Demystify the Blackbox Model of Automated Detection of Lung and Kidney Diseases from X-ray and CT Radiographs

by

Md. Nazmul Islam
19166020

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
M.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
BRAC University
December 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Md. Nazmul Islam
19166020
md.nazmul.islam1@g.bracu.ac.bd

Approval

The project titled "Demystify the Blackbox Model of Automated Detection of Lung and Kidney Diseases from X-ray and CT Radiographs"

Submitted by:
Md. Nazmul Islam (19166020)

Of Fall, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of M.Sc. in Computer Science and Engineering on December 23, 2021.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
Brac University
rabiul.alam@bracu.ac.bd

Examiner:
(External)

Prof. Mohammad Shamsul Arefin, Ph.D.
Professor
Department of Computer Science and Engineering
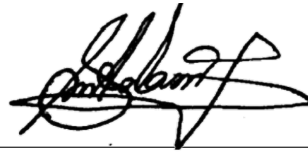Chittagong University of Engineering and Technology (CUET)
sarefin@cuet.ac.bd

Examiner:
(Internal)

Dr. Md Khalilur Rhaman
Associate Professor
Department of Computer Science and Engineering
Brac University
khalilur@bracu.ac.bd

Examiner:
(Internal)

Dr. Md. Ashraful Alam
Assistant Professor
Department of Computer Science and Engineering
Brac University
ashraful.alam@bracu.ac.bd

Program Coordinator:
(Member)

Dr. Amitabha Chakrabarty
Associate Professor
Department of Computer Science and Engineering
Brac University
amitabha@bracu.ac.bd

Chairperson:
(Chair)

Sadia Hamid Kazi, Ph.D.
Associate Professor
Department of Computer Science and Engineering
Brac University
skazi@bracu.ac.bd

Abstract

Radiology imaging, such as magnetic resonance imaging (MRI), computed tomography (CT-scan), X-ray imaging (X-ray), and ultrasound imaging (US), is used to diagnose a variety of disorders, including brain disease, whole-body difficulties, kidney disease, COVID 19, dental status and many more. Out of those diseases, one of the lung diseases, the coronavirus epidemic has spread to virtually every country on the globe, inflicting enormous health, financial, and emotional devastation, as well as the collapse of healthcare systems in some countries. Any automated COVID detection system that allows for fast detection of the COVID-19 infection might be highly beneficial to the healthcare system and people around the world. Molecular or antigen testing along with radiology X-ray imaging and CT scan radiographs are now utilized in clinics to diagnose COVID-19. Nonetheless, due to a spike in coronavirus and the overwhelming workload of doctors, developing an AI-based auto-COVID detection system with high accuracy has become imperative. Given the enigmatic nature of COVID-19 visual markers, lung opacity, and viral pneumonia, diagnosis can be challenging as the three diseases look quite similar in nature on x-ray images. Past studies related to lung disease auto detection were mostly performed on small datasets and the majority of the studies did not reveal the blackbox of models. Furthermore, despite the fact that there are a large number of infected people around the world, the amount of COVID data sets needed to build an AI system is limited and dispersed. Moreover, renal failure, a public health concern, and the scarcity of nephrologists around the globe have necessitated the development of an AI-based system to auto-diagnose kidney diseases. For this thesis, we chose two imaging modalities to study: x-ray and CT scan images. We automated and assessed our established models for lung illnesses (COVID 19, lung opacity, and viral pneumonia). Furthermore, utilizing CT images, we constructed models to automate kidney disease (kidney tumor, cyst, and stone). This research utilized artificial intelligence (AI) to deliver high-accuracy automated COVID-19 detection from normal chest X-ray images. Further, this study was extended to differentiate COVID-19 from normal, lung opacity, and viral pneumonia x-ray pictures. We constructed six models and trained them with x-ray images. Out of those six models, three can be used to categorize normal and COVID diseases (i.e., binary class classification). The other three can be used to categorize normal, COVID-19, lung opacity, and viral pneumonia images (i.e., four-class classification). All models were trained and validated using the transfer learning approach and then tested with unseen x-ray pictures. Each of the binary class models was trained with different input picture resolutions, and it was found that greater input image resolution training contributed to the model's better performance and accuracy. With the employed two-class classification model, the best accuracy, precision, and recall are found to be 97.5%, 99.5% and 99.5%, respectively. The high accuracy of this test can significantly assist in reducing global suffering from COVID-19. Moreover, three models' decisions are verified and compared by visualizing all internal layers, including the final layer's heatmap utilizing GradCam. The best accuracy found for the multiclass model is 93% while testing the model with unseen data. Each of the three multiclass models is explained using explainable AI to uncover the blackbox of the models.In this letter, we also presented a chest CT scan data set for COVID and healthy patients considering a varying range of severity of COVID, which we

published on kaggle and that can assist other researchers to contribute to healthcare AI. We also developed three deep learning approaches for detecting COVID quickly and cheaply from chest CT radiographs. Our three transfer learning-based approaches, Inception v3, Resnet 50, and VGG16, achieve accuracy of 99.8%, 91.3%, and 99.3%, respectively, on unseen data. We delve deeper into the black boxes of those models to demonstrate how our model comes to a certain conclusion, and we found that, despite the low accuracy of the model based on VGG16, it detects the COVID spot in images well, which we believe may further assist doctors in visualizing which regions are affected.This research also deals with the three major renal diseases categories: kidney stones, cysts, and tumors, and gathered and annotated a total of 12,446 CT whole abdomen and urogram images in order to construct an AI-based kidney diseases diagnostic system and contribute to the AI community's research scope. The collected images were exposed to exploratory data analysis, which revealed that the images from all of the classes had the same type of mean color distribution. Furthermore, six machine learning models were built, three of which are based on the state-of-the-art variants of the Vision transformers EANet, CCT, and Swin transformers, while the other three are based on well-known deep learning models Resnet, VGG16, and Inception v3, which were adjusted in the last layers. While the VGG16 and CCT models performed admirably, the swin transformer outperformed all of them in terms of accuracy, with an accuracy of 99.30 percent. The F1 score and precision and recall comparison reveal that the Swin transformer outperforms all other models and that it is the quickest to train. The study also revealed the blackbox of the VGG16, Resnet50, and Inception models, demonstrating that VGG16 is superior than Resnet50 and Inceptionv3 in terms of monitoring the necessary anatomy abnormalities. We believe that the superior accuracy of our Swin transformer-based model and the VGG16-based model can both be useful in diagnosing kidney tumors, cysts, and stones.

Keywords: Kidney disease, Vision transformer, Transfer learning, Explainable AI, CT Imaging, X-ray Imaging, Lung opacity, Viral Pneumonia, COVID-19.

Dedication

This thesis is devoted to all individuals battling for their fundamental needs and survival around the world. The thesis is also dedicated to those who are suffering from disease, and I pray that their suffering will be alleviated.

Acknowledgement

My heartfelt gratitude to my supervisor, **Dr. Md. Golam Rabiul Alam** Associate Professor, Department of Computer Science and Engineering, BRAC University, for his unwavering support throughout this research endeavor.

I'd like to thank my parents and the rest of my family members for their support. I would not be in this position now if it weren't for them.

I'd want to express my gratitude to all of the faculty members, staff, students, and other BRAC University stakeholders for their direct and indirect aid during various events of my education and research.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AI   Artificial Intelligence

AP   Anterior-Posterior

AUC Area Under the Curve

CCT Compact Convolutional Transformer

CNN Convolution Neural Network

CoVs Coronaviruses

CT   Computed tomography

DL   Deep Learning

DNNs Deep Neural Networks

EANet External Attention Transformer

FPNs Feature Pyramid Networks

GAN Generative Adversarial Network

GradCam Gradient Weighted Class Activation Mapping

HRCT High Resolution Computed Tomography

KNN K-nearest neighbor

LN   Layer Normalization

MLP Multilayer Perceptron

MRI Magentic Resonance Imaging

MSA multi-head self-attention

NLP Natural Language Processing

NN   Neural Networks

PA   Posterior-Anterior

$PACS$ Picture archiving and communication system

PCA principal component analysis

PPV Positive Predictive Value

$R-CNN$ Region Based Convolutional Neural Networks

RCC Renal Cell Carcinoma

RDA residual dual-attention

RF Random Forest

ROC Receiver Operating Characteristic

$RT-PCR$ Reverse Transcription Polymearse Chain Reaction

$SwinTransformers$ Shifted Window Transformers

US Ultrasound

WHO World Health Organization

Chapter 1

Introduction

Radiology imaging like magnetic resonance imaging (MRI), computed tomography (CT-scan), X-ray imaging (X-ray), and ultrasound (US) is used to diagnose various kinds of diseases like brain disease, whole abdomen problems, kidney disease, COVID 19, and so much more. In this thesis, we picked up two modalities of imaging, X-ray and CT scan. We automated for lung disease (COVID 19, lungopacity, viral pneumonia) and kidney disease (kidney tumor, cyst, and stone) by using transfer learning and the current state-of-the art transformer model. We revealed the blackbox of the transfer learning models to show the models' performance reliability.

1.1 The Motivation behind Lung Disease Analysis of X-ray Radiographs using AI

The novel coronavirus, taxonomically known as SARS-CoV-2 and labeled by the World Health Organization (WHO) as COVID-19, appeared in Wuhan, Hubei Province, China, before the end of 2019 [62] and has sparked unprecedented global fear. The World Health Organization declared this a public health emergency of worldwide concern and dubbed it a global pandemic due to the virus's fast spread from human to human [63]. The COVID-19 reproductive number varied from 2.24 to 3.58 implies that 100 infected individuals on average may spread the disease to 200 or more persons [66]. As a result, Covid 19 has proliferated globally and is practically dispersed everywhere. Overall, 218,205,951 cases of COVID-19, including 4,526,583 deaths, have been declared by the WHO since 05:05 pm central European summer time (CEST) on 02 September 2021 [93]. Coronaviruses (CoVs) cause respiratory and intestinal diseases in human and animal populations [26]. The period of incubation (i.e. from commencement of symptoms) is between 2 and 14 days, a median of 4 and 5 days. Most persons (i.e., about 80%) who become infected do not have symptoms or mild illnesses. As a result, people who are infected with the virus but are unaware of it can infect others, allowing the infection to spread widely. In situations of age, smoking, or other major medical diseases such as cancer, heart, lungs, kidney, or liver illness, diabetes, immunocompromising conditions, Sickle cell disease, or obesity, the risk of serious disease and death increases with COVID-19 [70], [71]. Additionally, numerous forms of covid are spreading faster and having a negative impact on life and the economy, something humans have never experienced

in recent history[92].

Effective screening of affected patients is crucial in the fight against COVID-19, as it enables those infected to receive prompt treatment and care, as well as isolation to prevent the virus from spreading [64]. The most common clinical testing approach in COVID-19 patients using respiratory specimens for the purpose of testing is reverse transcription polymerase chain reaction (RT-PCR), which is expensive, less-sensitive and requires specialized medical personnel [61]. In addition to RT-PCR, chest X-rays can predict results in high likelihood scenarios in a patient with COVID-19 [91]. The RT-PCR test is paired with the chest XR in clinical practice, providing additional information regarding its severity.On the other hand, a high resolution computed tomography (HRCT) scan is another radiologic option to detect COVID-19, which assists clinicians in identifying the effect of COVID-19 on various organs at different phases of the condition.The management and prognosis of COVID-19 can help with an integrated approach based on first-line CXR and optional use of HRCT [89]. However, there is a scarcity of CT scan facilities and radiologists in remote regions [69], [83]. The norm has become for screening methods of Covid-19 and its severity is X-ray imaging together with RT-PCR, because X-ray imaging is rapid, accessible, and less expensive and less radioactively harmful to the human body than Computed Tomography (CT) imaging technology.

If chest X-rays are utilized to test for a diagnosis, expert radiologists must interpret the pictures. However, considering the cryptic nature of visual indicators of COVID-19, lung opacity, and viral pneumonia, it can be difficult to diagnose[39].

Moreover, doctors are overwhelmed by the rise of covid patients, and their workload has grown considerably [34], [75]. Considering the surge in Covid-19 cases worldwide and the high workload of doctors and healthcare employees, shortages of radiologists, the development of auto-detection systems based on Artificial Intelligence (AI) to detect Covid is a necessity. A number of methods have been reported for the detection of COVID-19 based on chest-X-Rays [86], [87] and different deep learning architectures [80], which are approximately 90% or higher in accuracy, but few solutions must be addressed before it is being used in medical environments. The research project in many papers is on the changes and accuracy of network design, but less attention is paid to explaining models, and most of the models are used to enhance the accuracy of models with small data sets.

1.2 Contribution towards Lung Disease Analysis of X-ray Radiographs using AI

This paper employs three deep learning algorithms for the classificaiton of two classes as well as four classes based on Xception, Resnet50, and VGG19. Explainability is added by using gradient weighted class activation mapping (GradCam), trained and validated on a large dataset considering the back history of publications. The primary contributions and proposals of this work are, in essence, as follows:

- Three CNN models were developed for COVID-19 mass screening (two classes: covid positive & covid negative) from chest x-ray images. Afterward, explainable AI was applied to demystify the black box of models.

- Three additional multi-class models were constructed to diagnose non-covid,

covid, lung opacity and viral pneumonia from chest x-ray radio graphs. Furthermore, explainable AI was used to validate and explain the performance of each generated multi-class model and to demystify the black box of individual CNN layers

- Presented a comprehensive performance study of the proposed binary and multi class systems in terms of the confusion matrix, accuracy, sensitivity, specificity, and F1-score. Additionally, we compared test accuracy for the implemented dual class CNN models for different training input image resolutions and investigated the impact of input image size on the models' accuracy.

1.3 The Motivation behind Lung Disease Analysis of CT Radiographs using AI

Although Xray imaging is less expensive than CT, CT radiographs give more information about the severity of COVID[90].
Several studies have been published in last few months in this domain. However, most of the studies use the COVID Xray data set, as the publicly available Covid CT dataset is sporadically distributed. A few studies have been published with CT datasets, but most of the studies are performed with small datasets and used the transfer learning approach to improve accuracy. However, this publicly available data set is needed to do more research in this domain, as well as require more analysis of the black box of the models that are built to detect Covid from CT radiographs.

1.4 Contribution towards Lung Disease Analysis of X-ray Radiographs using AI

Not only are High Resolution Computed Tomography (HRCT) chest datasets released in this paper, but three deep learning algorithms based on Inception v3, Resnet50, and VGG16 are also employed. Explainability is achieved by utilizing gradient weighted class activation mapping (GradCam). The following are the key contributions and proposals of this research:

- We have collected and annotated 6082 Chest CT images with the HRCT protocol from hospitals to train and validate the proposed model. We made the data set public for reproducibility to enable further research.

- We employed three CNN based deep learning models for COVID detection from chest CT radio-graphs and presented a thorough performance study of the proposed system.

- We adopted explainable AI to assess and demystify the black box of the proposed models to show affected COVID severity.

1.5 The Motivation behind Kidney Disease Analysis of CT Radiographs using AI

Kidney disease is a public health concern since the disease is spreading despite current control attempts. To tackle the rising health issues of Kidney disease, a broad and integrated public health approach is required as a complement to therapeutic measures to control it[6]. Chronic kidney disease affects more than 10% of the world population[7], and it was ranked 16th among the leading causes of death in 2016 and is expected to jump to 5th by 2040[20]. Cyst formation, nephrolithiasis (kidney stone), and renal cell carcinoma (kidney tumor) are the most frequent kidney illnesses that impede kidney function. A kidney cyst is a fluid-filled pocket that forms on the surface of the kidney and is enclosed by a thin wall. Within the kidneys, one or more cysts may develop with water density: From 0 to 20 Hounsfield units [25], [28], [29]. Kidney stone disease is characterized by the formation of crystal concretions within the kidneys, which affects about 12% of the world population[18]. Kidney tumor is abnormal growth in Kidney which is known as Renal cell carcinoma (RCC) is among the 10 most common cancers worldwide[15]. Kidney cysts, tumors, and stones may cause kidney failure and even death due to the severity of the kidney disease, hence the requirement for early detection of kidney abnormalities and generating public awareness among health-care practitioners and the general public[4]. The number of nephrologists, on the other hand, is severely limited. The number of nephrologists per million population varies by region, as shown in Fig. 1.1, indicating a global shortage of nephrologists[88]. In South Asia, there is barely one nephrologist per million people, where in Europe there are 25.3 nephrologists per million people.
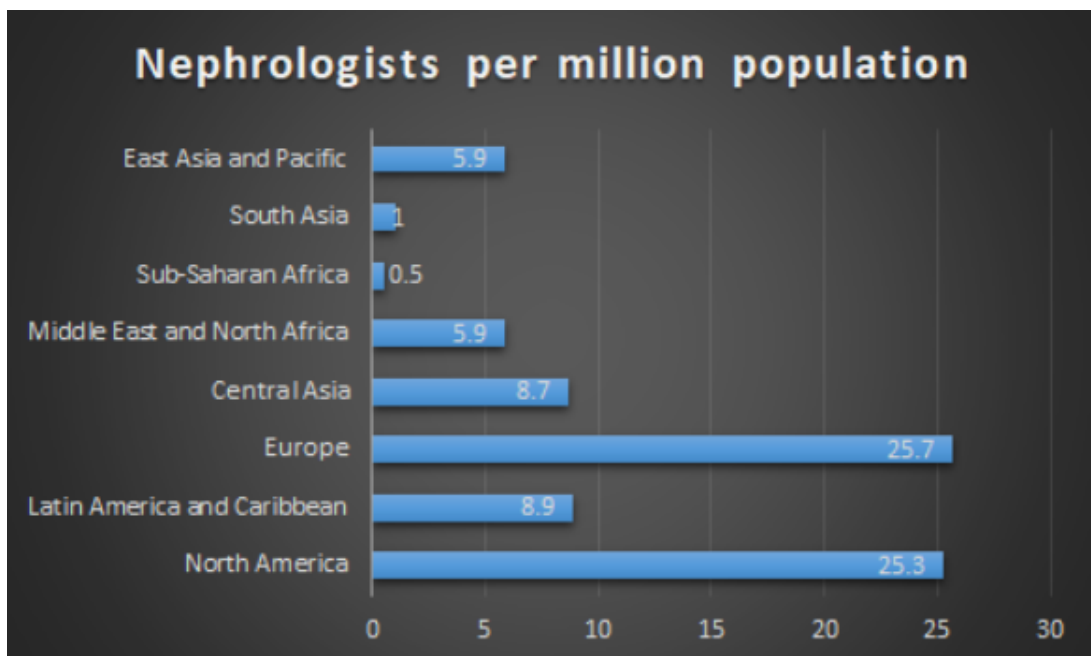
Figure 1.1: No. of Nephrologists per million population around the globe.

X-ray, Computed tomography (CT), B-ultrasound machines (US) and MRI (magnetic resonance imaging) machines are often used in conjunction with pathology tests to check for abnormalities in kidney diseases. The CT machine scans the

desired part of the human anatomy with X-ray beams to obtain a cross-sectional image which provides three-dimensional information about the desired anatomy[2]. CT scans in kidney examinations are ideal for study because they provide three-dimensional information and slice-by-slice images.

Considering the sufferings of the population due to kidney diseases, the shortage of nephrologists around the globe, and the advancement of deep-learning research in vision tasks, it has become imperative to build an AI (artificial intelligence) model to detect kidney radiological findings easily to assist doctors, and reduce the sufferings of people. A few studies have been published in recent years in this domain. However, the publicly available data set is scarce. In addition, most past studies have utilized traditional machine learning algorithms to classify single classes of disease only; either cysts, or either tumors, or either stones. Some studies utilised ultrasound (US) images.

1.6 Contribution towards Kidney Disease Analysis of CT Radiographs using AI

Here in this paper we created and annotated the "CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone" dataset[78], implemented a total of six models, and evaluated each of them to come to the conclusion which model is best suitable to use in realtime. No study that we are aware of has done an analysis based on a transformer model with renal cyst, tumor and stone auto detection. The following are the research's major contributions:

- We collected and annotated a dataset namely "CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone" with 12,446 images utilizing the whole abdomen and the eurogram protocol.

- We employed three CNN based deep learning models (VGG16, Resnet50, and Inception v3 to detect kidney abnormalities and presented a thorough performance study, including explanation of the black-box of the suggested models using gradient weighted class activation mapping (GradCam).

- We employed three recent state-of-the-art Vision transformer variants (EANet, CCT, and Swin transformers) in CT kidney datasets and presented the performance literature of the models using the confusion matrix, accuracy, sensitivity, specificity, and F1 score.

1.7 Organization of the Report

The following is the format of this report: The background study for this project was described in Chapter 2. In Chapter 3, Methodologies are addressed and briefly reviewed. Chapter 4 delves into the findings and result analysis. The thesis' principal conclusion is stated in Chapter 5.

Chapter 2

Literature Review

In this chapter, I'll discuss about some of the works that have direct or indirect impact on our study.

2.1 Existing Works related to Kidney Disease

Because of the advent of Deep Learning and its implementation in image processing and classification, a considerable amount of research has grown in deep learning applications, specifically in autodiagnosis of radiological findings and segmentation tasks. The Deep Neural Networks(DNNs) are cutting-edge deep learning models that are mostly used for image classification and have demonstrated their superiority in the ImageNet datasets[3]. DNNs are employed in MRI brain illness classification, covid detection, bio-informatics, robotics, and many more areas, which has made them popular, particularly when utilizing the transfer learning approach [24]. A variety of different DNN architectures have been implemented for classifying 1000 different categories of images in the ImageNet dataset, including 8-layered DNN AlexNet[5], GoogLeNet[10], 18-layered architecture SqueezeNet [13], ResNet and several variants of ResNet utilizing skip connections [11], and so on. In the classification task that employs a transfer learning technique, inception [10], exception[14], EfficientNet[30] networks have grown in prominence over time.

In recent times, popularly used transformer models for natural language processing are being introduced in vision tasks, which are showing supremacy and good results over other models while doing classification tasks. The Vision transformer (ViT)[42] and several variations of the Vision transformer, like the Big Transformer (BiT)[50], EANet (External Attention Transformer)[74], Compact Convolutional Transformer (CCT)[76], and Swin Transformer (Shifted Window Transformer)[81], are performing better in the imagent benchmark and CIFAR-10[22] dataset.

The Computer Assisted Diagnosis (CAD) system is built to classify tumors from MRI brain images on paper[36], [47], [72] to assist doctors and radiologists with the diagnosis using deep residual networks.Recently, a covid surge made it an urge for researchers to detect COVID 19 using AI, which is very promising as well[38], [60]. Similarly, numerous deep learning methods are employed in research on kidney disease classification as well. The renal ultrasound pictures are enhanced with a median filter, a Gaussian filter, and morphological operations in the article [17], and

then characteristics from the images are retrieved with principal component analysis (PCA) and the K-nearest neighbor (KNN) classifier.The paper [96] evaluated different traditional ML algorithms, such as Decision Trees (DT), Random Forest (RF), Support Vector,Machines (SVM), Multilayer Perceptron (MLP), K-Nearest Neighbor (kNN), Naive Bayes (BernoulliNB), and deep neural networks using CNN and got the highest F1 score of 85.3%. In paper [58], pre-trained DNN models such as ResNet-101, ShuffleNet, and MobileNet-v2 are used to extract features from kidney ultrasound pictures, which are then classified using a support vector machine, with final predictions made using the majority voting technique. This paper used ultrasound images for classification problem and got the highest accuracy of 95.58%. The residual dual-attention module (RDA module) was employed for the segmentation of renal cysts in CT images in the paper [73]. Paper [32] integrates the features of using conventional and deep transfer learning techniques, and finally, features are used by the SVM Classifier to classify normal and abnormal images using US images. In paper [27] two CNN models are used consecutively, where the first CNN is used to identify the urinary tract, and the second CNN is used to detect the presence of stone and got 95% accuracy. An automated detection of kidney stones (having/not having stone) in paper [94] using coronal computed tomography (CT) images and a deep learning (DL) technique yielded a detection accuracy of 96.82 percent. this paper used 1799 images total to train and validate the model and classified one class only. The paper [31] proposed two morphology convolution layers, modified feature pyramid networks (FPNs) in the faster RCNN and combined four IOU threshold cascade RCNNs to detect kidney lesions and got an area under the curve (AUC) value of 0.871. The kidney cyst image detection system for abdominal CT scan images using a fully connected convolution neural network was developed in the paper [19] and they got a true-positive rate of 84.3%.

In summary, the efforts utilizing machine learning and deep learning approaches to classify a few kidney radiological findings have provided promising results, but the majority of the tasks are performed on xray or ultrasound images, and the accuracy is around 95%. A few approaches were there with CT scan images only with dual class classification. Considering the scarcity of data and the above findings of research articles, we created a database of kidney stone, cyst and tumor CT images. We implemented three deep learning techniques (VGG16, Inceptionv3 and Resnet50) to classify four classes of kidney disease and demistified the blackbox of the models to show why our model came to a certain conclusion about a class. We also implemented the latest state-of-the-art innovations in vision learning (EANet, CCT, and Swin transformer algorithms) to classify the four classes and have shown that our model has promising accuracy which can reduce the suffering of the world population through early diagnosis of diseases.

2.2 Existing Works related to Lung Disease with X-ray images

Due to global demand and the advent of Artificial Intelligence (AI), a considerable body of research has evolved to apply Artificial Intelligence (AI) to diagnose different respiratory disorders, particularly those directly related to covid, utilizing basic

XR and CT images.

However, the studies in many articles focused on the modifications and accuracy of network design, but less on explaining models, and most models are used to improve the accuracy of models with small data sets. L. Wang, Z. Lin, and A. Wong proposed COVID-Net, a deep CNN [60], to detect COVID-19 instances from roughly 14k chest X-ray images, but the achieved accuracy was 83.5 percent. Khan et al. [49] used the transfer learning method on 310 cases of normal pneumonia, 330 cases of bacterial pneumonia, 327 cases of viral pneumonia, and 284 COVID-19 pneumonia photos and got 89.5% accuracy. However, as the models employed in this paper used a small number of images, they required detailed analysis. Explainable AI was not implemented to show model efficacy. Y. Oh, S. Park and J. Ye trained the Resnet 18 model and showed explainability [54], but the paper used a low number of images in training and testing and got an accuracy of 89%. Explainability in each layer of the model is desired. The Darknet model was developed and trained by T. Ozturk, M. Talo, E. Yildirim, U. Baloglu, O. Yildirim and U. Rajendra Acharya [55] with 127 covid images only, and got 87% accuracy, but model explainability was not explored. A. Altan and S. Karasu [35] trained EfficientNet B and got a promising 99% accuracy, but it was trained with only 219 covid images, and Explainable AI was not implemented. J. Civit-Masot, F. Luna-Perejón, M. Domínguez Morales and A. Civit [41] trained Vgg16 with 132 Covid images and got an averageF1 (avgF1) score of 85%, which is not promising in a real clinical scenario due to a low F1 score. Inception v3, InceptionResNetv2, Resnet50 were trained by A. Narin, C. Kaya and Z. Pamuk [82] and got a good accuracy of 98%, which is promising, but those models were trained with 50 normal and 50 covid images, and the explainability of the model was not depicted. Given the scarcity of COVID-19 images, some approaches have concentrated on generating artificial data in order to train stronger models. An auxiliary Generative Adversarial Network (GAN) was employed to generate artificial images [59]. The results showed that data augmentation boosted accuracy on the VGG16 net from 85% to 95% [52]. J. Arias-Londono, J. Gomez-Garcia, L. Moro-Velazquez and J. Godino-Llorente trained the covid-net model with large et of images but got 91.53% accuracy [37]. Before a model can be utilized in clinics, it must undergo extensive analysis. Summary results from different letters are given in Table 2.1.

In summary, numerous recent efforts to transfer learning approaches for detecting COVID-19 from a small dataset have been reported with promising results. Nevertheless, the process required verification on a large dataset. A few models have been developed that have low accuracy and need to be improved. Additionally, the majority of the research does not use explainable AI and does not demonstrate how the model diagnoses Covid-19 from photos. Although a few studies show Explainable AI, they are limited to the final layer of the model only.

In order to address the research's recurrent issues, we built three transfer learning-based Neural Networks in this study in order to categorize Covid-19 utilizing bigger X-ray data sets, exhibiting promising accuracy in an unseen dataset. Likewise, we leveraged Explainable AI to demistify the blackbox of the stated three models and illustrate how the system identifies covid from x-ray images across all blocks and layers by combining heatmap and original images. Our explainability not only serves

Table 2.1: Summary of the papers in the research field

Reference Number	Number of images			Architecture	performance matrix	Explainable
	Covid	Normal	Others			
[52]	69	79	158	AlexNet, Google Net, Resnet18	Acc=99%	NO
[41]	132	132	132	Vgg16	avF1=0.85	NO
[35]	219	1341	1345	EfficientNetB	Acc=99%	NO
[46]	25	50	-	VGG19, DenseNet	AvF1=0.90	NO
[82]	50	50	-	Inception v3, InceptionResNetv2, Resnet50	Acc=98%	NO
[37]	7716	45022	21707	CovidNet	Acc=91.53%	YES
[54]	180	191	131	Resnet18	Acc=89%	YES
[49]	310	284	657	CoroNet	Acc=89.5%	NO
[55]	127	500	600	DarkNet	Acc=87%	NO
[95]	100	1431	-	EfficientNet	se=96%,sp=70%	NO
[43]	48	11203	1591	Inception Resnetv2	Acc=92.2%	NO

as a responsible and transparent audit of our models, but also may assist physicians in improving the screening of Covid 19. Moreover, we trained all three models using images with different input resolutions and found that increasing the image resolution during training improves accuracy in the test condition.Additionally, to address the problems associated with diagnosing patients with Viral Pneumonia and Lung Opacity, we constructed a multiclass model (four classes), validated its accuracy, and explained it using Explainable AI.

2.3 Existing Works related to Lung Disease with CT images

Because of the advent of Deep Learning and the rapid spread of coronavirus, a considerable body of research has evolved to detect covid using basic X-ray and CT images. However, the dataset, which is the major feed to the CNN, is sparse[54], [79]. Due to the shortage of datasets, several investigations in many papers concentrated on network architecture alterations and accuracy. Some articles attempted to incorporate all publicly available data sources and obtained intermediate and praised accuracy. Some of the models that provided good accuracy require demistyfyig blackbox to gain insight into how the model predicts certain classes.

Covid-net was proposed at paper [60] but got an accuracy of 83.5 %.The author in [82], attempted to apply the transfer learning based approach using Renset 50, Inception v3, and Inception Resnet v2 models, and obtained an accuracy of 98% using Resnet50. In paper [48] the athor used a Densenet201 based transfer learning approach and got 96% praisable accuracy. However, models' explainability was not provided. Paper [33] constructs covid caps: A Capsule Network-based Framework for COVID-19 identification from x-ray images which gives an accuracy of 95.7% requires explaibaility. In paper [59], utilized artificial ways to generate new images by utilizing the Generative Adversarial Network (GAN) network and then trained the model. The Darknet model was developed and trained on paper [55] with 127 covid images only, and got 87% accuracy. The accuracy of paper [65] was 86.6 percent using 219 covid imaging data and 175 healthy data.It identifies the following three types of chest CT input image slices: COVID-19, influenza-A viral pneumonia (IAVP), and cases unrelated to infection The anticipated class of an image is the one to which the image has the highest probability of belonging. This AI system uses two different deep learning algorithms: a three-dimensional (3D) CNN segmentation model for lung segmentation and a ResNet-based model for picture classification[65]. The remaining 90 scans (30 COVID-19 patients plus 30 IAVP cases plus 30 healthy instances) were used as a testing dataset, while the imaging dataset of 528 (189 COVID-19 cases plus 194 IAVP cases plus 145 healthy cases) was used for training and validation. On the testing dataset, the AI system had an overall accuracy rate of 86.7 percent, with a f1-score of 83.9 percent for COVID-19 patients, 84.7 percent for IAVP cases, 91.5 percent for healthy cases, and a f1-score of 83.9 percent for COVID-19 cases[68].Paper [67] used 313 covid positive image data and 329 healthy chest CT data to classify and got an accuracy of 90.8%.Paper [84] used 224 covid and 504 healthy images of xray radiographs and got 96% accuracy.

In summary, recent efforts utilizing transfer learning approaches have provided promising results but require testing on large datasets and require explanation of the blackbox of the models.

We published a dataset of a total of 6082 images, proposed three CNN-based networks, and showed the model gives promising accuracy and provides the explainability of the models utilizing gradcam. We believe that our models' high accuracy based on CT images can assist medical physicians in not only identifying covid and normal individuals, but also in determining the severity of diseases.

Chapter 3

Methodology

3.1 Methodology behind Lung Disease Analysis of X-ray Radiographs using AI

In this section, the methodology is depicted in the following order; accumulating data to train the Neural Network, image preprocessing, the experiments and training the Neural Network and evaluation processes.

3.1.1 Dataset

The study was created utilizing an XR image dataset comprising the PA and AP views (Posterior-Anterior (PA) and Anterior-Posterior (AP) views. This study looked at two different pathological conditions: normal and Covid-19. The majority of the data for this investigation came from Kaggle's covid19-radiography-database [21], [23], [40], [44], [45], [51], [53], [85]. The covid19-radiography-database was developed by combining different databases [56], [57]. For example, images from the COVID-19 DATABASE of the Italian Society of Medical and Interventional Radiology (SIRM) [57], the Novel Corona Virus 2019 Dataset by Joseph Paul Cohen and Paul Morrison, and Lan Dao [56], 43 articles (radiography metadata contains references), Normal and Viral pneumonia images from the Chest X-Ray Images (pneumonia) database. The database includes 3616 COVID-19 positive cases along with 10,192 normal circumstances, 6012 Lung Opacity (Non-COVID lung infection), and 1345 Viral Pneumonia images. Fig. 3.1 shows sample image data for covid and normal class images. Fig. 3.2 shows sample images used for multiclass classification.

3.1.2 Image Processing

The X-ray images with Posterior-Anterior (PA) view and Anterior-Posterior (AP) view were scaled before being fed into the Neural Networks.We developed six models in total, three of which are utilized to classify normal and covid pictures. The remaining three models are utilized for multiclass categorization in order to determine whether normal, covid, lung opacity, and viral pneumonia are present in the image. We carried out two investigations in this paper for binary classification. For the first investigation into the Resnet50 and VGG19 models, images were resized to 224 by 224 pixels and normalized. Images are resized to 299 by 299 for the Xception model as per the standard requirement. For our second experiment, we scaled all of the
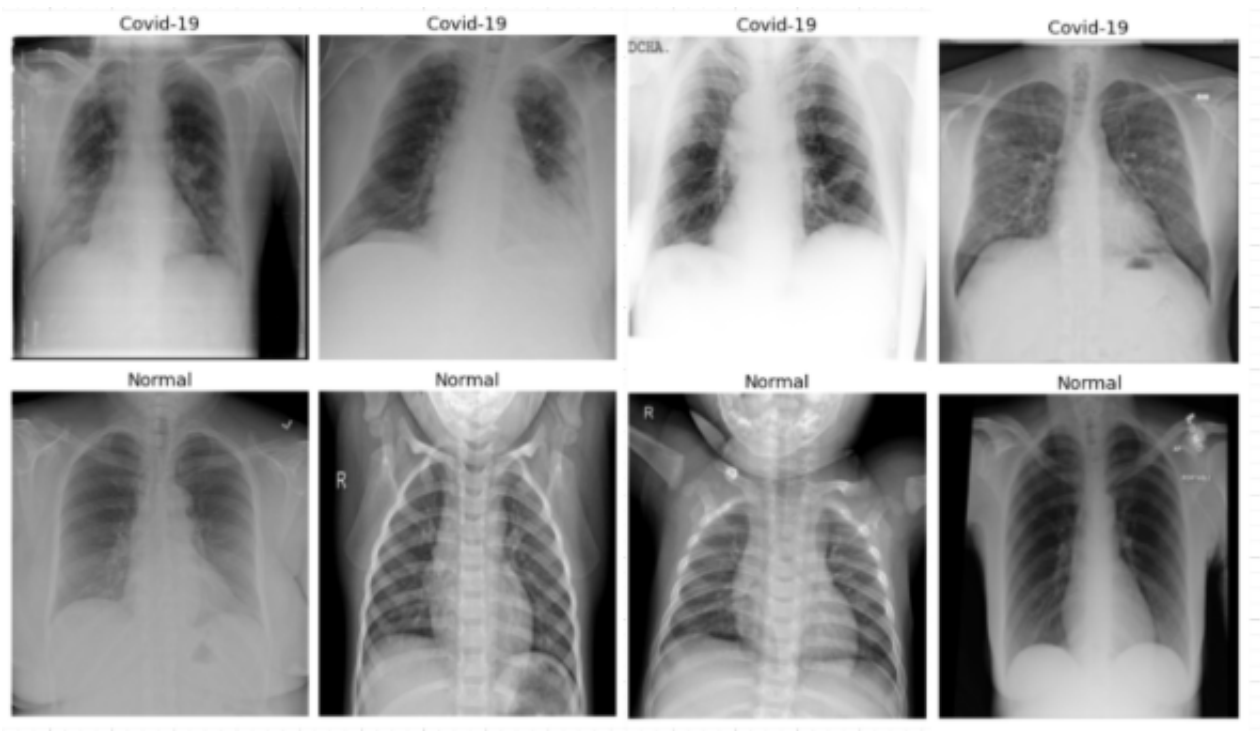
Figure 3.1: Sample image data(X-ray) for covid and normal class images for two class classification.
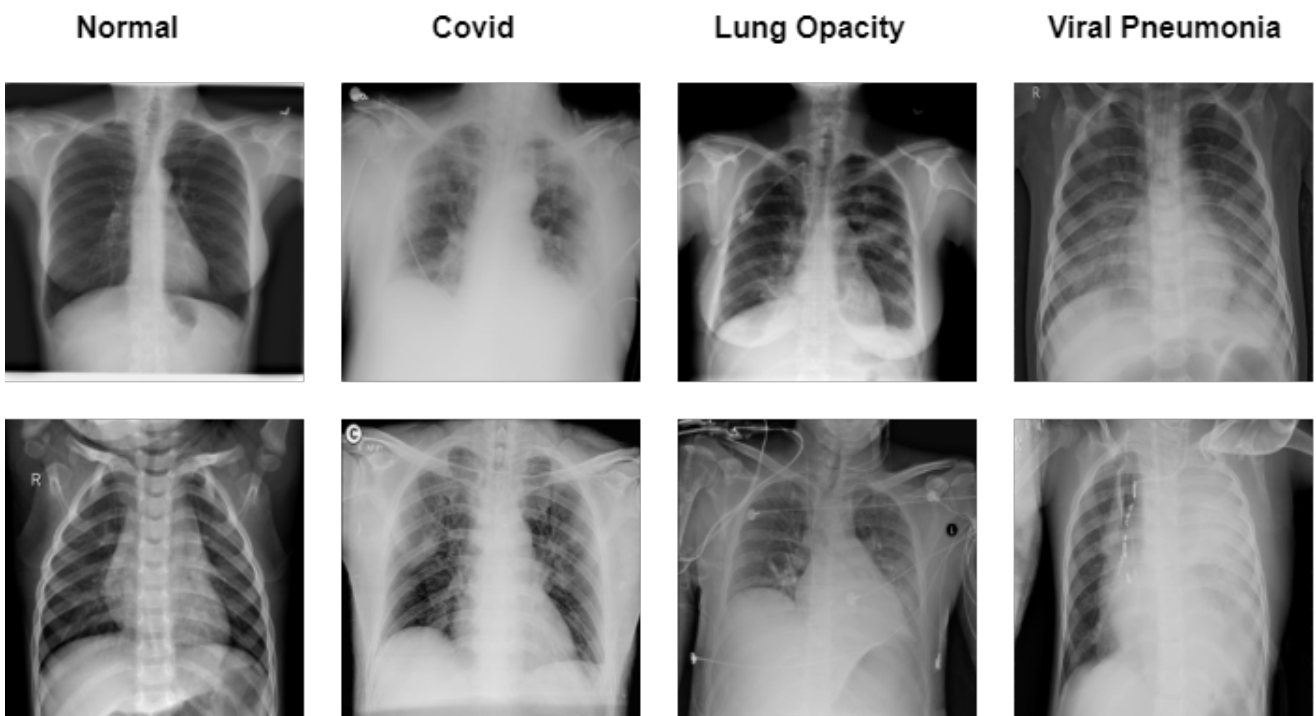

Figure 3.2: Sample image data (X-ray) for multi class classification to detect normal, covid, lung opacity and viral pneumonia.

pictures to 512×512 pixels for the three models we developed for two class classifications. The investigation chose 1000 images from positive and 1000 images from non-covid cases randomly from 3616 COVID-19 positive photographs and 10,192 normal-covid images from the database. The trials were assessed with a randomized 80/20 split for training and test splits, with 20% of the training data used as a validation set to avoid overfitting. For the second setup models for multiclass, we used 1000 images each of covid, normal, lung opacity, and viral Pneumonia instances. The images were resized as per models' standard requirements, i.e. for Resnet50 and VGG19 input images were resized to 224 by 224 pixels and normalized. For the Xception model, input images were resized to 299 by 299 pixels.

The dataset was also adjusted using Z-normalization. In machine learning methods, standardization helps to stabilize the model while also increasing the training pace. The following (3.1) applies to z normalization.

$$\hat{X} = \frac{X[:i] - \mu_i}{\sigma_i} \tag{3.1}$$

Here, μ_i is the mean value and σ_i is the standard deviation of the feature.

In this letter, we trained, validated, and tested the collected chest X-ray data in six different Convolution Neural Networks. The Xception, Resnet50, and Vgg19 models for both binary class and multiclass were adjusted slightly in the last two layers to achieve regularization.

3.1.3 Neural Network Models

For both the binary class and multiclass set up, three different Convolution Neural Networks were employed to train, validate, and test the accumulated chest X-ray data. The Xception, Resnet50, and Vgg19 models were used to train the dataset on Google Colab Pro edition with 26.3 GB of Gen RAM and 16160MB of GPU RAM. The CNN models were tweaked slightly in the last layers to accomplish regularization, and the pre-trained weight "image-net" was employed to aid the model's learning process. A weighted categorical loss function is applied to compensate for the class imbalance problem in the training data. Models are constructed with the Adam optimizer with default parameters to achieve computational efficiency and adaptive learning rate. Early stopping is employed to monitor validation loss and stop training the model when validation loss is stagnant with a patience parameter of 15. The model's last step training was fine-tuned using a factor of 0.25 and patience 15 when the validation loss reached its plateau. If the model's performance does not improve after 15 patience, the training is halted.

The final receiver operating characteristic (ROC) curve, confusion matrix, and evaluation matrices were obtained after training the models using a mini-batch size of 16 images and 200 epochs. Furthermore, heatmaps are created in different layers using the GradCam Technique. The original image and heatmap are superimposed to locate the crucial locations in the image to anticipate the pathological condition and model interpretability. Fig. 3.3 illustrates the overall system diagram used in this investigation. For two class classification only normal and covid image is detected whereas for multi class problem Normal,Covid, Lung Opacity and Viral Pneumonia is detected

Vgg19

In the Image Net Large Scale Visual Recognition Challenge (ILSVRC) in 2014, VGGNet [8] was a neural network that did exceptionally well. It came in first place for image localization and second place for image classification. Fig. 3.4 shows VGG19 in our classification experiment, which consists of 16 convolution layers, 5 Max-pooling layers, one average pooling, flattening, and a dense layer. VGG CNN has six major blocks which make the network deep, utilizing small 3x3 filters, with a stride of 1, and the same padding for the Conv Layers and 2x2 filters with a stride of 2 for the Maxpooling/Downsapling layers. In our tailored Vgg19 network, we froze the weight up to the max-pooling layer of the conv5 block, and then we added one average pooling, flattening, and dense layer to finetune the model.Fig. 3.4 shows VGG19 in our classification experiments for lung disease diagnosis. For the four class problems, Normal, Covid, Lung Opacity, and Viral Pneumonia were detected, whereas for two classes, only covid and normal images were detected.
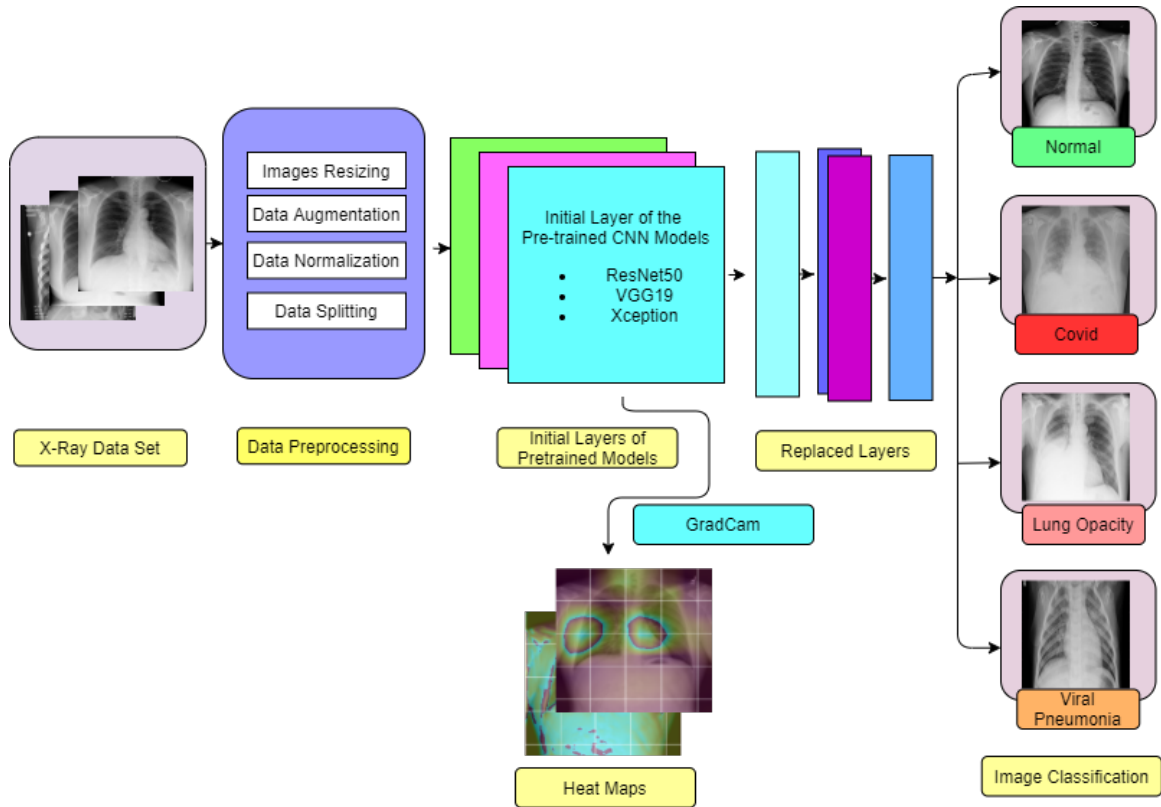
Figure 3.3: Complete block diagram of experiments for lung disease diagnosis.

Resnet50

ResNet [11] has emerged as a ground-breaking deep neural network (DNN) model for computer vision problems. It made its debut in 2015 when it won the ImageNet [3] competition.

In most situations, DNNs outperform neural networks (NN) with fewer layers. However, training a massively stacked NN is notorious for its vanishing gradient issue, which causes model performance to deteriorate. Identity shortcut connections or skip connections that bypass one or more layers – have been used to overcome this

Figure 3.4: VGG19 in our classification experiments for lung disease diagnosis.

issue in ResNet. The use of residual blocks, which consist of the main path and identify shortcut links, helps to solve the vanishing gradient problem, which is shown in Fig. 3.5.

The primary path consists of a sequence of Neural Networks, whereas the second path, known as skip connections, is a straight path from input to output. Skip links connect to the outp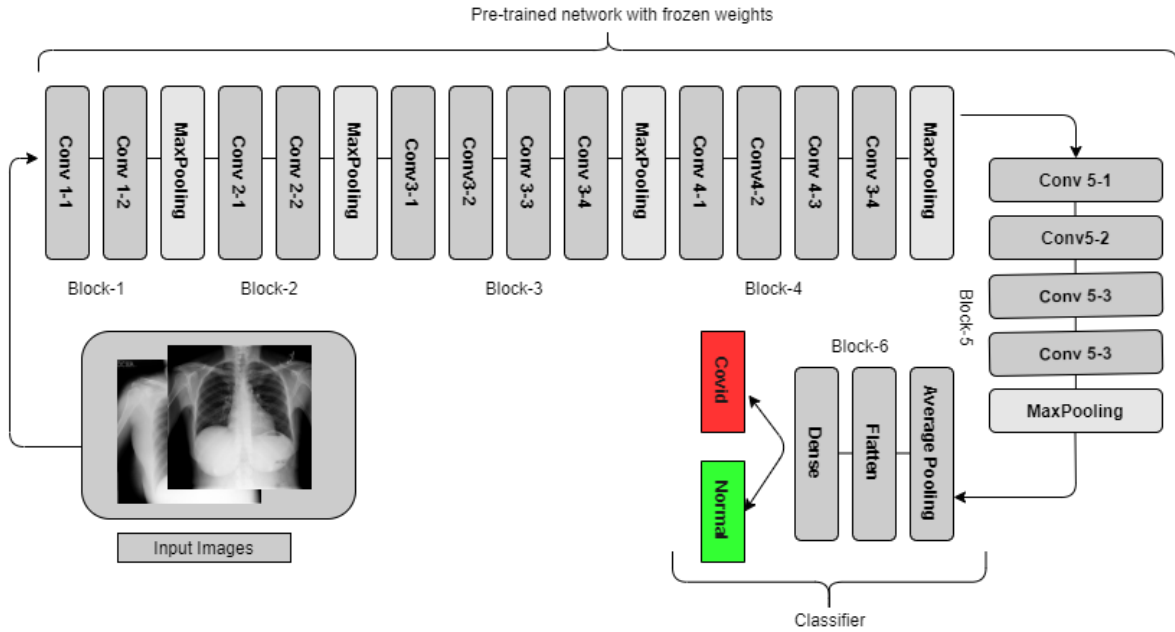ut of the network directly and the network provides output as $F(X)+X$. The Skip connection addresses the gradient vanishing problem associated with conventional neural networks.

We used the rasenet50 version of Resnet and tweaked it somewhat in the final stages. We utilized the weight of the image-net to create a classification detector for covid and normal instances by freezing the weight of the top layers.Our proposed structure for the ResNet-50, which is used to categorize chest X-rays, is shown in Fig. 3.6.

On top of the pre-trained model, more layers are added. To classify images, an average pooling layer with a pool size of (4,4), a flattening layer, a dense layer with Relu activation, a dropout layer with a dropout probability of 50% drops 50% of the parameters randomly and reduces overfitting, and finally, a dense layer is used.

Xception

Xception [14], a modified version of Inception-v3, is a Depthwise Separable Convolutions-based deep neural network architecture which Google researchers developed. Depthwise Convolution is a convolution in which each input channel is subjected to a single convolutional filter. Pointwise convolution is a kind of convolution that employs a 1x1 kernel, which iterates across each point. This kernel has a depth equal to the number of channels in the input picture. A pointwise convolution is used with depthwise convolutions to create depthwise-separable convolutions.

A pointwise convolution follows the depthwise convolution in the original depthwise separable convolution, and a depthwise convolution follows the pointwise convolution in the modified depthwise separable convolution. The Xception model used modified
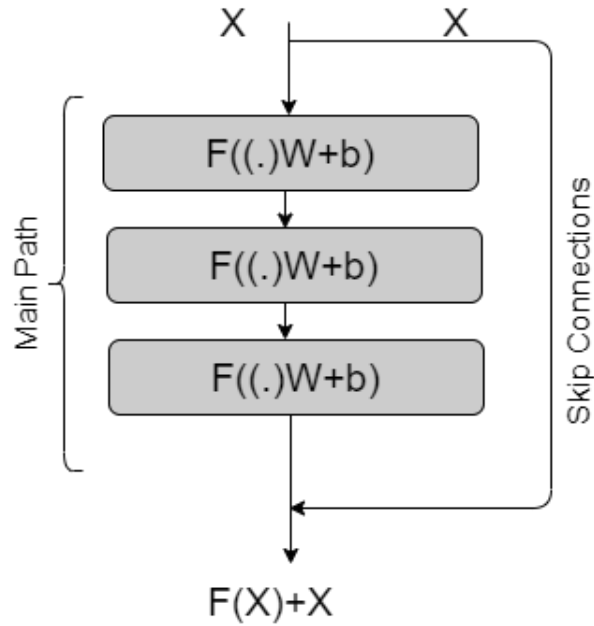
Figure 3.5: One Resblock in resnet50 with main and skip connection.

separable convolutions. Our tweaked Xception model to classify covid and normal images is depicted in Fig. 3.7.

As in the figure, Separable Convolutions are the modified depthwise separable convolutions, and there are residual connections in the middle flow. In the Xception model, the data initially passes via the input flow, then eight times through the middle flow, and lastly through the exit flow. We used pre-trained "image-net" weight for the system, and a few of the last layers of the model were modified to fine-tune and gain regularization. Average Pooling layer with a pool size of 4 by 4, a flattening layer, and a dense layer is added to classify images. For the four classes, similar architecture is applied and the final dense layer has an output size of four to detect normal, covid, Lung Opacity and Viral Pneumonia images.

3.1.4   The Experiments

Our tweaked and proposed three models were utilized in this investigation for two separate experiment setups.

Two class Setup

In two class experiment setup we utilized all the three proposed model utilizing last dense layer with an output size of 2. we did two investigation with this experiment setup.

In the first investigation, models were trained with standard image requirements as per Xception, Vgg19 and Resnet models. The standard inut size requirements for Xception, Vgg19 and Resnet are 299×299, 224×224 and 224×224 pixels respectively, for each of the three algorithms. In the second investigation, we first resized the input photos to 512×512 pixels, and then we trained all of the proposed models using the resized input images.

For both the investigation, model performances are evaluated using unseen data.
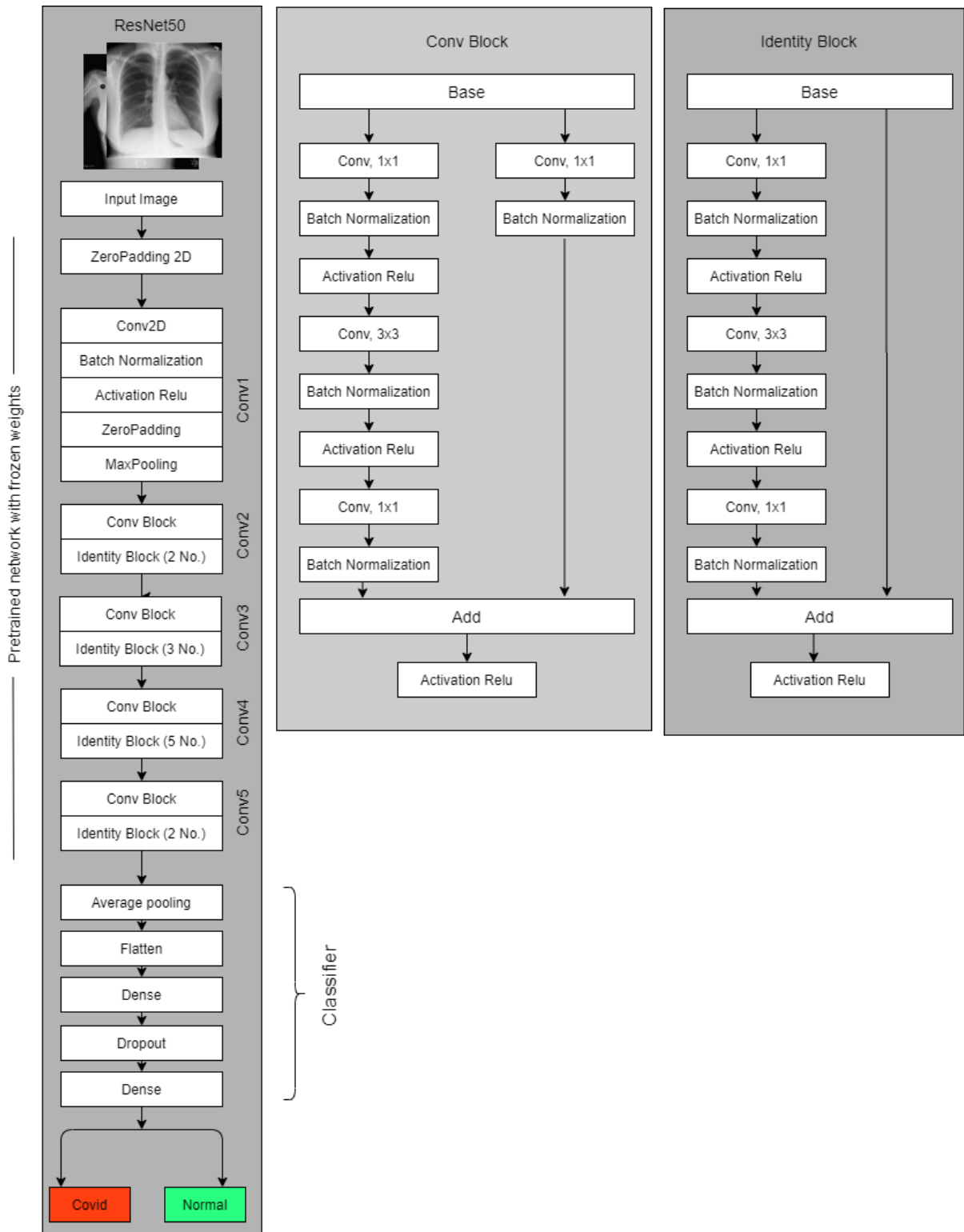
Figure 3.6: Modified Resnet50 to Classify lung disease.

The performance of both experimental setup with three proposed model is evaluated using data that has not been observed by the model while training.
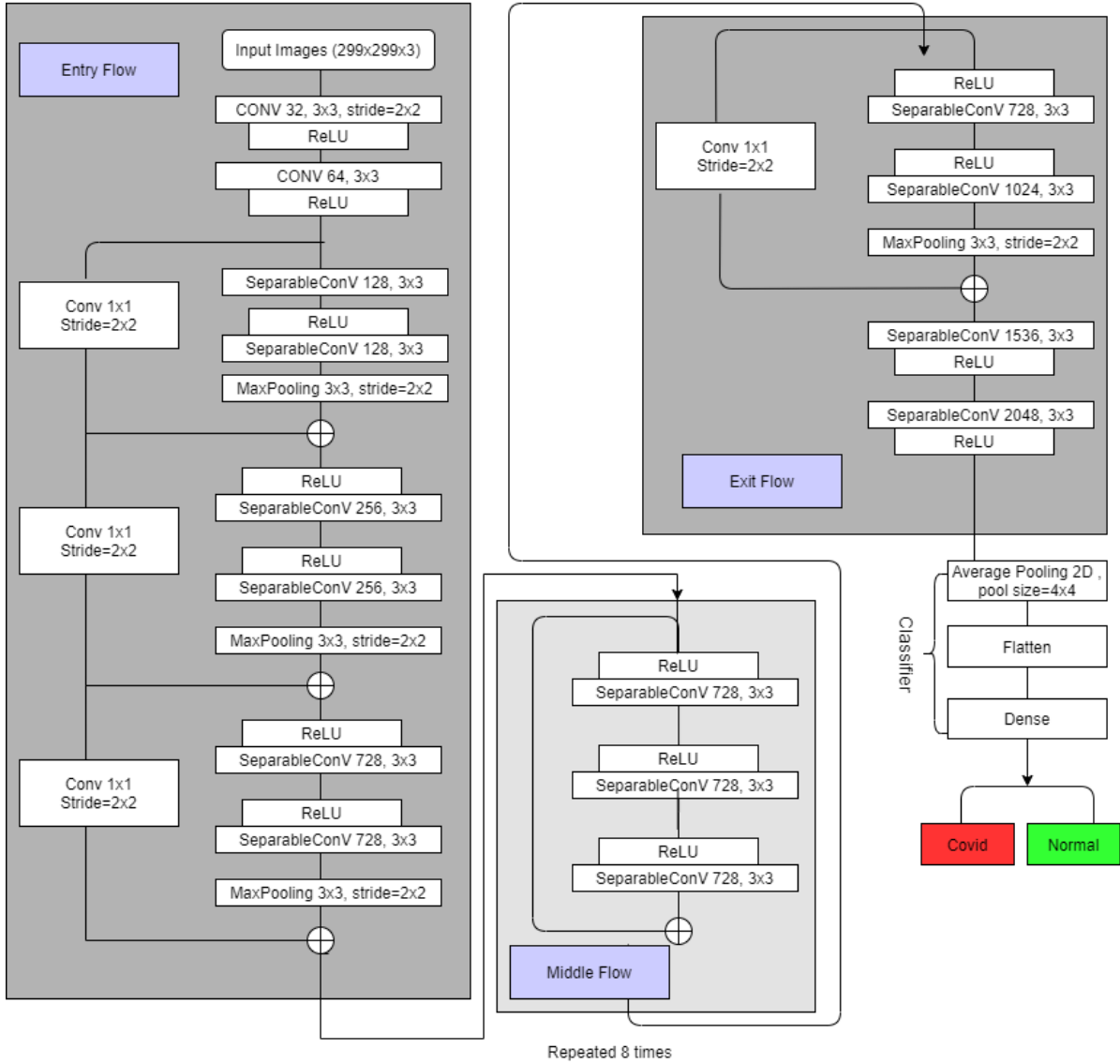
Figure 3.7: Modified Xception Network to Classify Lung Disease.

Multi/Four class Setup

In multi class experiment setup we used all the three proposed model utilizing last dense layer with an output size of 4. The Xception, Vgg19, and Resnet models were trained in the field of standard picture needs. The typical criteria for inut size are 299×299 for Xception, 224×224 for Vgg19 and Resnet accordingly.

3.1.5 Evaluating Model performances and deep layer feature investigation

We evaluated accuracy, sensitivity or recall, specificity, precision known as Positive Predictive Value (PPV), and F1 score to evaluate the performance of three deep

learning algorithms for categorizing X-ray pictures as per (3.2), (3.3), (3.4), (3.5), and (3.6). Precision[1] is defined as the ratio of correct positive identifications relative to all positive identifications. A low precision will result in a significant number of false positives, meaning patients will be mistakenly categorized as having a certain condition. The recall[1] is the number of true positives over the number of true positives plus the number of false negatives. Recall takes the false negative rate into consideration. Incorrect diagnosis may occur due to low recall rate. The F1-score(3.6) is the harmonic mean of precision and recall. A high F1 score is desired in the classification task.

$$\text{Accuracy}_i = \frac{\text{TP}_i + \text{TN}_i}{\text{TP}_i + \text{TN}_i + \text{FP}_i + \text{FN}_i} \tag{3.2}$$

$$\text{Precision}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \tag{3.3}$$

$$\text{Sensitivity}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \tag{3.4}$$

$$\text{Specificity}_i = \frac{\text{TN}_i}{\text{TN}_i + \text{FP}_i} \tag{3.5}$$

$$\text{F1\_score}_i = 2 \times \frac{\text{Precision}_i \times \text{Sensitivity}_i}{\text{Precision}_i + \text{Sensitivity}_i} \tag{3.6}$$

Where,

- i=Covid and Normal for classification problem.

- TP= True Positive

- FN= False Negative.

- TN=True Negative

This paper used the gradient weighted Class Activation Mapping (GradCAM)[16] algorithm to make models more transparent by visualizing the input areas crucial for model predictions and demystifying the model performances by visualizing heatmap in three different CNN networks layers. Fig. 3.8 describes complete process for Gradcam analysis in our paper for covid and non-covid images. For the four classes, the same architecture is applied except the output class is Covid, Normal, Lung Opacity, and Viral Pneumonia. By visualizing heatmap, it is often possible to depict why the model concluded a particular class. First, we passed an image through the model to generate a prediction, and from the prediction value, we developed the class prediction of the image. After that, we computed the gradient of the class with respect to Feature Map activation A^k

$$A^k = \frac{\partial y^c}{\partial A_{ij}^k} \tag{3.7}$$

To derive the neuron significance weights, these gradients flowing back are global-average-pooled across the width and height dimensions (indexed by i and j, respectively).

$$w_k^c = \frac{1}{Z} \sum_j \sum_i \frac{\partial y^c}{\partial A_{ij}^k} \qquad (3.8)$$

Then, using the equation x, we compute the Grad Cam.

$$L_{\mathrm{Grad_C AM}}^c = ReLU \left(\sum_k w_k^c A^k \right) \qquad (3.9)$$

We created visualization by superimposing the original image with the heatmap. For the binary class, this letter examined the GradCam visualization for all the Xception model layers and compared heatmap across all the models to find better efficacy and explain which part of the features affecting model to decide a particular class. First, we analyzed different layers and different blocks of a model to find which regions affect categorisation using GradCam. Second, we compared the last convolution layer of activation maps of three CNN models to demystify model performances. For the multiclass classification we analyzed last CNN layers heatmap for the three models and concluded with comparing the same.
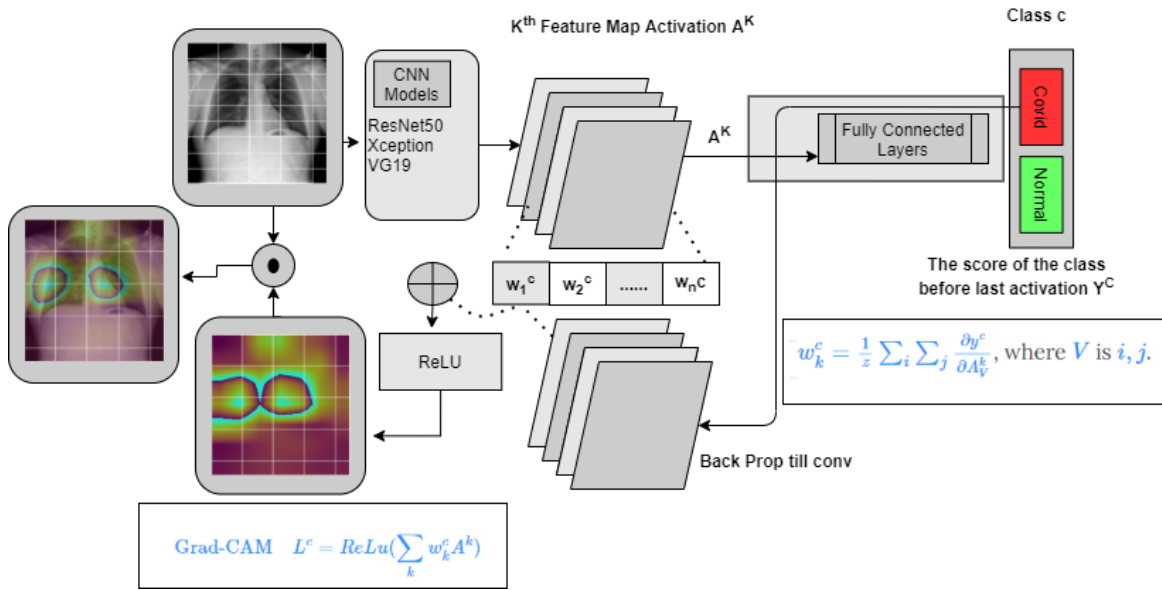
Figure 3.8: The complete process for Gradcam analysis for lung disease diagnosis.

3.2 Methodology behind Lung Disease Analysis of CT Radiographs using AI

Fig. 3.13 illustrates the overall system diagram used in this investigation. First we created a data set and then, from the dataset, we went through data preprocessing steps which included image resizing, data augmentation, data normalization, and data splitting from the train, test, and validation sets. After that, we trained the model with our modified Inception v3, Resnet 50, and VGG16 models where all the layers except the last 3 were frozen and we added a few extra layers on top of them.

The model was tested with unseen data. Gradcam analysis was used to depict the models' explainability.

3.2.1 Data Collection Procedure for the CT COVID Dataset named as "HRCT Chest Covid Data -CT SCAN"

The data was collected from different hospitals in Dhaka, Bangladesh and verified by the radiologist.

- With the help of medical technologists for CT and MRI, we first visited different hospitals and separated the Covid and non-Covid CT Dicom image data from the CT Scan machine.

- All the images with the HRCT chest protocol were selected for COVID and non-COVID patient images. The HRCT protocol uses the same 1.5 mm slice thickness and 1.5 mm increment.

- From the entire set of images of COVID data, we created a batch of images of the affected COVID lung region, and from the batch of images of each patient, we excluded the meta data and patient information.

- Following that, we converted the images to lossless jpg format so that we wouldn't lose any necessary information.

- In a similar way, we created batches of images for Healthy chest scan data and after removing patient information and meta data we converted the dicom images to lossless jpg format

- We created a data.csv file Fig. 3.9 which contains image id, image path, and classes. The dataset contains 63% healthy data and 37% COVID data.

- Finally, a radiologist reviewed the radiograph imaging data.

3.2.2 DataSet Description

Even though a large number of patients are affected worldwide, HRCT chest scan data sets are scattered and limited. To address this, we created an HRCT Chest Covid Data -CT SCAN[77] and made it publicly available so that other researchers could benefit from it.

All images were taken using the same protocol with a thickness of 1.5 mm and a 1.5 mm increment from different hospitals in Bangladesh. Initially, the dicom study was meticulously selected, anomized, and annotations were deactivated. Following diagnosis, a unique batch of dicom studies was generated for each patient who had COVID symptoms. In a similar way, a few batches of dicom images were generated from the HRCT chest scan for healthy patients. Following that, the dicom studies were converted to Jpeg format without shrinking the pictures to prevent losing research information. Our dataset contains a total of 6082 images, out of which 3840 non-Covid pictures and 2242 Covid images, and the images' diagnosis was verified by a panel of physicians. Fig. 3.10 depicts a selection of pictures from our dataset[77] .

Figure 3.9: Data.csv file for Dataset named "HRCT Chest Covid Data -CT SCAN".

A different category of affected covid data is chosen for our dataset, from mild to severe, which can be found through the mean and standard distribution of colours of the images for covid and normal class, shown in the Fig. 3.11 and Fig. 3.12

3.2.3   Image Processing

The CT images were scaled before to being input into the Neural Networks in accordance with the models' standard specifications. Images were reduced to 244 by 244 pixels and normalized for the initial investigation of the Resnet50 and VGG16 models. Images are downsized to 299 by 299 pixels for the Inception v3 model in accordance with industry standards. The investigation randomly selected 2000 photos from our database of positive and non-covid instances. Trials were evaluated using a randomized 80/20 split for training and test data, with 20% of training data used as a validation set to avoid overfitting. Additionally, the dataset is normalized using Z-normalization, which aids in model stabilization while increasing training speed. The following (3.10) applies to z normalization:

$$\hat{X} = \frac{X[:i] - \mu_i}{\sigma_i} \tag{3.10}$$

Here, μ_i is the mean and σ_i is the standard deviation value of the feature.

3.2.4   Neural Network Models

The dataset was trained using the Inception v3, Resnet50, and Vgg16 models on Google Colab Pro version with 26.3 GB of Gen RAM and 16160MB of GPU(Persistence-M) RAM. The CNN models were slightly altered in the last layers to achieve regularization, and the model's learning process was aided by the pre-trained weight "image-net"[3].
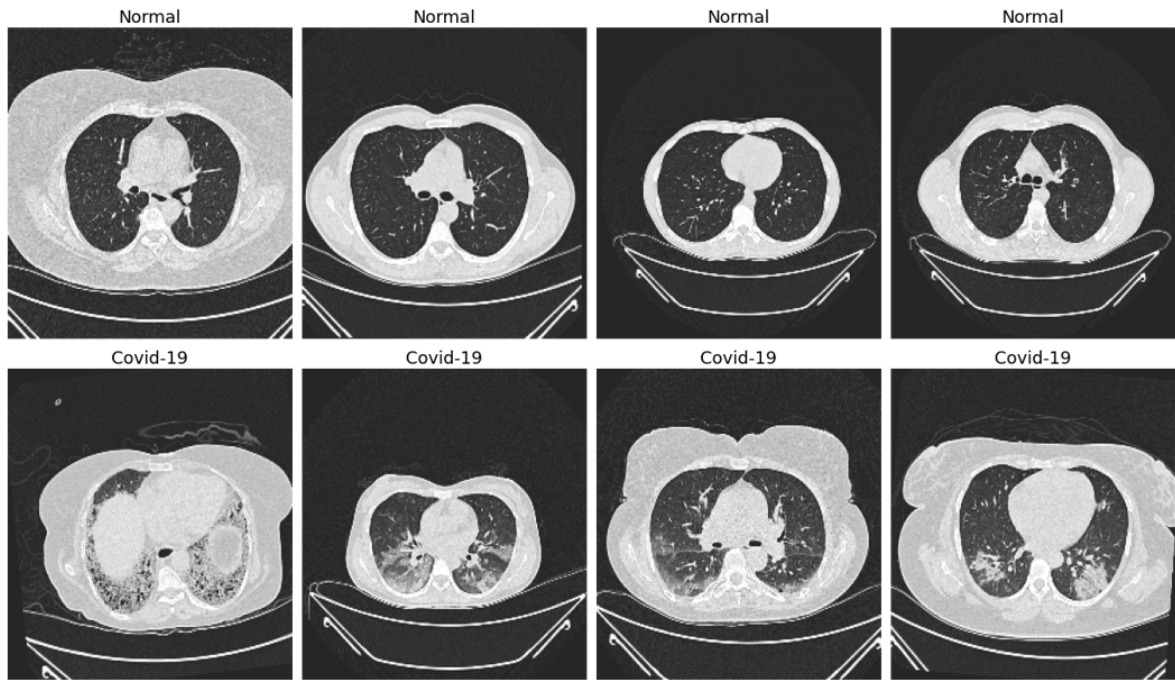
Figure 3.10: Sample image data for covid and normal class images.
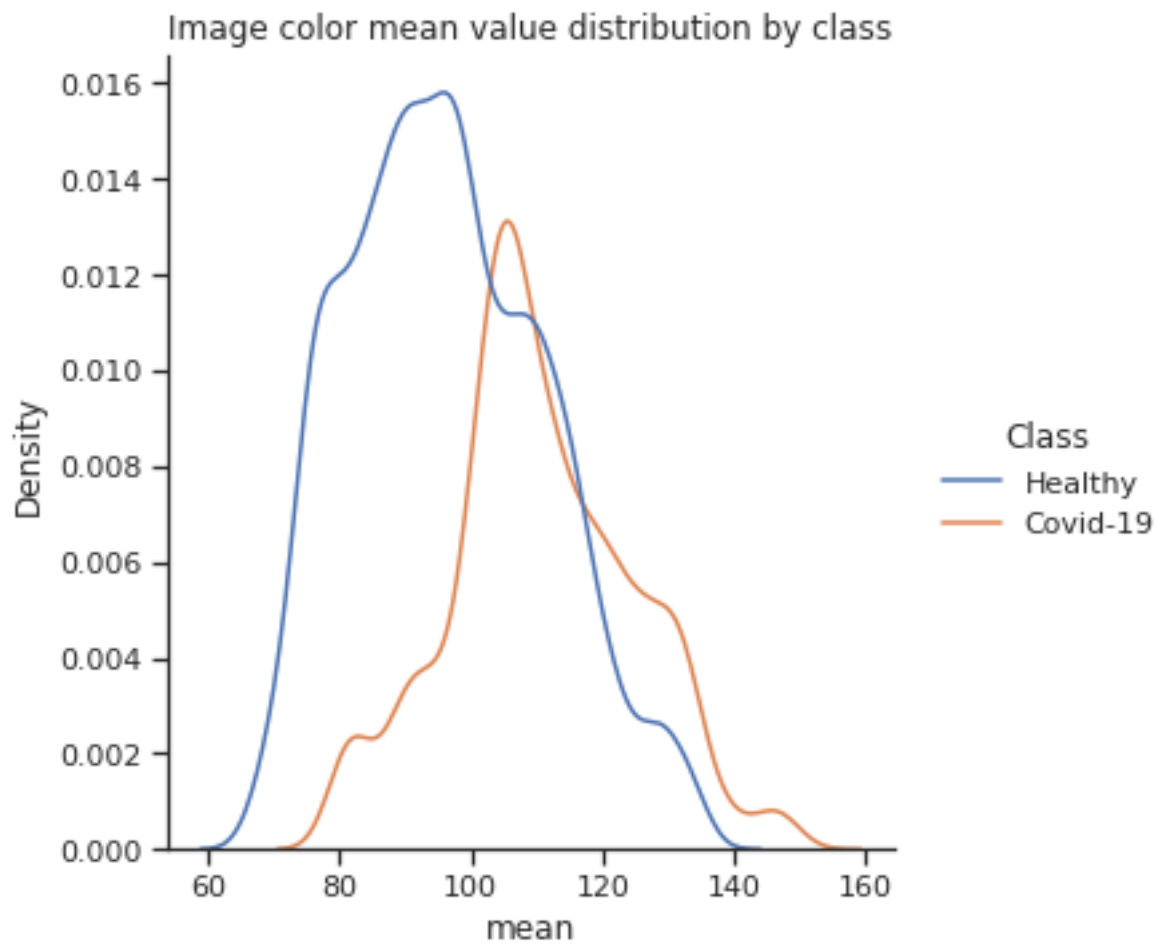

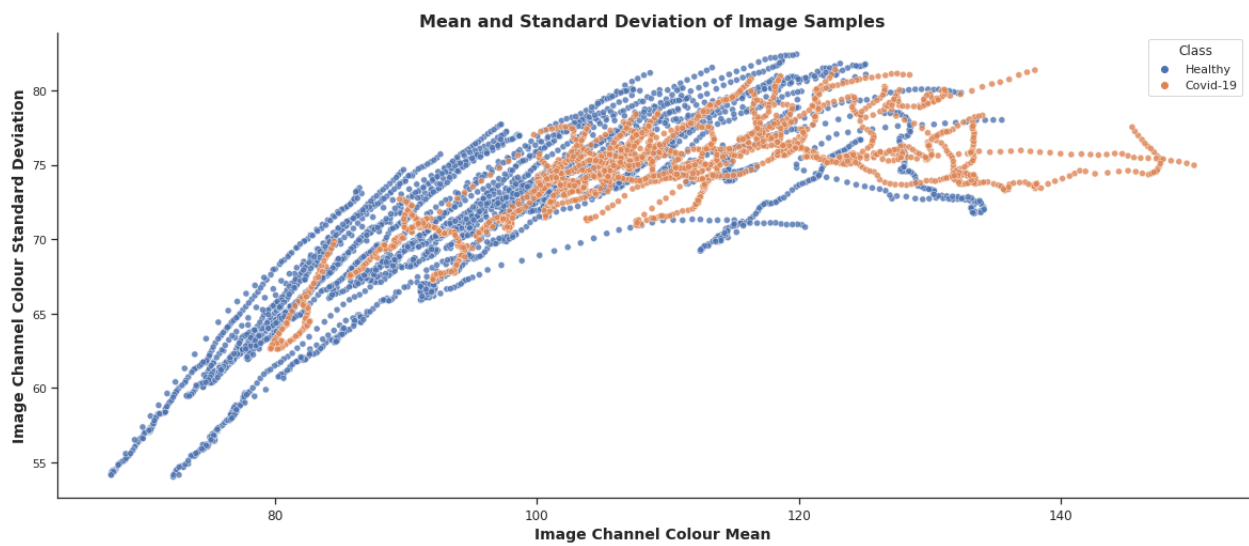Figure 3.11: Class-wise mean color value distribution of the CT images.

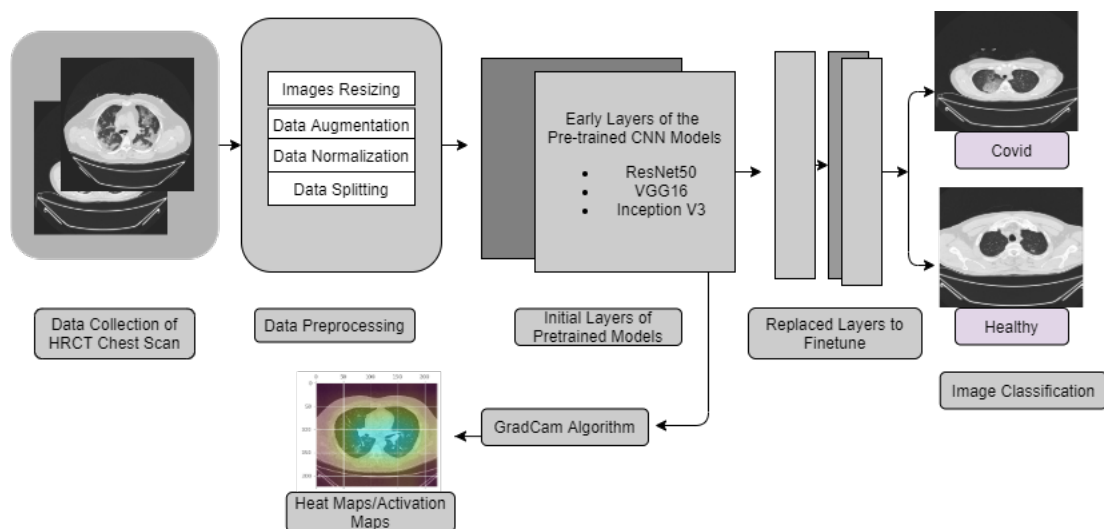Figure 3.12: Mean and standard deviation of Image samples.

Figure 3.13: The overview of the proposed system

To compensate for the class imbalance problem in the training data, a weighted categorical loss function is used. To maximize computational efficiency and adaptive learning rate, models are generated using the Adam optimizer with default parameters. With a patience level of 15, early stopping is used to monitor validation loss and to stop training the model when validation loss becomes stable. When the validation loss reached a plateau, the model's final step training was fine-tuned using a factor of 0.25 and patience 15.

After training the models with a mini-batch size of 16 pictures and 100 epochs, the final receiver operating characteristic (ROC) curve, confusion matrix, and evaluation matrix were obtained. Additionally, using the GradCam technique, heatmaps are formed in the last convolution layers of the three employed models. The original image and heatmap are layered to identify the image's critical regions and to demistify the blackbox of the models.

Vgg16

VGG16 is comprised of a 16-layer deep neural network[8], and we modified the VGG16 by taking all but the final three layers. We froze the weights of 13 Conv blocks and 5 MaxPooling blocks, and additionally we added average pooling, flattening, a dense layer with relu activation function, drop out, and finally another dense layer with sigmoid activation function to distinguish covid and non-covid chest ct radiographs. Fig. 3.14 depicts our modified version of VGG16 network.
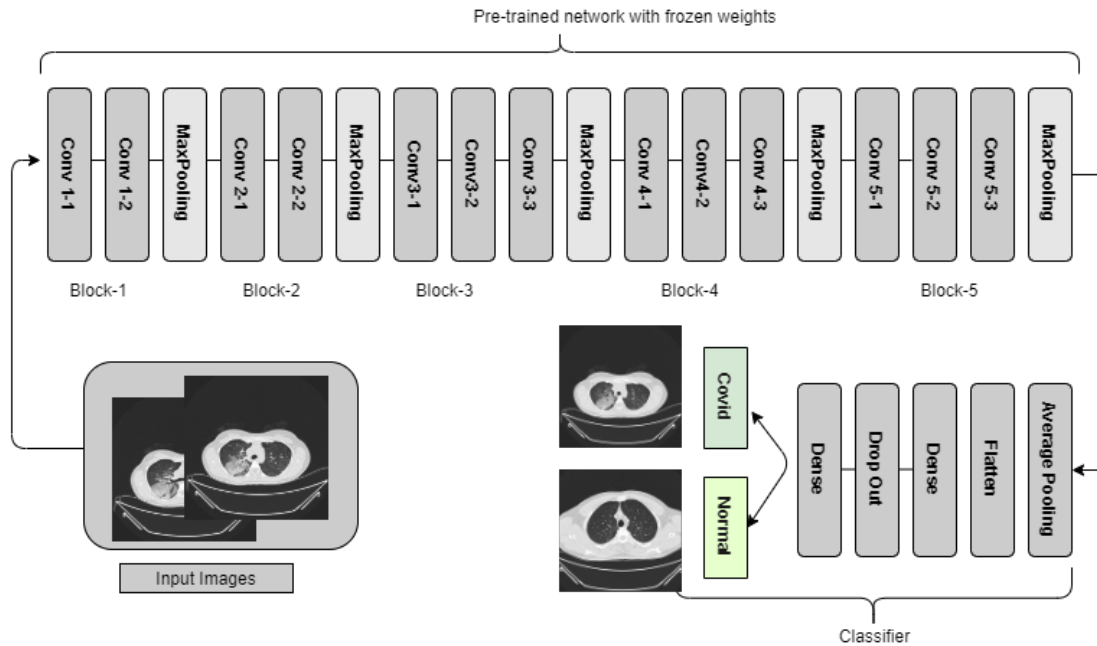

Figure 3.14: VGG16 in our classification experiments.

Resnet50

The resnet[11] gained popularity as a way to circumvent the drawbacks of very deep neural networks and avoid the vanishing gradient problem. By utilizing skip connections, Resnet was able to alleviate difficulties associated with standard deep

neural networks. We modified the original resnet50 by freezing all layers except the final three, and then adding average pooling, flattening, a dense layer with relu activation function, drop out (50 percent dropout rate), and finally another dense layer with sigmoid activation function to classify covid and non-covid chest ct radiographs.

Inception v3

Inception-v3[10], a deep neural network architecture a deep neural network architecture based on different filter size, is improved and tuned to categorize covid chest ct pictures. We used pre-trained "image-net" weights for the system, and we eliminated the model's final three layers in favor of a few additional layers to fine-tune and gain regularization. To finetune model performance, average pooling, flattening, a dense layer with relu activation function, a drop out (50 percent dropout rate) layer, and finally another dense layer with sigmoid activation function are added sequentially. Fig. 3.15 depicts our modified Inception v3 model.

3.2.5   Performance Evaluation Methods

We examined accuracy, sensitivity or recall, specificity, precision known as Positive Predictive Value (PPV), and F1 score to evaluate the performance of three deep learning models for categorizing CT pictures.

Moreover, we used gradcam[16] to examine the heatmap and de-mystify the blackbox of our models. Fig. 3.16 describes process for Gradcam analysis in our paper. We created a heatmap of the last convolution layer of the three models to see why our model concluded a certain way. A prediction was made for an image by first passing it through the model to get a prediction, then classifying it. Afterwards, we computed the class gradient with regard to Feature Map activation, A^k.

These gradients flowing back are global-average-pooled over the width and height dimensions to obtain the neuron significance weights.

To derive the neuron significance weights, these gradients flowing back are global-average-pooled across the width and height dimensions (indexed by i and j, respectively). Then, using the equation (3.11), we compute the Grad Cam.

$$L^c_{Grad_CAM} = ReLU\left(\sum_k w^c_k A^k\right) \tag{3.11}$$

3.3   Methodology behind Kidney Disease Analysis of CT Radiographs using AI

We first collected and annotated the datasets to create a database for Kidney Stone, Tumor, Normal, and Cyst findings. Data augmentation, image scaling and normalization, and data splitting are among the preprocessing techniques utilized. After that, we employed six models to investigate our data, including three Visual Transformer variants (EANet, CCT, and Swin Transformer), Inception v3, and Vgg16
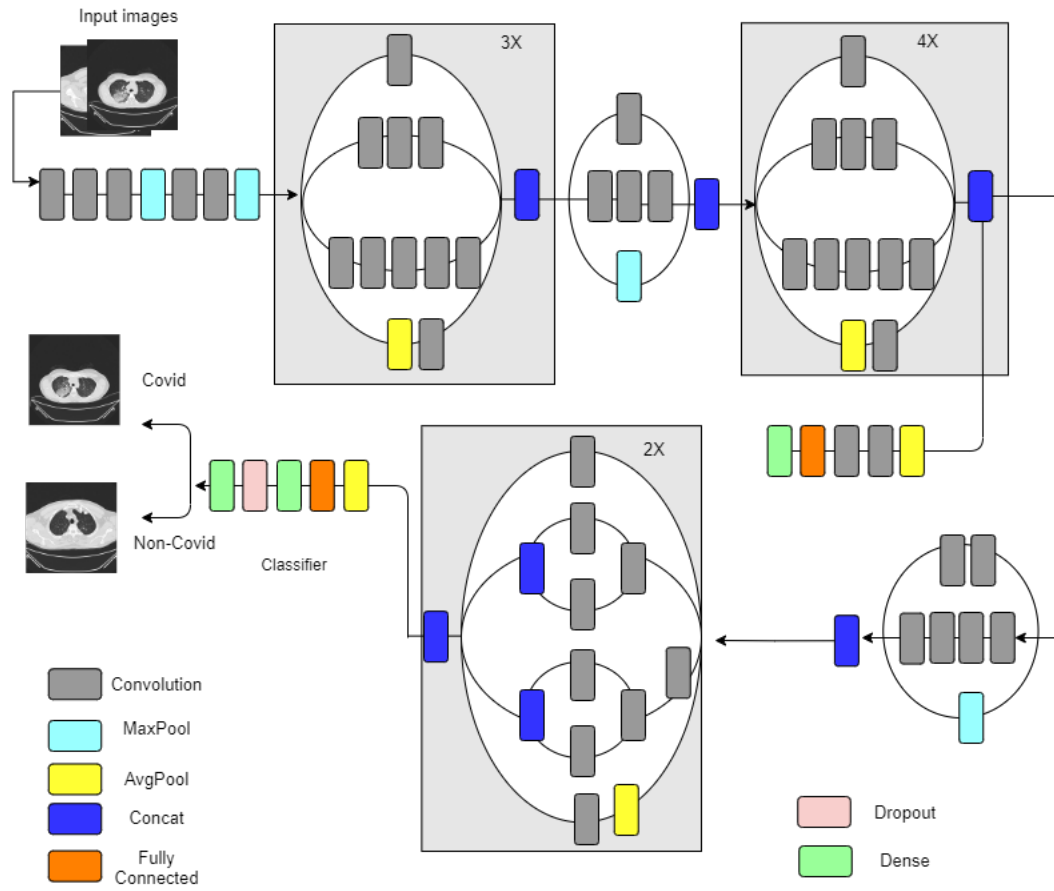
Figure 3.15: Modified Inception v3 Network to Classify Covid and Normal image
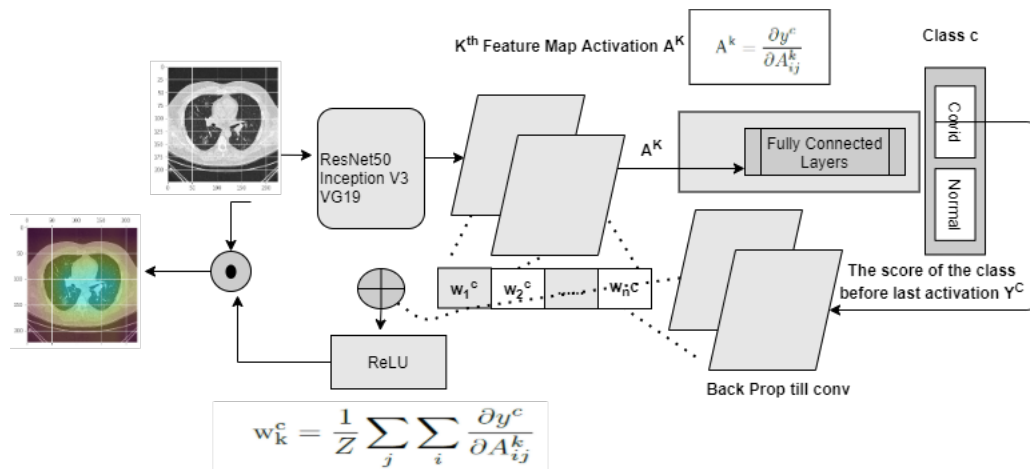
Figure 3.16: The overview of GradCam analysis for covid and non covid class

and Resnet 50. The model's performance was evaluated using previously unseen data. The Block contains details about our experiment's diagram can be found in Fig. 3.17
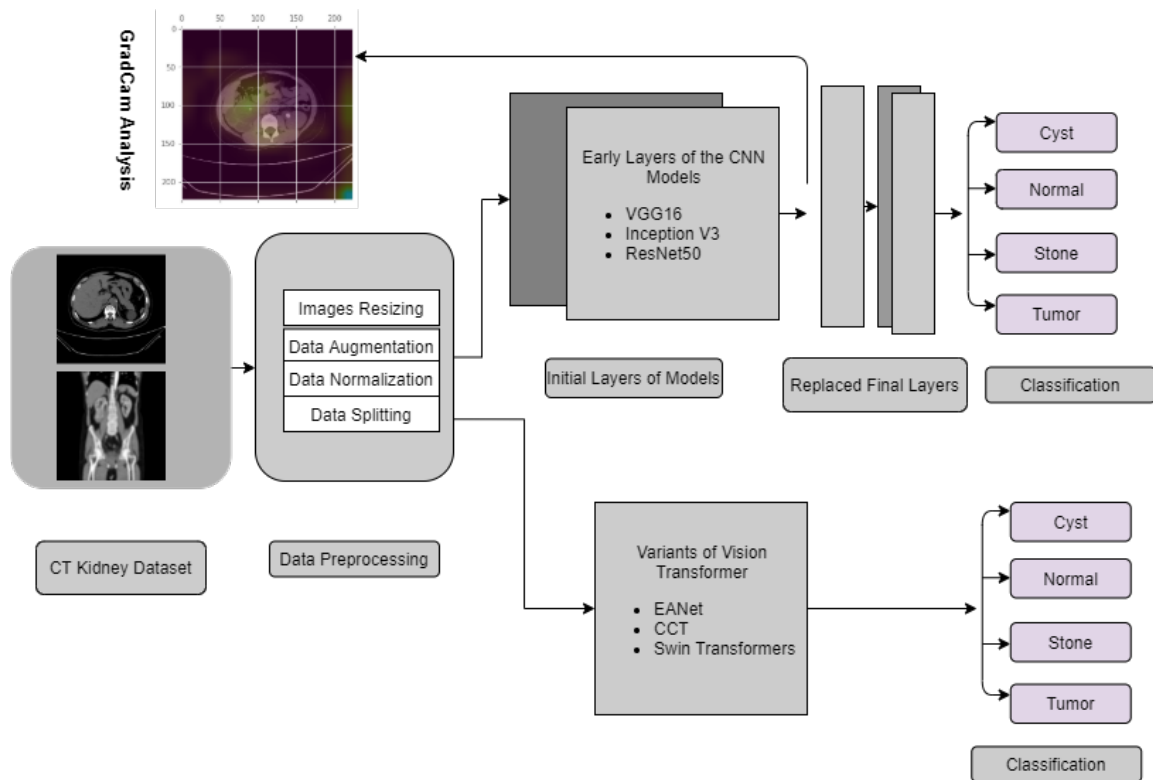
Figure 3.17: Complete Block Diagram of Experiments to diagnose Kidney tumor, cyst and stone

The methodology is presented in this part in the following order: dataset description, image preprocessing, neural network models, and evaluation strategies of the experiments.

3.3.1 Data Collection Procedure for CT Kidney Data named as "CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone"

- The data was gathered from PACS (picture archiving and communication system) systems at various hospitals in Dhaka, Bangladesh, where patients had been diagnosed with a kidney tumor, cyst, normal, or stone condition.

- Both the Coronal and Axial cuts were chosen from both contrast and non-contrast examinations for the complete abdomen and urogram, according to protocol.

- The Dicom study was then chosen, one disease diagnosis at a time, first kidney stone and from those a batch of Dicom images of the region of interest for each radiological finding was made.

- We excluded the meta data and patient information and then converted the dicom images to lossless jpg without shrinking the size of the images.

- In a similar manner, we collected Data for Kidney Normal, Tumor and Cyst findings and converted to jpg.

- Finally, radiograph imaging data was verified by a radiologist and a nephrologist.

- We created a data.csv file Fig. 3.18 which contains image id, image path, and classes. The data set contains 12,446 unique data points. Within it, the cyst contains 3,709, normal 5,077, stone 1,377, and tumor 2,283 images.

Figure 3.18: Data.csv file for Dataset named "CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone".

3.3.2 DataSet Description

The dataset was collected from PACS (Picture archiving and communication system) from different hospitals in Dhaka, Bangladesh where patients were already diagnosed with having a kidney tumor, cyst, normal or stone findings. Both the Coronal and Axial cuts were selected from both contrast and non-contrast studies with protocol for the whole abdomen and urogram. The Dicom study was then carefully selected, one diagnosis at a time, and from those we created a batch of Dicom images of the region of interest for each radiological finding. Following that, we excluded each patient's information and meta data from the Dicom images and converted the Dicom images to a lossless jpg image format. After the conversion, each image finding was again verified by a radiologist and a medical technologist to reconfirm the correctness of the data.

Our created dataset contains 12,446 unique data within it in which the cyst contains 3,709, normal 5,077, stone 1,377, and tumor 2,283. The dataset was uploaded to Kaggle and made publicly available so that other researchers could reproduce the result and further analyze it. Fig. 3.19 depicts a sample selection of our datasets. The red marks represent the finding area or region of interest that a radiologist uses to reach a conclusion for specific diagnosis classes.

Fig. 3.20 and Fig. 3.21 show the image color mean value distribution and the image color mean value distribution by class for our dataset. From both these distributions, it can be concluded that the whole dataset is very similar to the distribution of individual normal, stone, cyst, and tumor images. The mean and standard deviation of the image samples plot show that most of the images are centered, whereas stones and cysts have lower mean and standard deviation which can be visualized in Fig. 3.22. From the plot, we can conclude various types of images are being collected and images are not very similar in nature, which shows the variety and severity of different classes in our datasets.

3.3.3 Image Processing

After converting DICOM images into jpg images, we preprocessed scaled the images as per the standard size requirement of neural network models. For the all transformer variant algorithm, we resized it to 168 by 168 pixels. Images for Inception v3 were rescaled to 299 by 299 pixels, while images for VGG16 and Resnet were reduced to 224 by 224 pixels. We then randomized all the images and took 1300 examples of each diagnosis for the models' consideration. 80% of the images were taken to train the model and 20% to test the data. Within 80% of the train images, we took 20% to validate the model to avoid overfitting. Additionally, the dataset is normalized using Z-normalization[9] using following (3.12):

$$\hat{X} = \frac{X[:i] - \mu_i}{\sigma_i} \tag{3.12}$$

Here, μ_i is the mean and σ_i is the standard deviation value of the feature.

3.3.4 Neural Network Models

From the dataset, i. e., the CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone, we randomly chose 1300 images of each class and trained our six models. All the neural network models were trained on Google Colab Pro Edition with 26.3 GB of GEN Ram and 16160 MB of GPU RAM using Cuda version 11.2. All the models were trained with a batch size of 16 and up to 100 epochs.

Vgg16

In our experiment, the 16-layer VGG 16[8] model was tweaked in the last few layers by using the first 13 layers of the original VGG16 model, and we added average pooling, flattening, and a dense layer with a relu activation function. A dropout and finally another dense layer is added to classify the normal kidney as well as cysts, tumors, and stones. The total number of parameters in our modified VGG16 is 14747780, out of which 4752708 are the trainable parameters and 9995072 are
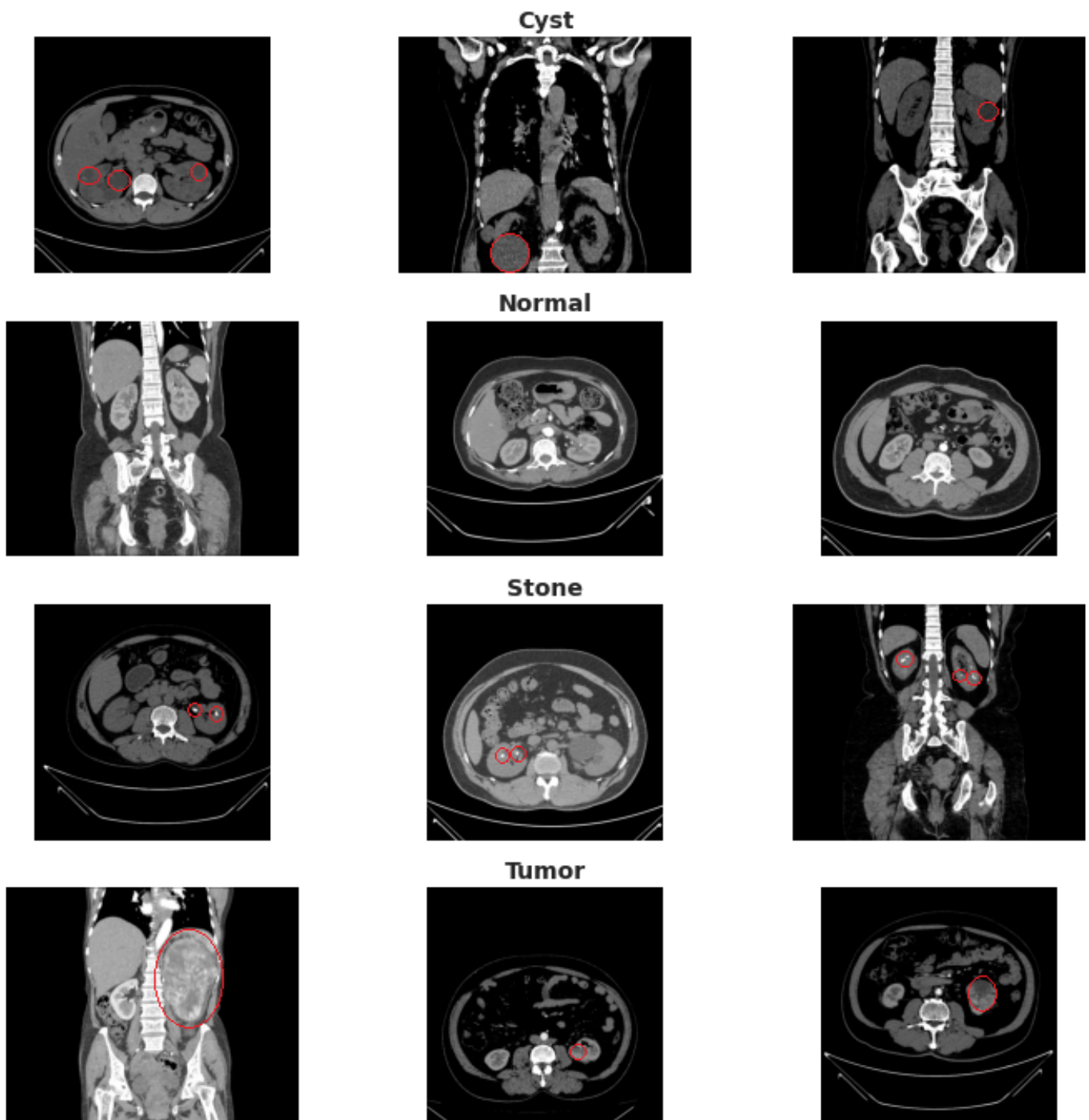
Figure 3.19: Sample image data of kidney cysts, normal, stone and tumor findings.
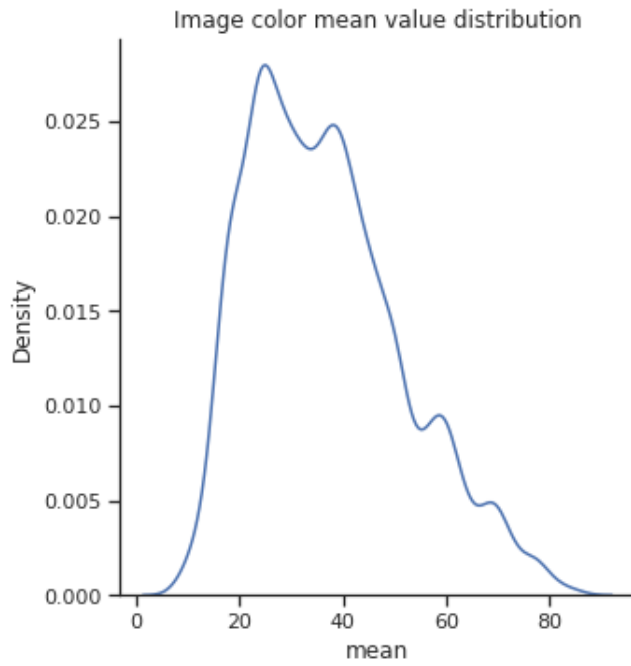
Figure 3.20: Colour mean value distribution of images for kidney disease analysis..
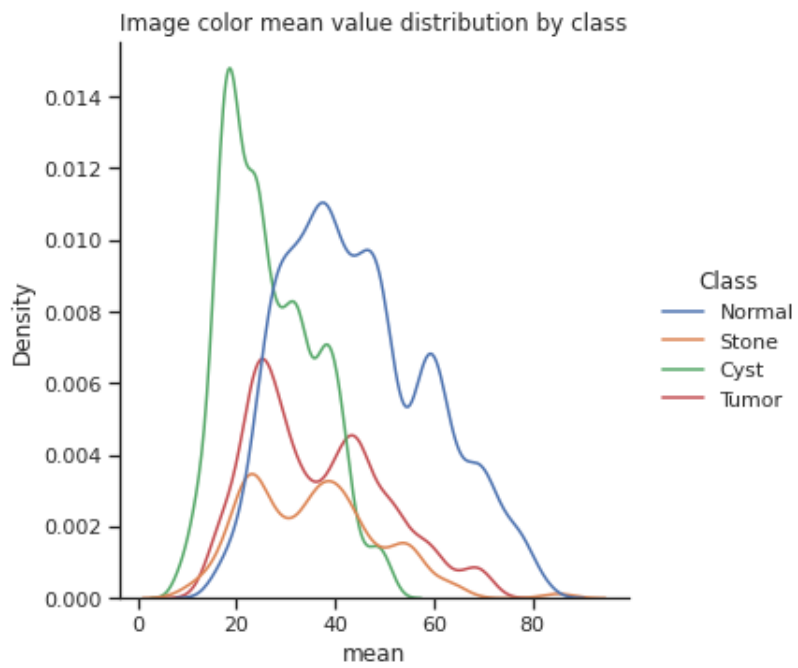
Figure 3.21: Image colour mean value distribution by class for kidney disease analysis..
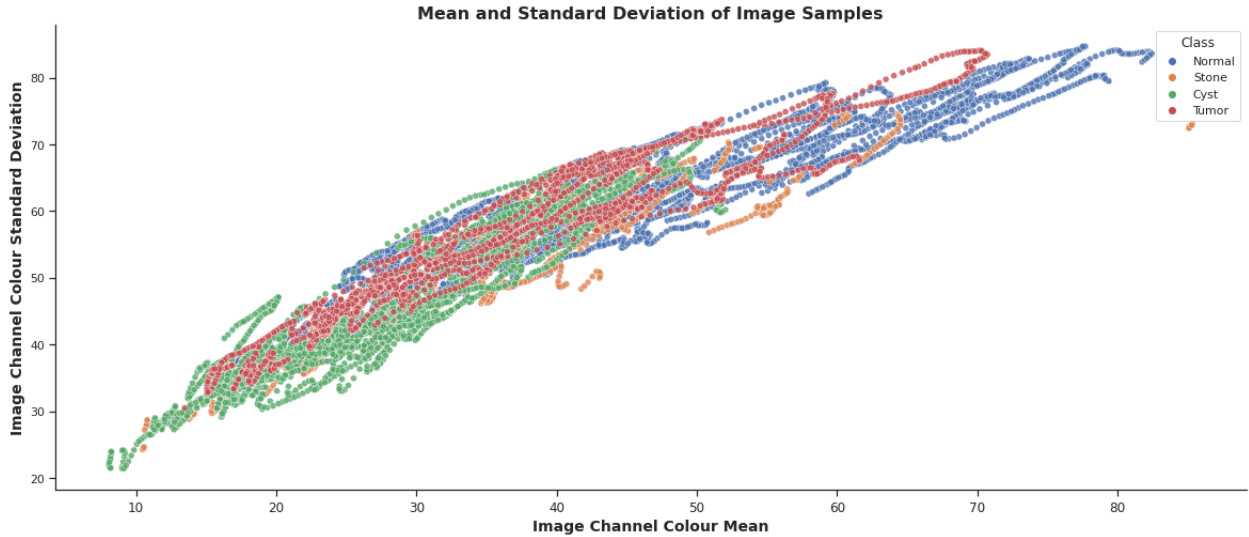
Figure 3.22: Mean and standard deviation of Image samples for kidney disease analysis.

the non-trainable parameters. Table 3.1 shows the number of parameters of the different models used in our study.

Resnet50

To avoid the vanishing gradient problem, and performance degradation of deep neural networks, skip connections are being used in the original Resent model. We utilized 50-layer resnet50 [11] models and modified them as the same as the Vgg16 and Inception v3 layers in the final few layers to achieve the classification task. The total number of parameters in our modified Resnet 50 model is 23719108. Trainable and nontrainable parameters are 135492 and 23583616 respectively.

Inception v3

A variant of the Inception family neural network, Inception v3 based on Depthwise Separable Convolutions, is used in our study to classify images. Similar to VGG 16, we modified the original Inception v3[10] model in the last few layers, by keeping all the layers except the last three. We added average pooling, flattening, a dense layer, a dropout, and finally a dense layer to do the classification task. The total number of parameters in inception v3 is 22327396 with 524612 trainable parameters. The total number of non-trainable parameters is 21802784.

3.3.5 External Attention Transformer(EANet)

Though the transformer-based models were popular in Natural Language Processing, the recent advent of the vision transformer is gaining popularity over time, which utilizes the transformer architecture that uses self-attention to sequences of image patches[42]. The sequence of image patches is the input to the multiple transformer block in this case, which uses the multihead attention layer as a self-attention mechanism. A tensor of batch_size, num_patches, and projection_dim is produced by transformer blocks, which which may subsequently be passed to the classifier head

Table 3.1: No of parameters of different models

Model	Total Parameter	Trainable Parameter
VGG16	14,747,780	4,752,708
Inception v3	22,327,396	524,612
Resnet50	23,719,108	135,492
EANet	600,907	600,900
Swin Transformers	412,788	396,372
CCT	407,365	407,365

using softmax to generate class probabilities. One variant of the Vision Transformer EANet is shown in Fig. 3.23 . EANet[74] utilizes external attention, based on two external, small, learnable, and shared memories, M_k and M_v.The purpose of EANet is to drop patches that contain redundant and useless information and hence improve performance and computational efficiency. External attention is implemented using two cascaded linear layers and two normalization layers. EANet computes attention between input pixels and external memory unit via following formulas (3.13),(3.14)-

$$A = Norm\left(FM_k^T\right) \tag{3.13}$$

Finally, input features are updated from M_v by the similarities in Attention A.

$$F_{out} = AM_v \tag{3.14}$$

We utilized TensorFlow Addons packages to implement EANet. After doing data augmentation with random rotation at scale 0.1, random contrast with a factor of 0.1, and random zoom with a height and width factor of 0.2, we implemented the patch extraction and encoding layer. Following that, we implemented an extraneous attention block, and transformer block. The output of the transformer block is then provided to the classifier head to produce class probabilities to calculate the probabilities of kidney normality, stone, cyst, and tumor findings.

3.3.6   Compact Convolutional Transformer (CCT)

Convolution and transformers are combined on CCT to maximize the benefits of convolution and transformers in vision. Instead of using non overlapping patches, which are used by the normal vision transformer in CCT[76], the convolution technique is used where local information is well-exploited. Fig. 3.24 illustrates the CCT procedure-
CCT is run using TensorFlow Addons, where first data is augmented using random rotation at scale 0.1, random contrast with a factor of 0.1, and random zoom with a height and width factor of 0.2.To avoid gradient vanishing problems in CCT, a stochastic depth [12] regularization technique is used, which is very much similar to dropout except, in stochastic depth, a set of layers is randomly dropped. In CCT, In CCT, after doing convolution tokenization, data is fed to a transformer encoder and then sequence pooling. Following the sequence pooling MLP head gives the probabilities of different classes of the kidney diagnosis. The total number of parameters in our proposed CCT model has 407365 parameters and all the parameters are trainable.
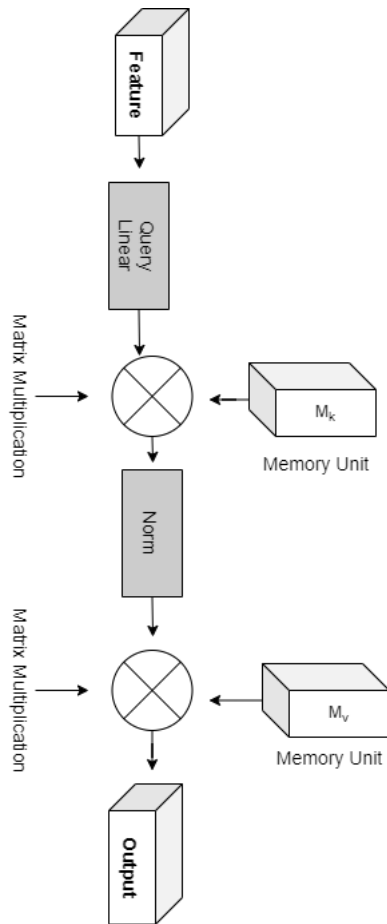
Figure 3.23: External attention of EANet model
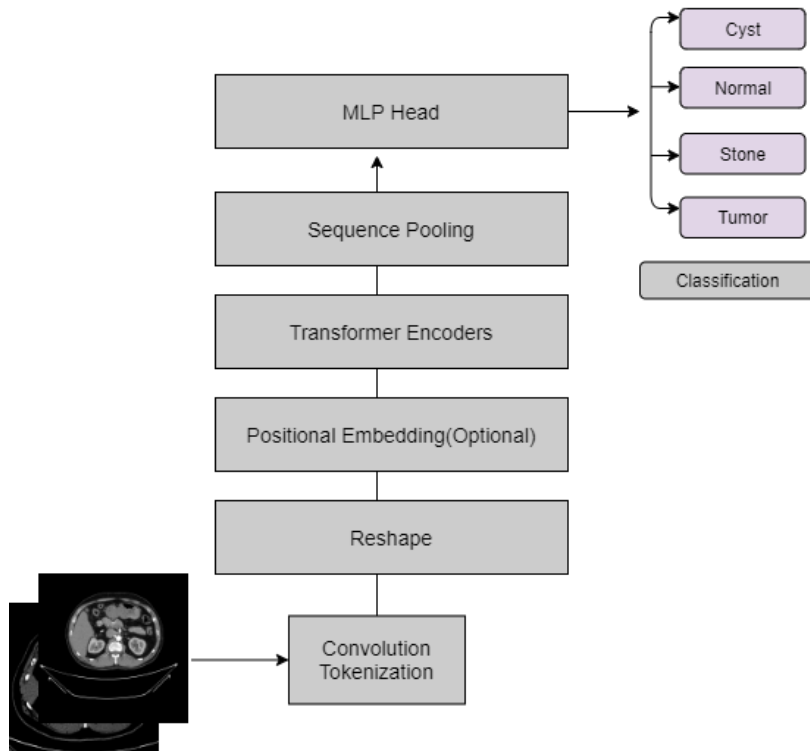
Figure 3.24: Compact Convolutional Transformer (CCT) used in the study

3.3.7   Shifted Window Transformers(Swin Transformers)

Another variant of the Vision Transformer is the Swin Transformer[81], which is another powerful tool in computer vision. Detailed block diagram of the Swin transformer is shown in Fig. 3.25. In the picture, we can see four unique building blocks. First, the input image is split into patches by the patch partition layer. The patch is then passed to the linear embedding layer and the swin transformer block. The main architecture is divided into four stages, each of which contains a linear embedding layer and a swin transformer block multiple times. The Swin transformer is built on a modified self-attention and a block that includes multi-head self-attention (MSA), layer normalization (LN), and a 2-Layer Multi-Layer perceptron (MLP). In this paper, we utilized the swin transformer to tackle the classification problem and diagnose kidney cysts, tumors, stones, and normal findings.
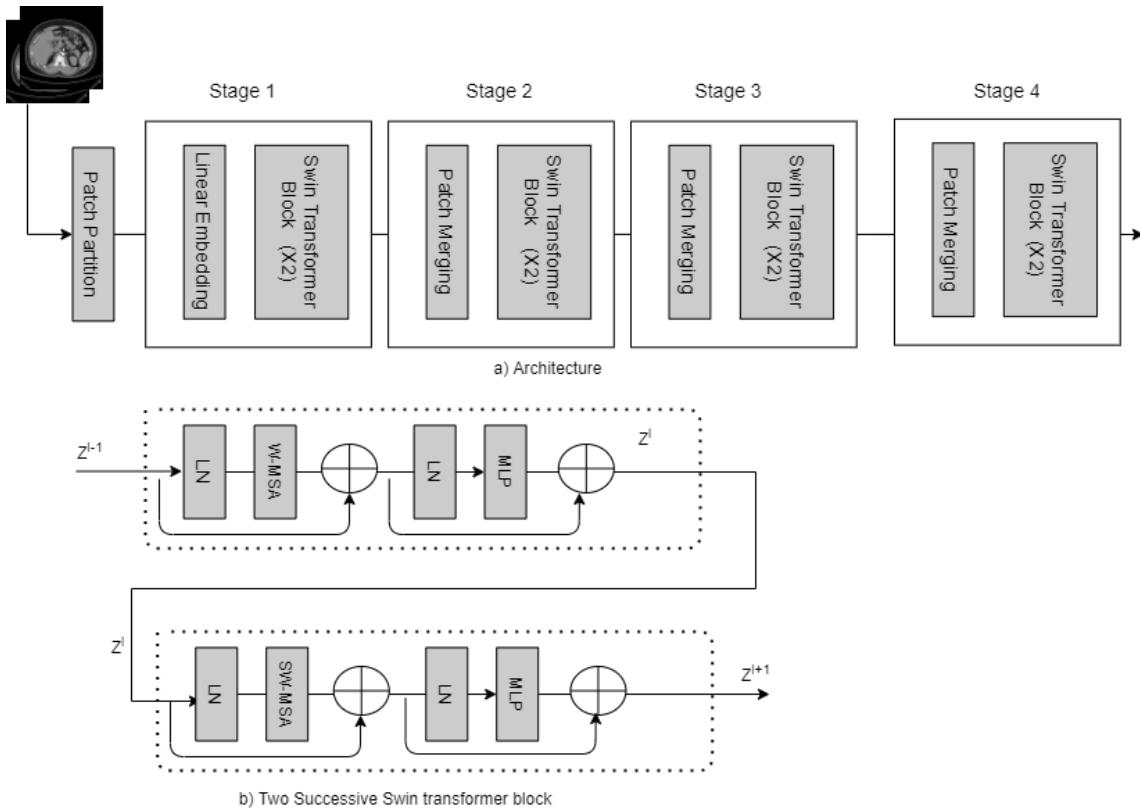
Figure 3.25: Shifted Window Transformer(Swin Transformer) diagram used in the study

3.3.8   Performance Evaluation Methods

The quantitative evaluation of all the six models is calculated based on the parameters of accuracy, sensitivity or recall, specifity and precision. True positive(T_p), false positive(F_p), true negative(T_n), and false negative(F_p) samples are used to calculate the accuracy(3.15), precision(3.16), sensitivity(3.17) and specificity(3.18). The F1 score(3.19) of all the models is calculated by using those models' sensitivity and specificity. The following formulas are applied to accuracy, precision, sensitivity, specificity, and F1 score.

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{3.15}$$

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{3.16}$$

$$\text{Recall/Sensitivity} = \frac{T_p}{T_p + F_n} \tag{3.17}$$

$$\text{Specificity} = \frac{T_n}{T_n + F_p} \tag{3.18}$$

$$\text{F1\_score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{3.19}$$

The area under the curve is also calculated for each developed model, and finally, all the models are compared to take a decision on which model is superior compared to other models.

Moreover, we used Gradcam[16] using (3.20) in the last convolution layer of the three models (VGG16, Resnet50, and Inception v3) to create a heatmap and then we superimposed the heatmap with the original image to see why our model came to a certain conclusion about a class.

$$L^c_{\text{Grad}_C\text{AM}} = ReLU \left(\sum_k w^c_k A^k \right) \tag{3.20}$$

Where,

- A^k= feature map activation

- w^c_k= neuron significance weights

Summary of the all the models implemented in all the disease are shown in table3.2 below-

Table 3.2: Summary of the all the models implemented in all the disease

Disease	Modality	Model
Lung	Xray	ResNet50 VGG19 Xception
Lung	CT	ResNet50 VGG16 Inception v3
Kidney	CT	ResNet50 Vgg16 Inception v3 EANet CCT Swin transformer

Chapter 4

Implementation & Result Analysis

In this section of the thesis, we will describe the findings and conclusions of our research.

4.1 Result Analysis of Lung Disease Analysis of X-ray Radiographs using AI

The results of the experiments are broken down into two sections. In the first phase, results were analyzed statistically, and in the second phase, feature extraction in different layers of different models was visualized using GradCam and analyzed why our model came to the conclusion of detecting Covid and non-Covid images.

4.1.1 Statistical Analysis

We assessed the model's performance using a separate test set that models were not exposed to during training. The models were evaluated statistically by computing the test precision, recall, F1-score (F1), accuracy (Acc), Positive Predictive Value (PPV), and the Area Under the ROC Curve (AUC).

Table 4.1 summarizes the performance of the three CNN networks studied in this paper for two class setup. performance is also shown for the implemented three networks for two different input image sizes. Similarly, Fig. 4.1 shows the performance comparison of all three models at normal picture size as per model's requirement and 512×512 input images.

From Table 4.1, we can examine the accuracy of the Xception and Vgg19 is 97% while training the model with standard image size requirements as per model, while the Resnet50 provides 92.5% accuracy. The F1 score for detecting normal and Covid images for Xception and Vgg19 is 0.971, 0.969, 0.967, and 0.972 respectively, where as for the resnet model, it is 0.923 and 0.972. Clearly, Xception and Vgg are outperforming the Resnet 50 considering accuracy and F1 score. The sensitivity to detecting covid is higher than the other two models and it is 97.9%, where high precision of 97.7% is observed in Vgg19 for the covid class. For the normal class, recall is highest in vgg 19 and precision is highest in the Xception model. For the 512×512 image size, again, accuracy is higher in Xception and Vgg19 and both provide an accuracy of 97.5%. The F1 score for covid and the normal class for the Xception model is 0.976 and 0.973, where as in the Vgg19 model, it is 0.975 and

Table 4.1: Performance measures for the three models for two class classification(X-ray Lung).
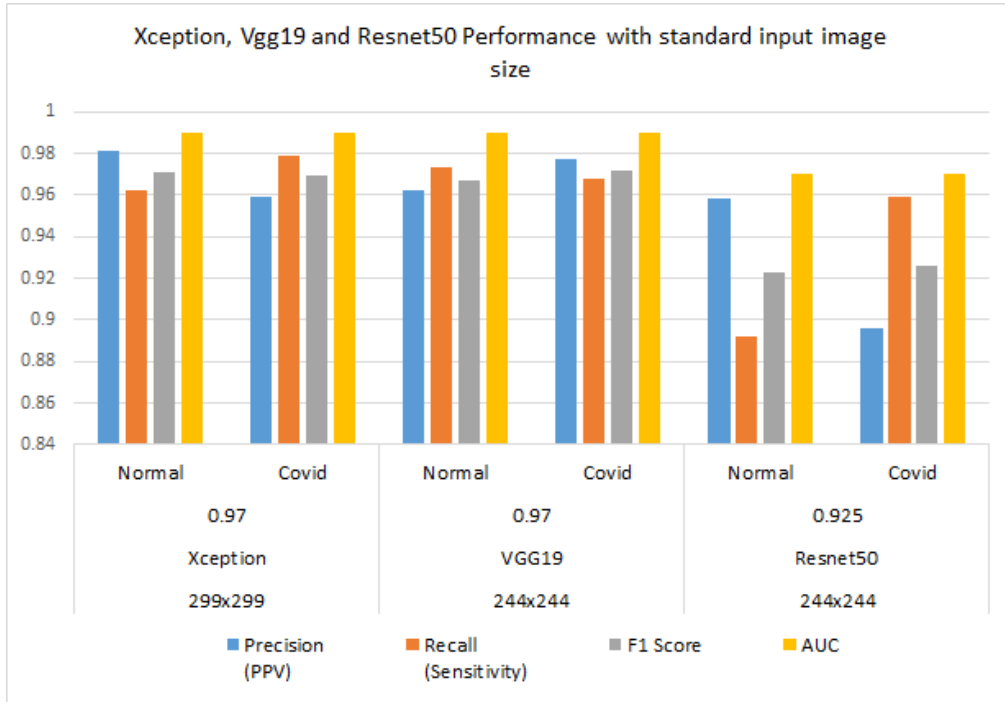
Input Image Size	Models	Accuracy	Class	Precision(PPV)	Recall(Sensitivity)	F1 Score	AUC	Explainable AI
299x299	Xception	0.97	Normal	0.981	0.962	0.971	0.99	Y
			Covid	0.959	0.979	0.969	0.99	
224x224	VGG19	0.97	Normal	0.962	0.973	0.967	0.99	Y
			Covid	0.977	0.968	0.972	0.99	
224x224	Resnet50	0.925	Normal	0.958	0.892	0.923	0.97	Y
			Covid	0.896	0.959	0.926	0.97	
512x512	Xception	0.975	Normal	0.995	0.953	0.973	1	Y
			Covid	0.958	0.995	0.976	0.99	
	VGG19	0.975	Normal	0.961	0.99	0.975	0.99	Y
			Covid	0.99	0.961	0.975	0.99	
	Resnet50	0.933	Normal	0.983	0.881	0.93	0.98	Y
			Covid	0.89	0.985	0.935	0.98	

0.975. So, it can be concluded that training with increased image size improves the accuracy and performance of the model. This may be because, while resizing to a lower resolution, we may lose some vital information from the image which may take part in decision making processes. Fig. 4.1 shows precision, recall, F1 score and AUC bar bar chart comparison for three models: Xception, Vgg19 and Resnet50 while training the model with standard input image size and 512×512 input image size. Precision, recall, F1 score and AUC bar bar chart for three models when model is trained with standard image size (a).Precision, recall, F1 score and AUC bar bar chart for three models when models are trained with the images size of 512×512 (b). It is evident that Resnet is performing poorly compared to Vgg19 and Xception. It is also evident that training with increased image resolution helps to achieve better performance of the models.

Fig. 4.3 shows ROC curves and confusion matrices for each of the models, considering each of the classes individually while training the model with standard image size. Top: ROC curves. Bottom: Normalized confusion matrices. Left: ROC curve and confusion matrices for the Xception model. Center: ROC curve and confusion matrices for the Vgg19. Right: ROC curve and confusion matrices for the Resnet model, and Fig. 4.2 shows the ROC curves and confusion matrices for each of the models, considering each of the classes individually while training the model with 512x512 input image resolution. Top: ROC curves. Bottom: Normalized confusion matrices. Left: ROC curve and confusion matrices for the Xception model. Center: ROC curve and confusion matrices for the Vgg19. Right: ROC curve and confusion matrices for the Resnet model. it is undeniable that training models with higher picture quality results in enhanced performance of the models after they have been trained.

Table 4.2 and Fig. 4.4 summarizes the performance of the three CNN networks studied in this paper for multi class setup. We can see the Xception model is outperforming VGG19 and Resnet50 and giving accuracy 93%. F1 score for Normal, Covid, Lung Opacity, and Viral Pneumonia is 0.92, 0.92, 0.96 respectively. Vgg19 is quite close performance wise and gives 92% accuracy.

(a) Bar graph with standard image resolution
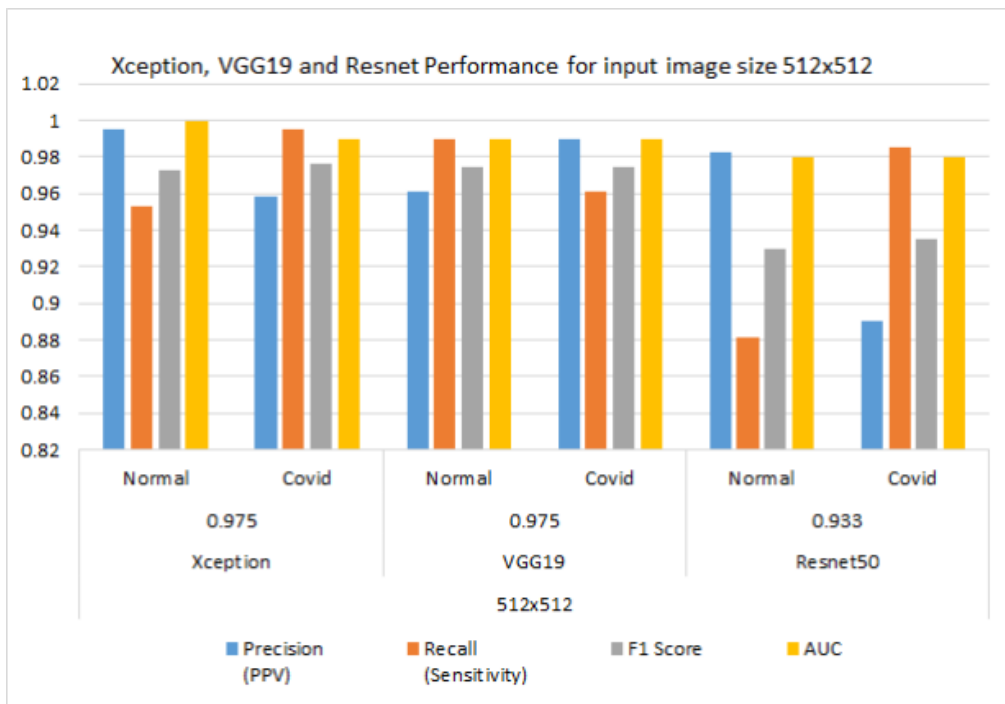

(b) Bar Graph 512×512

Figure 4.1: Precision, recall, F1 score and AUC bar chart comparison for three models.

Table 4.2: Performance measures for the three models for multi class classification(X-ray Lung).

Input Image Size	Models	Accuracy	Class	Precision(PPV)	Recall(Sensitivity)	F1 Score	AUC	Explainable AI
299x299	Xception	0.93	Normal	0.89	0.95	0.92	0.99	Y
			Covid	0.93	0.91	0.92	0.99	
			Lung Opacity	0.91	0.92	0.92	0.99	
			Viral Pneumonia	0.98	0.94	0.96	0.99	
224x224	VGG19	0.92	Normal	0.87	0.97	0.92	0.99	Y
			Covid	0.94	0.88	0.91	0.99	
			Lung Opacity	0.87	0.92	0.89	0.99	
			Viral Pneumonia	1	0.89	0.94	0.99	
224x224	Resnet50	0.75	Normal	0.8	0.9	0.85	0.96	Y
			Covid	0.65	0.78	0.7	0.89	
			Lung Opacity	0.73	0.49	0.59	0.87	
			Viral Pneumonia	0.85	0.85	0.85	0.94	
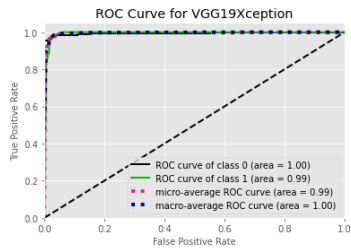
(a) ROC curve for Xception model at input image resolution 512×512
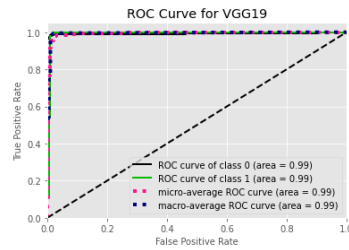
(b) ROC curve for Vgg19 at input image resolution 512×512
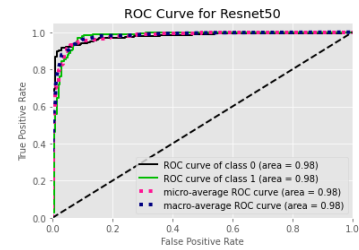
(c) ROC curve for Resnet50 at input image resolution 512×512

(d) Confusion Matrix for Xception model at input image resolution 512×512

(e) Confusion Matrix for VGG19 at input image resolution 512×512

(f) Confusion Matrix for Resnet50 at input image resolution 512×512

Figure 4.2: ROC curves and confusion matrices for dual class models with 512x512 input image resolution(xray-lung).

(a) ROC curve for Xception model at input image resolution 299 × 299

(b) ROC curve for Vgg19 at input image resolution 224 × 224
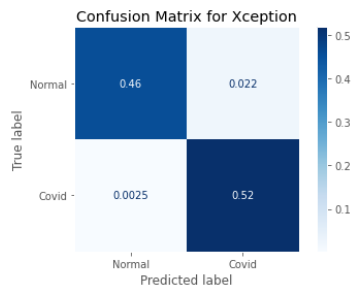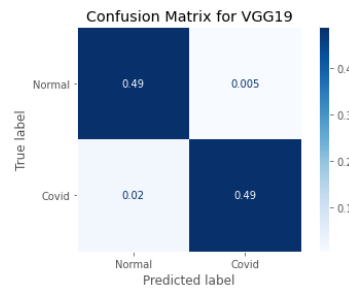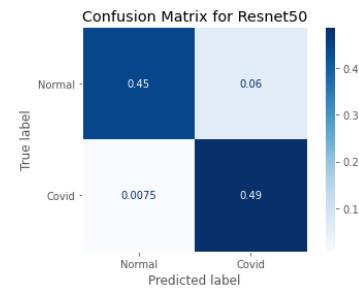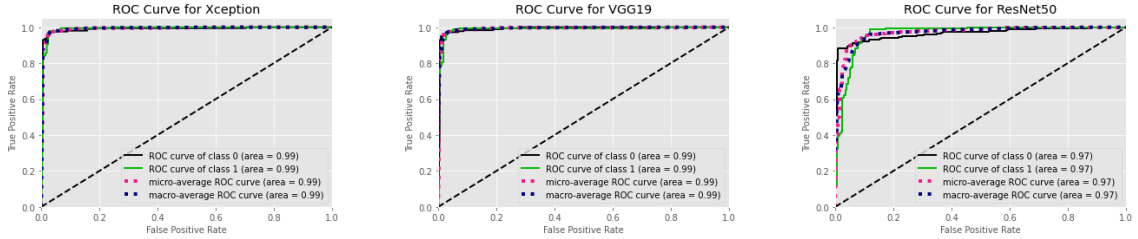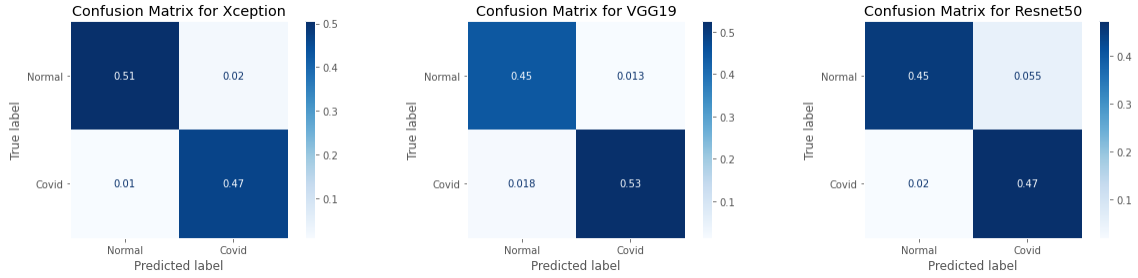

(c) ROC curve for Resnet50 at input image resolution 224 × 224

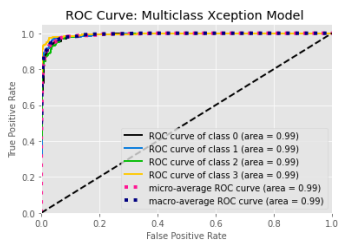(d) Confusion Matrix for Xception model at input image resolution 299 × 299

(e) Confusion Matrix for VGG19 at input image resolution 224 × 224

(f) Confusion Matrix for Resnet50 at input image resolution 224 × 224

Figure 4.3: ROC curves and confusion matrices dual class models with standard image size(xray-lung).

4.1.2 Model's Explainability and Interpretability

It is commonly feasible to show why the model determined a specific class using a heatmap. We qualitatively examined network-identified regions of interest using Grad-CAM activation maps and by displaying heatmaps at different layers of three proposed networks. To anticipate the pathological condition and model interpretability, and to further analyze, the original image and heatmap are superimposed to locate the critical spots in the image. Although the last CNN layer is generally used in GradCAM literature, we visualized all the network layers to study the learning process of the models, since the last layer contains high-level information. In this letter, we utilized the GradCAM approach to 1) visualize and compare the different layers of a model to identify the model's decision processes and 2) identify the network sections that have the most significant impact on categorization. 3) to compare activation maps of three CNN models' last convolution layer to demystify all the model performances. For the Xception model, we used GradCAM to display the activation maps for all the CNN blocks for covid and noncovid class images. We visulaized the activation map for all the layers in the same block for the Xception model. Furthermore, we implemented expainable AI to the Resnet50 and Vgg19 models and compared the activation maps to see which model should perform better in clinic conditions. We approached GradCam analysis for four class classification to show the last conv layer output of the Xception models since the accuracy of Xception model is higher than the Vgg19 and Resnet50.

Fig. 4.5 depicts the visualization results for the Separation Convolution layers in Block 14,12,10,8,4 and 1 for the Xception model for a ground truth normal image.

(a) ROC curve for Multi Class Xception model at input image resolution 299×299

(b) ROC curve for Multi Class Vgg19 at input image resolution 224×224
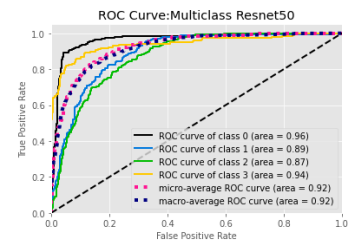
(c) ROC curve for Multi Class Resnet50 at input image resolution 224×224
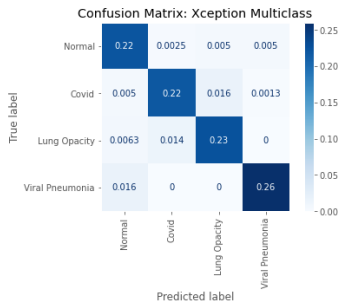
(d) Confusion Matrix for Multi Class Xception model at input image resolution 299×299
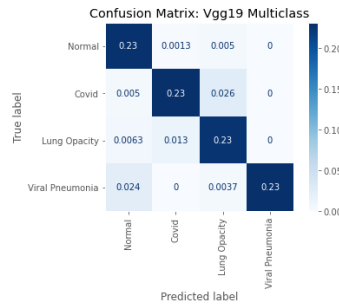
(e) Confusion Matrix for Multi Class VGG19 at input image resolution 224×224
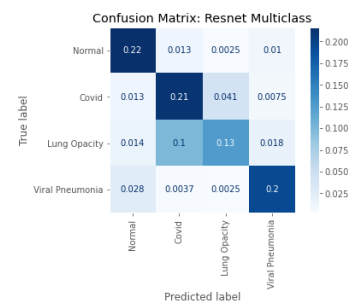
(f) Confusion Matrix for Multi Class Resnet50 at input image resolution 224×224

Figure 4.4: ROC curves and confusion matrices for Multi Class models with standard image size(xray-lung).
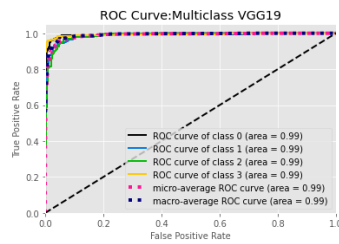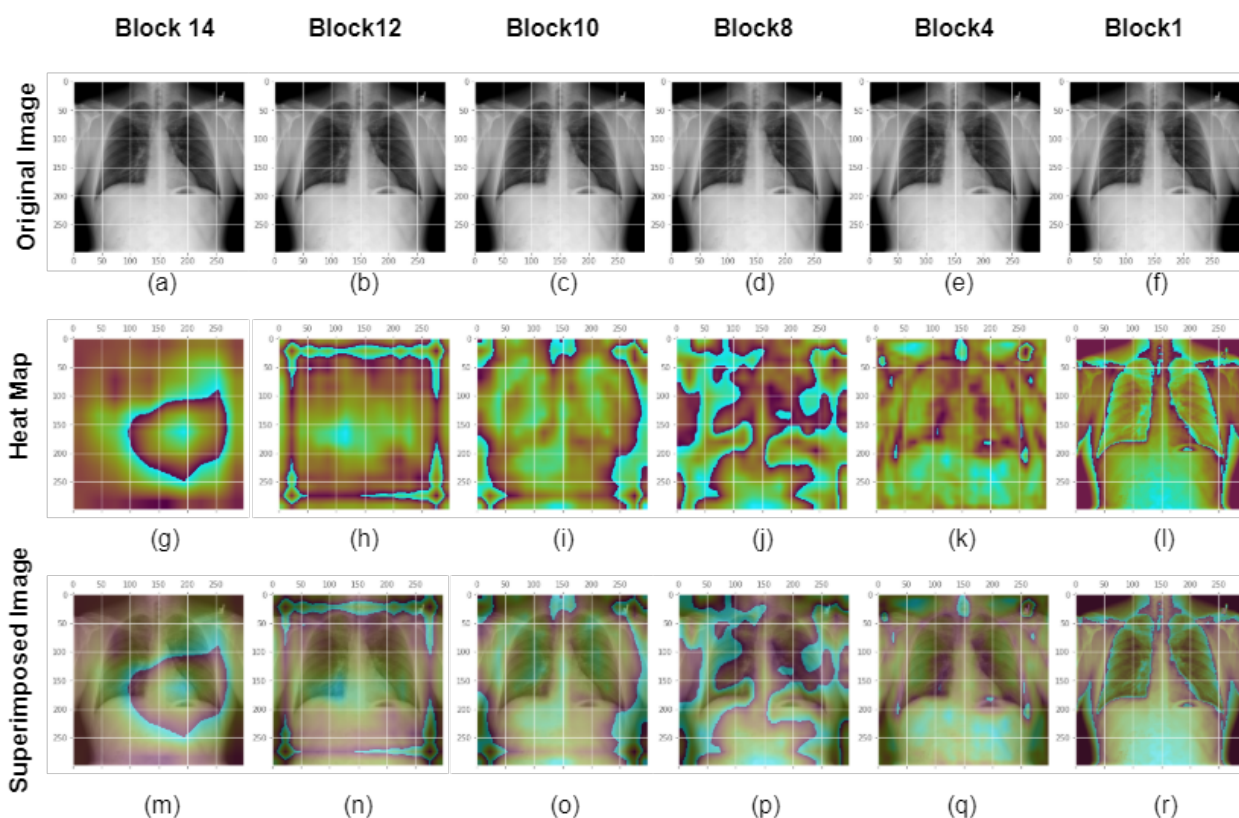
Figure 4.5: GradCam Analysis for the Xception model for normal class images at different layers.
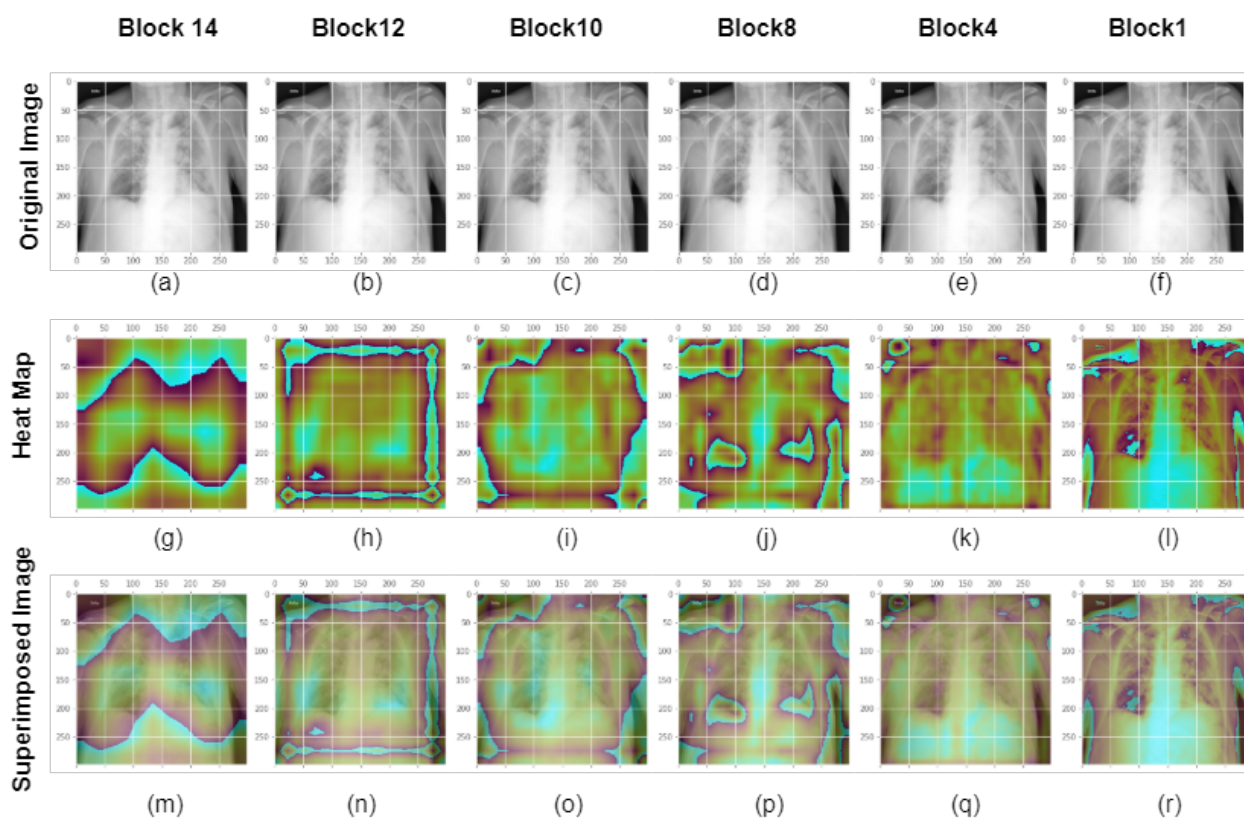
Figure 4.6: GradCam analysis of covid class photos at various layers using the Xception model.
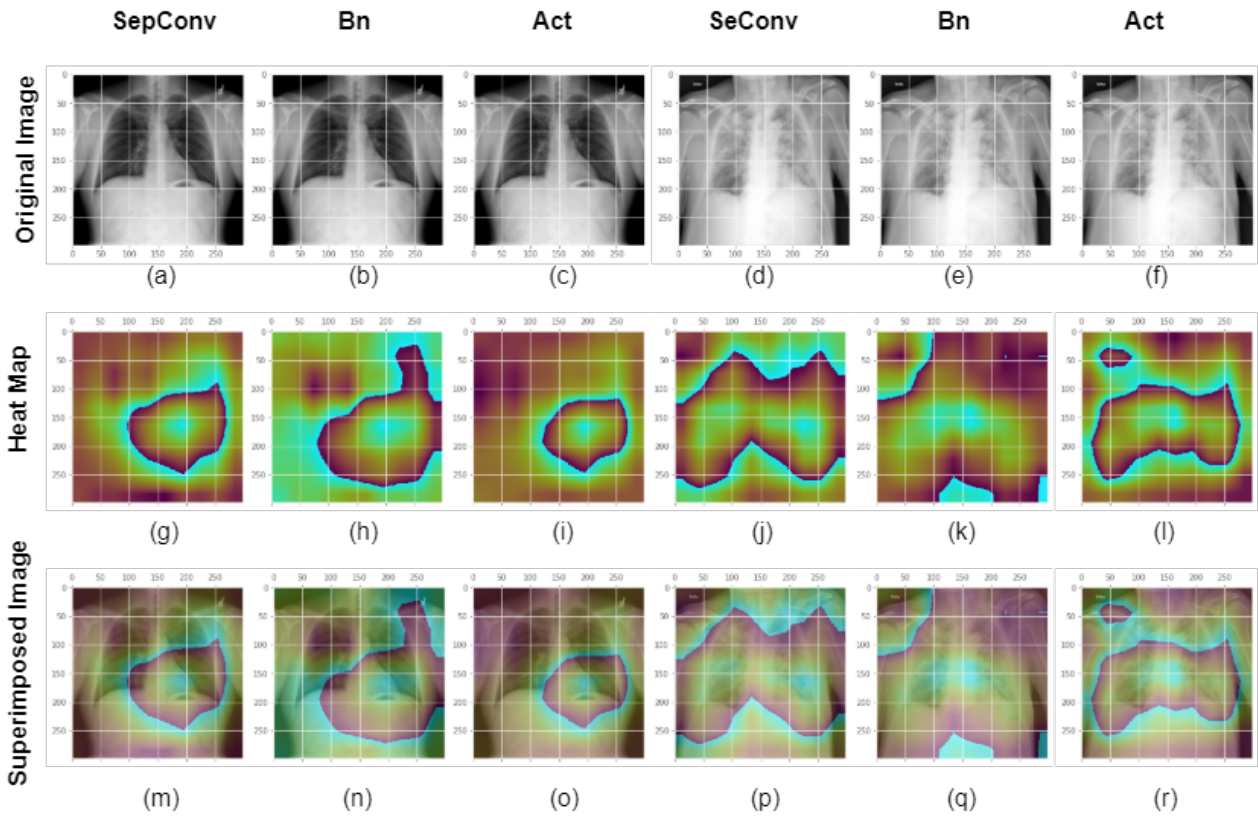
Figure 4.7: GradCam analysis of covid and normal class photos at various layers in a single block using the Xception model.
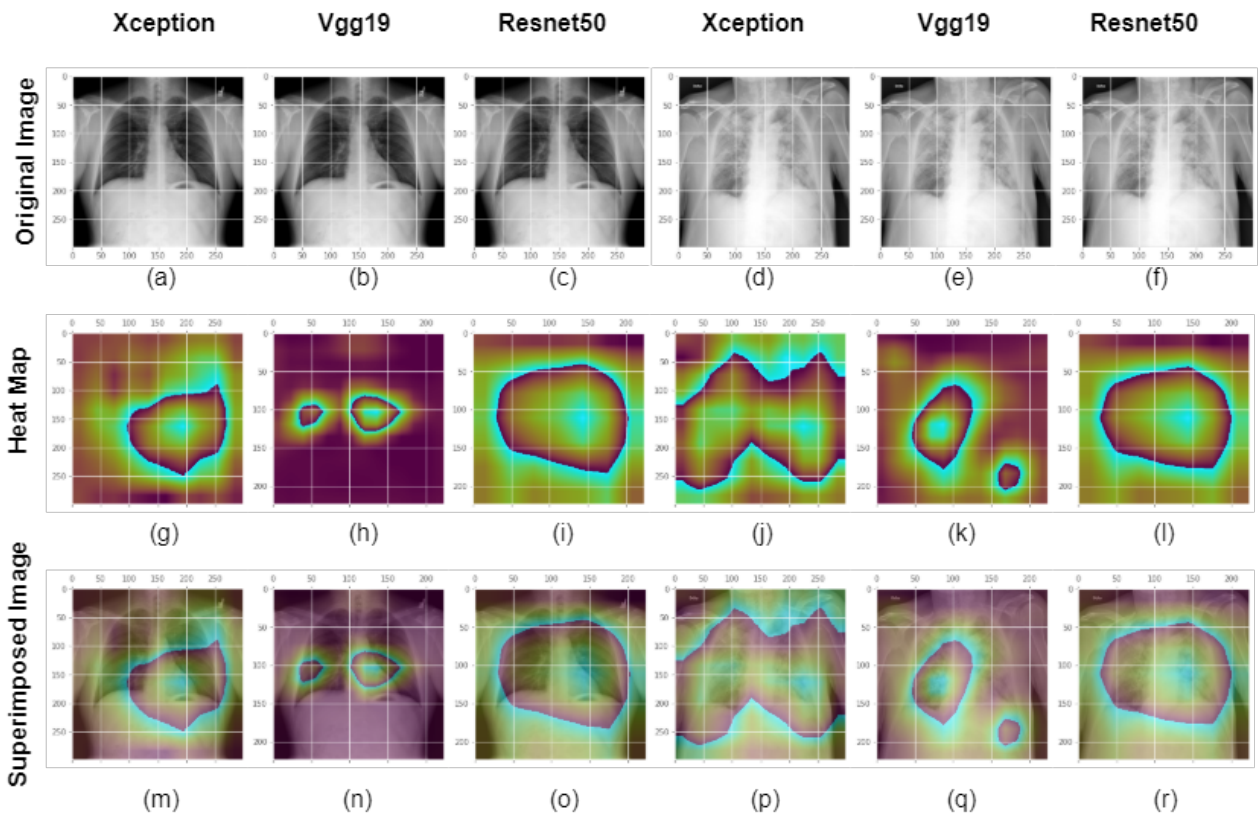
Figure 4.8: GradCam analysis of covid and normal class photos at the final convolution layer in the Xception, Vgg19, and Resnet models.
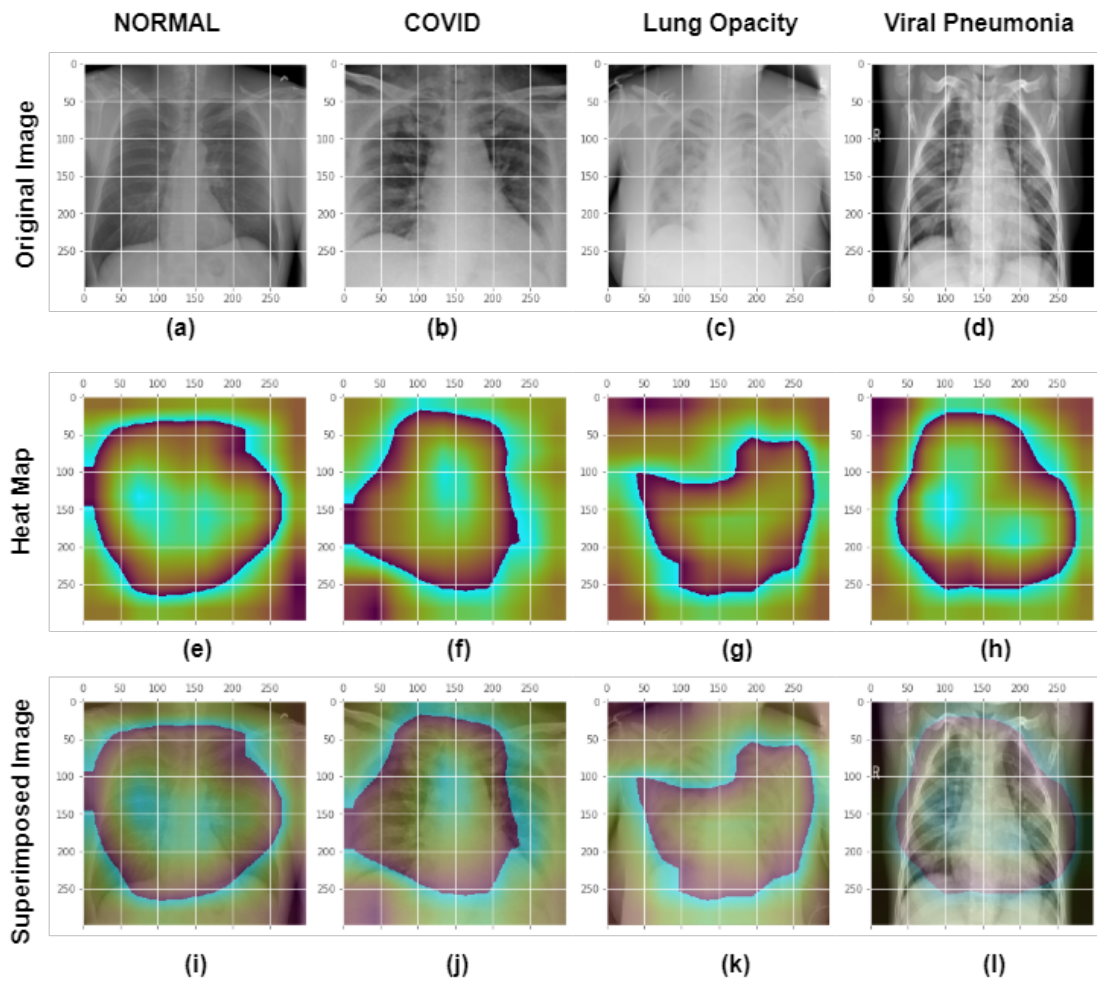
Figure 4.9: GradCam analysis of four class photos at the final convolution layer in the Xception model.
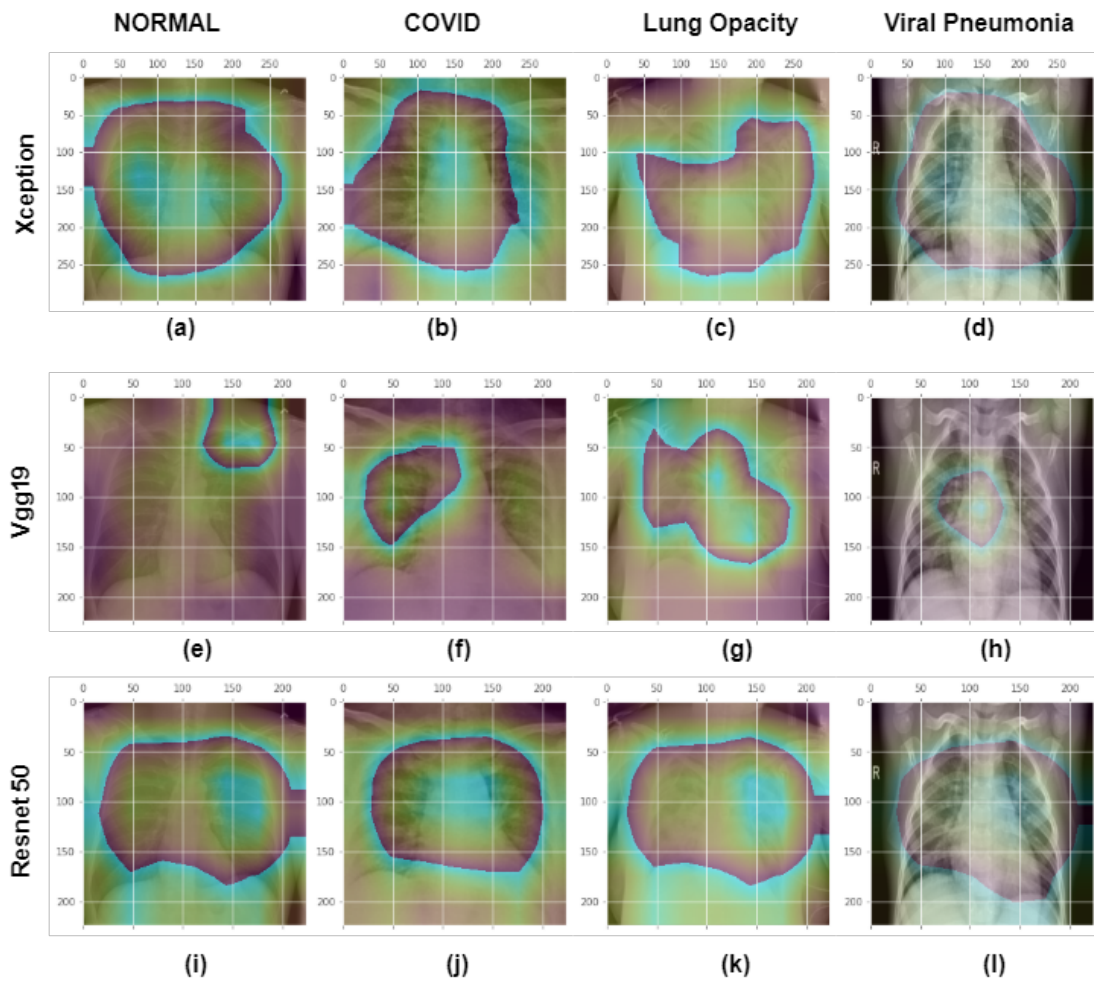
Figure 4.10: GradCam analysis of four class photos at the final convolution layer in the Xception model, Vgg19, and Resnet50 models.

First row: chest xray image belonging to the normal class. Second row: GradCam activation mapping for the xray image. Third Row: superimposed image of Xray image and Activation map. The first column shows the original image, heatmap and superimposed image of the block14 separation convolution 2 in the Xception model. The second, third, fourth, fifth, and sixth columns are for Block 12 sepconv2, Block 10 sepconv2, Block8 sepconv2, Block4 sepconv2, and Block1 conv1 respectively. Block 14 detects high-level features and impacts decision making, where as block 1 detects lower-level features such as colors and edges. From the original x-ray images, heatmap, and superimposed images, we can visualize that Separation Convolution block 14 has the most significant impact on image classification as it can see the high-level complicated features. Block 14 builds up features by combining all the features that were detected in the early layers. On the other hand, block 1 just impacts the model performance by detecting low-level features such as edges and colours. The Deep layers feature mainly contributed to better offering an explanation of the failure or success of a deep learning network in a particular decision. In this case, the Xception model detected the image truly and detected it as a class of image which is normal.

Similarly, Fig. 4.6 shows GradCam analysis of covid class photos at various layers using the Xception model. First row: shows a chest xray image from the covid class. The GradCam activation mapping for the xray image is shown in the second row. Third row: Xray picture placed on an activation map to show the usperimposed image. The first column displays the original image, heatmap, and overlay image of the Xception model's block14 separation convolution 2. The second, third, fourth, fifth, and sixth columns correspond to Block 12 sepconv2, Block 10 sepconv2, Block 8 sepconv2, Block 4 sepconv2, and Block 1 conv1. Block 14 detects high-level features and influences decision-making, whereas block 1 detects lower-level features like colors and edges.From Fig. 4.6, we ca say that x-ray images with ground truth Covid diagnosis has a comparable impact, indicating that Block 14 strongly influences finding classification areas. It can be seen that in the last block convolution layer, the model is looking into the chest region, which is the region of interest for detecting covid, and the model almost sees all the regions of the covid affected area.The Xception model was flawless in this case, predicting it. In summary, the last block contributes more to the classification than the whole block.

In the search for determining which layer in a convolutoin block provides more decisiveness, for covid and normal classes, we checked the sep convolution layer, batchnormalization (bn) layer and activation (act) layer heatmap along with the original and superimposed image, shown in Fig. 4.7. The first row displays covid and normal chest xray images. In the second line, the xray image's GradCam activation mapping is displayed. The third row displays superimposed images. The first three columns display the original image, heatmap, and overlay image of the Xception model's block14 separation convolution 2, batchnormalization, and ReLu activation layers for a normal class image. The latter third column displays the original image, activation mapping, and superimposed image of block14 separation convolution 2, batchnormalization, and ReLu activation layer for a covid class image for the Xception model. The activation layer model there observes a more narrow region in search for covid and normal classes. Although the convolution layer watches

the spread region, the activation contributes more to the decision making process. Since the activation layer is closest to the output, the activation layer must detect relevant features and in our Xception model it does the same.

In the final stage of our model visulization for two class classification, deep layer feature investigation, and model explainability, we checked the last convolution layer heatmap of all our implemented models.The GradCam analysis of covid and normal class photos at the final convolution layer in the Xception, Vgg19, and Resnet models is shown in Fig. 4.8. The first row shows a chest xray image from the covid and normal class. The GradCam activation mapping for the xray image is shown in the second row. Xray picture placed on an activation map to show the usperimposed image in the third row. The first three columns display the original image, heatmap, and overlay image at the final convolution layer of the Xception, Vgg19, and Resnet50 models for a normal class image. The last third column displays the original image, activation mapping, and superimposed image from the final convolution layer of Xception, Vgg19, and Resnet50 for a covid class image. The final convolution layers for the Xception, Vgg19 and Resnet models are block14 sep conv 2, block5 conv 4 and conv5 block 3_3 conv respectively. From the image shown, the Vgg19 and Xception models are more looking into the high lever features in search of desired regions. For normal class, all three models look into the desired region and all three models predict it perfectly. However, in covid class Vgg19 misclassified the image and from the heatmap we can see that although it is watching in the chest region, it came into prediction by watching in the other portion rather than the chest.

GradCam analysis of four class photos is shown in Fig. 4.9 and Fig. 4.10. Explainable AI at the final convolution layer in the Xception model is visualized in Fig. 4.9.First row: shows chest xray images of different classes. The GradCam activation mapping for the xray image is shown in the second row. Third row: Xray picture placed on an activation map to show the usperimposed image. The first column displays the original image, heatmap, and overlay image of the Normal class at the final convolution layer of the Xception model. The second, third, and fourth columns display the original image, activation mapping, and superimposed image for the Covid, Lung Opacity, and Viral Pneumonia classes respectively. All the images were correctly classified in this case and we can see the model is watching the region of interest to differentiate Normal, Covid, Lung Opacity and Viral Pneumonia classes. From Fig. 4.10 it is possible to visualize that Vgg19 misclassified Covid and Viral Pneumonia classes as the last layer is watching some other points rather than the desired region. However, the Xception and Resnet models detected image classes perfectly in this case.

4.2 Result Analysis of Lung Disease Analysis of CT Radiographs using AI

Table 4.3 and Fig. 4.11 summarizes the performance of the three CNN networks studied in this paper. We tested the results of the model using a different test set which the models did not view while they were trained.

From the 4.3 and Fig. 4.11, we can visualize that our proposed Incepion v2, Vgg16, and Resnet give 0.993, 0.998, and 0.915 accuracy consecutively. Inception and Vgg16

(a) ROC curve for Inception v3 model

(b) ROC curve for Vgg16

(c) ROC curve for Resnet50

(d) Confusion Matrix for Inception v3 model

(e) Confusion Matrix for VGG19

(f) Confusion Matrix for Resnet 50
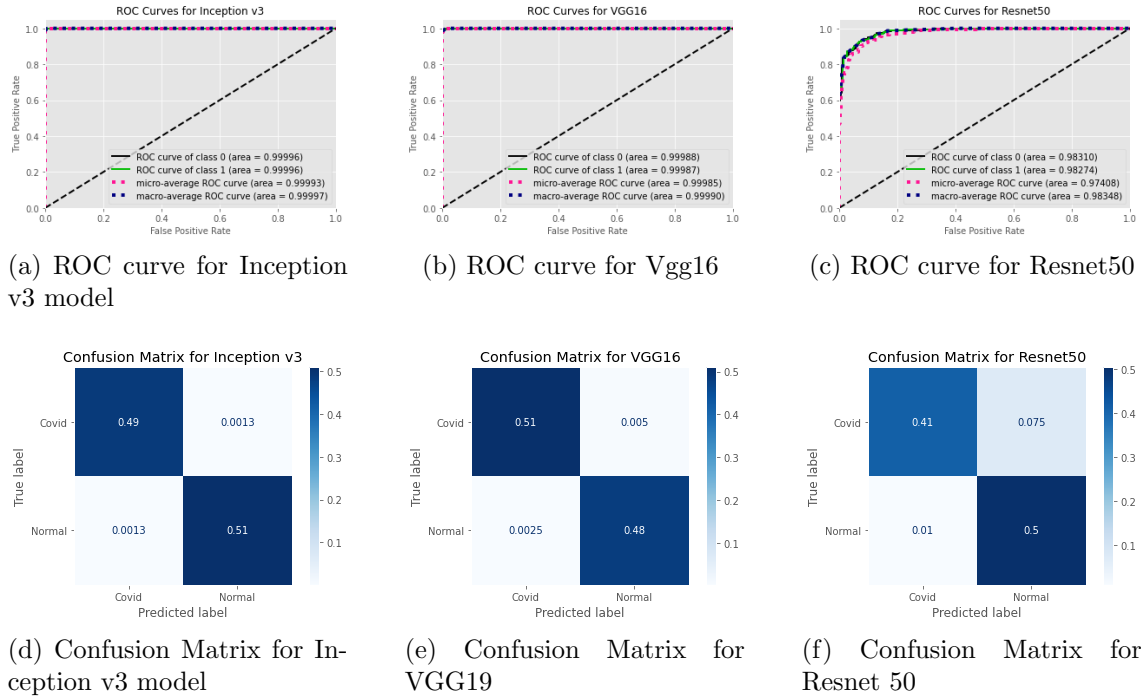
Figure 4.11: ROC curves and confusion matrices for each of the models(CT-Covid)

are performing better than Resnet here. The F1 score and area under the curve are also higher in Inception v3 and Vgg16. For the healthy class, precision and recall are higher in the Inception v3 model, whereas in Vgg16, pression and recall are similar for healthy and normal class pictures.

Table 4.3: Measures of performance for the three models studied in the research(CT-Covid).

Models	Accuracy	Class	Precision(PPV)	Recall(Sensitivity)	F1 Score	AUC	Explainable AI
VGG16	0.993	Healthy	0.99	0.995	0.992	0.99	Y
		Covid	0.995	0.99	0.993	0.99	
Inception v3	0.998	Healthy	0.998	0.998	0.998	0.99	Y
		Covid	0.977	0.977	0.977	0.99	
Resnet50	0.915	Healthy	0.87	0.981	0.922	0.98	y
		Covid	0.976	0.846	0.906	0.98	

We produced activation maps of the last convolution layer of the three models and compared them for covid and non covid images. Furthermore, to analyze, we superimposed the activation map with the original image to see which region is more critical to determining a certain class. The Fig. 4.12 shows the gradcam visualization of the last conv layers of Inception v3, Resnet50 and Vgg16 models.First row: shows a chest xray image from the normal and covid class. The GradCam activation mapping for the xray image is shown in the second row. Third row: Xray picture placed on an activation map to show the usperimposed image. The first three columns display the original image, heatmap, and overlay image at the final convolution layer of the Inception v3, Vgg16, and Resnet50 models for a normal class image. The last third column displays the original image, activation mapping, and superimposed image from the final convolution layer of Inception v3, Vgg16,

Figure 4.12: GradCam analysis of covid and normal class photos at the final convolution layer in the Inception v3, Vgg16, and Resnet models.

and Resnet50 for a covid class image. From the figure we can see that all the models are observing the chest region to conclude a certain class. However, Vgg16 is watching more high-level features and the covid region while deciding a covid and normal cl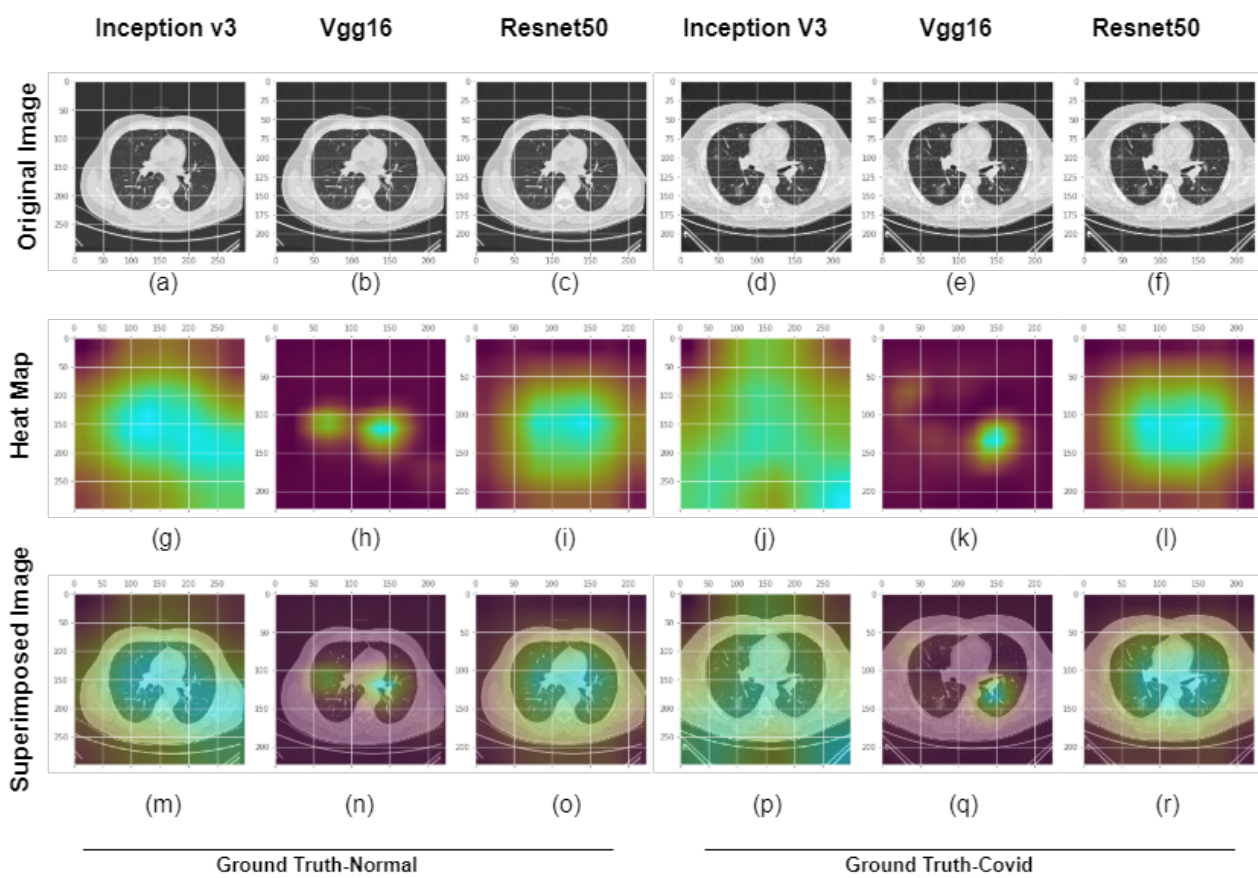ass images. Despite the fact that the Inception v3 model achieved great accuracy, gradcam analysis suggests that Vgg16 is the preferred model in this circumstance. Vgg16 can help doctors not only to find covid or normal class, but can also assist in detecting which regions are affected.

4.3 Result Analysis of Kidney Disease Analysis of CT Radiographs using AI

The results of the implemented six models using different tests are evaluated by calculating the accuracy, recall, F1 score (F1), accuracy (Acc), positive predictive value (PPV), and ROC curve area of interest (AUC). Table 4.4, Fig. 4.13 and Fig. 4.14 summarizes the performance of the six networks studied in this paper. Fig. 4.16 presents us with the gradcam analysis of the Inception v3, Resnet50, and Vgg16 models, and Fig. 4.15 provides us with a time comparison of the different models' training times.

From the table 4.4, we can see that the InceptionV3 model performed worse with our dataset and gave an accuracy of 61.60%. EANet and Resnet 50 performed moderately by giving accuracy of 77.02% and 73.80%. CCT, VGG16 and Swin Transformers provided accuracy of 96.54%, 98.20% and 99.30% accuracy respectively. The Swin transformer, which is a transformer-based model, is outperforming all the other models in respect of accuracy. For the cyst, normal, stone, and tumor classes, the highest F1 score is provided by the swin transformer also, and the numbers are 0.996, 0.998, 0.985, and 0.996 consecutively. The Swin transformer also provides the highest precision for Stone and Tumor classes, and readings are 0.981 and 0.993. For the cyst class, the Swin transformer and VGG 16 are providing the same value of 0.996, whereas for the normal class, the Swin transformer is performing better and giving a reading of 0.996. Considering the above, the Swin transformer is superior and outperforms all the models, and can be of great use in kidney medical imaging diagnosis.

From Fig. 4.13 and Fig. 4.14, we can see that the Area Under the Curve and F1 Score are found promising in the case of CCT, VGG16, and SWin Transformers, compared to Resnet50, EANet, and Inception v3. The F1 score is the highest of all the classes-cyst, normal, stone and tumor in the case of the Swin transformer model.

When it comes to model training time, the CCT and Swin transformers take less time to train the same amount of epochs. Because transformer models have a small number of parameters, training them takes less time. As a result, we can conclude that the Swin transformer requires less training time and has higher accuracy, making it superior in renal image diagnosis.

We analyzed the activation maps of the three models' final convolution layer for and analyzed the activation map by superimposing it on top of the original image to identify which region is more important in determining a specific class. From the figures, it can be concluded that VGG 16 is watching more regions of interest in the desired anatomy than the other models, and its accuracy is better than Resnet50

Table 4.4: Measures Of Performance For The Six models studied in the research(kidney Disease).

| Models	Accuracy	Class	Precision(PPV)	Recall(Sensitivity)	F1 Score	AUC
EANet	77.02%	Cyst	0.593	1	0.745	0.98
		Normal	0.896	0.848	0.871	0.98
		Stone	0.845	0.495	0.624	0.91
		Tumor	0.93	0.777	0.847	0.97
Swin Transformers	99.30%	Cyst	0.996	0.996	0.996	0.99993
		Normal	0.996	0.981	0.988	0.9998
		Stone	0.981	0.989	0.985	0.99975
		Tumor	0.993	1	0.996	1
CCT	96.54%	Cyst	0.968	0.923	0.945	0.99605
		Normal	0.989	0.975	0.982	0.99841
		Stone	0.94	1	0.969	0.99924
		Tumor	0.964	0.964	0.964	0.99723
VGG16	98.20%	Cyst	0.996	0.968	0.982	0.99856
		Normal	0.985	0.973	0.979	0.99844
		Stone	0.966	0.988	0.977	0.99908
		Tumor	0.982	0.996	0.989	0.99902
Inception v3	61.60%	Cyst	0.645	0.826	0.724	0.92689
		Normal	0.584	0.898	0.708	0.90642
		Stone	0.568	0.462	0.509	0.78185
		Tumor	0.76	0.295	0.425	0.8029
Resnet50	73.80%	Cyst	0.735	0.641	0.685	0.90721
		Normal	0.77	0.79	0.78	0.95069
		Stone	0.745	0.692	0.717	0.9314
		Tumor	0.706	0.827	0.762	0.94447
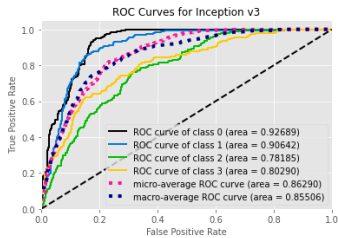
(a) ROC curve for Inception v3 model
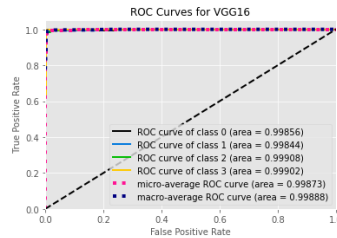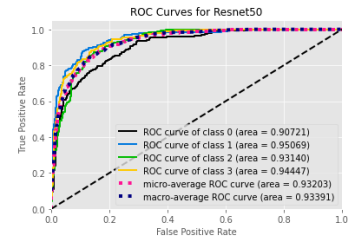
(b) ROC curve for Vgg16

(c) ROC curve for Resnet50
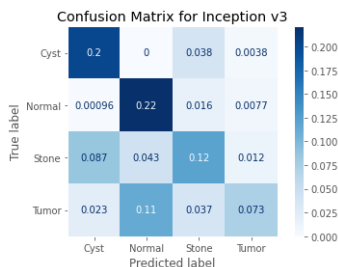
(d) Confusion Matrix for Inception v3 model
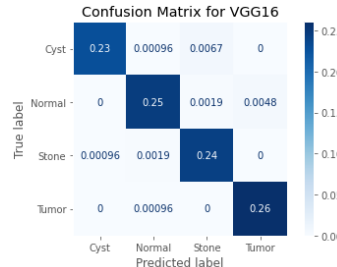
(e) Confusion Matrix for VGG16
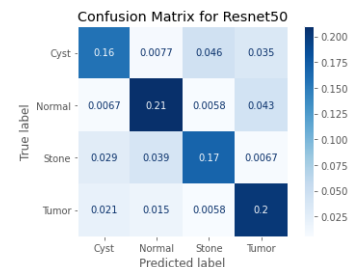
(f) Confusion Matrix for Resnet 50

Figure 4.13: ROC curves and confusion matrices for Inception v3, VGG16 and Resnet 50.

(a) ROC curve for EANet

(b) ROC curve for Swin transformer

(c) ROC curve for CCT

(d) Confusion Matrix for EANet

(e) Confusion Matrix for Swin transformer

(f) Confusion Matrix for CCT
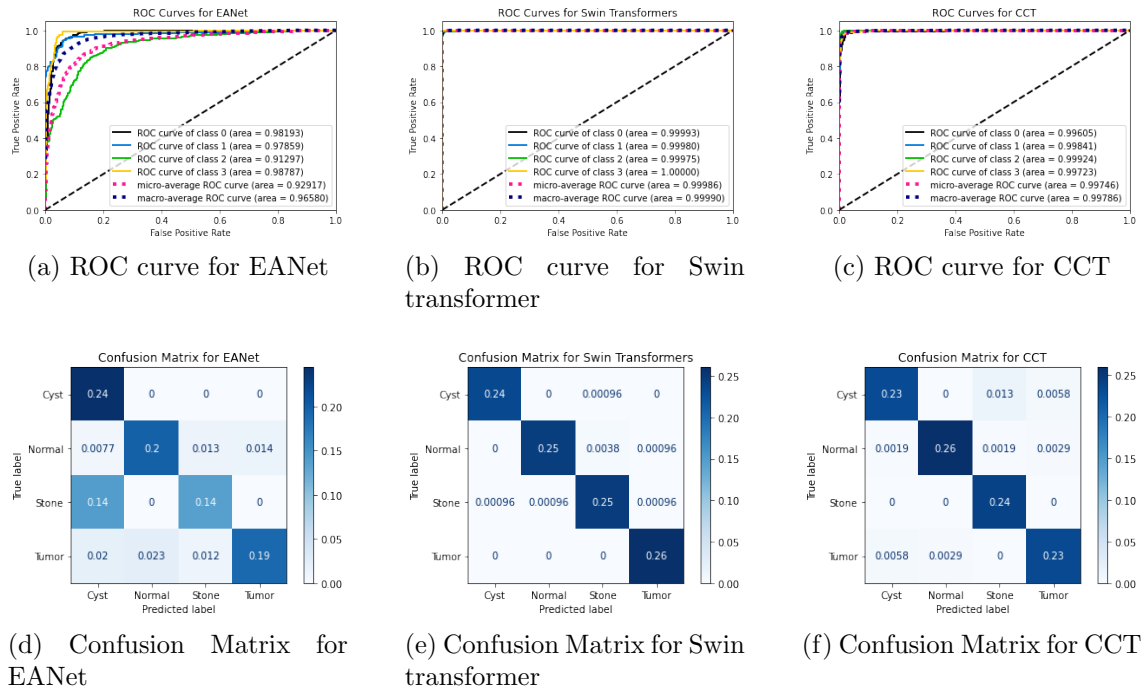
Figure 4.14: ROC curves and confusion matrices for EaNet, Swin transformer and CCT.
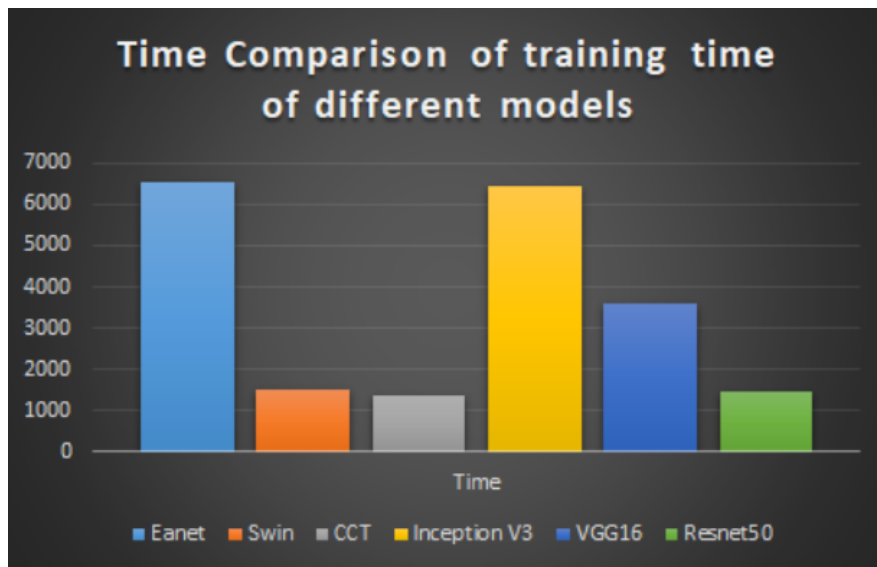
Figure 4.15: Training time comparison for 100 epochs of different models.
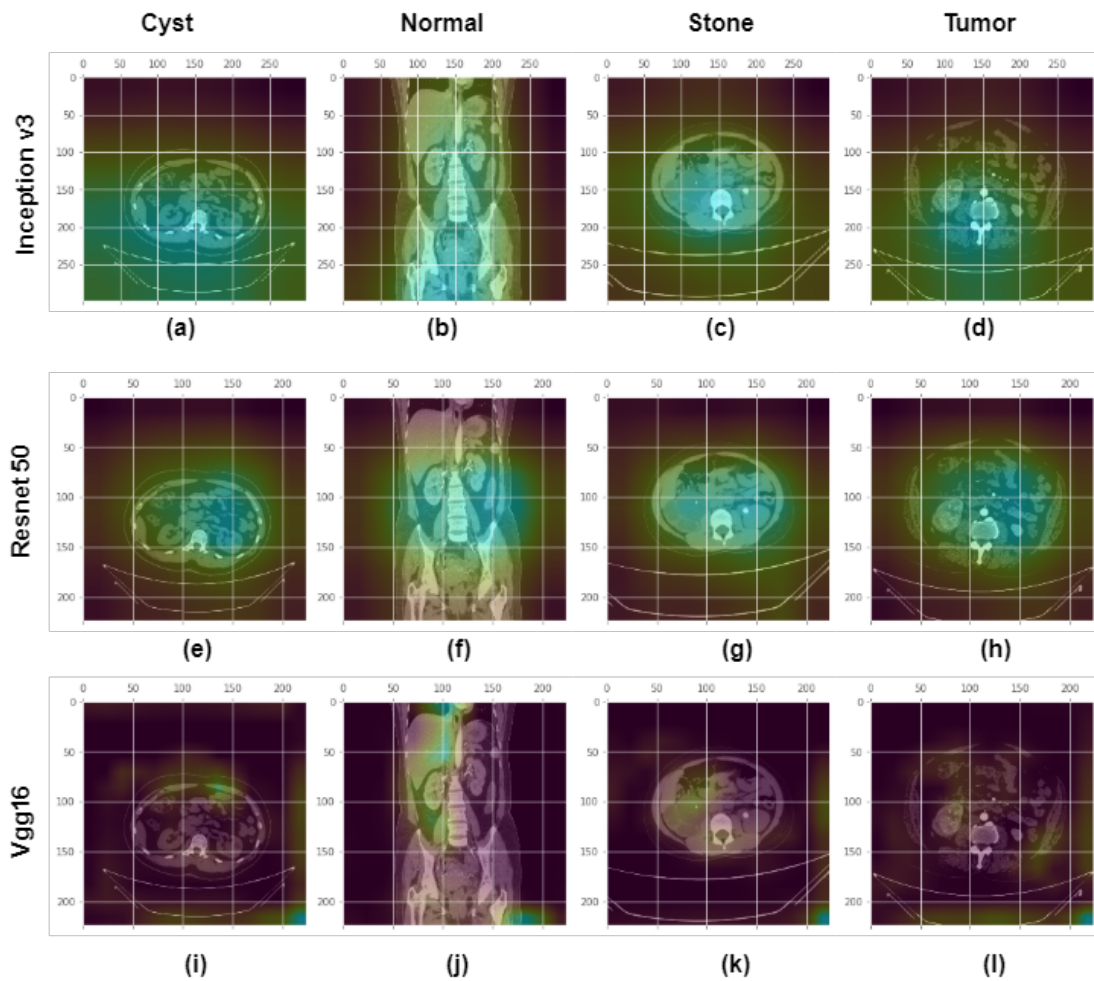
Figure 4.16: GradCam analysis of kidney Cyst, Normal, Stone and Tumor class photos at the final convolution layer in the Inception v3, Vgg16, and Resnet models.

and Inception v3, which are also resonant with the analysis. In our case, VGG16 is predicting all the images as correct class and watching the region of interest region perfectly, whereas Resnet is predicting normal findings as tumors and stones as normal in this case and also not watching where the model should watch to make a decision. Inception V3 is also not watching the region of interest perfectly and watching more low-level features, and in this case, it predicated the tumor class as the normal class.

Chapter 5

Conclusion

This study presents a total of six deep CNN-based transfer learning approaches . Three approaches are used for automatically recognizing COVID-19 disease from X-ray images, and the other three are used to detect normal, covid, lung opacity, and viral pneumonia images. Using chest X-ray pictures, three prominent and previously published CNN-based deep learning systems were modified to better distinguish the images for dual and four class images. Models were trained with pre-trained image-net weight, verified, and tested with an unseen dataset. Experiments were carried out in two ways for dual class classification; first, models were trained with the image size as per the standard requirements of the model; 299×299 for the Xception model, 224×224 for the VGG19, and the Resnet50 model. Second, the proposed three models were trained with 512×512 pixel input images. It is concluded that image input of 512×512 pixels delivers greater accuracy, which might be because image resizing loses some essential information from photos, lowering accuracy. It is critical to train the model with an image size that does not lose any meaningful information, as this is a healthcare-related sensitive diagnostic. With standard picture input, the accuracy of Xception, VGG19, and Resnet is 0.97, 0.97, and 0.925, respectively. With 512 by 512 pixels, the accuracy is 0.975, 0.975, and 0.933. Furthermore, Explainable AI was applied to three networks, and it was revealed that the Xception model was watching the chest region where Covid was present, suggesting that Xception was the superior of the three models. It was also evident from the experiment that the last block of all models watches relevant information for the classification, and the last block activation layer is responsible for detecting more complicated features as it is close to the model output. For the multi class classification, the Xception model gives an accuracy of 93% and the Xception model watches the relevant portion of the chest while detecting the four class images. COVID-19 has already posed a danger to the global healthcare system, with respiratory failure being the primary cause of death. Because doctors' time is restricted due to the enormous number of patients seen outside or in an emergency, the Xception model-based diagnosis presented in this study has the potential to save lives through early screening and appropriate therapy. Not only can our explainability serve as a responsible and transparent audit of our models, but it may also aid physicians in enhancing covid screening.

Furthermore, This study presents a total of 6082 chest HRCT images and made them public to contribute to the healthcare AI. Exploratory analysis of the images

performed and shown data varied from highly affected to covid to low symtomp images. Furthermore, this study developed three deep learning based models to classify COVID and non-COVID from the CT images. Three popularly known deep learning models, Resnet, Vgg16, and Inception v3, are modified and tweaked in the last layers. Models were trained with pretrained imagenet weight and tested with unknown data, and it was found that all three models were performing reasonably. With standard picture input, the accuracy of Inception v3, VGG16, and Resnet is 0.998, 0.993, and 0.915, respectively. Models' blackboxes are demistified and found VGG16 is watching more high level features and watching the covid affected region more while taking any decision about the class. We believe our models' superior accuracy and models' region of interest while detecting certain classes can help medical doctors not only detect COVID but also decide how severe the condition is, thus having the potential to reduce sufferings around the world.

Moreover, This thesis collected, annotated, and published a total of 12,446 whole abdomen and urogram CT scan images containing cysts, tumors, normal, and stone findings. Exploratory data analysis of the images was performed and showed that the images from all the classes had the same type of mean colour distribution. Furthermore, this study developed six models, out of which three models are based on recent state-of-the-art variants of the Vision transformers EANet, CCT, and Swin transformers, and the other three are based on popularly known deep learning models, Resnet, Vgg16, and Inception v3, which are tweaked in the last few layers. A comparison of all the models performed revealed that, while VGG16 and CCT performed well, the swin transformer outperformed all the models in terms of accuracy, providing an accuracy of 99.30%. The F1 score and precision and recall comparison provide evidence that the Swin transformer is outperforming all the models, and also, compared to all the models, the Swin transformer is taking less time to train the same number of epochs. The study also revealed the blackbox of VGG16, Resnet50, and Inception models and showed that VGG16 is better compared to Resnet50 and Inceptionv3 and watches the desired abnormalities in the anatomy better. We believe the superior accuracy of our model based on the Swin transformer and the VGG16-based model can both be of great use in detecting kidney tumors, cysts, and stones, and can reduce the pain and suffering of patients.

Chapter 6

Limitations and Future Direction

The most challenging task is to get data as healthcare data is scarce and limited. We feel more data is required to do more extensive study and to build better model that can see and diagnosis the diseases very early stages. Different anatomy diseases can be targeted in future and can be approached 3D model. We dream to see a combined system to detect all the diseases that is required for human being so that peoples pain can be alleviated significantly.

Bibliography

[1] C. J. van Rijsbergen, "An algorithm for information structuring and retrieval," *Comput. J.*, vol. 14, no. 4, pp. 407–412, Jan. 1971, ISSN: 0010-4620. DOI: 10.1093/comjnl/14.4.407. [Online]. Available: https://academic.oup.com/comjnl/article/14/4/407/325200.

[2] K. C. Saw, J. A. McAteer, A. G. Monga, G. T. Chua, J. E. Lingeman, and J. C. Williams Jr, "Helical ct of urinary calculi: Effect of stone composition, stone size, and scan collimation," *American Journal of Roentgenology*, vol. 175, no. 2, pp. 329–332, 2000.

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.

[4] S. Das and P. Dutta, "Chronic kidney disease prevalence among health care providers in bangladesh.," *Mymensingh medical journal: MMJ*, vol. 19, no. 3, pp. 415–421, 2010.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[6] S. Jacobson, "Chronic kidney disease–a public health problem?" *Lakartidningen*, vol. 110, no. 21, pp. 1018–1020, 2013.

[7] V. Jha, G. Garcia-Garcia, K. Iseki, *et al.*, "Chronic kidney disease: Global dimension and perspectives," *The Lancet*, vol. 382, no. 9888, pp. 260–272, 2013.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[9] S. Patro and K. K. Sahu, "Normalization: A preprocessing stage," *arXiv preprint arXiv:1503.06462*, 2015.

[10] C. Szegedy, W. Liu, Y. Jia, *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[12] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep networks with stochastic depth," in *European conference on computer vision*, Springer, 2016, pp. 646–661.

[13] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[14] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[15] J. J. Hsieh, M. P. Purdue, S. Signoretti, *et al.*, "Renal cell carcinoma," *Nature reviews Disease primers*, vol. 3, no. 1, pp. 1–19, 2017.

[16] 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," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[17] J. Verma, M. Nath, P. Tripathi, and K. Saini, "Analysis and identification of kidney stone using k th nearest neighbour (knn) and support vector machine (svm) classification techniques," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 574–580, 2017.

[18] T. Alelign and B. Petros, "Kidney stone disease: An update on current concepts," *Advances in urology*, vol. 2018, 2018.

[19] N. Blau, E. Klang, N. Kiryati, M. Amitai, O. Portnoy, and A. Mayer, "Fully automatic detection of renal cysts in abdominal ct scans," *International journal of computer assisted radiology and surgery*, vol. 13, no. 7, pp. 957–966, 2018.

[20] K. J. Foreman, N. Marquez, A. Dolgert, *et al.*, "Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: Reference and alternative scenarios for 2016–40 for 195 countries and territories," *The Lancet*, vol. 392, no. 10159, pp. 2052–2090, 2018.

[21] Paul Mooney, *Chest X-Ray Images (Pneumonia) — Kaggle*, 2018. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia (visited on 09/04/2021).

[22] T. Ho-Phuoc, "Cifar10 to compare visual recognition performance between deep neural networks and humans," *arXiv preprint arXiv:1811.07270*, 2018.

[23] RSNA, *RSNA Pneumonia Detection Challenge*, 2018. [Online]. Available: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data (visited on 09/04/2021).

[24] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *International conference on artificial neural networks*, Springer, 2018, pp. 270–279.

[25] A. J. Brownstein, S. U. B. Mahmood, A. Saeyeldin, *et al.*, "Simple renal cysts and bovine aortic arch: Markers for aortic disease," *Open Heart*, vol. 6, no. 1, e000862, 2019.

[26] J. Cui, F. Li, and Z. L. Shi, "Origin and evolution of pathogenic coronaviruses," *Nat. Rev. Microbiol.*, vol. 17, no. 3, pp. 181–192, Mar. 2019, ISSN: 17401534. DOI: 10.1038/s41579-018-0118-9. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30531947/.

[27] A. Parakh, H. Lee, J. H. Lee, B. H. Eisner, D. V. Sahani, and S. Do, "Urinary stone detection on ct images using deep convolutional neural networks: Evaluation of model performance and generalization," *Radiology: Artificial Intelligence*, vol. 1, no. 4, e180066, 2019.

[28] C. Rediger, L. Guerra, M. Keays, *et al.*, "Renal cyst evolution in childhood: A contemporary observational study," *Journal of pediatric urology*, vol. 15, no. 2, 188–e1, 2019.

[29] E. Sanna, S. Loukogeorgakis, T. Prior, *et al.*, "Fetal abdominal cysts: Antenatal course and postnatal outcomes," *Journal of perinatal medicine*, vol. 47, no. 4, pp. 418–421, 2019.

[30] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, PMLR, 2019, pp. 6105–6114.

[31] H. Zhang, Y. Chen, Y. Song, Z. Xiong, Y. Yang, and Q. J. Wu, "Automatic kidney lesion detection for ct images using morphological cascade convolutional neural networks," *IEEE Access*, vol. 7, pp. 83 001–83 011, 2019.

[32] Q. Zheng, S. L. Furth, G. E. Tasian, and Y. Fan, "Computer-aided diagnosis of congenital abnormalities of the kidney and urinary tract in children based on ultrasound imaging data by integrating texture image features and deep transfer learning image features," *Journal of pediatric urology*, vol. 15, no. 1, 75–e1, 2019.

[33] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images," *Pattern Recognition Letters*, vol. 138, pp. 638–643, 2020.

[34] Alicia Victoria Lozano, *Doctors, nurses 'can't take much more' amid Covid-19 surge in Southern California*, Dec. 2020. [Online]. Available: https://www.nbcnews.com/news/us-news/doctors-nurses-can-t-take-much-more-amid-covid-19-n1252582 (visited on 09/03/2021).

[35] A. Altan and S. Karasu, "Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique," *Chaos, Solitons and Fractals*, vol. 140, p. 110 071, Nov. 2020, ISSN: 09600779. DOI: 10.1016/j.chaos.2020.110071.

[36] J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes, "A distinctive approach in brain tumor detection and classification using mri," *Pattern Recognition Letters*, vol. 139, pp. 118–127, 2020.

[37] J. D. Arias-Londono, J. A. Gomez-Garcia, L. Moro-Velazquez, and J. I. Godino-Llorente, "Artificial Intelligence applied to chest X-Ray images for the automatic detection of COVID-19. A thoughtful evaluation approach," *IEEE Access*, 2020, ISSN: 21693536. DOI: 10.1109/ACCESS.2020.3044858. arXiv: 2011.14259.

[38] J. D. Arias-Londoño, J. A. Gomez-Garcia, L. Moro-Velázquez, and J. I. Godino-Llorente, "Artificial intelligence applied to chest x-ray images for the automatic detection of covid-19. a thoughtful evaluation approach," *IEEE Access*, vol. 8, pp. 226 811–226 827, 2020.

[39] H. X. Bai, B. Hsieh, Z. Xiong, *et al.*, "Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT," *Radiology*, vol. 296, no. 2, E46–E54, Aug. 2020, ISSN: 15271315. DOI: 10.1148/radiol.2020200823. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32155105/.

[40] M. E. Chowdhury, T. Rahman, A. Khandakar, *et al.*, "Can AI Help in Screening Viral and COVID-19 Pneumonia?" *IEEE Access*, vol. 8, pp. 132 665–132 676, Mar. 2020, ISSN: 21693536. DOI: 10.1109/ACCESS.2020.3010287. arXiv: 2003.13145. [Online]. Available: https://arxiv.org/abs/2003.13145v3.

[41] J. Civit-Masot, F. Luna-Perejón, M. D. Morales, and A. Civit, "Deep learning system for COVID-19 diagnosis aid using X-ray pulmonary images," *Appl. Sci.*, vol. 10, no. 13, p. 4640, Jul. 2020, ISSN: 20763417. DOI: 10.3390/app10134640. [Online]. Available: https://www.mdpi.com/2076-3417/10/13/4640/htm%20https://www.mdpi.com/2076-3417/10/13/4640.

[42] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[43] K. El Asnaoui and Y. Chawki, "Using X-ray images and deep learning for automated detection of coronavirus disease," *J. Biomol. Struct. Dyn.*, pp. 1–12, 2020, ISSN: 15380254. DOI: 10.1080/07391102.2020.1767212. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/07391102.2020.1767212.

[44] A. Haghanifar, M. Majdabadi, and S. Ko, *COVID-19 Chest X-Ray Image Repository*, 2020. [Online]. Available: https://web.archive.org/save/https://figshare.com/articles/dataset/COVID-19%7B%5C_%7DChest%7B%5C_%7DX-Ray%7B%5C_%7DImage%7B%5C_%7DRepository/12580328 (visited on 09/04/2021).

[45] A. Haghanifar, M. M. Majdabadi, Y. Choi, S. Deivalakshmi, and S. Ko, "COVID-CXNet: Detecting COVID-19 in Frontal Chest X-ray Images using Deep Learning," Jun. 2020. arXiv: 2006.13807. [Online]. Available: http://arxiv.org/abs/2006.13807.

[46] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," Mar. 2020. arXiv: 2003.11055. [Online]. Available: http://arxiv.org/abs/2003.11055.

[47] S. A. A. Ismael, A. Mohammed, and H. Hefny, "An enhanced deep learning approach for brain cancer mri images classification using residual networks," *Artificial intelligence in medicine*, vol. 102, p. 101 779, 2020.

[48] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the covid-19 infected patients using densenet201 based deep transfer learning," *Journal of Biomolecular Structure and Dynamics*, pp. 1–8, 2020.

[49] A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, ISSN: 18727565. DOI: 10.1016/j.cmpb.2020.105581. arXiv: 2004.04931. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32534344/.

[50] A. Kolesnikov, L. Beyer, X. Zhai, *et al.*, "Big transfer (bit): General visual representation learning," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, Springer, 2020, pp. 491–507.

[51] Liob, *Covid-19-image-repository/png at master · ml-workgroup/covid-19-image-repository*, 2020. [Online]. Available: https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png (visited on 09/04/2021).

[52] M. Loey, F. Smarandache, and N. E. M. Khalifa, "Within the lack of chest COVID-19 X-ray dataset: A novel detection model based on GAN and deep transfer learning," *Symmetry (Basel).*, vol. 12, no. 4, p. 651, Apr. 2020, ISSN: 20738994. DOI: 10.3390/SYM12040651. [Online]. Available: https://www.mdpi.com/2073-8994/12/4/651/htm%20https://www.mdpi.com/2073-8994/12/4/651.

[53] Medical Imaging Databank of the Valencia region BIMCV, *Bimcv-Covid19 – Bimcv*, 2020. [Online]. Available: https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/%7B%5C%%7D7B%7B%5C%%7D5C%7B%5C#%7D%7B%5C%%7D7D1590859488150-148be708-c3f3 (visited on 09/04/2021).

[54] Y. Oh, S. Park, and J. C. Ye, "Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets," *IEEE Trans. Med. Imaging*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020, ISSN: 1558254X. DOI: 10.1109/TMI.2020.2993291. arXiv: 2004.05758.

[55] 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," *Comput. Biol. Med.*, vol. 121, Jun. 2020, ISSN: 18790534. DOI: 10.1016/j.compbiomed.2020.103792. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32568675/.

[56] J. Paul Cohen, P. Morison, and Lan Dao, *Open database of COVID-19 cases with chest X-Ray or CT Images*, 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset (visited on 09/03/2021).

[57] I. S. o. M. Radiology and Interventional, *Covid-19 Database — Sirm*, 2020. [Online]. Available: https://www.sirm.org/category/%20senza-categoria/covid-19/ (visited on 09/03/2021).

[58] S. Sudharson and P. Kokil, "An ensemble of deep neural networks for kidney ultrasound image classification," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105 709, 2020.

[59] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection," *IEEE Access*, vol. 8, pp. 91 916–91 923, Mar. 2020, ISSN: 21693536. DOI: 10.1109/ACCESS.2020.2994762. arXiv: 2103.05094. [Online]. Available: https://arxiv.org/abs/2103.05094v1.

[60] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[61] W. Wang, Y. Xu, R. Gao, *et al.*, "Detection of SARS-CoV-2 in Different Types of Clinical Specimens," *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 18, pp. 1843–1844, May 2020, ISSN: 15383598. DOI: 10.1001/jama.2020.3786. [Online]. Available: /pmc/articles/PMC7066521/%20/pmc/articles/PMC7066521/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7066521/.

[62] WHO, *WHO/Europe — Coronavirus disease (COVID-19) outbreak - About the virus*, 2020. [Online]. Available: https://doi.org/10.1080/23120053.2018.1504532%7B%5C%%7D0Ahttp://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/novel-coronavirus-2019-ncov (visited on 09/03/2021).

[63] ——, *World Health Organization. 2020a. Novel Corona-virus (2019-nCoV). Situation Report - 1. 21 January 2020.* 2020. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/330760/nCoVsitrep21Jan2020-eng.pdf?sequence=3%7B%5C&%7DisAllowed=y (visited on 09/03/2021).

[64] World Health Organization, *Criteria for releasing COVID-19 patients from isolation*, 2020. [Online]. Available: https://www.who.int/publications/i/item/criteria-for-releasing-covid-19-patients-from-isolation (visited on 09/03/2021).

[65] X. Xu, X. Jiang, C. Ma, *et al.*, "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020.

[66] S. Zhao, Q. Lin, J. Ran, *et al.*, "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," *Int. J. Infect. Dis.*, vol. 92, p. 214, Mar. 2020. DOI: 10.1016/J.IJID.2020.01.050. [Online]. Available: /pmc/articles/PMC7110798/%20/pmc/articles/PMC7110798/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7110798/.

[67] C. Zheng, X. Deng, Q. Fu, *et al.*, "Deep learning-based detection for covid-19 from chest ct using weak label," *MedRxiv*, 2020.

[68] M. Abdulkareem and S. E. Petersen, "The promise of ai in detection, diagnosis, and epidemiology for combating covid-19: Beyond the hype," *Frontiers in Artificial Intelligence*, vol. 4, p. 53, 2021.

[69] Carolyn-Admin, *How Long is Too Long for Emergent CT Imaging in Rural Communities? : BC Emergency Medicine Network*, Jun. 2021. [Online]. Available: https://www.bcemergencynetwork.ca/research/how-long-is-too-long-for-emergent-ct-imaging-in-rural-communities/ (visited on 09/03/2021).

[70] CDC, *Different Groups of People — CDC*, 2021. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/index.html (visited on 09/03/2021).

[71] Centers for Disease Control and Prevention, *Symptoms of COVID-19 — CDC*, 2021. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html (visited on 09/03/2021).

[72] S. Chakrabarty, A. Sotiras, M. Milchenko, P. LaMontagne, M. Hileman, and D. Marcus, "Mri-based identification and classification of major intracranial tumor types by using a 3d convolutional neural network: A retrospective multi-institutional analysis," *Radiology: Artificial Intelligence*, vol. 3, no. 5, e200301, 2021.

[73] X. Fu, H. Liu, X. Bi, and X. Gong, "Deep-learning-based ct imaging in the quantitative evaluation of chronic kidney diseases," *Journal of Healthcare Engineering*, vol. 2021, 2021.

[74] M.-H. Guo, Z.-N. Liu, T.-J. Mu, and S.-M. Hu, "Beyond self-attention: External attention using two linear layers for visual tasks," *arXiv preprint arXiv:2105.02358*, 2021.

[75] K. Harmon, *COVID-Overwhelmed Hospitals Strain Staff and Hope to Avoid Rationing Care*, Jan. 2021. [Online]. Available: https://www.scientificamerican.com/article/covid-overwhelmed-hospitals-strain-staff-and-hope-to-avoid-rationing-care1/ (visited on 09/03/2021).

[76] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li, and H. Shi, "Escaping the big data paradigm with compact transformers," *arXiv preprint arXiv:2104.05704*, 2021.

[77] M. N. Islam, *Hrct chest covid data -ct scan*, 2021. [Online]. Available: https://www.kaggle.com/nazmul0087/hrct-chest-covid-data-ct-scan.

[78] ——, *Ct kidney dataset: Normal-cyst-tumor and stone*, 2021. [Online]. Available: https://www.kaggle.com/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone.

[79] S. Jadon, "Covid-19 detection from scarce chest x-ray image data using few-shot deep learning approach," in *Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications*, International Society for Optics and Photonics, vol. 11601, 2021, p. 116010X.

[80] J. Liu, B. Dong, S. Wang, *et al.*, "Covid-19 lung infection segmentation with a novel two-stage cross-domain transfer learning framework," *Medical Image Analysis*, vol. 74, p. 102 205, 2021.

[81] Z. Liu, Y. Lin, Y. Cao, *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," *arXiv preprint arXiv:2103.14030*, 2021.

[82] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Anal. Appl.*, vol. 24, no. 3, pp. 1207–1220, Mar. 2021, ISSN: 1433755X. DOI: 10.1007/s10044-021-00984-y. arXiv: 2003.10849. [Online]. Available: https://arxiv.org/abs/2003.10849v3.

[83] Nick Klenske, *Rural Radiology Equity*, Jun. 2021. [Online]. Available: https://www.rsna.org/news/2021/june/Rural%20Radiology%20Equity (visited on 09/03/2021).

[84] M. K. Pandit, S. A. Banday, R. Naaz, and M. A. Chishti, "Automatic detection of covid-19 from chest radiographs using deep learning," *Radiography*, vol. 27, no. 2, pp. 483–489, 2021.

[85] Rahman Tawsifur, *COVID-19 Radiography Database — Kaggle*, 2021. [Online]. Available: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database/activity%7B%5C%%7D0Ahttps://www.kaggle.com/tawsifurrahman/covid19-radiography-database (visited on 09/03/2021).

[86] B. G. Santa Cruz, M. N. Bossa, J. Soelter, and A. D. Husch, "Public covid-19 x-ray datasets and their impact on model bias-a systematic review of a significant problem," *medRxiv*, 2021.

[87] P. Soda, N. C. D'Amico, J. Tessadori, *et al.*, "Aiforcovid: Predicting the clinical outcomes in patients with covid-19 applying ai to chest-x-rays. an italian multicentre study.," *Medical image analysis*, p. 102 216, 2021.

[88] S. M. Sozio, K. A. Pivert, F. J. Caskey, and A. Levin, "The state of the global nephrology workforce: A joint asn–era-edta–isn investigation," *Kidney international*, 2021.

[89] N. Sverzellati, C. J. Ryerson, G. Milanese, *et al.*, "Chest x-ray or CT for COVID-19 pneumonia? Comparative study in a simulated triage setting," *Eur. Respir. J.*, p. 2 004 188, Feb. 2021, ISSN: 0903-1936. DOI: 10.1183/13993003.04188-2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33574070/.

[90] N. Sverzellati, C. J. Ryerson, G. Milanese, *et al.*, "Chest x-ray or ct for covid-19 pneumonia? comparative study in a simulated triage setting," *European Respiratory Journal*, 2021.

[91] Whitney J. Palmer, *Chest X-ray Can Predict Outcomes for Patients With COVID-19*, Apr. 2021. [Online]. Available: https://www.diagnosticimaging.com/view/chest-x-ray-can-predict-outcomes-for-patients-with-covid-19 (visited on 09/03/2021).

[92] WHO, *Tracking SARS-CoV-2 variants*, 2021. [Online]. Available: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (visited on 09/03/2021).

[93] World Health Organization, *WHO Coronavirus (COVID-19) Dashboard. WHO Coronavirus (COVID-19) Dashboard With Vaccination Data*, 2021. [Online]. Available: https://covid19.who.int/ (visited on 09/03/2021).

[94] K. Yildirim, P. G. Bozdag, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Deep learning model for automated kidney stone detection using coronal ct images," *Computers in Biology and Medicine*, p. 104 569, 2021.

[95] J. Zhang, Y. Xie, G. Pang, *et al.*, "Viral Pneumonia Screening on Chest X-Rays Using Confidence-Aware Anomaly Detection," *IEEE Trans. Med. Imaging*, vol. 40, no. 3, pp. 879–890, Mar. 2021, ISSN: 1558254X. DOI: 10.1109/TMI.2020.3040950. arXiv: 2003.12338. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33245693/.

[96] I. AKSAKALLI, S. KAÇDIOĞLU, and Y. S. HANAY, "Kidney x-ray images classification using machine learning and deep learning methods," *Balkan Journal of Electrical and Computer Engineering*, vol. 9, no. 2, pp. 144–151,