure 4.12 respectively.

	Precision	Recall	F1-Score
COVID-19	98.3%	97.9%	98.1%
Normal	97.9%	98.3%	98.1%

Accuracy 98.1%

TABLE 4.22: Attention UNet Model Classification Report

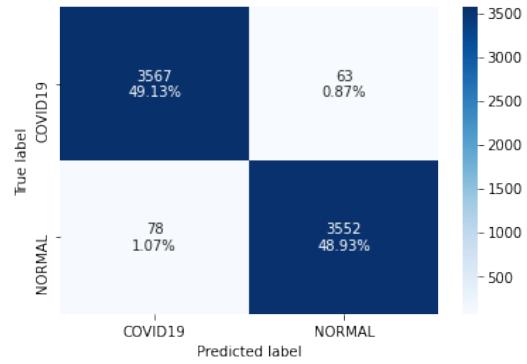


FIGURE 4.11: Attention UNet Model Confusion Matrix

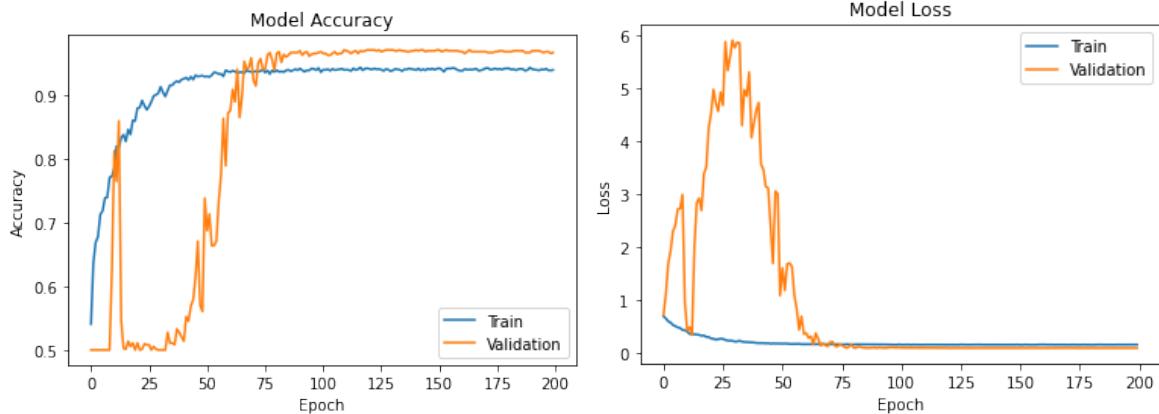


FIGURE 4.12: Attention UNet Model Training Performance Trends

The last model we have evaluated is UNet++. We have achieved an average accuracy of 90.2% after 10-fold cross validation on the balanced dataset. The average and best accuracy obtained on the unseen data are 88.0% and 91.2% respectively. We have displayed the classification report, confusion matrix, and trends in loss and accuracy in Table 4.25, Figure 4.13, and Figure 4.14 respectively.

	Precision	Recall	F1-Score
COVID-19	92.9%	88.1%	90.5%
Normal	88.1%	92.9%	90.5%

Accuracy	90.2%
-----------------	-------

TABLE 4.25: UNet++ Model Classification Report

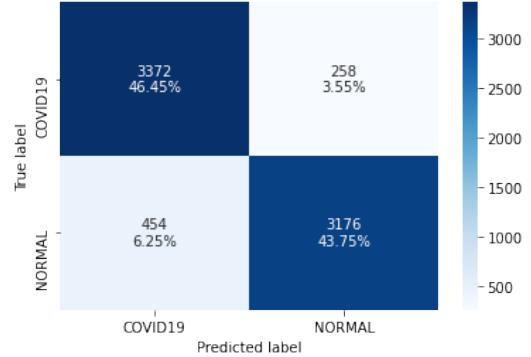


FIGURE 4.13: UNet++ Model Confusion Matrix

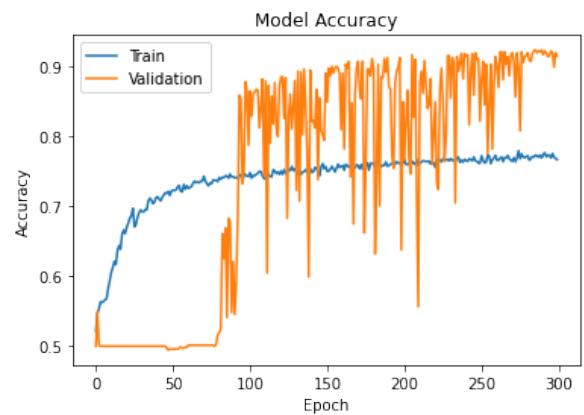
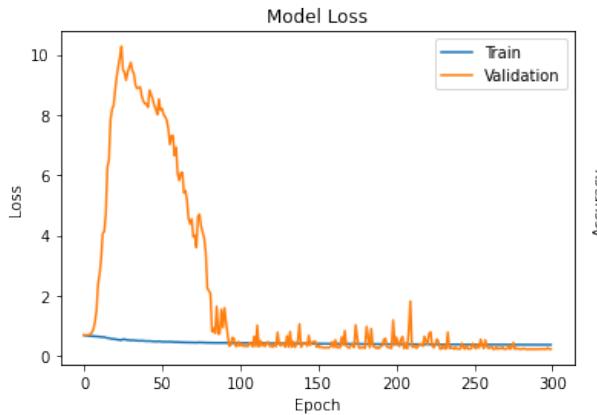


FIGURE 4.14: UNet++ Model Training Performance Trends

4.2.1.2 Ensemble Model

We have ensembled each of our base CT models to further improve classification performance.

We have achieved an average accuracy of 98.7% after cross validation of the balanced dataset.

The average and best accuracy obtained on unseen data are 97.9% and 97.8%. Table 4.28, Figure 4.15, and Figure 4.16 displays the classification report, confusion matrix, and trends in loss and accuracy respectively.

	Precision	Recall	F1-Score
COVID-19	98.8%	98.5%	98.7%
Normal	98.5%	98.8%	98.7%

Accuracy	98.7%
-----------------	-------

TABLE 4.28: CT Ensemble Model Classification Report

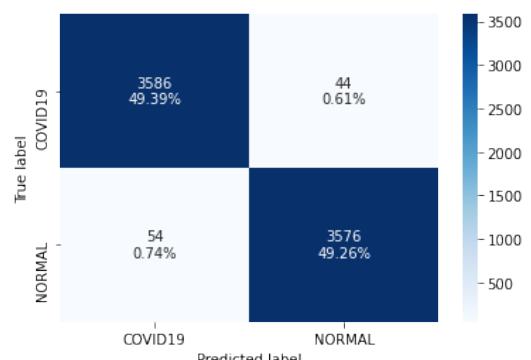


FIGURE 4.15: CT Ensemble Model Confusion Matrix

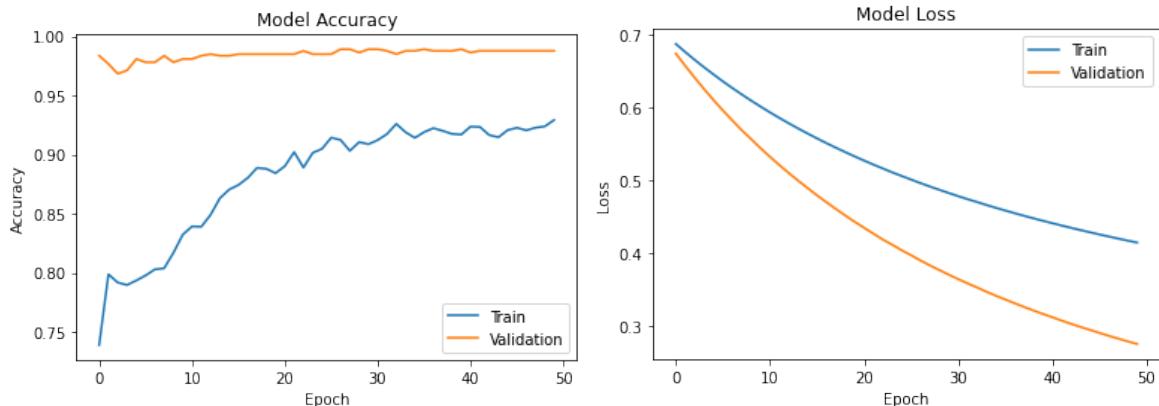


FIGURE 4.16: CT Ensemble Model Training Performance Trends

4.2.2 Performance Comparison

In the following sections, we compare our models amongst each other and determine the best model. We also compare the best model with the RT-PCR test and CNN, followed by similar studies from literature. Finally, we compare the heatmaps obtained with findings observed by a professional Radiologist.

4.2.2.1 Comparison of Trained Models

We compare each of our base CT models along with the ensemble model across various parameters such as Accuracy, Precision, Recall, F1-Score, Specificity, and Sensitivity respectively. The average metric scores are tabulated in Table 4.29.

	Accuracy	Precision	Recall	F1-Score	Specificity	Sensitivity
UNet	94.9%	94.9%	94.9%	94.9%	94.9%	94.9%
Attention UNet	98.1%	98.1%	98.1%	98.1%	98.1%	98.1%
UNet++	90.2%	90.5%	90.5%	90.5%	90.3%	90.5%
Ensemble Model	98.7%	98.7%	98.7%	98.7%	98.7%	98.7%

TABLE 4.29: CT Models Performance Comparison

We observe that out of the base models, Attention UNet outperforms the rest. As expected, a slight performance increase is also observed from the Ensemble model.

4.2.2.2 Comparison with RT-PCR and CNN

We compare the results obtained by our Ensemble model with traditional RT-PCR test and a basic CNN model, only for **COVID-19 class**. Once again, we use the study conducted by the Infectious Diseases Society of America (IDSA) [84] for comparison purposes. Table 4.30 compares the results.

	Accuracy	Sensitivity	Specificity
RT-PCR	94.0%	84.2%	98.9%
CNN	82.3%	80.2%	84.6%
Our Approach	98.7%	98.5%	98.8%

TABLE 4.30: Comparison with RT-PCR and CNN for X-ray scans

We observe that our Ensemble model outperforms CNN across all metrics. We also observe that our approach performs better than traditional RT-PCR test across most metrics.

4.2.2.3 Comparison with Related Work

We compare the performance of our Ensemble model to related work discussed in Section 2 across various metrics. The results are summarized in Table 4.31.

	Accuracy	Precision	Recall	F1-Score	Specificity	Sensitivity
Wang et al. [24]	79.3%	55.0%	83.0%	66.2%	67.0%	83.0%
Jin et al. [23]	80.4%	95.0%	73.5%	82.9%	92.9%	73.5%
Song et al. [25]	86.0%	81.0%	93.0%	86.6%	76.7%	93.0%
Zheng et al. [9]	90.1%	84.0%	90.7%	87.2%	91.1%	90.7%
Chen et al. [15]	91.6%	84.6%	100%	91.6%	93.6%	100%
Li et al. [14]	94.0%	89.7%	89.7%	89.7%	95.7%	89.7%
Ensemble Model	98.7%	98.7%	98.7%	98.7%	98.7%	98.7%

TABLE 4.31: Comparison with Related Work

Our Ensemble model performs significantly better when compared other approaches across all metrics. Once again, the data pre-processing and processing workflow such as splitting the lung parenchyma, augmentation, transfer learning, and ensemble learning respectively, seems to play a major factor to the models optimal performance.

4.2.2.4 Comparison with Radiologist Findings

We once again compare the generated heatmaps with that of the findings observed by a professional Radiologist. From Table 4.32, we can observe that the heatmaps generated are in direct conformance with the senior Radiologist's finding². The heatmaps identify at least one or more of the critical regions highlighted by the Radiologist. We also observe that all areas highlighted by the Radiologist are shades of red, purple, or green and not blue.

Furthermore, similar to our observations from Section 4.1.2.4, we have verified that our heatmaps correlate to the studies conducted by Harmon et al., Li et al., and Chung et al., therefore ensuring the reliability of our results. We observe that consolidations, GGO's, and lesions are major contributing factors to COVID-19 detection.

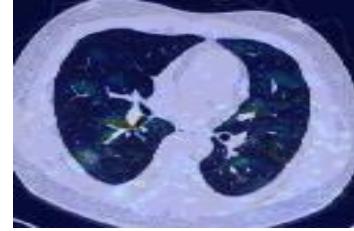
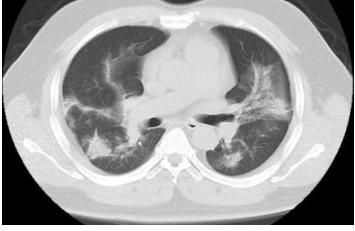
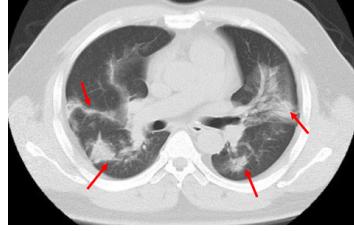
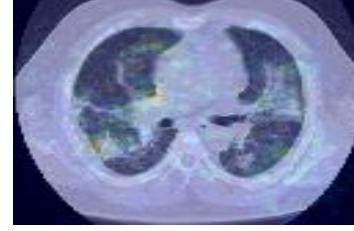
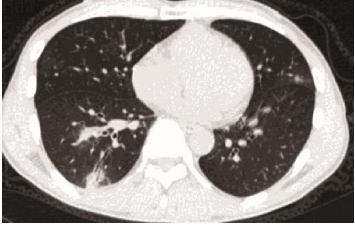
Original	Radiologist Finding	Heatmap
	 COVID-19	 COVID-19
	 COVID-19	 COVID-19
	 COVID-19	 COVID-19

TABLE 4.32: Comparison with Radiologist Finding

²The authors thank Dr. G.R. Mahadevan (MD, IDCCM,D.Diab, Specialist in COVID-19) from Maya Nursing Home,Tamil Nadu for annotating the X-ray and CT images.

4.3 Web Interface

In this section, we highlight the results obtained by the Usability Study we have conducted. It is of utmost importance to understand and measure the usability of the COVID-19 and Pneumonia Diagnosis Platform. The feedback received would be really beneficial for future iterations of this application in terms of design improvements and feature additions.

4.3.1 Hypothesis

Through this experiment, Alister George Luiz (the investigator) is attempting to prove that the COVID-19 and Pneumonia Diagnosis Platform is usable in the sense that it is user-friendly, consistent, familiar, responsive, and intuitive.

4.3.2 Target Audience

Any user, with or without technical experience, in any age group was eligible for this study. A consent form was provided virtually to comply with the compulsory policy of no face-to-face interactions due to the outbreak of the pandemic. All the terms and conditions were presented before the participant undertook the survey and was given freedom to accept or decline participation. The consent form was taken from the resources provided by the F20PA course page.

4.3.3 Experiment Design

The main motive behind designing a Usability Study was due to the fact that the COVID-19 and Pneumonia Diagnosis Platform is a user-facing product. The purpose of this user study was to **quantitatively** and **qualitatively** understand the perspective of the user towards our final product. Our experiment is composed of three activities. The investigator ensured that the participants viewed the video demonstration of the Diagnosis Portal and also provided Screenshots covering all aspects of the website. We have utilized Microsoft Forms³ to conduct this study.

4.3.4 Activity 1: Tasks

In the first activity, we have asked the participants several questions after they have viewed the video demonstration. The goal of this activity was to verify that the basic system functionality

³The Microsoft Forms Link containing the Survey can be found here: <https://forms.office.com/r/vFuJ0VCxe5>

is easily understandable and that they could perform these tasks without intervention from the investigator or any similar secondary help.

The outcome of each question was either a yes or no to evaluate the participants understandability. Besides this, the investigator also welcomed feedback on the existing systems, which may be considered in future versions of this Web Portal. It is worth to note that participants were given a brief explanation of the project in the beginning of the study, and any questions were welcome via email or other messaging sources.

We have asked this question to participants before the activity - **How frequently do you use a Computer?**

All users provided 'Daily' as their response. The main purpose of this question was to understand the influence of familiarity with computers in their respective ease of understanding of the system. We found that participants with daily usage of computers are usually more technically adept and can easily comprehend the system workflow. Table 4.33 displays the questions and results obtained for Activity 1.

Q.No	Question	Yes	No
1	Would you be able to utilize the functionality of this website independently?	15	0
2	Were you able to comprehend the Disclaimer message easily?	15	0
3	Were you able to identify where to upload the scan easily?	15	0
4	Were you able to understand that the slideshow showcased previous diagnosis results obtained from the portal for both COVID-19 and Pneumonia?	13	2
5	Were you able to identify the use of the toggle button between X-ray and CT scan?	15	0

TABLE 4.33: Activity 1 Results

The results obtained indicates that most of the participants easily understood the basic functionality and workflow of this Diagnosis Portal. Three participants could not recognize that the Gallery section showcased previous diagnosis results obtained from the portal. We believe adding a more explicit description would help in better understanding of the Gallery section.

Besides these questions, the participant was also asked to provide feedback if they felt so regarding the interface or any general suggestions. Table 4.34 highlights some of the top suggestions provided, which would be definitely considered in future versions of this Diagnosis Portal.

S.No	Suggestions for Improvement
1	Improvements to UI, Multi-page Site as in separate page for COVID-19 and Pneumonia.
2	Just keep adjusting the layout of the website so rather than scrolling all site content are seen together. Also since it's a diagnosis web portal, for future works integrating different medical issues related to COVID-19 can be trained in the AI, for more efficient outcomes.
3	If the result of the diagnosis indicates COVID - 19, guidelines and measures (according to region) being presented along with it would be helpful.
4	The options for switching between X-ray and CT could be renamed as "Click here to view X-ray" and "Click here to view CT Scan".

TABLE 4.34: Feedback on the Diagnosis Portal

It is indeed possible to extract relevant functionality improvements from the general feedback. Revision in User Interface, adding more information related to COVID-19, providing guidelines and measures pertaining to the region and so on are high in our priority list. We thank the participants for providing such valuable feedback.

4.3.5 Activity 2: System Usability Scale (SUS) Survey

Next, we have utilized the System Usability Scale which is a renowned and widely used tool to measure the usability of the system [85], [86]. We successfully completed the usability testing with fifteen participants.

The System Usability Scale (SUS) comprised of ten questions. Each participant had to provide a rating between 1 to 5 which indeed correlated to **Strongly Disagree** and **Strongly Agree**. One interesting aspect of using the SUS was that we could easily average out the results from each participant and obtain the general consensus. The standard SUS formula was used to compute the final scores [87] for each participant. The questions are displayed in Table 4.35 and the results obtained from each participant along with the scores is provided in Table 4.36.

Q.No	Question
1	I think that I would like to use this system frequently. Given that feedback on the results obtained by the portal must be validated by a medical professional.
2	I found this system unnecessarily complex.
3	I thought this system was easy to use.
4	I think that I would need assistance to be able to use the system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found this system very cumbersome/awkward to use.
9	I feel very confident using this system.
10	I needed to learn a lot of things before I could get going with this system.

TABLE 4.35: System Usability Scale (SUS) Questions

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score
P1	4	1	4	1	5	1	5	1	4	1	92.5
P2	5	1	5	2	4	1	4	1	5	1	92.5
P3	5	1	5	1	5	1	5	1	5	1	100
P4	5	1	5	1	5	1	4	2	5	1	95
P5	5	1	5	1	3	1	5	1	5	1	95
P6	5	2	5	1	5	1	5	2	5	1	95
P7	3	1	4	1	3	1	3	1	3	1	77.5
P8	5	1	5	1	5	1	5	1	5	1	100
P9	4	2	4	1	3	1	4	1	4	1	95
P10	5	1	5	1	4	2	5	1	5	1	95
P11	5	1	5	2	5	1	4	1	4	1	92.5
P12	5	1	5	1	5	2	5	1	5	1	97.5
P13	2	1	4	2	3	1	4	1	4	1	77.5
P14	4	2	5	1	4	2	4	1	3	2	80
P15	3	2	4	2	4	2	4	2	3	2	70

TABLE 4.36: System Usability Scale (SUS) Scores

The average SUS score is **90.3**. As per Jeff Sauro, a SUS score above 68 is **Above Average**. A score above 80.3 is **Excellent**. As per our this criteria, we are pleased to know that our Diagnosis Portal falls into the latter category. Figure 4.17 displays our SUS scores and the average score. We also believe that the Mode statistical quantity is in fact really useful in understanding the participants general mindset. Therefore, we have highlighted the Mode, Maximum and Minimum per question in Table 4.37.

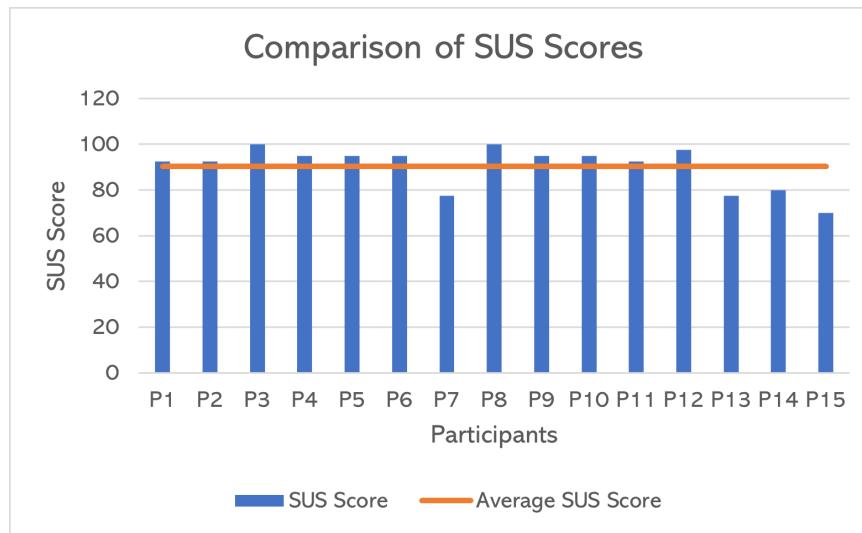


FIGURE 4.17: Comparison of SUS Scores

Statistic	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Mode	5	1	5	1	5	1	5	1	5	1
Maximum	5	2	5	2	5	2	5	2	5	2
Minimum	2	1	4	1	3	1	3	1	3	1

TABLE 4.37: System Usability Scale (SUS) Statistical Analysis

4.3.6 Activity 3: Metrics Test

The last activity involves explicitly testing the certain metrics stated in Section 4.3.1 and Table A.3. We have asked the participants to fill in a Likert Scale with each metric having scores ranging from Very Poor to Excellent. Table 4.38 summarizes the results.

Metric	Very Poor	Poor	Fair	Good	Excellent
User-Friendliness	0	0	1	7	7
Consistency	0	0	1	8	6

Familiarity	0	0	5	7	3
Responsiveness	0	0	0	7	8
Intuitiveness	0	0	1	11	3
Overall User Experience	0	0	0	8	7

TABLE 4.38: Activity 3 Likert Scale Results

We are pleased to note that all of the participants ranked each metric above 'Fair'. We also note that system familiarity as higher 'Fair' number, therefore, we plan to improve this metric by scouting the user interface of web applications with similar functionality. We believe having a user interface similar to popular pre-existing web applications would improve the familiarity of the Diagnosis Portal with its users.

4.4 Requirements Validation

In this section, we confirm that all the requirements stated at the inception of this project is satisfied. Table 4.39 tabulates our requirements and their status of completion.

Functional Requirements				
ID	Description	Priority	Status	
FR-1	X-ray Segmentation The system shall be able to accept and segment X-ray scans.	M	Completed	
FR-2	CT Segmentation The system shall be able to accept and segment CT scans.	S	Completed	
FR-3	COVID-19 Diagnosis using X-ray Scans The system shall be able to classify positive COVID-19 patients from others given test X-ray scans	M	Completed	
FR-4	COVID-19 Diagnosis using CT Scans The system shall be able to classify positive COVID-19 patients from others given test CT scans.	S	Completed	

FR-5	Visualize Lung Region of Interest's The system shall be able to interpret and visualize classification results by highlighting lung ROIs.	M	Completed
FR-6	Multi-class Diagnosis The system shall be able to differentiate COVID-19 and Pneumonia (Viral or Bacterial) patients.	W	Completed for X-rays
FR-7	Web Interface The system shall have an interface which presents diagnosis results after segmentation for visualization and analysis purposes.	S	Completed

TABLE 4.39: Requirements Validation

All the requirements stated from FR-1 to FR-4 has been satisfied through our implementation workflow. The segmentation is carried out during data preparation, especially in the case of CT scans where we extract the lung parenchyma and by our state-of-the-art deep learning models during model training. Indeed, the diagnosis is carried out when our models return their prediction after processing the input scan. Therefore, satisfying these requirements.

Through our Grad-CAM visualization, we are able to interpret and visualize classification results by viewing the highlighted ROI's, therefore satisfying FR-5. For multi-class diagnosis, that is, FR-6, in addition to COVID-19 diagnosis, we have also been able to successfully diagnose patients with Pneumonia using X-ray scans. This was indeed possible due to the availability of an open-source dataset consisting of both COVID-19 and Pneumonia X-ray scans. Unfortunately, as we were not able to find an open source dataset consisting of CT scans from Pneumonia patients, we have performed binary classification utilizing COVID-19 and Normal samples.

The last requirement, FR-7, was satisfied through our Flutter web application, which allows users to input their chest X-ray or CT scans and receive real-time diagnosis along with the critical regions highlighted via the heatmaps generated.

Chapter 5

Conclusion and Future Work

This chapter concludes this manuscript and comments on the limitations of this project, future works we wish to undertake and finally my thoughts and reflections on the experiences throughout this year-long project.

5.1 Limitations

Despite our best efforts there are indeed some limitations to this project implementation which could be overcome in future iterations. The first being the availability of a reasonably sized dataset. We have noticed that there have only been a handful of researchers and institutions actively collecting chest X-rays and CT scans with most of the data not being open-source. Therefore, most of the datasets on Kaggle and other forums usually cite the same root source. Although we have performed Data Augmentation to generate more samples, we believe having more COVID-19 samples would indeed increase the reliability of our model.

The free-to-use tier of Google Colab was another major limitation in training our models. Due to the limited RAM (12GB) and runtime duration (12 hours per day), we had to incorporate memory optimization strategies such as reducing the batch size and image dimensions in order to be able to fit the data into memory and conduct training. This was especially for the case of CT scans. Furthermore, we also had a ceiling for the number of augmented images generated due to the same memory limitation.

As we have utilized the free credit offered by Microsoft Azure for students, the Ubuntu Virtual Machine used to host our Python Scripts and Flutter Web Application on their platform would only run for around 2.5 months, after which a monthly fee of \$30 apply for continued use.

5.2 Future Work

We have a lot of future plans for this project. Provided we gain access to a powerful machine well beyond the capabilities of a standard Google Colab environment, we believe we can further increase the reliability and performance of our models. Here are a list of proposed ideas:

1. **Data Sources** - Identify more data sources to increase the size of our existing dataset.
2. **Image Dimensions** - Given a machine with higher RAM, perform model training with larger image dimensions.
3. **Data Augmentation** - Utilize Data Augmentation to generate more samples per class.
4. **Multi-class CT Classification** - Obtain CT scans from Pneumonia patients and build multi-class classification model.
5. **Research Papers** - Publish research papers on our X-ray and CT scan implementation, along with a survey paper.
6. **Radiologist Validation** - Validate more model results and heatmaps with a Radiologist and receive their feedback on the web application and thereby make suitable changes.

5.3 Reflections

Given that the results obtained are validated by medical professionals, our Diagnosis Portal would provide them a safer working environment by minimizing the rate of virus exposure. As our system enables highly accurate rapid diagnoses of COVID-19, medical professionals would be able to save valuable time and effectuate quarantine and treatment protocols more efficiently.

Through this year-long project, I was able to gain valuable experience and improve both technical and soft skills through various phases of this project. More importantly, being my first ever major Data Science project, I was able to learn from my mistakes and overcome all the roadblocks faced with the guidance and support of my project supervisor, Dr. Hani Ragab, who was always available to clear my doubts, provide fruitful insights and resources.

I truly believe that my commitment and hard work coupled with the guidance and encouragement provided by my project supervisor, professors, and parents have led to this project achieve its potential and successful completion.

Bibliography

- [1] J. Elflein. (2021). "Covid-19 tests by country," [Online]. Available: <https://www.statista.com/statistics/1028731/covid19-tests-select-countries-worldwide/>.
- [2] World Health Organization. (2020). "Coronavirus," [Online]. Available: <https://www.who.int/health-topics/coronavirus> (visited on 09/20/2020).
- [3] CDC. (2020). "About covid-19," [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/cdcresponse/about-COVID-19.html>.
- [4] CDC. (2020). "Social distancing," [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html>.
- [5] J. Laguarta, F. Hueto, and B. Subirana, "Covid-19 artificial intelligence diagnosis using only cough recordings," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 275–281, 2020.
- [6] N. Jawerth. (2020). "How is the covid-19 virus detected using real time rt-pcr?" [Online]. Available: <https://www.iaea.org/newscenter/news/how-is-the-covid-19-virus-detected-using-real-time-rt-pcr> (visited on 10/02/2020).
- [7] K. LM, L. SA, L. O, B. D, and L. J, "Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based sars-cov-2 tests by time since exposure," *Annals of Internal Medicine*, vol. 173, no. 4, pp. 262–267, 2020, PMID: 32422057. DOI: [10.7326/M20-1495](https://doi.org/10.7326/M20-1495). eprint: <https://doi.org/10.7326/M20-1495>. [Online]. Available: <https://doi.org/10.7326/M20-1495>.
- [8] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, J. Cui, W. Xu, Y. Yang, Z. A. Fayad, and et al., "Ct imaging features of 2019 novel coronavirus (2019-ncov)," *Radiology*, vol. 295, no. 1, 202–207, 2020, ISSN: 0033-8419, 1527-1315. DOI: [10.1148/radiol.2020200230](https://doi.org/10.1148/radiol.2020200230).
- [9] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for covid-19 from chest ct using weak label," *medRxiv*, 2020.

- [10] Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu, and H. Shi, "Longitudinal assessment of covid-19 using a deep learning-based quantitative ct pipeline: Illustration of two cases," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, e200082, 2020, ISSN: 2638-6135. DOI: [10.1148/ryct.2020200082](https://doi.org/10.1148/ryct.2020200082).
- [11] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, and L. Xia, "Serial quantitative chest ct assessment of covid-19: Deep-learning approach," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, e200075, 2020, ISSN: 2638-6135. DOI: [10.1148/ryct.2020200075](https://doi.org/10.1148/ryct.2020200075).
- [12] H. Yue, Q. Yu, C. Liu, Y. Huang, Z. Jiang, C. Shao, H. Zhang, B. Ma, Y. Wang, G. Xie, and et al., "Machine learning-based ct radiomics method for predicting hospital stay in patients with pneumonia associated with sars-cov-2 infection: A multicenter study," *Annals of Translational Medicine*, vol. 8, no. 14, 859–859, 2020, ISSN: 23055839, 23055847. DOI: [10.21037/atm-20-3026](https://doi.org/10.21037/atm-20-3026).
- [13] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. Siegel, "Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," *arXiv preprint arXiv:2003.05037*, 2020.
- [14] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, and et al., "Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct," *Radiology*, 2020, ISSN: 0033-8419. DOI: [10.1148/radiol.2020200905](https://doi.org/10.1148/radiol.2020200905). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233473/>.
- [15] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng, and et al., *Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study*. 2020. DOI: [10.1101/2020.02.25.20021568](https://doi.org/10.1101/2020.02.25.20021568). [Online]. Available: <http://medrxiv.org/lookup/doi/10.1101/2020.02.25.20021568>.
- [16] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng, and et al., *AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks*. 2020. DOI: [10.1101/2020.03.19.20039354](https://doi.org/10.1101/2020.03.19.20039354). [Online]. Available: <http://medrxiv.org/lookup/doi/10.1101/2020.03.19.20039354>.
- [17] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, and Y. Shi, "Lung infection quantification of covid-19 in ct images with deep learning," *arXiv preprint arXiv:2003.04655*, 2020.

- [18] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *arXiv:1505.04597 [cs]*, 2015, arXiv: 1505.04597. [Online]. Available: <http://arxiv.org/abs/1505.04597>.
- [19] M. H. Soomro, M. Coppotelli, S. Conforto, M. Schmid, G. Giunta, L. Del Secco, E. Neri, D. Caruso, M. Rengo, and A. Laghi, “Automated segmentation of colorectal tumor in 3d mri using 3d multiscale densely connected convolutional neural network,” *Journal of Healthcare Engineering*, vol. 2019, 2019, ISSN: 2040-2295. DOI: [10.1155/2019/1075434](https://doi.org/10.1155/2019/1075434). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6374810/>.
- [20] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” *arXiv:1606.04797 [cs]*, 2016, arXiv: 1606.04797. [Online]. Available: <http://arxiv.org/abs/1606.04797>.
- [21] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. Tavares, A. Bradley, J. P. Papa, V. Belegiannis, and et al., Eds., Springer International Publishing, 2018, 3–11, ISBN: 978-3-030-00889-5.
- [22] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, and et al., “Attention u-net: Learning where to look for the pancreas,” *arXiv:1804.03999 [cs]*, 2018, arXiv: 1804.03999. [Online]. Available: <http://arxiv.org/abs/1804.03999>.
- [23] C. Jin, W. Chen, Y. Cao, Z. Xu, X. Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi, and J. Feng, “Development and evaluation of an ai system for covid-19 diagnosis,” *medRxiv*, 2020.
- [24] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, et al., “A deep learning algorithm using ct images to screen for corona virus disease (covid-19),” *MedRxiv*, 2020.
- [25] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, Y. Jie, R. Wang, et al., “Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images,” *medRxiv*, 2020.
- [26] L. Tang, X. Zhang, Y. Wang, and X. Zeng, “Severe covid-19 pneumonia: Assessing inflammation burden with volume-rendered chest ct,” *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, e200044, 2020, ISSN: 2638-6135. DOI: [10.1148/ryct.2020200044](https://doi.org/10.1148/ryct.2020200044).

- [27] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, “Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19,” *IEEE Reviews in Biomedical Engineering*, 1–1, 2020, ISSN: 1941-1189. DOI: [10.1109/RBME.2020.2987975](https://doi.org/10.1109/RBME.2020.2987975).
- [28] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, “Covid-19 screening on chest x-ray images using deep learning based anomaly detection,” Mar. 2020.
- [29] A. Narin, C. Kaya, and Z. Pamuk, “Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks,” Mar. 2020.
- [30] B. Ghoshal and A. Tucker, “Estimating uncertainty and interpretability in deep learning for coronavirus (covid-19) detection,” *arXiv:2003.10769 [cs, eess, stat]*, 2020, arXiv: 2003.10769. [Online]. Available: <http://arxiv.org/abs/2003.10769>.
- [31] L. Wang and A. Wong, “Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images,” *arXiv:2003.09871 [cs, eess]*, 2020, arXiv: 2003.09871. [Online]. Available: <http://arxiv.org/abs/2003.09871>.
- [32] A. Mahmood, A. G. Ospina, M. Bennamoun, S. An, F. Sohel, F. Boussaid, R. Hovey, R. B. Fisher, and G. A. Kendrick, “Automatic hierarchical classification of kelps using deep residual features,” *Sensors*, vol. 20, no. 2, p. 447, 2020, ISSN: 1424-8220. DOI: [10.3390/s20020447](https://doi.org/10.3390/s20020447).
- [33] H. Y. F. Wong, H. Y. S. Lam, A. H.-T. Fong, S. T. Leung, T. W.-Y. Chin, C. S. Y. Lo, M. M.-S. Lui, J. C. Y. Lee, K. W.-H. Chiu, T. Chung, *et al.*, “Frequency and distribution of chest radiographic findings in covid-19 positive patients,” *Radiology*, p. 201 160, 2020.
- [34] J. P. Cohen, P. Morrison, and L. Dao, “Covid-19 image data collection,” *arXiv:2003.11597 [cs, eess, q-bio]*, 2020, arXiv: 2003.11597. [Online]. Available: <http://arxiv.org/abs/2003.11597>.
- [35] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, “Sanity checks for saliency maps,” in *Advances in Neural Information Processing Systems*, 2018, pp. 9505–9515.
- [36] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” *arXiv:1311.2901 [cs]*, 2013, arXiv: 1311.2901. [Online]. Available: <http://arxiv.org/abs/1311.2901>.
- [37] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization,” *arXiv:1512.04150 [cs]*, 2015, arXiv: 1512.04150. [Online]. Available: <http://arxiv.org/abs/1512.04150>.

- [38] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” pp. 618–626, 2017.
- [39] 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. 103 792, 2020, ISSN: 00104825. DOI: [10.1016/j.compbiomed.2020.103792](https://doi.org/10.1016/j.compbiomed.2020.103792).
- [40] S. A. Harmon, T. H. Sanford, S. Xu, E. B. Turkbey, H. Roth, Z. Xu, D. Yang, A. Myronenko, V. Anderson, A. Amalou, and et al., “Artificial intelligence for the detection of covid-19 pneumonia on chest ct using multinational datasets,” *Nature Communications*, vol. 11, no. 1, p. 4080, 2020, ISSN: 2041-1723. DOI: [10.1038/s41467-020-17971-2](https://doi.org/10.1038/s41467-020-17971-2).
- [41] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., “Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai,” *Information Fusion*, vol. 58, pp. 82–115, 2020.
- [42] J.-M. Fellous, G. Sapiro, A. Rossi, H. S. Mayberg, and M. Ferrante, “Explainable artificial intelligence for neuroscience: Behavioral neurostimulation,” *Frontiers in Neuroscience*, vol. 13, p. 1346, 2019.
- [43] M. Noroozi, A. Vinjimoor, P. Favaro, and H. Pirsiaavash, “Boosting self-supervised learning via knowledge transfer,” pp. 9359–9367, 2018.
- [44] C. Doersch, A. Gupta, and A. A. Efros, “Unsupervised visual representation learning by context prediction,” pp. 1422–1430, 2015.
- [45] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A survey on deep transfer learning,” pp. 270–279, 2018.
- [46] B Ganglia. (2020). “Covid-19 image data collection,” [Online]. Available: <https://github.com/ieee8023/covid-chestxray-dataset>.
- [47] P. Mooney. (2018). “Chest x-ray images (pneumonia),” [Online]. Available: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>.
- [48] A. G. Chung. (2020). “Open source covid-19 dataset,” [Online]. Available: <https://github.com/agchung>.
- [49] K. Murphy, H. Smits, A. J. Knoops, M. B. Korst, T. Samson, E. T. Scholten, S. Schalekamp, C. M. Schaefer-Prokop, R. H. Philipsen, A. Meijers, et al., “Covid-19 on chest radiographs:

- A multireader evaluation of an artificial intelligence system," *Radiology*, vol. 296, no. 3, E166–E172, 2020.
- [50] Keras. (2021). "Keras documentation: Image data pre-processing," [Online]. Available: <https://keras.io/api/preprocessing/image/>.
- [51] ImageNet. (2021). "Imagenet," [Online]. Available: <http://www.image-net.org/>.
- [52] G. Huang, Z. Liu, K. Weinberger, and L. Van Der Maaten, *Densely Connected Convolutional Networks*. 2016. [Online]. Available: <https://arxiv.org/pdf/1608.06993v3.pdf>.
- [53] B. Er, "Microsoft presents : Deep residual networks - baki er - medium," *Medium*, 2016. [Online]. Available: <https://medium.com/@bakiiii/microsoft-presents-deep-residual-networks-d0ebd3fe5887>.
- [54] V. Feng, "An overview of resnet and its variants - towards data science," *Medium*, 2017. [Online]. Available: <https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035>.
- [55] N. Radwan. (2019). "Figure 2.6: A schematic illustration of the densenet-121 architecture [82].," [Online]. Available: https://www.researchgate.net/figure/A-schematic-illustration-of-the-DenseNet-121-architecture-82_fig5_334170752.
- [56] Neurohive. (2018), [Online]. Available: <https://neurohive.io/en/popular-networks/vgg16/>.
- [57] K. Simonyan and A. Zisserman, *Published as a conference paper at ICLR 2015 VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION*. 2015. [Online]. Available: <https://arxiv.org/pdf/1409.1556.pdf>.
- [58] J. Wei. (2019). "Vgg neural networks: The next step after alexnet - towards data science," [Online]. Available: <https://towardsdatascience.com/vgg-neural-networks-the-next-step-after-alexnet-3f91fa9ffe2c>.
- [59] Keras. (2021). "Keras documentation: Keras applications," [Online]. Available: <https://keras.io/api/applications/>.
- [60] Keras. (2021). "Keras documentation: Callbacks api," [Online]. Available: <https://keras.io/api/callbacks/>.
- [61] Keras. (2021). "Keras documentation: Earlystopping," [Online]. Available: https://keras.io/api/callbacks/early_stopping/.
- [62] Keras. (2021). "Keras documentation: Reducelronplateau," [Online]. Available: https://keras.io/api/callbacks/reduce_lr_on_plateau/.

- [63] Keras. (2021). “Keras documentation: Modelcheckpoint,” [Online]. Available: https://keras.io/api/callbacks/model_checkpoint/.
- [64] P. Bhavsar. (2019). “An ultimate guide to transfer learning in nlp,” [Online]. Available: <https://www.topbots.com/transfer-learning-in-nlp/>.
- [65] R. Polikar. (2009). “Ensemble learning,” [Online]. Available: http://www.scholarpedia.org/article/Ensemble_learning#:~:text=Ensemble\%20learning\%20is\%20the\%20process,\%2C\%20function\%20approximation\%2C\%20etc..
- [66] E. Lutins. (2017). “Ensemble methods in machine learning: What are they and why use them?” [Online]. Available: <https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f>.
- [67] Keras. (2021). “Keras documentation: Concatenate layer,” [Online]. Available: https://keras.io/api/layers/merging_layers/concatenate/.
- [68] SCRUM. (2018). “What is scrum?” [Online]. Available: <https://www.scrum.org/resources/what-is-scrum>.
- [69] Scikit-learn. (2020), [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html.
- [70] Keras, *Keras documentation: Computer vision*, 2021. [Online]. Available: <https://keras.io/examples/vision/>.
- [71] Keras. (2020). “Keras documentation: Grad-cam class activation visualization,” [Online]. Available: https://keras.io/examples/vision/grad_cam/.
- [72] Numpy. (2019), [Online]. Available: <https://numpy.org/>.
- [73] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, Z. Minqing, L. Xin, D. Xueyuan, C. Shucheng, and et al., “Covid-19 ct lung and infection segmentation dataset,” *Zenodo*, 2020. DOI: [10.5281/zenodo.3757476](https://doi.org/10.5281/zenodo.3757476). [Online]. Available: <https://zenodo.org/record/3757476>.
- [74] W. Ning, S. Lei, J. Yang, Y. Cao, P. Jiang, Q. Yang, J. Zhang, X. Wang, F. Chen, Z. Geng, *et al.*, “Ictcf: An integrative resource of chest computed tomography images and clinical features of patients with covid-19 pneumonia,” 2020.
- [75] H. Sankesara, “Unet - towards data science,” *Medium*, 2019. [Online]. Available: <https://towardsdatascience.com/u-net-b229b32b4a71>.

- [76] J. Zhang, *Unet — line by line explanation - towards data science*, 2019. [Online]. Available: <https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5>.
- [77] Z. Zongwei, R. Mahfuzur, T. Nima, and L. Jianming. (2018). “A nested u-net architecture for medical image segmentation,” [Online]. Available: <https://paperswithcode.com/paper/unet-a-nested-u-net-architecture-for-medical>.
- [78] J. H. Jing. (2019). “Biomedical image segmentation: Attention u-net - towards data science,” [Online]. Available: <https://towardsdatascience.com/biomedical-image-segmentation-attention-u-net-29b6f0827405>.
- [79] J. H. Jing, “Biomedical image segmentation: Unet++ - towards data science,” *Medium*, 2019. [Online]. Available: <https://towardsdatascience.com/biomedical-image-segmentation-unet-991d075a3a4b>.
- [80] R. Vinod. (2020). “A detailed explanation of the attention u-net - towards data science,” [Online]. Available: <https://towardsdatascience.com/a-detailed-explanation-of-the-attention-u-net-b371a5590831>.
- [81] PyPi. (2021), [Online]. Available: <https://pypi.org/project/keras-unet-collection/>.
- [82] A. Luiz. (2020). “The best all-in-one responsive and scalable code structure for flutter,” [Online]. Available: <https://medium.com/@calisterluiz/the-best-all-in-one-responsive-and-scalable-code-structure-for-flutter-717a384f9eda>.
- [83] M. I. of Technology Press (MIT Press, *How can i create a single confusion matrix after k fold cross validation ?* 2016. [Online]. Available: https://www.researchgate.net/post/How_can_I_create_a_single_confusion_matrix_after_K_fold_cross_validation.
- [84] IDSociety. (2021), [Online]. Available: <https://www.idsociety.org/covid-19-real-time-learning-network/diagnostics/RT-pcr-testing/#:~:text=The%20sensitivity%20and%20specificity%20of,CI%2095.3%25%E2%80%9399.7%25..>
- [85] A. S. for Public Affairs. (2021). “System usability scale (sus) | usability.gov,” [Online]. Available: <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>.
- [86] J. Brooke, “Sus-a quick and dirty usability scale. digital equipment co ltd,” *Reading, United Kingdom*, 1986.

- [87] N. Thomas. (2015). "How to use the system usability scale (sus) to evaluate the usability of your website - usability geek," [Online]. Available: <https://usabilitygeek.com/how-to-use-the-system-usability-scale-sus-to-evaluate-the-usability-of-your-website/>.
- [88] Wuhan Municipal Health Commission. (2020). "Wellness commission on the current pneumonia outbreak in our city," [Online]. Available: <https://web.archive.org/web/20200109215413/http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989> (visited on 09/28/2020).
- [89] World Health Organization, *Situation Report - 1*, ser. Novel Coronavirus (2019-nCoV). 2020. [Online]. Available: https://www.who.int/docs/default-source/coronavirus/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4.
- [90] E. Edwards. (2020). "Coronavirus u.s. case confirmation: 1st coronavirus case is in seattle," [Online]. Available: <https://www.nbcnews.com/health/health-news/1st-case-coronavirus-china-confirmed-u-s-n1119486> (visited on 09/28/2020).
- [91] E. Edwards. (2020). "New coronavirus outbreak: Who declares it a global public health emergency," [Online]. Available: <https://www.nbcnews.com/health/health-news/new-coronavirus-outbreak-declared-global-public-health-emergency-who-says-n1126836>.
- [92] R. Muccari, D. Chow, and J. Murphy. (2020). "Coronavirus timeline: Tracking the critical moments of covid-19," [Online]. Available: <https://www.nbcnews.com/health/health-news/coronavirus-timeline-tracking-critical-moments-covid-19-n1154341> (visited on 09/28/2020).
- [93] M. C. White and L. Bayly. (2020). "Dow closes down 1,000 points as coronavirus fears slam wall street," [Online]. Available: <https://www.nbcnews.com/business/markets/dow-plunges-950-points-fears-coronavirus-will-tank-global-economic-n1141546> (visited on 09/28/2020).
- [94] R. Shabad. (2020). "Senate passes \$8.3 billion emergency bill to combat coronavirus," [Online]. Available: <https://www.nbcnews.com/politics/congress/senate-passes-8-3-billion-emergency-bill-combat-coronavirus-n1150521> (visited on 10/02/2020).
- [95] Jason. (2020). "Who declares covid-19 disease to be a pandemic," [Online]. Available: <https://www.nbcnews.com/health/health-news/live-blog/coronavirus-updates->

- [live-u-s-cases-top-1-000-spread-n1155241/ncrd1155646#blogHeader](#) (visited on 10/02/2020).
- [96] L. Givetash. (2020). "U.s. surpasses china with coronavirus cases as global total tops 500,000," [Online]. Available: <https://www.nbcnews.com/news/world/u-s-surpasses-china-coronavirus-cases-global-total-tops-500-n1170136> (visited on 10/02/2020).
- [97] M. Phillips. (2020). "Coronavirus cases top 1 million worldwide, researchers say," [Online]. Available: <https://www.foxnews.com/health/coronavirus-cases-top-1-million-worldwide-johns-hopkins> (visited on 10/02/2020).
- [98] *Coronavirus (covid-19)*. [Online]. Available: <https://news.google.com/covid19/map?hl=en-US&gl=US&ceid=US:en>.
- [99] E. C. for Disease Prevention and Control. (2020). "The daily number of new reported cases of covid-19 by country worldwide," [Online]. Available: <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide> (visited on 10/02/2020).
- [100] *Coronavirus cases distribution*. [Online]. Available: <https://www.worldometers.info/coronavirus/>.
- [101] A. J. Rodriguez-Morales, J. A. Cardona-Ospina, E. Gutiérrez-Ocampo, R. Villamizar-Peña, Y. Holguin-Rivera, J. P. Escalera-Antezana, L. E. Alvarado-Arnez, D. K. Bonilla-Aldana, C. Franco-Paredes, A. F. Henao-Martinez, and et al., "Clinical, laboratory and imaging features of covid-19: A systematic review and meta-analysis," *Travel Medicine and Infectious Disease*, vol. 34, p. 101623, 2020, ISSN: 14778939. DOI: [10.1016/j.tmaid.2020.101623](https://doi.org/10.1016/j.tmaid.2020.101623).
- [102] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, and et al., "Clinical features of patients infected with 2019 novel coronavirus in wuhan, china," *The Lancet*, vol. 395, no. 10223, 497–506, 2020, ISSN: 01406736. DOI: [10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- [103] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, and et al., "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: A descriptive study," *The Lancet*, vol. 395, no. 10223, 507–513, 2020, ISSN: 01406736. DOI: [10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7).

- [104] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, and et al., “Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china,” *JAMA*, vol. 323, no. 11, 1061–1069, 2020, ISSN: 0098-7484. DOI: [10.1001/jama.2020.1585](https://doi.org/10.1001/jama.2020.1585).
- [105] K. Liu, Y.-Y. Fang, Y. Deng, W. Liu, M.-F. Wang, J.-P. Ma, W. Xiao, Y.-N. Wang, M.-H. Zhong, C.-H. Li, and et al., “Clinical characteristics of novel coronavirus cases in tertiary hospitals in hubei province:” *Chinese Medical Journal*, vol. 133, no. 9, 1025–1031, 2020, ISSN: 0366-6999. DOI: [10.1097/CM9.0000000000000744](https://doi.org/10.1097/CM9.0000000000000744).
- [106] D. Chang, M. Lin, L. Wei, L. Xie, G. Zhu, C. S. D. Cruz, and L. Sharma, “Epidemiologic and clinical characteristics of novel coronavirus infections involving 13 patients outside wuhan, china,” *Jama*, vol. 323, no. 11, p. 1092, 2020. DOI: [10.1001/jama.2020.1623](https://doi.org/10.1001/jama.2020.1623).
- [107] Y. Pan, H. Guan, S. Zhou, Y. Wang, Q. Li, T. Zhu, Q. Hu, and L. Xia, “Initial ct findings and temporal changes in patients with the novel coronavirus pneumonia (2019-ncov): A study of 63 patients in wuhan, china,” *European Radiology*, vol. 30, no. 6, 3306–3309, 2020, ISSN: 0938-7994, 1432-1084. DOI: [10.1007/s00330-020-06731-x](https://doi.org/10.1007/s00330-020-06731-x).
- [108] M. Zhang, X. Wang, Y. Chen, K. Zhao, Y. Cai, C. An, M. Lin, and X. Mu, “[clinical features of 2019 novel coronavirus pneumonia in the early stage from a fever clinic in beijing],” *Zhonghua jie he he hu xi za zhi = Zhonghua jiehe he huxi zazhi = Chinese journal of tuberculosis and respiratory diseases*, vol. 43, no. 3, 215—218, 2020, ISSN: 1001-0939. DOI: [10.3760/cma.j.issn.1001-0939.2020.03.015](https://doi.org/10.3760/cma.j.issn.1001-0939.2020.03.015).
- [109] L. United Imaging Healthcare Co. (2020). “United imaging’s emergency radiology departments support mobile cabin hospitals, facilitate 5g remote diagnosis,” [Online]. Available: <https://www.prnewswire.com/news-releases/united-imaging-s-emergency-radiology-departments-support-mobile-cabin-hospitals-facilitate-5g-remote-diagnosis-301010528.html>.
- [110] V. Singh, K. Ma, B. Tumeroy, Y.-J. Chang, A. Wimmer, T. O’Donnell, and T. Chen, “Darwin: Deformable patient avatar representation with deep image network,” *Lecture Notes in Computer Science*, vol. 10434, 497–504, 2017, ISSN: 0302-9743.
- [111] V. Singh, Y.-j. Chang, K. Ma, M. Wels, G. Soza, and T. Chen, “Estimating a patient surface model for optimizing the medical scanning workflow,” *Lecture Notes in Computer Science*, vol. 8673, 472–479, ISSN: 0302-9743.

- [112] Siemens. (2020). “Somatom edge plus,” [Online]. Available: <https://www.siemens-healthineers.com/en-us/computed-tomography/single-source-ct/somatom-edge-plus>.
- [113] J. Li, U. K. Udayasankar, T. L. Toth, J. Seamans, W. C. Small, and M. K. Kalra, “Automatic patient centering for mdct: Effect on radiation dose,” *American Journal of Roentgenology*, vol. 188, no. 2, 547–552, 2007, ISSN: 0361-803X, 1546-3141. DOI: [10.2214/AJR.06.0370](https://doi.org/10.2214/AJR.06.0370).
- [114] C. Martin, “Optimisation in general radiography,” *Biomedical Imaging and Intervention Journal*, vol. 3, no. 2, 2007, ISSN: 1823-5530. DOI: [10.2349/bijj.3.2.e18](https://doi.org/10.2349/bijj.3.2.e18). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3097657/>.
- [115] F. Achilles, A.-E. Ichim, H. Coskun, F. Tombari, S. Noachtar, and N. Navab, “Patient mocap: Human pose estimation under blanket occlusion for hospital monitoring applications,” Lecture Notes in Computer Science, vol. 9900, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds., 491–499, 2016. DOI: [10.1007/978-3-319-46720-7_57](https://doi.org/10.1007/978-3-319-46720-7_57). [Online]. Available: http://link.springer.com/10.1007/978-3-319-46720-7_57.
- [116] Siemens. (2020). “Künstliche Intelligenz: How artificial intelligence can improve ct scans,” [Online]. Available: <https://new.siemens.com/global/en/company/stories/research-technologies/artificial-intelligence/artificial-intelligence-imaging-techniques.html>.
- [117] Y. Wang, X. Lu, Y. Zhang, X. Zhang, K. Wang, J. Liu, X. Li, R. Hu, X. Meng, S. Dou, and et al., “Precise pulmonary scanning and reducing medical radiation exposure by developing a clinically applicable intelligent ct system: Toward improving patient care,” *EBioMedicine*, vol. 54, 102724–102724, 2020, ISSN: 2352-3964.
- [118] *The scrum framework poster*. [Online]. Available: <https://www.scrum.org/resources/scrum-framework-poster>.

Appendix A

Requirements Analysis and Evaluation Strategy

This chapter provides an insight into the project requirements followed by demonstrating the project implementation workflow, UI wireframe, data collection, and processing methodology, and concludes with an evaluation strategy.

A.1 Requirements Analysis

This section identifies and outlines the functional and non-functional requirements pertaining to this project. The priority of each of these requirements are ranked according to the MSCW prioritization technique. The following color-scheme indicates the respective ranking order:

- **Must Have (M)** - These requirements are fundamental to achieve the aim of this project.
- **Should Have (S)** - These requirements are important to the project but not vital and could be achieved in the long run.
- **Could Have (C)** - These requirements are not fundamental to achieve the aim of this project but would be an added benefit if accomplished.
- **Want To Have (W)** - These requirements would be prioritized in later releases of the project.

A.1.1 Functional and Non-Functional Requirements

The Functional Requirements (FR's) describes the essential components, purpose, and objectives of this project. Table A.1 tabulates the FR's along with a brief description for the same and its

priority according to MSCW prioritization technique. An evaluation for each of these functional requirements is specified in section [A.3](#).

Non-Functional Requirements (NFR's) emphasizes the system's operation which includes its performance, usability, portability, and so on. Table [A.2](#) tabulates each of these NFR's and applies the MSCW prioritization technique.

Functional Requirements		
ID	Description	Priority
FR-1	X-ray Segmentation The system shall be able to accept and segment X-ray scans.	M
FR-2	CT Segmentation The system shall be able to accept and segment CT scans.	S
FR-3	COVID-19 Diagnosis using X-ray Scans The system shall be able to classify positive COVID-19 patients from others given test X-ray scans	M
FR-4	COVID-19 Diagnosis using CT Scans The system shall be able to classify positive COVID-19 patients from others given test CT scans.	S
FR-5	Visualize Lung Region of Interest's The system shall be able to interpret and visualize classification results by highlighting lung ROIs.	M
FR-6	Multi-class Diagnosis The system shall be able to differentiate COVID-19 and Pneumonia (Viral or Bacterial) patients.	W
FR-7	Web Interface The system shall have an interface which presents diagnosis results after segmentation for visualization and analysis purposes.	S

TABLE A.1: Functional Requirements

Non-Functional Requirements		
ID	Description	Priority
NFR-1	Environment The system shall be deployed on a cloud platform such as IBM Cloud.	C
NFR-2	User Interface The system shall have an intuitive and user-friendly interface evaluated via a System Usability Study where the user can input scans and receive diagnosis results which could be confirmed by a medical professional.	S
NFR-3	Extensibility The system shall be flexible to extensions, this includes bug fixes, updated features and performance improvement.	S
NFR-4	Version Control All versions of the code shall be published on GitHub which enables version control and be open source.	M
NFR-5	Documentation The code shall include relevant comments and contain an instructions guide to setup the environment and run the code.	M
NFR-6	Modular Programming The program functionality shall be separated into independent components to emphasize the scalability of software.	S
NFR-7	Reusability The code developed shall be reusable by other researchers and developers.	C

TABLE A.2: Non-Functional Requirements

A.2 Implementation Workflow

The activity diagram displayed in Figure A.1 provides a blueprint of the workflow that would be followed for the development phase commencing next semester.

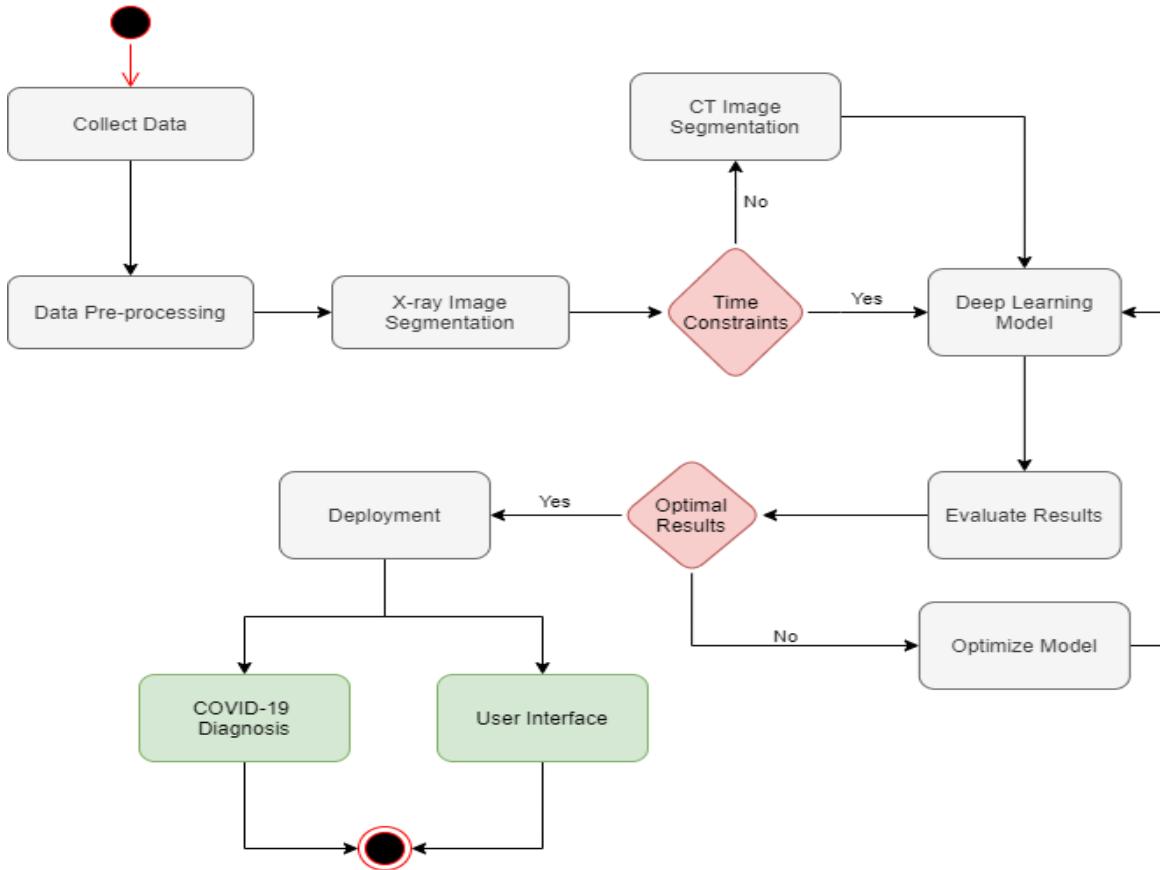


FIGURE A.1: Activity Diagram displaying the implementation workflow.

A.3 Evaluation Strategy

Table A.3 summarizes the methods and strategies used to evaluate each of the Functional Requirements described in Chapter A.

Evaluation Strategy	
ID	Description
FR-1	X-ray Segmentation The lung segments produced from X-ray scans shall be evaluated based on its capability to identify possible ROIs.
FR-2	CT Segmentation The lung segments produced from CT scans shall be evaluated based on its capability to identify possible ROIs.
FR-3	COVID-19 Diagnosis using X-ray Scans The diagnosis results obtained after achieving FR-1 shall be evaluated against various statistical metrics such as Accuracy, Precision and Recall.

FR-4	COVID-19 Diagnosis using CT Scans The diagnosis results obtained after achieving FR-2 shall be evaluated against various statistical metrics such as Accuracy, Precision and Recall.
FR-5	Visualize Lung Region of Interest's The ROIs visualized correlates to the observed lung characteristics in COVID-19 patients.
FR-6	Multi-class Diagnosis The diagnosis results obtained shall be evaluated against various statistical metrics such as Accuracy, Precision and Recall.
FR-7	Web Interface The user interface developed shall evaluated based on the following metrics, that is, user-friendliness, consistency, familiarity, responsiveness, and intuitiveness

TABLE A.3: Evaluation Strategy

A.4 Implementation Methodology

A brief summary of the methodology that would be carried out in semester 2 as well as their evaluation is described in this section.

A.4.1 Data Collection

The two primary sources for collecting open-source anonymized data would be the COVID-19 Kaggle datasets and the scans provided by Cohen et al. [34] as seen from the literature review. The images obtained shall be separated into X-ray and CT scans separately and furthermore, on the basis of their labels.

- **Testing** - Datasets which correspond to other lung diseases would also be used to analyze the generalizability of the model.
- **Evaluation** - Training would be conducted using 10-fold cross-validation, followed by rigorous verification to prevent the model from either overfitting or underfitting.

A.4.2 Data Pre-processing

Before carrying out image segmentation for both X-ray and CT scans, all images in the training dataset shall undergo the same data pre-processing pipeline as per the model's input parameters.

- **Testing** - The model would be tested on images with different dimensions to ensure no input errors are caused.
- **Evaluation** - An input verification technique shall be applied to validate the provided training images before the deep learning workflow commences.

A.4.3 Image Segmentation

The provided training images, both X-ray and CT scans shall undergo segmentation such that the ROIs would be highlighted and lead to effective COVID-19 diagnosis

- **Testing** - Multiple image segmentation techniques as suggested by the literature shall be used during the development phase to identify the one that yields the best results.
- **Evaluation** - All models used for experimentation purposes would be documented and the best would be utilized for final demonstration purposes.

A.4.4 Model Optimization

As observed in the literature review, multiple studies [9]–[17] indicate variants of the U-Net architecture to be the best suited for CT scan segmentation whereas, for X-rays, variants of ResNet and CNN seem to be ideal as per the research conducted [28]–[31].

- **Testing** - Each of the proposed model variations shall be experimented with for testing purposes, and the results obtained shall be compared to existing literature.
- **Evaluation** - The results obtained after tweaking the various hyper-parameters for each of the proposed models shall be tracked, thus being able to identify the most optimal set of values.

Appendix B

Project Management

This chapter highlights the possible ethical issues and risks that could affect the development phase of this project. Furthermore, the proposed development methodology and project plan are also included in this section.

B.1 Risk Management

To plan out a balanced and well-thought project development schedule, it is essential to take into account the inevitable risks that would occur during the phase and therefore, draw out efficient risk mitigation strategies to reduce the impact of the risk on the completion of this project. Risk Management comprises of four steps:

- **Risk Identification** - Recognizing various risks and labeling their type as follows.

Tools	People	Estimation	Requirements	Organization	Technology
-------	--------	------------	--------------	--------------	------------

- **Risk Analysis** - Analyzing the probability of risk occurrence and their impact if it does infact occur. The following color scheme maps the likelihood and impact of a risk.

Color	Likelihood	Impact
Red	High	Catastrophic
Yellow	Medium	Serious
Green	Low	Tolerable

- **Risk Planning** - Devising risk mitigating strategies and thus reduce its impact on the development phase. These strategies could belong to either **Avoidance**, **Minimization**, and **Contingency** categories respectively.
- **Risk Monitoring** - Regularly assessing each of the possible risks according to its probability of occurrence.

Table B.3 summarizes the identified risks and describes possible mitigating strategies.

Risk Management		
	Planning and Monitoring	
Risk, Type	Monitoring and Avoidance	Contingency and Minimization
Insufficient Data Type: Requirements	Monitoring: Review if the learning model underfits on the training data. Avoidance: Explore and store all possible open source COVID-19 data sources such as Kaggle.	Contingency: Use datasets for other lung diseases and experiment with Transfer Learning. Minimization: Ensure sufficient data from open source COVID-19 repositories are utilized.
Inadequate Computation Power Type: Technology	Monitoring: Track and ensure reasonable time taken for model training. Avoidance: Set up GPU powered deep learning libraries such as Tensorflow-GPU before development phase commences, and ensure sufficient storage on local machine.	Contingency: Utilize GPU powered cloud computing services such as IBM Cloud or MACS computers. Minimization: Ensure GPU powered deep learning libraries are installed and configured on local machine.

Difficulty in Implementation Type: Requirements	Monitoring: Rank the proposed models based on ease of implementation, model performance and documentation. Avoidance: Thorough literature review and research into the proposed models.	Contingency: Utilize alternative models which are well documented and are easily usable. Minimization: Allow sufficient time for project development phase after thorough research into existing studies.
Loss of Data Type: Tools	Monitoring: Perform regular checks on backup sources and processes. Avoidance: Set up version control and frequently commit changes to the project repository on GitHub.	Contingency: Restart project development phase from scratch. Minimization: Ensure that all phases in development are well documented, and utilize it in an unfortunate scenario of data loss.
Slow Project Progression Type: Estimation	Monitoring: Ensure that the devised project development schedule is followed precisely. Avoidance: Ensure all tools and datasets are set up before commencing development phase and start as early as possible.	Contingency: Aim to complete the tasks with the highest priority given limited development time. Minimization: Create a rank based priority list for each of the requirements as per project supervisor's recommendation.
Personal Health and Deadlines Type: People	Monitoring: Keep track of the deadlines announced and update the project plan accordingly. Avoidance: Achieve weekly targets and ensure that the project supervisor is aware of the announced deadlines.	Contingency: Complete tasks as per its importance after discussion with project supervisor. Minimization: Adhere to healthy living practices, follow social distancing protocols and keep flexible plans as per new deadlines.

TABLE B.3: Risk Analysis

B.2 Project Plan

The various tasks involved in this project and their proposed completion dates for next semester has been illustrated using a Gantt Chart which is displayed in Figure B.1. TeamGantt was used for designing the Gantt Chart, Microsoft Planner would be used to keep track of the weekly requirements.

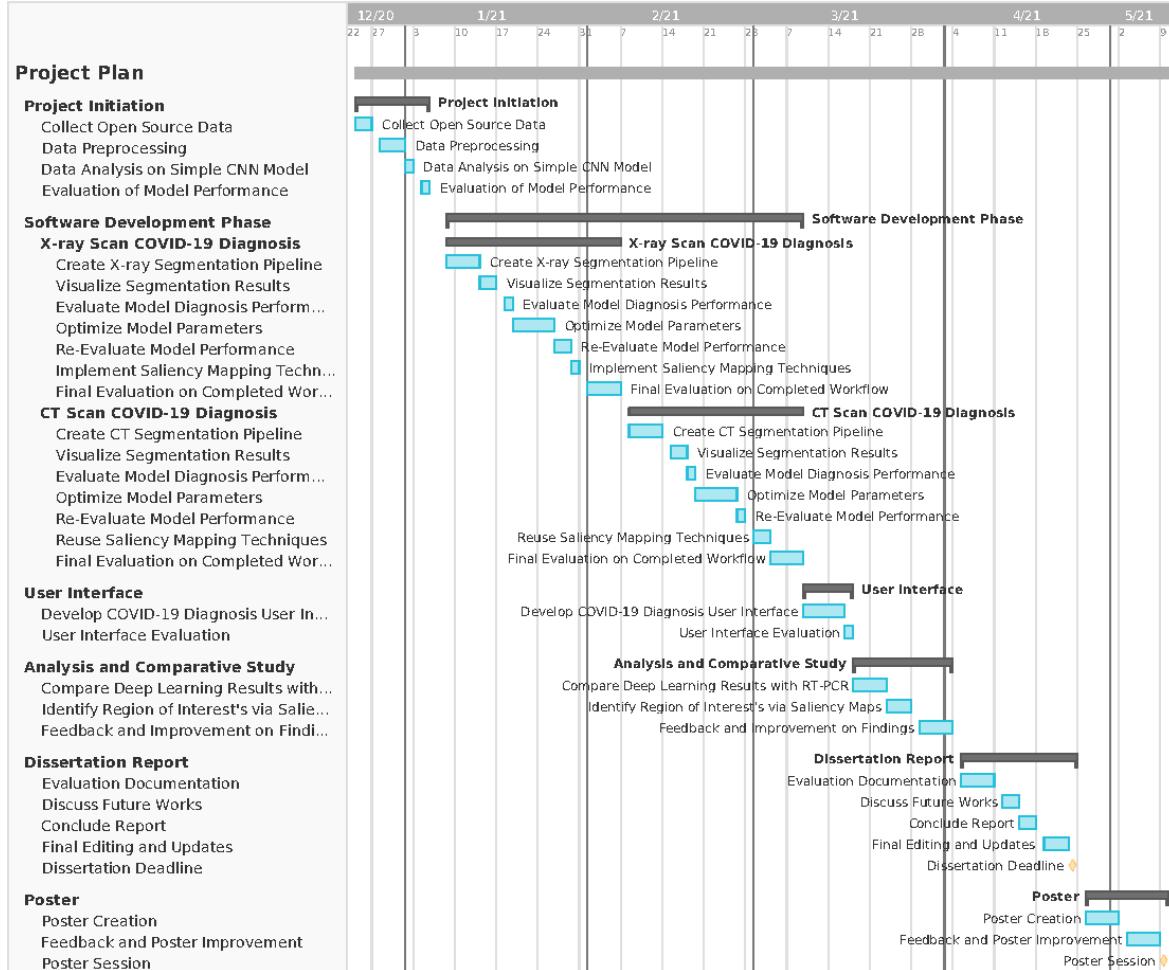


FIGURE B.1: Gantt Chart Displaying the proposed Project Plan.

B.3 Development Methodology

For this project which involves building a deep learning model accompanied by a suitable interface for user interaction, an iterative development approach is ideal. **SCRUM** [68], a subset of Agile development methodology proves to be a viable approach.

Implementing each of the components involved in this project independently and adhering to the weekly plan and supervisor's suggestions would lead to the deadlines being met and therefore

guarantee a successful project completion.

Further discussion into SCRUM development methodology can be found in Appendix [G](#).

B.4 Professional, Legal, Ethical and Social Issues

B.4.1 Professional and Legal Issues

All research papers discussed in this document are referenced accordingly. Any code used from external sources in this project shall acknowledge the original authors and would be well documented. The research papers referenced in this project are either open source or have been granted access. The code used for this project shall be open source and would be published under the MIT License. Our project shall comply with relevant standards such as GDPR and HIPAA. Any output from this research will be provided with sufficient warnings and disclaimers about lack of clinical studies of the model.

B.4.2 Ethical and Social Issues

No human subjects are involved in this dissertation project. The datasets used would be open source and will not contain any personal user information and therefore anonymized. Furthermore, all data utilized in this project shall be referenced. This project shall not cause any harm to the surrounding environment or to any observers. The ultimate aim of this project is to develop a fully automated and rapid COVID-19 diagnosis mechanism that reduces virus exposure between patients and medical professionals.

Appendix C

The COVID-19 Pandemic Era

Rise of the Global Pandemic

COVID-19 which is now officially declared as a global pandemic by the WHO was initially discovered in late 2019 emerging from Wuhan, People's Republic of China. A media statement was released by the Wuhan Municipal Health Commission confirming multiple cases of "Viral Pneumonia from an unknown cause" on December 31st 2019 [88].

On January 7th 2020, this unknown disease was identified as the novel coronavirus by the WHO. Three days later, the first known death caused by the coronavirus was reported [89]. The spread of the virus continued rapidly within China and on January 20th 2020, WHO reports the first confirmed cases outside China in Thailand, Japan, and South Korea. The very next day The United States reports its first confirmed coronavirus case [90].

Following these set of events saw the introduction of quarantine protocols via lockdowns. Starting with Wuhan, many other cities across the world also adopted the same orders to suspend the spread of the coronavirus. This prompted the WHO to declare the outbreak a global public health emergency [91].

Within the span of a month, the death toll from COVID-19 surpassed that of SARS and the WHO gives the official name for the disease caused by the coronavirus "COVID-19" [92]. The adverse effects of COVID-19 on various industries and the stock market began to show.

Travel bans, high-profile event cancellations and activation of emergency funds were implemented across different countries as the number of positive cases rose above 100,000 [94].

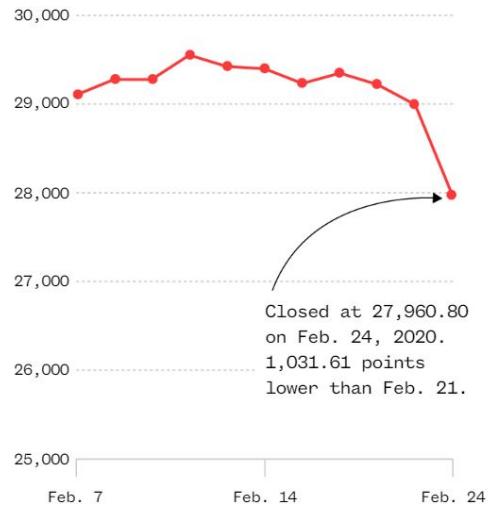


FIGURE C.1: Dow Jones Industrial Average experienced the worst day in two years.
[93]

Due to the rapid spread of the virus worldwide, on March 11th 2020, the WHO declares the coronavirus outbreak as a global pandemic [95]. Over the next few months, nationwide lockdowns were enforced in countries such as The United Kingdom, India, South Africa, Italy, Belgium, and so on.

At the end of March 2020, The United States coronavirus cases officially surpassed China, with the former reporting 82,474 cases and the latter 81,961 cases.

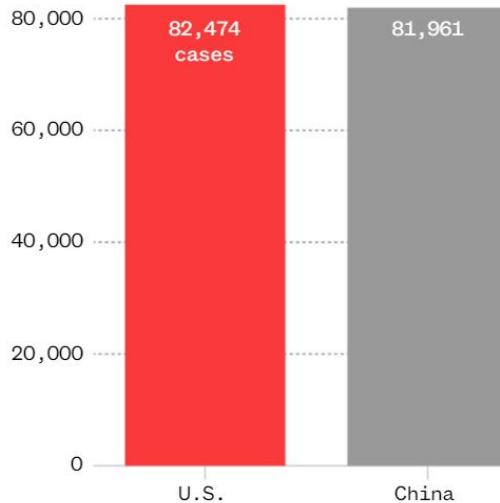


FIGURE C.2: COVID-19 cases comparison between The United States and China
[96]

On April 2nd 2020, the number of coronavirus cases worldwide surpassed 1 million with the number of deaths exceeding 51,000 [97]. The number of positive cases doubled in April and this trend continued to follow for the next few months were as of October 2nd 2020, the total number of

worldwide COVID-19 cases stand at 34,312,510, and the death toll surpassing over a million, to be precise 1,023,243 [98].

Among the countries affected by COVID-19, The United States, India and Brazil share the highest percentages of positive cases.

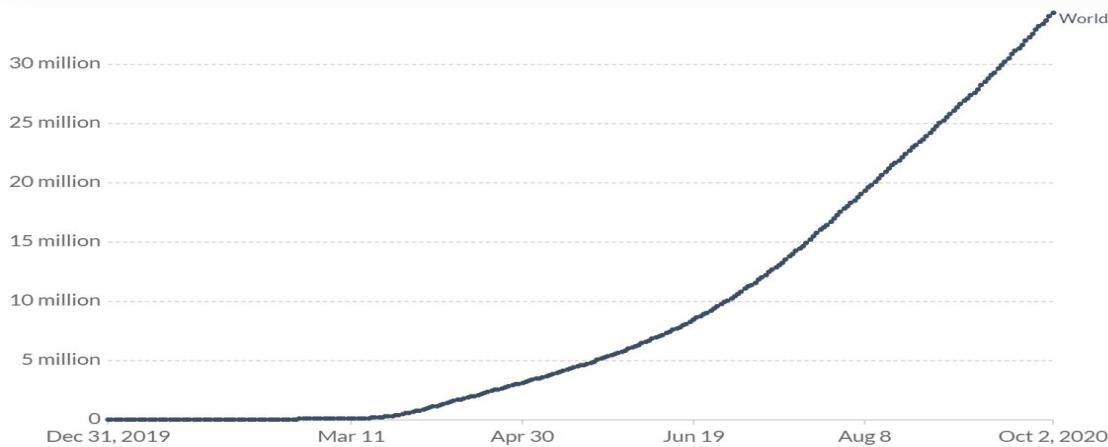


FIGURE C.3: Cumulative confirmed COVID-19 cases [99]

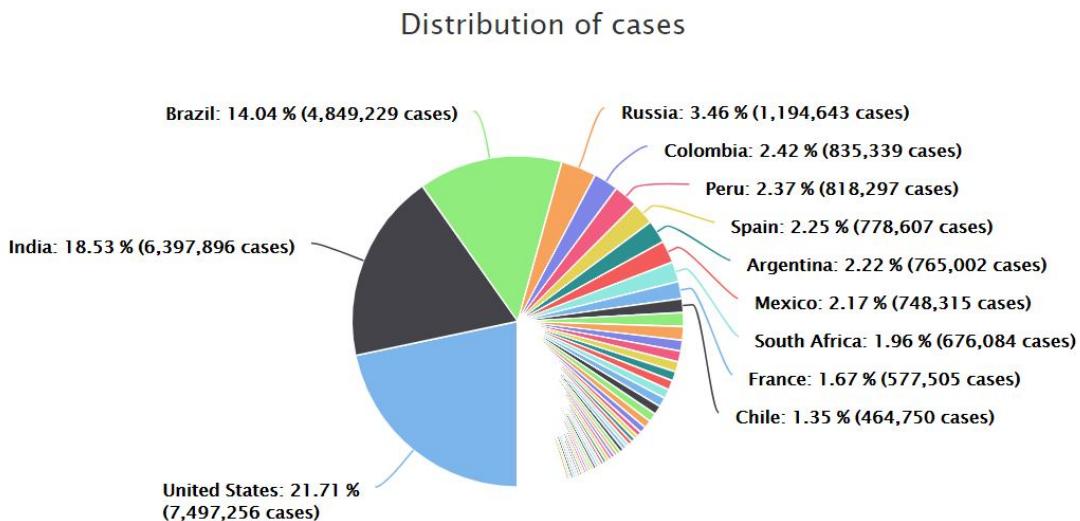


FIGURE C.4: Country wise distribution of COVID-19 cases [100]

As we have now seen the major trends and events that took place in the rapid spread of COVID-19, it is upon us to work together as a community, follow the recommendations suggested by national and international healthcare agencies through which we would be able to overcome this pandemic and successfully curb the spread of the virus.

Appendix D

COVID-19 Lung Characteristics

Additional Study on COVID-19 Patients

Morales et al. conducted a systematic literature review with meta-analysis of the imaging features of COVID-19 also observed similar lung characteristics such as GGO's from X-ray scans from patients diagnosed with COVID-19 from multiple studies [101], the results are tabulated in Table D.1.

Study	Unilateral Pneumonia	Bilateral Pneumonia	Ground-glass Opacity
Huang et al. [102]	-	40 (97.6%)	12 (29.3%)
Chen et al. [103]	25 (25.3%)	74 (74.7%)	14 (14.1%)
Wang et al. [104]	0 (0%)	138 (100.0%)	138 (100.0%)
Liu et al. [105]	-	36 (26.3%)	55 (40.1%)
Chang et al. [106]	1 (7.7%)	-	6 (46.2%)
Pan et al. [107]	-	38 (60.3%)	14 (22.2%)
Zhang et al. [108]	2 (22.2%)	5 (55.6%)	7 (77.8%)

TABLE D.1: Chest X-ray imaging characteristics from multiple studies

Appendix E

Automated Diagnosis Workflow

AI Powered Medical Scanning

Chest X-ray and CT scans are the primary screening mechanisms for diagnosing COVID-19. The limitation to these being the viral exposure between the technician and the patient, as assistance is required to perfect patient positioning for these scans to yield satisfactory results. Therefore, an automated and contact-less workflow is the need of the hour.

Modern CT and X-ray systems include cameras to allow technicians to remotely monitor the patient. But especially since the outbreak of the pandemic, this would not be sufficient as technicians would not be able to determine scanning parameters such as scan range [27].

Fortunately, a fully automated scanning workflow is indeed possible through AI. Patient pose and shape could be acquired via visual sensors such as RGB, Time-of-Flight (TOF) pressure imaging or thermal (FIR) cameras. Therefore, optimal scanning parameters could be determined [109]–[116].

Through this automated scanning workflow, radiation exposure can be reduced significantly as well as increasing the overall efficiency of the diagnosis procedure [117]. Especially during this period of the pandemic, using the same workflow would prevent virus exposure between technicians or medical practitioners and patients.

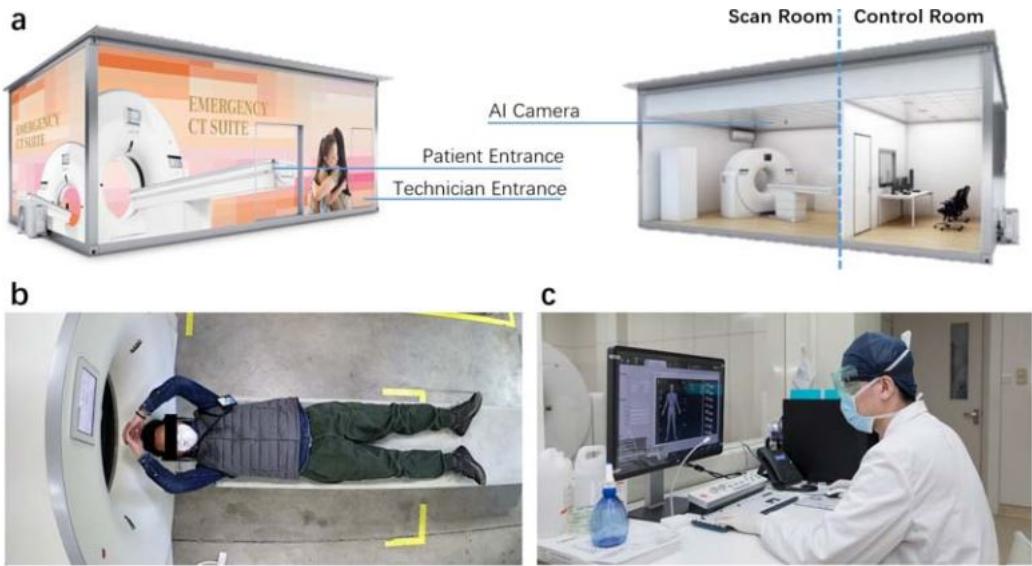


FIGURE E.1: AI-empowered automated image acquisition workflow [27].

Utilizing this workflow we can procure high-quality X-ray and CT scans suitable for segmentation, but more importantly, we can ensure the safety of both patients and technicians by reducing the radiation and virus exposure respectively.

Appendix F

CT Image Segmentation Techniques

Extended CT Scan COVID-19 Applications

Table F1 below is an extended version of Table 2.2 which summarizes existing studies and research, both quantification and diagnosis COVID-19 applications conducted on CT scans.

Study	Method	Application	Target ROI
Zheng et al. [9]	U-Net	Diagnosis	Lung
Cao et al. [10]	U-Net	Quantification	Lung
			Lesion
Huang et al. [11]	U-Net	Quantification	Lung
			Lung Lobes
			Lesion
Yue et al. [12]	U-Net	Quantification	Lung Lobes
			Lesion
Gozel et al. [13]	U-Net	Diagnosis	Lung
			Lesion
Shan et al. [17]	VB-Net	Diagnosis	Lung
			Lung Lobes
			Lung Segments
			Lesion
Li et al. [14]	U-Net	Diagnosis	Lesion
Chen et al. [15]	UNet++	Diagnosis	Lesion

Jin et al. [16]	UNet++	Diagnosis	Lung
			Lesion
Tang et al. [26]	Commercial Software	Quantification	Lung
			Lesion
			Trachea
			Bronchus

TABLE F.1: CT Image Segmentation Techniques in COVID-19 Applications [27]

Appendix G

SCRUM

SCRUM Development Methodology

SCRUM involves building software iteratively coupled with a continuous integration and development pipeline if set up, this enables frequent feedback from the project supervisor. This development methodology allows us to perform weekly increments where a reasonable part of the functionality is developed or modified and these updates could be reviewed by the project supervisor. Therefore, it would be easier to detect and mitigate risks involved such as slow progression in model development or misinterpretation of the project requirements.

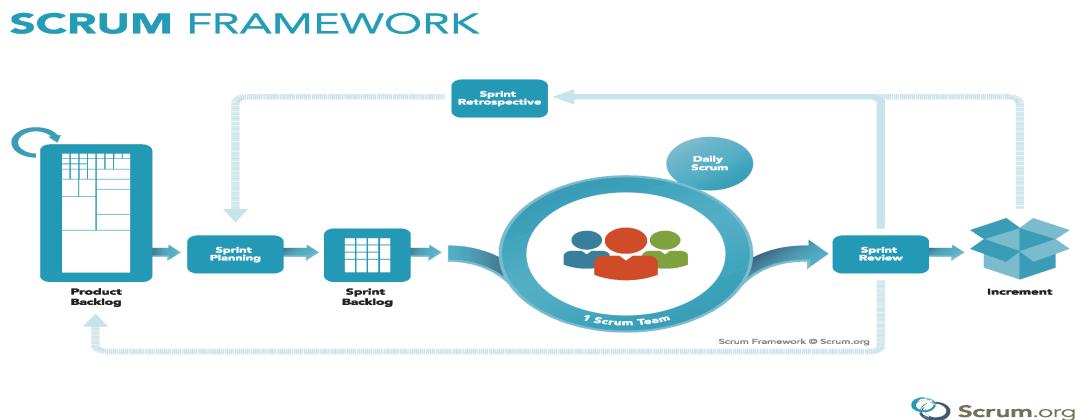


FIGURE G.1: SCRUM Development Methodology [118]

Appendix H

Paper: "A Deep Learning Method for COVID-19 Detection from X-ray Scans"

We intend to publish the following paper in *Computers in Biology and Medicine*,¹. The attached paper is still under review and minor changes are expected.

¹Link to the Journal - <https://www.journals.elsevier.com/computers-in-biology-and-medicine>

A Deep Learning Method for COVID-19 Detection from X-ray Scans

Alister Luiz, Hani Ragab, Kayvan Karim and Hadj Batatia

MACS, Heriot-Watt University

ARTICLE INFO

Keywords:

COVID-19, SARS-CoV-2, Chest X-ray, Classification, Deep Learning, Convolutional Neural Networks

ABSTRACT

RT-PCR is currently the most widely used diagnosis test for COVID-19. However, it has some limitations including shortage of test kits, delays in receiving results, and a relatively low accuracy of COVID-19 detection. In this paper, we propose an automated method for quick, efficient, and affordable COVID-19 diagnosis based on X-Ray scans. Our proposed method is based on a deep learning classification model for identifying lung patterns that are characteristic of COVID-19 infection.

In contrast to existing methods, we adopt a DenseNet121 architecture. We used transfer learning from an existing model to speed up the training phase. We used a dataset of a total size of 1716 X-ray scans that we did split into eleven folds. Ten folds were used in 10-fold cross validation training and testing, whereas the eleventh fold was kept aside and only used to choose the best out of the ten models produced by the 10-fold cross validation. The dataset contained 572 samples of each of the three classes: COVID-19, Pneumonia, and Negatives.

We obtained 94.6% accuracy with excellent sensitivity, specificity and F1-score for the three classes. Our method gives better diagnosis when compared with the RT-PCR test results. Further, we did implement a CNN-based model and DenseNet121 provided better results. We deployed an experimental web application (www.aiinnov.com) to allow live testing of our model. After thorough clinical validation, such application could help reduce workload and limit virus exposure for health-care professionals.

1. Introduction

COVID-19, which was declared a global pandemic by the World Health Organization on March 11, 2020, has affected millions of people worldwide, and had important consequences on health, social life, with a severe global economic impact.

Over a billion tests have been carried across the world for diagnosing patients with COVID-19 [7]. This additional activity creates higher workload on professionals and requires complex strict organization at healthcare providers, with consequences on ordinary processes, especially for chronic diseases. Despite the ongoing vaccination efforts undertaken by many countries, there is still a need for efficient and quick testing mechanisms. These would alleviate the burden on health professionals and limit contact cases.

X-rays and CT scans have proven to be viable alternatives to the RT-PCR test for COVID-19 detection. Especially when considering (i) their availability in countries and regions where RT-PCR kits are not available, (ii) faster diagnosis time, and (iii) their reduced exposure risks.

Chung et al. [5] conducted a study on 21 symptomatic patients infected with coronavirus admitted to three hospitals in provinces of Guangdong, Jiangxi, and Shandong respectively in China from January 18th, 2020 to 27th, 2020. Their aim was to identify potential imaging features of COVID-19 from CT scans with the help of two experienced fellowship-trained cardiothoracic radiologists. The degree of lobe involvement was assessed, and a "Total Severity Score" was assigned by summing up each of the individual lobe scores. Patients were also re-evaluated to study the progression of features by the same two radiologists. Table 1 shows the common observed characteristics reported by the study. Pri-

Table 1
Findings from Chest CT Examination in 21 Patients [5].

Finding	Value
Ground-glass opacities and consolidation	
Absence of both glass opacities and Consolidation	3 (14%)
Presence of either ground-glass opacities or consolidation	18 (86%)
Presence of ground-glass opacities without consolidation	12 (57%)
Presence of ground-glass opacities with consolidation	6 (29%)
Presence of consolidation without ground-glass opacities	0 (0%)
More than two lobes affected	15 (71%)
Total lung severity score	
Mean	9.9
Range	0-19

mary observations include Ground-Glass Opacities (GGO) symptoms found in 12 patients and consolidation symptoms in 6 others. As shown in Figure 1, the virus affects more than two lobes with bilateral involvement in most cases. Other observations include rounded morphology detected in 7 patients, reticulation in 3, and crazy paving in 4 [5].

One obvious limitation of this study is the relatively low number of patients, with only 8 out of 21 carrying a follow-up CT scan. As this study was conducted during the dawn of coronavirus, this number is certainly a very respected amount. The study also shows it is feasible, by looking at chest medical imagery, to identify traits that highly indicate the presence of COVID-19. This conclusion has been reinforced by medical practice.

A systematic literature review with meta-analysis of the imaging features of COVID-19 also observed similar lung characteristics such as GGO's from patients diagnosed with COVID-19 from multiple studies [24]. Its findings are summarized in Table 2. We observe that GGO is a common symptom among COVID-19 patients. An example being Wang et al. [28] whose study identified GGO's in 100% of the patients under examination. Another study from Zhang

ORCID(s):

Table 2

Chest X-ray imaging characteristics from multiple studies.

Study	Ground-Glass Opacities (GGO)
Huang et al. [12]	29.3%
Chen et al. [3]	14.1%
Wang et al. [28]	100.0%
Liu et al. [17]	40.1%
Chang et al. [2]	46.2%
Pan et al. [21]	22.2%
Zhang et al. [33]	77.8%

et al. [33] also observe GGO's in majority of the patients (77.8%).

Deep learning can be used to detect objects and patterns in images with hight accuracy. As such, it can also be used to detect COVID-19 from X-ray scans by identifying related patterns. We propose a deep learning model that can analyze chest X-rays and detect presence of COVID-19 infections. Our model achieves higher sensitivity and specificity than the RT-PCR test. Furthermore, we provide heatmap images to identify areas of concern or interest for human experts.

While achieving higher diagnosis performance, the proposed solution has the advantages of being quick and affordable. It can be deployed in areas with limited medical facilities and allows telemedicine procedures. For illustration purpose, we deployed our model at www.aiinnov.com.

The remainder of this paper is organized as follows. We review notable COVID-19 diagnosis methods based on deep learning in section 2. We introduce our deep learning-based approach in section 3. We report our experimental results and discuss their interpretability for potential usage by medical professionals in section 4. Finally, we review our work and discuss potential future extensions and improvements in section 5.

2. Related work

The rapid widespread of the coronavirus global pandemic across the world forced the scientific and medical community to identify alternative diagnosis mechanisms to supplement the RT-PCR test and possibly overcome its limitations. As described in the previous section, medical imagery such as X-rays and CT scans have the potential to play a vital role in combating the rising numbers by saving valuable time in diagnosis and reducing virus exposure. We will focus on X-rays, as they are most often the first imaging modality used on suspected patients, due to their wide availability in most clinics and medical facilities.

Automatic X-ray image processing is required in order to further assist medical professionals in diagnosing COVID-19. Image segmentation is a preliminary stage to delimit the lung regions of interest (ROI). This should overcome the ambiguities due to the ribs being projected onto soft tissues in such images.

2.1. Parsing Medical Imagery

Feeding images into classical machine learning algorithms could be challenging. Representation learning is a set of

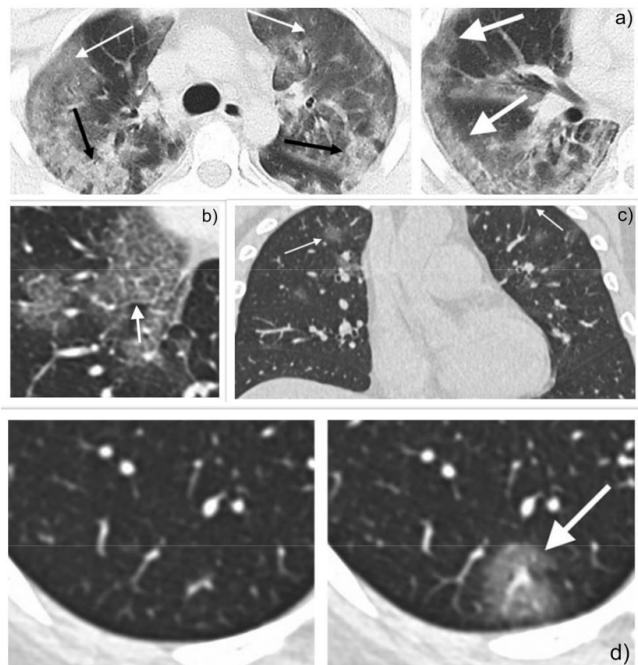


Figure 1: Observed lung CT Scan characteristics. a) White arrow indicates patchy GGO, and black arrow shows consolidative pulmonary opacities. b) White arrow indicates GGO's with a rounded morphology. c) White arrow shows crazy-paving pattern, GGO and interlobular septal thickening with intralobular lines. d) Follow-up CT scan progression. White arrow indicates new solitary, rounded, peripheral ground-glass lesion [5].

methods that allow a machine to be fed with raw data, including images, and to automatically discover the representations needed for the machine learning task at hand. Deep-learning methods are known to perform representation learning with multiple levels, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level.

In medical image processing, using deep learning has grown very fast in recent years, with the potential of providing efficient methods for diagnosis, prognosis and follow-up that allow coping with the ever-growing volumes of medical big data [23].

X-rays have been used for various groundbreaking AI-empowered medical research applications. Deep learning techniques applied on X-ray scans have allowed scientists to detect lung cancer at an early stage. Indeed, Sim et al. conducted a multi-center study where a deep convolutional neural network (DCNN) was used to assist 12 radiologists to detect malignant pulmonary nodules on chest radiographs. The average sensitivity improved from 65.1% to 70.3% when the radiologists re-reviewed radiographs with the DCNN software [27].

RetinaNet, a popular deep learning technique, was used for early detection of breast cancer allowing patients to have proper treatment and consequently reduce rate of morbidity

Table 3
X-ray image segmentation techniques in COVID-19 diagnosis applications [26]

Proposal	Method	COVID-19	Bacterial Pneumonia	Viral Pneumonia	Normal	Accuracy	Sensitivity	Specificity	F1 Score
Ghoshal [9]	CNN	70	-	-	-	92.9%	85.7%	99.4%	75.0%
Zhang [32]	ResNet	70	-	-	1008	96.0%	96.0%	70.7%	73.5%
Narin [20]	RestNet50	50	-	50	-	98.0%	91.8%	96.6%	83.5%
Wang [29]	CNN	45	931	660	1203	83.5%	91.0%	99.5%	94.7%
Karakanis [14]	ResNet8	175	100	-	100	98.3%	99.3%	98.1%	99.1%

[10]. Deep learning has also been used to detect both viral and bacterial Pneumonia that share similar lung characteristics with COVID-19 [16, 29].

2.2. COVID-19 Diagnosis using Deep Learning

Despite being the first imaging modality for patients suspected with COVID-19, X-ray images are less sensitive than 3D chest CT scans. Anomalous chest radiographs are found in 69% of the patients initially during admission and this number increases to 80% after a certain period once hospitalized [30].

Four popular deep learning architectures are used across various studies to detect the abnormalities commonly found in COVID-19 patients with lung X-ray scans, namely ResNet [32], ResNet-50 [20], ResNet-8 [14], and CNN [9, 29]. We present a few notable works from the four categories in the rest of the section.

Ghoshal et al. proposed a Bayesian CNN to estimate the uncertainty in COVID-19 prediction [9]. X-rays of 70 COVID-19 patients were obtained from Cohen et al. [6] and others from Kaggle’s chest X-ray images. The authors reported that Bayesian inference improves the detection accuracy of the model from 85.7% to 92.9%.

Narin et al. experimented with three deep learning models, namely ResNet-50, InceptionV3 and Inception-ResNetV2, with the main objective of detecting COVID-19 from X-ray images [20]. Their dataset includes X-rays from 50 COVID-19 patients and 50 normal scans. The reported results indicate that the ResNet-50 model achieves the highest accuracy with 98% followed by InceptionV3 which attains 97%.

Zhang et al. also suggest a ResNet based model for COVID-19 detection [32]. Their proposed model considers COVID-19 cases as abnormal situations given a large dataset of normal X-ray images. Precisely, they experimented with a dataset containing 70 X-rays from COVID-19 patients and 1008 normal, and report 96% accuracy but a relatively low 73.5% F1 score.

Wang et al. proposed the COVID-net, a deep CNN based model, which achieves a testing accuracy of 83.5% [29]. Their dataset includes X-rays from 45 COVID-19 positive, 931 Bacterial Pneumonia, 660 viral-pneumonia, and 1203 normal X-rays.

Karakanis et al. implemented a ResNet8 model to perform multi-class classification for COVID-19, pneumonia, and normal X-rays [14]. The X-ray dataset comprised of 175 COVID-19, 100 normal, and 100 pneumonia scans. The model achieved competitive performance with 98.3% accu-

racy, 99.3% sensitivity and 98.1% specificity.

Table 3 summarizes the results obtained by the studies discussed in this section. They all aimed at detecting COVID-19 cases while discriminating them from Pneumonia. Despite the good results reported, the major limitation relates to the low number of COVID-19 cases included in the experiments. Therefore, generalizability and stability of the proposed models are yet to be evaluated.

3. Our Approach

We propose a multiclass classifier that can categorize an X-Ray as either containing COVID-19, Pneumonia, or a healthy chest. Our classifier uses a deep learning DenseNet121 architecture [13]. The motivation of using this architecture arises from the complex nature of X-ray images that exhibit multiple tissue layers and anatomic structures projected into the 2D image.

Our hypothesis is that reusing features from different network layers allows capturing characteristic COVID-19 patterns from layered lung tissues. Architectures such as CNN do not allow such reuse of features from different layers. DenseNet121 is precisely designed for this purpose. It allows features propagation and reuse. Therefore, it is best suited given our objective.

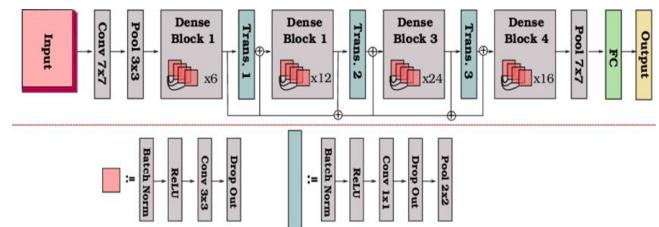


Figure 2: Architecture of DenseNet121 [22].

Figure 2 depicts the architecture of a DenseNet121 with 5 layers. Its main characteristic is the direct connections between all of them. In addition, DenseNet121 requires fewer parameters when compared to an equivalent traditional CNN. We implemented the ADAM optimization algorithm for training the model. We review our X-Ray dataset in 3.1, present our methodology in 3.2, and discuss how used transfer learning in 3.3.

Table 4
Dataset used in our study.

Disease	Total Images	Non-augmented Training	Augmented Training	Test (not-augmented)	Unseen Data
COVID-19	572	468	1070	52	52
Pneumonia	572	468	1070	52	52
Normal	572	468	1070	52	52

3.1. X-ray Dataset Preparation

Our model was trained and tested using images from a Kaggle dataset¹. The dataset is compiled from multiple sources [8, 18, 4]. It is an open-source database of COVID-19 cases which includes X-ray and CT scans. It includes COVID-19 and Pneumonia cases as well as healthy chest scans.

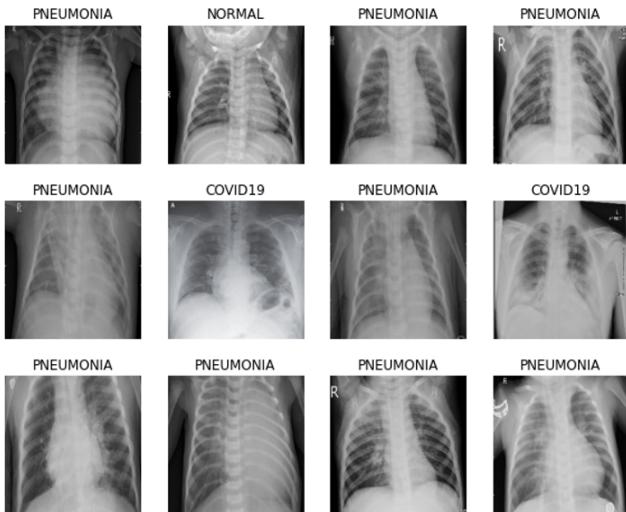


Figure 3: Sample of X-ray scans with labels for model training.

Furthermore, given the objective of our classification task to identify COVID-19 patterns, only the posterior-anterior view of the lungs was considered for training and testing. Indeed, this view visualizes lungs, bony thoracic cavity, mediastinum, and great vessels [19]. This choice allowed us to reduce the number of training instances and work with a less unbalanced dataset. Figure 3 contains the images and shows their corresponding labels.

3.2. Methodology

We plan to use 10-fold cross validation and leave an eleventh fold to test the 10 models on it as unseen data. Indeed, 10-fold cross validation produces 10 models that, collectively, saw all the training dataset. The original Kaggle dataset contained 575 COVID-19 cases, 4273 Pneumonia cases, and 2583 Negatives. We built a balanced dataset of 572 samples of each of the classes (572 is a multiple of 11). The size of the eleventh fold, to which we refer to from now on as *unseen data*, is 52.

At each of the 10 iterations of the 10-fold cross validation, we take 10% of the dataset out to be used as test-

¹ Available at <https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia>

Table 5
Results of our approach

Class	COVID-19	Normal	Pneumonia
Sensitivity	97.7%	92.7%	93.5%
Specificity	98.9%	96.8%	96.3%
F1 Score	97.7%	93.1%	93.0%

ing dataset. The remaining 90% undergoes an augmentation step. The augmentation step uses three operators to produce slightly modified image, namely rotation, zooming and shearing. Images rotation rotates an image by 5°. Zooming applies a +2% zoom. Shearing, which distorts an image along an axis (essentially converting rectangles into parallelograms), applies a 2° counter-clockwise distortion. The augmentation of each of the 468 samples of each class (90% of 520) resulted in creating 1070 samples of that class. Since we have three classes, COVID-19, Pneumonia and Negatives, the total number of samples in the final training dataset is of 3210. Only the training dataset was augmented; no augmentation was applied to the test dataset. Table 4 shows the size of the data we used.

3.3. Training and transfer learning

Transfer learning is a machine-learning technique where a model trained for an application with some dataset is reused on another dataset. The pre-trained model approach to transfer learning consists of using the source model as a starting point for training the target model.

In order to reduce training-time and improve accuracy, we adopted the pre-trained transfer learning approach. We used an ImageNet model trained with large volume of images for classification tasks. Using the pre-trained model, we trained only the final layer. All other layers used fixed weights obtained from the prior training using the ImageNet dataset.

We implemented our deep learning solution using the Keras framework. We utilized several functions from the Callback API to control the training stage. We particularly used them for monitoring the loss metric, assessing the evolution of the learning rate to avoid plateau phenomena, and selecting models with the best performance. Practically, this led to significantly reducing the training time.

4. Results

Table 5 shows the results of applying our method on the dataset we described in 3.1. The table reports class-wise sensitivity, specificity and F1-scores. Our cross validation resulted in 94.6% average accuracy on the balanced dataset.

Table 6

Comparison of the results of our approach with and without Data Augmentation.

	With Augmentation	Without Augmentation
Accuracy	94.6%	93.84%
Sensitivity	94.6 %	94.0%
Specificity	97.3 %	96.0%

Table 7

Comparison of our approach with RT-PCR and CNN.

Metric	Our Approach	RT-PCR	CNN	Narin et al. [20]
Accuracy	98.5%	-	93.9%	97.0%
Sensitivity	97.7%	84.2%	85.2%	91.1%
Specificity	98.9%	98.9%	98.2%	99.7%

Furthermore, each of the 10 models were saved after cross validation and tested on completely unseen data that were not part of the 10 folds used in cross-validation. Table 6 compares the performance of our model when augmentation is not used to when we use it. It is trivial that augmentation leads to better performance.

Comparison to Existing Approaches

We compared the results given by our model, only for the COVID-19 class, with the RT-PCR test according to the study conducted by the Infectious Diseases Society of America (IDSA). We also did implement a CNN-based model, on the same dataset, to compare our model with deep learning models that use state of the art techniques. Furthermore, we did apply the approach proposed by Narin et al. [20] to our dataset. We contacted the authors of [14] and [32] asking for their source codes but did not get them.

Table 7 summarizes the comparison and uses three metrics, namely sensitivity, specificity and accuracy (relatively to the COVID-19 class). The table shows the high quality of our model. It can be noted that our method scores higher than RT-PCR and better than CNN. The difference is more significant for the accuracy, in favor of our method.

5. Discussion and Interpretation

Deep learning models are often regarded as black boxes. Interpretability and understandability of such models are needed to comprehend or discern their functioning and results. Interpretability becomes a requirement when deep learning is applied to science and medicine. In our study, we investigated means of understanding the reasoning underlying diagnosis of COVID-19 given by our model.

One way of achieving interpretability is by providing radiologists with means to cross-check their traditional distinct features obtained from lung image segmentation. For example, superpose a visualization of the output of a deep learning outcome on the original figure features (such as the GGO in Figure 1).

Various interpretation methods exist for deep-learning. For example, saliency methods are a set of tools used to analyze deep learning results [1]. Despite having been criticized

due to their lack of reliability, they can provide analysts with insights about their models [15]. The most used saliency tools include i) *Saliency Map* which estimates specific parts of the image that contribute to highest layer activation [31]; ii) *Class Activation Mapping (CAM)* which averages and adds the activations of each feature map (Global Average Pooling) and uses this to highlight important regions [34]; and iii) *Gradient-Weighted CAM (Grad-CAM)* which consists in calculating the gradient of the classification score with respect to the convolutional features [25].

We calculated the Grad-CAM heatmaps of the COVID-19 class. Grad-CAM heatmaps should highlight regions in the lungs which, as discussed in section 1, exhibit the most common characteristics in patients diagnosed with COVID-19. These include features such as GGO's, consolidations, lesions, and crazy-paving patterns, which are some of the most contributing features to diagnosis. This technique provides interpretation means that would assist radiologists in identifying the same lung characteristics as with traditional segmentation-based methods.

Grad-CAM Heat Maps for COVID-19

For the purpose of building a more interpretable model, we generated heatmaps of the COVID-19 scans. The heatmaps highlight the regions that led to the classification by our model. Heatmaps obtained for a sample of 15 previously unseen scans is displayed in Figure 4. Regions in red are the most influential in making the decision, whereas blue regions are the least influential.

We verified that the heat maps we obtained correspond to the regions highlighted from studies conducted by Harmon et al. [11] and Li et al. [16]. The lower lung lobes seem to be the most affected in patients diagnosed with COVID-19, this correlates with the observations from Chung et al. [5] and therefore, provides additional validation that lung characteristics such as GGO's, consolidations, and lesions are major contributing factors to COVID-19 detection.

To further verify that the heatmaps produced by the model are in conformance with the same observed characteristics, we have provided a subset of five X-rays scans to two senior radiologists and asked them to highlight the critical regions and provide their diagnosis results.

Figure 5 compares the radiologist's annotations and diagnosis with those produced by our model along with the heatmap result. We can observe that the heatmaps are in direct compliance with the radiologists' findings. Precisely, visually, the heatmaps show that our predictive model highlights at least one or more of the critical regions identified by the radiologist. Furthermore, all areas highlighted by the radiologist are shades of either red or purple, but never blue. We can conclude that if the model is integrated into a seamless COVID-19 or Pneumonia diagnosis workflow in which X-Ray scans are overlayed with our heatmaps, then we could considerably increase the efficiency and effectiveness of diagnosis due to faster processing and affirmation for the results obtained.

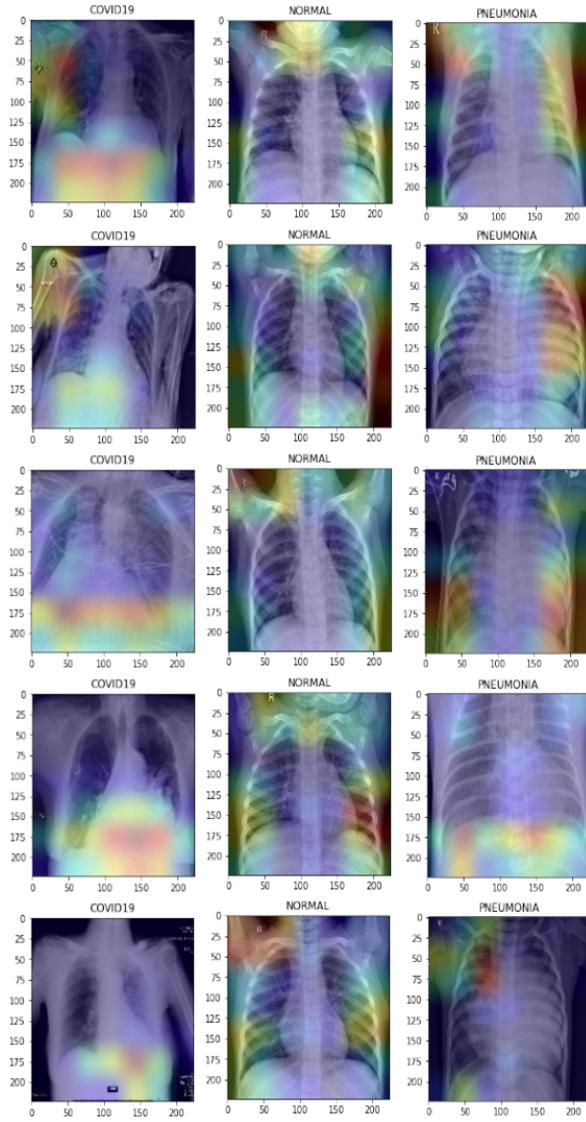


Figure 4: Heat map visualization for DenseNet121 results.

6. Conclusion

The use of machine and deep learning methods in medicine has a great potential. It is the means to achieve predictive and personalized medicine, which is the next paradigm shift in the domain. This paper builds on results from many prior studies and presents a case for predictive medicine in COVID-19. It proposes a deep learning model for COVID-19 diagnosis based on X-ray images. More specifically, the proposed classification model detects lung patterns that are characteristics of COVID-19 infections using X-ray images.

The proposed model is based on the DenseNet121 architecture and was trained on a dataset from trustworthy public datasets. It consists of balanced dataset composed of 1716 images from patients with COVID-19, Pneumonia and Normal scans. Experimental results gave 94.6% accuracy, showing better performance than the RT-PCR and CNN-based classification methods.

We discussed our results and used interpretability tech-

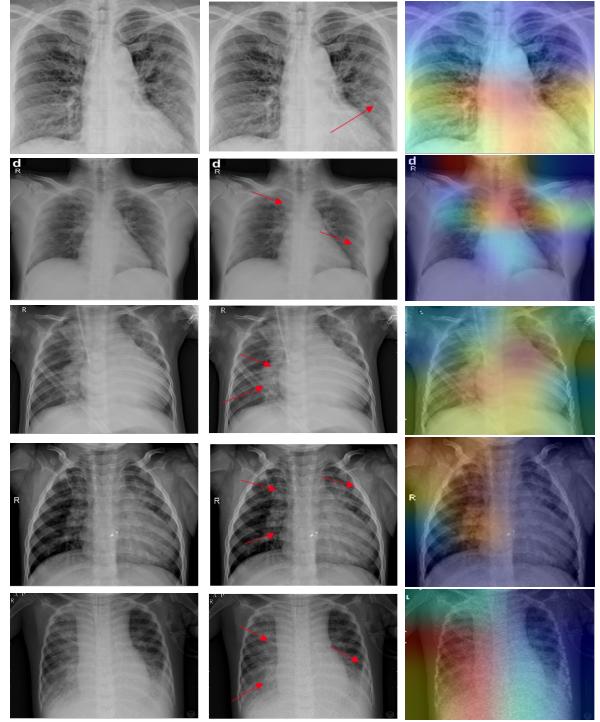


Figure 5: Comparison of radiologist's annotations and heatmap results. Annotations are in accordance with our classification. One can also notice the good overall matching of radiologist's annotation with high heat areas. The upper two rows correspond to COVID-19 and the lower three to pneumonia.

niques for understanding the underlying decisions. More specifically, salient maps were used to highlight regions in the lungs which exhibit the most common characteristics of COVID-19. We also implemented heatmaps visualization to illustrate the regions of lungs that contributed the most to the classification decision.

The proposed method is non-invasive, quick, cost effective and can lead to semi-automated telemedicine procedures. Future work will be twofold. First, a large-scale clinical study will be implemented to validate the results and derive a medical protocol for routine clinics. Second, the method will be generalized to COVID-19 CT images. The objective will be to derive prognosis using short time series of CT images. Perspective work consists also in ensuring robustness of the proposed model. One possible approach is to assess the quality of input images based on criteria to be defined.

Acknowledgement

The authors thank Dr. G.R. Mahadevan (MD, IDCCM, D.Diab, Specialist in COVID-19) from Maya Nursing Home, Tamil Nadu for annotating the X-ray images.

References

- [1] Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., Kim, B., 2018. Sanity checks for saliency maps. arXiv preprint

- arXiv:1810.03292 .
- [2] Chang, D., Lin, M., Wei, L., Xie, L., Zhu, G., Cruz, C.S.D., Sharma, L., 2020. Epidemiologic and clinical characteristics of novel coronavirus infections involving 13 patients outside wuhan, china. *Jama* 323, 1092–1093.
- [3] Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., et al., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study. *The lancet* 395, 507–513.
- [4] Chung, A.G., . Open source covid-19 dataset. URL: <https://github.com/agchung>.
- [5] Chung, M., Bernheim, A., Mei, X., Zhang, N., Huang, M., Zeng, X., Cui, J., Xu, W., Yang, Y., Fayad, Z.A., et al., 2020. Ct imaging features of 2019 novel coronavirus (2019-ncov). *Radiology* 295, 202–207.
- [6] Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q., Ghassemi, M., 2020. Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988 .
- [7] Elflein, J., 2021. Covid-19 tests by country. URL: <https://www.statista.com/statistics/1028731/covid19-tests-select-countries-worldwide/>.
- [8] Ganglia, B., . Covid-19 image data collection. URL: <https://github.com/ieee8023/covid-chestxray-dataset>.
- [9] Ghoshal, B., Tucker, A., 2020. Estimating uncertainty and interpretability in deep learning for coronavirus (covid-19) detection. arXiv preprint arXiv:2003.10769 .
- [10] Hamed, G., Marey, M.A.E.R., Amin, S.E.S., Tolba, M.F., 2020. Deep learning in breast cancer detection and classification, in: Joint European-US Workshop on Applications of Invariance in Computer Vision, Springer. pp. 322–333.
- [11] Harmon, S.A., Sanford, T.H., Xu, S., Turkbey, E.B., Roth, H., Xu, Z., Yang, D., Myronenko, A., Anderson, V., Amalou, A., et al., 2020. Artificial intelligence for the detection of covid-19 pneumonia on chest ct using multinational datasets. *Nature communications* 11, 1–7.
- [12] Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al., 2020. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The lancet* 395, 497–506.
- [13] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
- [14] Karakanis, S., Leontidis, G., 2021. Lightweight deep learning models for detecting covid-19 from chest x-ray images. *Computers in Biology and Medicine* 130, 104181.
- [15] Kindermans, P.J., Hooker, S., Adebayo, J., Alber, M., Schütt, K.T., Dähne, S., Erhan, D., Kim, B., 2019. The (un)reliability of saliency methods, in: Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Springer, pp. 267–280.
- [16] Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., et al., 2020. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology* .
- [17] Liu, K., Fang, Y.Y., Deng, Y., Liu, W., Wang, M.F., Ma, J.P., Xiao, W., Wang, Y.N., Zhong, M.H., Li, C.H., et al., 2020. Clinical characteristics of novel coronavirus cases in tertiary hospitals in hubei province. *Chinese medical journal* .
- [18] Mooney, P., 2018. Chest x-ray images (pneumonia). URL: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>.
- [19] Murphy, K., Smits, H., Knoops, A.J., Korst, M.B., Samson, T., Scholten, E.T., Schalekamp, S., Schaefer-Prokop, C.M., Philipsen, R.H., Meijers, A., et al., 2020. Covid-19 on chest radiographs: a multireader evaluation of an artificial intelligence system. *Radiology* 296, E166–E172.
- [20] Narin, A., Kaya, C., Pamuk, Z., 2020. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849 .
- [21] Pan, Y., Guan, H., Zhou, S., Wang, Y., Li, Q., Zhu, T., Hu, Q., Xia, L., 2020. Initial ct findings and temporal changes in patients with the novel coronavirus pneumonia (2019-ncov): a study of 63 patients in wuhan, china. *European radiology* 30, 3306–3309.
- [22] Radwan, N., 2019. Leveraging sparse and dense features for reliable state estimation in urban environments. Ph.D. thesis. University of Freiburg, Freiburg im Breisgau, Germany.
- [23] Razzak, M.I., Naz, S., Zaib, A., 2018. Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps* , 323–350.
- [24] Rodriguez-Morales, A.J., Cardona-Ospina, J.A., Gutiérrez-Ocampo, E., Villamizar-Peña, R., Holguin-Rivera, Y., Escalera-Antezana, J.P., Alvarado-Arnez, L.E., Bonilla-Aldana, D.K., Franco-Paredes, C., Henao-Martinez, A.F., et al., 2020. Clinical, laboratory and imaging features of covid-19: A systematic review and meta-analysis. *Travel medicine and infectious disease* 34, 101623.
- [25] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE international conference on computer vision, pp. 618–626.
- [26] Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., Shen, D., 2020. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE reviews in biomedical engineering* .
- [27] Sim, Y., Chung, M.J., Kotter, E., Yune, S., Kim, M., Do, S., Han, K., Kim, H., Yang, S., Lee, D.J., et al., 2020. Deep convolutional neural network-based software improves radiologist detection of malignant lung nodules on chest radiographs. *Radiology* 294, 199–209.
- [28] Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., et al., 2020a. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in wuhan, china. *Jama* 323, 1061–1069.
- [29] Wang, L., Lin, Z.Q., Wong, A., 2020b. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports* 10, 1–12.
- [30] Wong, H.Y.F., Lam, H.Y.S., Fong, A.H.T., Leung, S.T., Chin, T.W.Y., Lo, C.S.Y., Lui, M.M.S., Lee, J.C.Y., Chiu, K.W.H., Chung, T.W.H., et al., 2020. Frequency and distribution of chest radiographic findings in patients positive for covid-19. *Radiology* 296, E72–E78.
- [31] Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks, in: European conference on computer vision, Springer. pp. 818–833.
- [32] Zhang, J., Xie, Y., Li, Y., Shen, C., Xia, Y., 2020a. Covid-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv preprint arXiv:2003.12338 .
- [33] Zhang, M., Wang, X., Chen, Y., Zhao, K., Cai, Y., An, C., Lin, M., Mu, X., 2020b. Clinical features of 2019 novel coronavirus pneumonia in the early stage from a fever clinic in beijing. *Zhonghua jie he hu xi za zhi= Zhonghua jiehe he huxi zazhi= Chinese journal of tuberculosis and respiratory diseases* 43, 215–218.
- [34] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2921–2929.