DETECTING COVID-19 IN X-RAY IMAGES WITH DEEP LEARNING

ETHEL AW MIAO HAN

UNIVERSITI TUNKU ABDUL RAHMAN

DETECTING COVID-19 IN X-RAY IMAGES WITH DEEP LEARNING

ETHEL AW MIAO HAN

A project report submitted in partial fulfilment of the requirements for the award of Bachelor of Science Software Engineering with Honours

Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

May 2023

DECLARATION

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

Signature :

Name : ETHEL AW MIAO HAN

ID No. : 1904486

Date : 17/5/2023

APPROVAL FOR SUBMISSION

I certify that this project report entitled **"DETECTING COVID-19 IN X-RAY IMAGES WITH DEEP LEARNING"** was prepared by **ETHEL AW MIAO HAN** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Software Engineering with Honours at Universiti Tunku Abdul Rahman.

Approved by,

Signature	:	kckhor
Supervisor	:	Khor Kok Chin
Date	:	17/5/2023

ACKNOWLEDGEMENTS

In the preparation, production and completion of this final year project, a number of people who in one way or another have stretched out their helping hands to me. It is with fervent gratitude I would like to sincerely acknowledge their assistance and participation. This research might not have seen the light of the day with the absence of any of them.

First and foremost, I would like to express my utmost appreciation and warmest thanks to my supervisor Dr. Khor Kok Chin for enlightening me at the first glance of research. Without her dedicated guidance, invaluable advice, genuine enthusiasm and tremendous understanding at every stage throughout the research, this final year project would have never been accomplished. It is truly an honour and pleasure for me to have my final year project under supervision of Dr. Khor. His immense knowledge, professional expertise and plentiful experience in the field of deep learning have been indispensable in enabling me to develop a deeper understanding of this intricate and fast-paced area of research. His patience, generosity, and willingness to go above and beyond in helping me navigate the challenges of this project have been truly inspiring.

Furthermore, my profound and eternal thanks go to my family members for their unconditional love, affection and motivation. I am indebted to my parents Aw Siew Keon and Ta Sock Cher, for giving birth to me in the first place and raising me up with constant devotion and dedication. Earnest thanks also convey to my siblings who support me morally throughout the execution of this project and daily life. Without them walking side-by-side in this journey, this project would not have been the same.

Moreover, I would like to extend my sincere thanks to the online sources who made their datasets publicly available, which were instrumental in the success of this project. Their contributions to the scientific community have been immensely valuable, and I am grateful for their willingness to share their data.

Last but not least, I would like to thank all the researchers and practitioners in the field of deep learning who have contributed to the development of this exciting area of research. Their work has inspired me and paved the way for this project.

ABSTRACT

The Corona Virus Disease-2019 (COVID-19) has had a profound impact on the world and thus creates awareness of the need for a fast and accurate diagnosis if a similar outbreak occurs again. Chest X-Ray (CXR) is widely used to detect COVID-19 manually, but it is time-consuming and prone to errors, especially when the outbreak is severe. Deep Learning (DL) algorithms, i.e., Convolutional Neural Networks (CNNs), have shown promising results in automatically detecting COVID-19. This project used (i) single CNNs, (ii) incrementally learned CNNs, and (iii) incrementally learned multiple CNNs with majority voting to extract features from CXR images. Then, an XGBoost classifier was used with each of these CNNs to detect COVID-19. A dataset consisting of 22,900 images was used for training (66.67%), validation (16.67%), and testing (16.67%). The results show that using XGBoost classifier with incrementally learned and incrementally learned multiple CNNs gave good and comparable detection accuracy (94.56% and 94.58%). The best performer – incrementally learned multiple CNNs with majority voting used ResNet152, DenseNet201, and VGG16.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS / ABBREVIATIONS

+ssRNA	single-stranded RNA
AdaBoost	adaptive boosting
Adam	adaptive moment estimation
AI	artificial intelligence
AP	anterior-to-posterior
AR	augmented reality
ARDS	acute respiratory distress syndrome
AUC	area under the ROC Curve
BIMCV	Valencian Region Medical ImageBank
BoW	bag of words
CNN	convolutional neural network
COVID-19	coronavirus disease 2019
CT	computed tomography
CXR	chest X-ray
DBNs	deep belief networks
DL	deep learning
DNNs	deep neural networks
DT	decision trees
FFNNs	feedforward neural networks
FN	false negative
FNR	false negative rate
FP	false positive
FPR	false positive rate
HOG	histogram of oriented gradients
ICU	intensive-care unit
IR 4.0	fourth industrial revolution
kNN	k-nearest neighbour
LR	logistic regression
LSTM	long short-term memory
MAE	mean absolute error
MC	montgomery county
MCWS	marker-controlled watershed segmentation

NB	naive bayes
NLP	natural language processing
PA	posterior-to-anterior
PPV	positive precision value
ReLU	rectified linear unit
RF	random forest
RMSE	root mean squared error
RMSprop	root mean square propagation
RNA	ribonucleic acid
RNNs	recurrent neural networks
RT-PCR	reverse transcriptase-polymerase chain reaction
SARS-Cov-2	severe acute respiratory syndrome coronavirus 2
SDGs	sustainable development goals
SIFT	scale invariant feature transform
SNR	signal-to-noise ratio
SPV	shared prosperity vision
TN	true negative
TP	true positive
UMLS	unified medical language system
WHO	world health organization
XGB	XGBoost

CHAPTER 1

INTRODUCTION

The Coronavirus Disease 2019 (COVID-19) pandemic continues to have a devastating effect on the health and well-being of the global population, caused by the infection of individuals by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). A critical step in the fight against COVID-19 is effective screening of infected patients, such that those infected can receive timely treatment and care, as well as be isolated to prevent the virus from spreading. The primary screening method used for detecting COVID-19 cases is reverse transcriptase-polymerase chain reaction (RT-PCR) testing, which can detect SARS-CoV-2 ribonucleic acid (RNA) from respiratory specimens (collected through a variety of means such as nasopharyngeal or oropharyngeal swabs) (Wang et al., 2020). While RT-PCR testing is the gold standard as it is highly specific, it is a very time-consuming, tedious, and complicated manual process that is in short supply. Besides, the sensitivity of RT-PCR testing is highly variable and has yet to be reported in a clear and consistent manner, and preliminary findings in China show rather poor sensitivity (West et al., 2020; Fang et al., 2020). Furthermore, subsequent findings showed highly variable positive rates depending on how the specimen was collected as well as decreasing positive rates with time after symptom onset (Yang et al., 2020; Wikramaratna et al., 2020).

Radiography examination is an alternative screening method that has also been used for COVID-19 screening, where radiologists conduct and analyse chest radiography imaging (e.g., chest X-ray (CXR) or computed tomography (CT) imaging) to look for visual indicators associated with SARS-CoV-2 viral infection. Early investigations discovered that patients with COVID-19 infection have anomalies in chest radiography images, with some suggesting that radiography examination could be utilised as a primary tool for COVID-19 screening in epidemic areas (Ng et al., 2020; Huang et al., 2020; Guan et al., 2020; Ai et al., 2020). Motivated by this and inspired by the open source and open access efforts of the research community and intrigued in exploring the efficacy of AI systems leveraging the more readily available and accessible CXR imaging modality, in this project I develop deep convolutional neural network models tailored for the detection of COVID-19 cases from chest X-ray (CXR) images that is open source and available to the general public. Thus, this project initiative finally contributes to the national development planning in terms of Shared Prosperity Vision (SPV) 2030 and 10-10 MySTIE. In the view of SPV 2030, the development of an automated AI screening system is in line with KEGA 3: Fourth Industrial Revolution (IR 4.0), which includes big data, artificial intelligence, augmented reality (AR) and machine learning. In the view of 10-10 MySTIE, this project clearly contributes to both Science & Technology Drivers and Socio-Economic Drivers on Advanced Intelligent Systems and Medical & Healthcare, respectively. In the international view, this project aligns with Sustainable Development Goals (SDGs), a universal call to achieve a better and more sustainable future for all. Through the lens of SDG, this project aligns with SDG Goal #3: Good Health and Well-being by providing a technological solution to the health issue – the spread of communicable disease. It also has an education dimension (SDG 4) to the reader on how deep learning is helping to fight COVID-19.

1.1 Problem Statement

1.1.1 Limited X-ray images related to COVID-19

Deep learning techniques require a large amount of data for training and testing. Deep learning models trained on limited datasets are not generalized, and thus, such models are not reliable (Alzubaidi et al., 2021). Most existing approaches for classifying COVID-19 cases depend on pre-trained deep classification networks like ResNet50, InceptionV3, etc. One of the main issues with these approaches is that they do not consider the limited dataset of COVID-19 cases (Calli et al., 2021). In addition, these off-the-shelf models are prone to over-fitting issues in a limited dataset regime which is the case for the task of properly detecting COVID-19 from existing (limited) lung CT/X-ray images. Such an issue arises in deep learning-based models when the capacity of the networks (number of trainable parameters) is much larger than the amount of information at hand (Bejani and Shatee, 2021).

However, acquiring datasets for training deep learning models is not easy, as no high-quality public database of X-ray images for COVID-19 patients is currently available. Considering the fast growth of the disease, public high-quality and well-annotated datasets are non-existent at this point (Cohen et al., 2020). In fact, even in private or hospital-owned cases where datasets are compiled, the datasets are very limited and still under development. This is exacerbated by the limited number of studies and expertise in properly labelling and annotating the existing data, which in turn directly affects the performance of the underlying model in both the training and testing phase, making it difficult to train a high-capacity network and to properly assess its performance in real-world applications (Robinson et al., 2021). As a result, model-based or domain-knowledge-aware methods must be considered to cope with such dataset scarcities as well as deficiencies. Thus, in this project, I collect a large-scale multi-class dataset from various online sources comprising 33,920 chest X-ray images: 11,956 images from confirmed COVID-19 cases, 11,263 images with confirmed bacterial or viral pneumonia cases, and 10,701 images of healthy people.

1.1.2 Slow detection by medical experts

The outbreak of COVID-19 has placed immense pressure on imaging departments, which are tasked with reading thousands of cases daily. Typically, patients and clinicians must wait for hours to receive imaging results, making it difficult to immediately screen and diagnose suspected patients, particularly in settings with limited medical resources (Rubin et al., 2020). Therefore, the development and deployment of automated screening tools that can accelerate large-scale screening and improve clinical diagnosis efficiency are crucial. Computational imaging-based procedures, such as chest X-ray, can provide more rapid diagnosis and limit the spread of COVID-19, especially since test kit results are not instant (Jacofsky et al., 2020). Thus, advanced AI-aided chest X-ray diagnosis systems are urgently needed to accurately confirm suspected cases, conduct virus surveillance, and screen patients for further diagnosis and treatment (Song et al., 2021).

The manual screening process for COVID-19 is further complicated by the fact that some features are difficult to detect by human eyes. This is where deep learning models can play a critical role, as they can identify complex patterns and subtle features that are not easily discernible by humans. The development of various deep learning models for chest X-ray image analysis can help radiologists in triaging, analysing, and assessing cases associated with the disease, ultimately enhancing the efficiency of clinical diagnosis.

1.1.3 Limited computing resources for training deep learning models using large datasets

Training deep learning models on large datasets can be a challenging task, primarily due to the significant computational resources required. The process can be excessively expensive and time-consuming, particularly for researchers or organizations with limited resources.

Moreover, the process of training a deep learning model on a large dataset is iterative, requiring multiple passes through the entire dataset to optimize the model's performance. This process can take days, weeks, or even months, making it difficult to iterate quickly and efficiently.

Another challenge is that the memory requirements for training deep learning models on large datasets can be substantial. This can limit the size of the model or the size of the dataset that can be used, leading to suboptimal performance or biased results.

Overall, the main challenges in training deep learning models on large datasets are the high computational cost and time required, as well as the potential limitations in model size and complexity. Thus, in this project, I proposed an approach to overcome these challenges by breaking down the dataset into subsets and incrementally training the deep learning model while leveraging the power of majority voting to combine their outputs. This allows for more efficient use of computational resources while also providing a more robust representation of the data.

1.2 Aim and Objectives

The project aims to propose a deep learning solution for automated COVID-19 detection to reduce the burden on healthcare systems and professionals.

The objectives of the project are as follows:
- to collect a minimum of 30,000 x-ray images related to COVID-19 from various online datasets.
- to train deep learning models: single models, incrementally learned single models, and incrementally learned multiple models with majority voting, using the collected x-ray images.
- to evaluate the deep learning models and select the best performer using performance metrics such as accuracy, precision, recall, and F1 score.

1.3 Scope and Limitation of Study

This project served as a comprehensive guide for researchers, healthcare professionals, and developers involved in the development of deep learning solutions for automated COVID-19 detection using chest X-ray (CXR) images. It focused on exploring different training techniques in deep learning models for COVID-19 detection, specifically: (i) single CNNs, (ii) incrementally learned single CNNs, and (iii) incrementally learned multiple CNNs with majority voting.

This project contributed to the understanding of training techniques and their impact on the accuracy and effectiveness of deep learning models for automated COVID-19 detection. By analyzing and evaluating the performance of these approaches using performance metrics such as accuracy, precision, recall, and F1 score, the study aimed to identify the most effective approach for automated COVID-19 detection. The insights gained from this study served as a guide for researchers and practitioners in making informed decisions when developing accurate and reliable COVID-19 detection systems using CXR images.

In summary, this project provided an exploration of different training techniques in deep learning models for automated COVID-19 detection using

CXR images. By evaluating their performance and discussing their capabilities, the study aimed to facilitate informed decision-making in the development of accurate and efficient COVID-19 detection systems.

CHAPTER 2

LITERATURE REVIEW

The World has experienced outbreaks of coronavirus infections during the last two decades: (i) the severe acute respiratory syndrome (SARS)-CoV outbreak in 2002–2003 from Guangdong, China; (ii) the Middle East respiratory syndrome (MERS)-CoV outbreak in 2011 from Jeddah, Saudi Arabia; and (iii) coronavirus disease 2019 (COVID-19) or SARS-CoV-2 outbreak from Wuhan, China, in December 2019. According to the Centers for Disease Control and Prevention (CDC), SARS was first reported in Asia in February 2003. The illness spread to 29 countries, where 8,096 people got SARS and 774 of them died (10%). The SARS global outbreak was contained in July 2003. MERS is a viral respiratory disease that was first reported in Saudi Arabia in September 2012 and has since spread to 27 countries. According to the World Health Organization (WHO), as of March 2023, there are 604 confirmed MERS cases and 36% (936) of these patients have died. Even though all three diseases are from the same family of coronavirus, the genomic sequence of COVID-19 showed similar but distinct genome composition from its predecessors SARS and MERS (Prompetchara, Ketloy and Palaga, 2020; Kumar et al., 2020). Despite a lower fatality rate of COVID-19, i.e., around 3%, when compared to SARS (10%) and MERS (36%), as depicted in Table 1, COVID-19 has resulted in many fold deaths (> 6.3 million already) than combined deaths of MERS and SARS (Mahase, 2020). The recent outbreak of COVID-19 was and still is an extremely infectious disease that has spread all over the World, forcing the WHO to declare it a pandemic on 11th March 2020.

Table 2.1: Details of coronavirus.

CoV	Year	Origin	Mortality rate
SARS	2002	Guangdong province, China	10%
MERS	2013	Saudi Arabia	34%
COVID-19	2019	Wuhan, China	3.4%

Among the causing pathogens for respiratory diseases, CoV has become the most dangerous one because of its serial interval (5 to 7.5) and reproductive rate (2 to 3) (Nishiura, Linton and Akhmetzhanov, 2020). The CoV belongs to single-stranded RNA viruses (+ssRNA) family mostly observed in animals (Perlman and Netland 2009; Chan et al., 2013). However, they can be transmitted to humans, which can cause severe and often fatal respiratory diseases in their new host. Severe cases of coronavirus disease result in acute respiratory distress syndrome (ARDS) or complete respiratory failure, which requires support from mechanical ventilation and an intensive-care unit (ICU). People with a compromised immune system or the elderly are more likely to develop serious illnesses, including heart and kidney failures and septic shock (Pormohammad et al., 2020). In general, COVID-19 spreads more quickly than SARS and has symptoms like other coronaviruses. Figure 2.1 shows the distribution of COVID-19 cases and deaths worldwide, as of 3rd July 2022 (European Centre for Disease Prevention and Control, 2022).

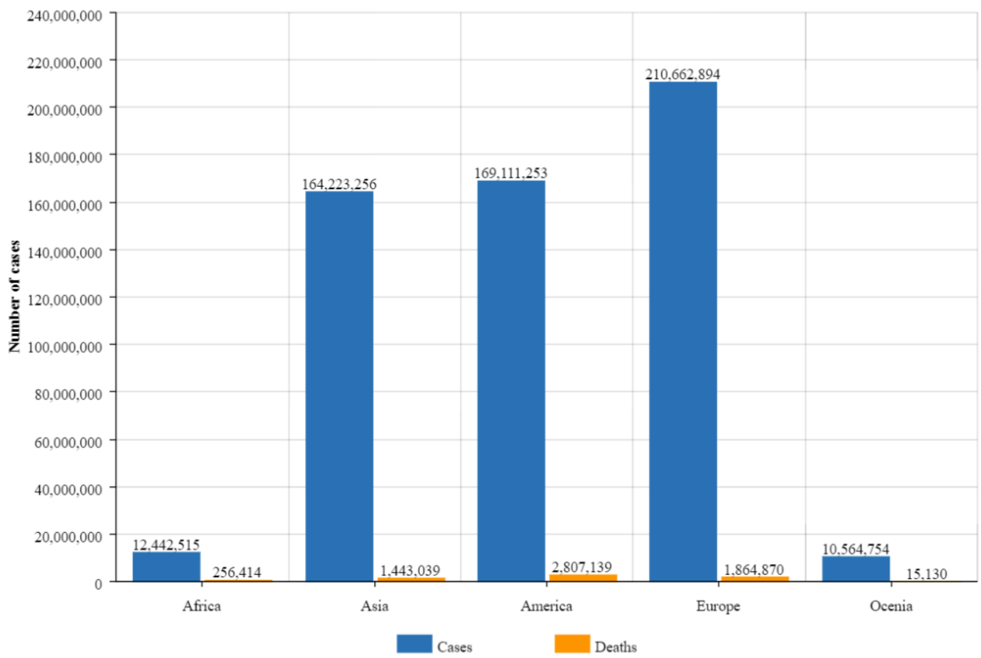

Figure 2.1: COVID-19 cases and deaths worldwide, as of 17th July 2022.

The diagnostic tools to detect COVID-19 are currently reverse transcription of polymerase chain reaction (RT-PCR) assays and chest imaging techniques, such as Computed Tomography (CT) and X-ray imaging.

Primarily, RT-PCR has become the gold standard in the diagnosis of COVID-19 (Kakodkar, Kaka and Baig, 2020; Li et al., 2020). However, RT-PCR arrays have a high false alarm rate which may be caused by the virus mutations in the SARS-CoV-2 genome, sample contamination, or damage to the sample acquired from the patient (Tahamtan and Ardebili, 2020; Xia et al., 2020). In fact, it is shown in hospitalized patients that RT-PCR sensitivity is low, and the test results are highly unstable (Li et al., 2020; Xiao, Tong and Zhang, 2020; Yang et al., 2020; World Health Organization, 2020). Therefore, it is recommended to perform chest CT imaging initially on the suspected COVID-19 cases since it is a more reliable clinical tool in the diagnosis with higher sensitivity compared to RT-PCR (Salehi et al., 2020). Hence, several studies suggest performing CT on the negative RT-PCR findings of the suspected cases (Salehi et al., 2020; Fang et al., 2020; Ai et al., 2020). However, there are several limitations of CT scans. Their sensitivity is limited in the early COVID-19 phase groups, and they are limited to recognising only specific viruses, slow in image acquisition, and costly (Bernheim et al., 2020; Li and Xia, 2020). On the other hand, X-ray imaging is faster, cheaper, and less harmful to the body in terms of radiation exposure compared to CT (Narin, Kaya and Pamuk, 2021; Brenner and Hall, 2007). Moreover, unlike CT devices, X-ray devices are easily accessible; hence, reducing the risk of COVID-19 contamination during the imaging process (Rubin et al., 2020). Currently, Chest X-ray (CXR) imaging is widely used as an assistive tool in COVID-19 prognosis, and it has been reported to have a potential diagnosis capability in recent studies (Shi et al., 2020).

In order to automate COVID-19 detection from CXR images, many studies have proposed to use deep learning, especially Convolutional Neural Networks (CNNs) (Narin, Kaya and Pamuk, 2020; Chowdhury, Rahman and Kabir, 2020; Pham, 2021; Chowdhury et al., 2020; Apostolopoulos and Mpesiana, 2020; Hall et al., 2020; Wang, Lin and Wong, 2020; Sethy and Behera, 2020; Zhang et al., 2020; Afshar et al., 2020). However, the main limitation of these studies is that the data is scarce for the target COVID-19 class. Such a limited amount of data degrades the learning performance of the deep networks. Thus, in the following section, I will discuss how a large-scale multi-class dataset is collected from various online sources in our project.

2.1 Dataset

The dataset used to train and test the proposed deep learning models comprises 33,920 CXR images with 11,956 COVID-19 cases, 11,263 non-COVID infections (viral or bacterial pneumonia) cases, and 10,701 Normal (healthy) cases. To generate the dataset, I combined nine different publicly available data repositories: (1) COVID-19 image data collection, (2) SIRM COVID-19 DATABASE, (3) BIMCV-COVID-19+ (Spain), (4) COVID-19 Image Repository, (5) RSNA pneumonia detection challenge dataset, (6) Chest X-Ray Images (pneumonia), (7) PadChest dataset, (8) Montgomery County chest X-ray dataset, and (9) Shenzhen chest X-ray dataset.

The choice of these nine datasets from which to create the dataset is guided by the fact that all nine datasets are open source and fully accessible to the research community and the general public. Section 2.1.1 overviews COVID-19 oriented datasets, whereas Section 2.1.2 overviews non-COVID-19 oriented datasets.

2.1.1 Chest X-ray datasets containing COVID-19 samples

2.1.1.1 COVID-19 Image Data Collection

The COVID-19 Image Data Collection is a collection of anonymized COVID-19 images acquired from websites of medical and scientific associations and research papers (Cohen et al., 2020; Giovagnoni, 2020; Società Italiana di Radiologia Medica e Interventistica, 2020). The dataset was created by researchers from the University of Montreal with the help of the international research community to ensure that it will be continuously updated. Nowadays, the dataset includes 646 X-ray images of patients affected by COVID-19 and other diseases, such as MERS, SARS, and ARDS. Each image is assigned a diagnosis of respiratory disease, with a strong focus on COVID-19 (currently 468 out of 646) and very few cases of no finding (20). Additionally, the dataset contains global severity scores for 100 images created in a post hoc analysis of images according to a severity scheme (Cohen et al., 2020).

2.1.1.2 Italian Society of Medical and Interventional Radiology (SIRM) COVID-19 DATABASE

The SIRM COVID-19 database reports 384 COVID-19 positive radiographic images (CXR and CT) with varying resolutions. Out of 384 radiographic images, 94 are chest X-ray images, and 290 are lung CT images. This database is updated in a random manner and until 10th May 2020, there were 71 confirmed COVID-19 cases were reported in this database.

2.1.1.3 BIMCV-COVID-19+ (Spain)

BIMCV COVID-19+ (Spain) is a large dataset from the Valencian Region Medical ImageBank (BIMCV) containing chest X-ray images CXR (CR, DX) and computed tomography (CT) imaging of COVID-19+ (positive) patients along with their radiological findings and locations, pathologies, radiological reports (in Spanish) and other data (Vayá et al., 2020). The images provided are 16 bits in png format. The new iteration of the database includes 7,377 CX, 9,463 DX and 6,687 CT studies from 1,311 COVID-19+ patients. I am using a total of 11,177 COVID-19 chest X-ray images from this dataset.

2.1.1.4 COVID-19 Image Repository

This dataset is from the Institute for Diagnostic and Interventional Radiology (Hannover, Germany), contains 240 chest-Xray images (CR, DX) from 71 patients at different timepoints in the course of COVID-19 (Winther et al., 2020). It contains metadata including scanning view (AP vs PA), patient master data, laboratory data and longitudinal information on admission, ICU-admission, and death. The dataset contains raw, unprocessed, gray value image data as Nifti files.

2.1.2 Chest X-ray datasets without COVID-19 samples

2.1.2.1 RSNA Pneumonia Detection Challenge Dataset

The RSNA Pneumonia Detection Challenge is a competition that aims to locate lung opacities on chest radiographs (Radiological Society of North America, 2020). This dataset is a subset of 30k images from the ChestXray-NIH dataset with an enrichment of images with a Pneumonia related diagnosis. This dataset offers images for two classes: Normal and Pneumonia (non-normal). All images are 8-bit grayscale with 1024 x 1024 resolution in DICOM format. I am using a total of 8,311 images from this dataset, of which 4,954 are from the normal class and 3,357 are from the pneumonia class.

2.1.2.2 Chest X-Ray Images (pneumonia)

Kaggle chest X-ray database is a very popular database, which has 5,247 chest X-ray images of normal, viral and bacterial pneumonia with resolution varying from 400p to 2000p (Kermany et al., 2018). Out of 5,247 chest X-ray images, 3,906 images are from different subjects affected by pneumonia (2,561 images for bacterial pneumonia and 1,345 images for viral pneumonia) and 1,341 images are from normal subjects.

2.1.2.3 PadChest Dataset

The PadChest dataset has 160,868 CXRs from 109,931 studies and 67,000 patients (Bustos et al., 2020). The dataset includes six different position views of CXR and additional information regarding image acquisition and patient demography. The images are stored as 16-bit grayscale images with full resolution. 27,593 of the reports were manually labelled by physicians. Using these labels, an RNN was trained and used to label the rest of the dataset from the reports. The reports were used to extract 174 findings, 19 differential diagnoses, and 104 anatomic locations. The labels conform to a hierarchical taxonomy based on the standard Unified Medical Language System (UMLS) (Bodenreider, 2004).

2.1.2.4 Montgomery County Chest X-ray Dataset (MC)

This pre-COVID-19 pandemic dataset contains 138 frontal chest images collected in the department of health and human services, Montgomery Country, Maryland, USA (Jaeger et al., 2014). To each image, a short radiological report and a disease diagnosis are assigned (58 images with tuberculosis manifestations and 80 normal), as well as a lung segmentation annotation automatically generated under the supervision of a radiologist using anatomical landmarks (Candemir et al., 2013). The images themselves contain written markings of the scanning view (AP and PA) and there is additional metadata about gender and age.

2.1.2.5 Shenzhen Chest X-ray Dataset

This dataset, released together with the Montgomery dataset, and the CXR images are collected from Shenzhen No.3 Hospital in Shenzhen, Guangdong province, China in September 2012. It contains 662 frontal X-rays, of which 326 are normal and 336 contain TB manifestations. The images, including some pediatric images, are distributed as 8-bit grayscale with full resolution and are annotated for signs of tuberculosis. Additionally, metadata includes sex, age, and a short radiological description (Jaeger et al., 2014).

2.1.3 Summary

To summarise, I created this dataset by combining nine separate publicly available data repositories, as depicted in table 2. In this study, only posterior-to-anterior (PA) or anterior-to-posterior (AP) chest X-rays were considered, as this view of radiography is preferred and widely used by the radiologist, whereas a lateral image is usually taken to complement the frontal view. Thus, they were excluded from this study. This dataset was created by utilizing nine publicly available datasets and repositories, all of which are scattered, and with varying formats. The quality of the dataset was ensured through a rigorous quality control process where duplicates, extremely low-quality, and over-exposed images were identified and removed. The resulting dataset thus comprises images of high interclass dissimilarity with few varying resolutions, quality, and signal-to-noise ratio (SNR) levels.

Table 2.2: Summary of the datasets used, mainly focusing on COVID-19, Pneumonia, and Normal. Other classes are stated in the remark.

Dataset Name	COVID-19	Pneumonia	Normal	Total	Remarks
COVID-19 Image Data Collection	468	-	-	468	Other classes: MERS, SARS, ARDS, and Normal
SIRM COVID-19 DATABASE	71	-	-	71	Other classes: Normal
BIMCV-COVID-19+	11,177	-	-	11,177	Other classes: N/A
COVID-19 Image Repository	240	-	-	240	Other classes: N/A
RSNA Pneumonia Detection Challenge	-	3,357	4,954	8,311	Other classes: N/A
Chest X-Ray Images (pneumonia)	-	3,906	1,341	5,247	Other classes: N/A Viral and bacterial

					pneumonia are added as pneumonia.
PadChest dataset	-	4,000	4,000	8,000	Other classes: 18 other differential diagnoses. Only AP and PA chest X-ray images are included.
MC chest X-ray dataset	-	-	80	80	Other classes: Tuberculosis manifestations
Shenzhen chest X-ray dataset	-	-	326	326	Other classes: Tuberculosis manifestations
Total	11,956	11,263	10,701	33,920	

2.2 Deep Learning

2.2.1 From Traditional Machine Learning to Deep Learning

In recent years, machine learning has become more and more popular in research and has been incorporated into a large number of applications, including multimedia concept retrieval, image classification, video recommendation, social network analysis, text mining, and so forth. Among various machine learning algorithms, “deep learning,” also known as representation learning, is widely used in these applications (Deng, 2014). The explosive growth and availability of data and the remarkable advancement in hardware technologies have led to the emergence of new studies in distributed and deep learning. Deep learning, which has its roots in conventional neural networks, significantly outperforms its predecessors. It utilizes graph technologies with transformations among neurons to develop many-layered learning models. Many of the latest deep learning techniques have been presented and have demonstrated promising results across different kinds of applications, such as Natural Language Processing (NLP), visual data processing, speech and audio processing, and many other well-known applications (Yan et al., 2017; Yan et al., 2015).

Traditionally, the efficiency of machine-learning algorithms highly relied on the goodness of the representation of the input data. A bad data representation often leads to lower performance compared to a good data representation. Therefore, feature engineering has been an important research direction in machine learning for a long time, which focuses on building features from raw data and has led to lots of research studies. Furthermore, feature engineering is often very domain specific and requires significant human effort. For example, in computer vision, different kinds of features have been proposed and compared, including Histogram of Oriented Gradients (HOG), Scale Invariant Feature Transform (SIFT), and Bag of Words (BoW) (Dalal and Triggs, 2005; Lowe, 2019). Similar situations have happened in other domains including speech recognition and NLP.

Comparatively, deep learning algorithms perform feature extraction in an automated way, which allows researchers to extract discriminative features with minimal domain knowledge and human effort (Najafadabi et al., 2015). These algorithms include a layered architecture of data representation, where the high-level features can be extracted from the last layers of the networks while the low-level features are extracted from the lower layers. These kinds of architectures were originally inspired by Artificial Intelligence (AI) simulating its process of the key sensorial areas in the human brain. Our brains can automatically extract data representations from different scenes. The input is the scene information received from the eyes, while the output is the classified objects. This highlights the major advantage of deep learning, i.e., it mimics how the human brain works.

In conclusion, deep learning is currently a popular research direction that utilizes convolutional neural networks to extract relevant features through convolution, pooling, and fully connected layers. These basic structures allow the network to learn and improve its performance. Many software industries and relevant areas of research are moving towards deep learning due to its strong feature extraction ability and learning ability not acquired by traditional machine learning methods. This feature provides many conveniences for many studies, eliminating the need for a very complex modelling process. In addition, deep learning is now showing substantial results and advances in image classification, object detection, image segmentation, and other areas. The depth and versatility of learning applications can continue to be expanded to other applications.

2.2.2 Deep Convolutional Neural Network Architecture

Deep learning techniques have become the main parts of various state-of-the-art multimedia systems and computer vision (Ha et al., 2015). More specifically, CNNs have shown significant results in different real-world tasks, including image processing, object detection, and video processing. This section discusses the deep CNN architectures with top-5 error rates proposed over the past few years for visual data processing.

In 1998, LeCun et al. presented the first version of LeNet-5 (LeCun et al., 1998). LeNet-5 is a conventional CNN that includes two convolutional layers along with a subsampling layer and finally ends with a full connection in the last layer, as depicted in Figure 2.2. Although, since the early 2000s, LeNet-5 and other CNN techniques were greatly leveraged in different problems, including the segmentation, detection, and classification of images, they were almost forsaken by data mining and machine-learning research groups. More than one decade later, the CNN algorithm started its prosperity in computer vision communities. Specifically, AlexNet is considered the first CNN model that substantially improved the image classification results on a very large dataset (e.g., ImageNet) (Krizhevsky, Sutskever and Hinton, 2012). Figure 2.3 depicts the architecture of the model. It was the winner of the ILSVRC 2012 and improved on the best results from the previous years by almost 10% regarding the top five test errors. To improve the efficiency and the speed of training, a GPU implementation of CNN is utilized in this network. Data augmentation and dropout techniques are also used to substantially reduce the overfitting problem. Nevertheless, there are two major drawbacks of this model: 1) it requires a fixed resolution of the image; 2) there is no clear understanding of why it performs so well.

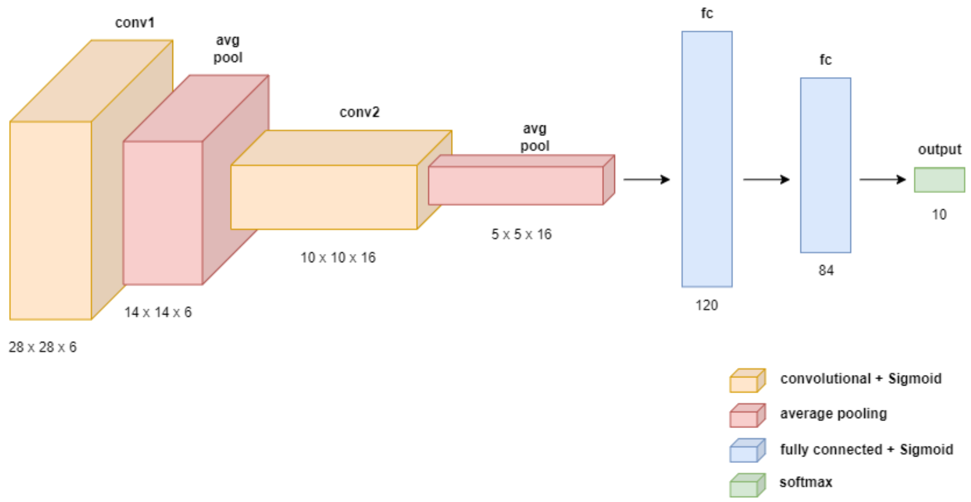

Figure 2.2: Architecture of LeNet-5.

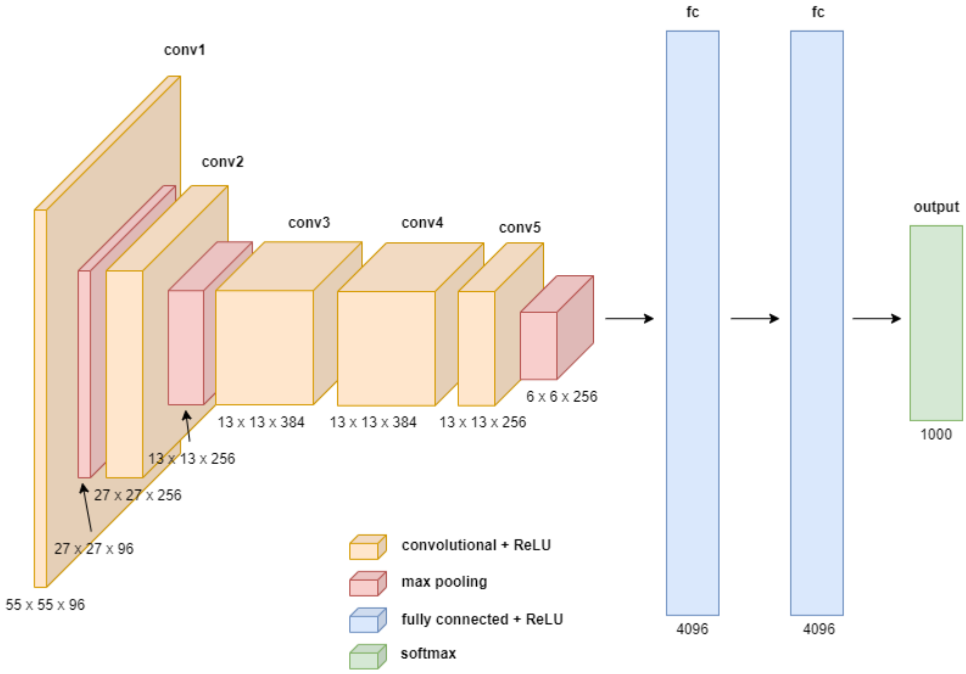

Figure 2.3: Architecture of AlexNet.

Since then, a variety of CNN methods have been developed and submitted to the ILSVRC competition. In 2014, two influential but different models were presented that mostly focused on the depth of neural networks. The first one, known as VGGNet, includes a very simple 16-layer CNN, as shown in Figure 2.4 (Simonyan and Zisserman, 2014). In this network, at each layer, the spatial size of the input is reduced, while the depth of the network is increased to achieve a more effective and efficient model. Although VGGNet was not the winner of the ILSVRC 2014, it still shows a significant improvement (7.3% top five error) over the previous top models that came from its two major specifications: simplicity and depth. In contrast to VGGNet, GoogleNet, the winner of this competition (6.7% error), proposed a new complex module named “Inception,” allowing several operations (pooling, convolutional, etc.) to work in parallel (Szegedy et al., 2014).

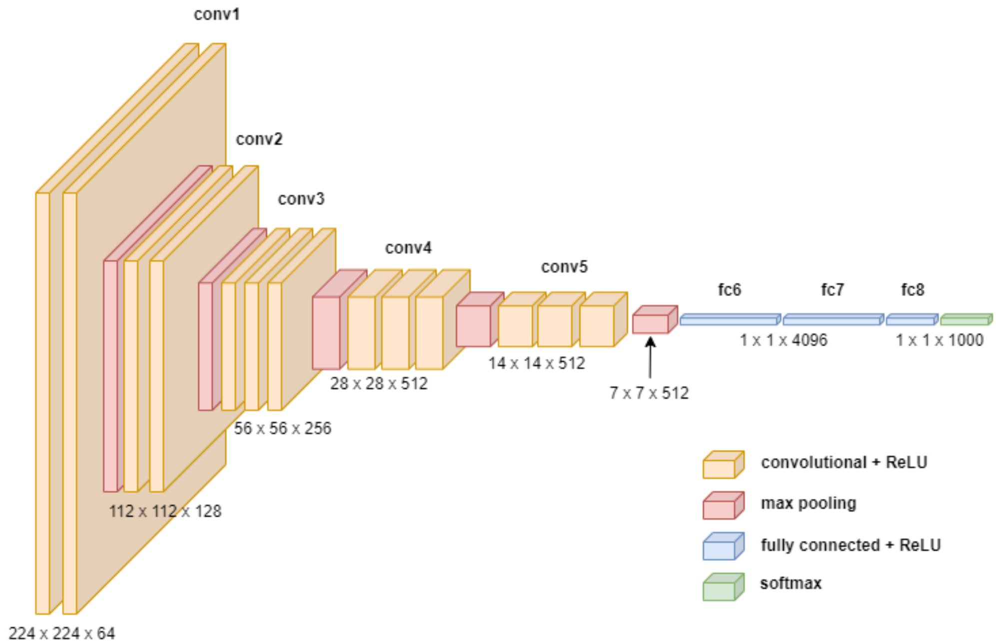

Figure 2.4: Architecture of VGG16.

The Microsoft deep residual network (known as ResNet) took the lead in the 2015 competitions including ILSVRC 2015 and in COCO detection and segmentation tasks by introducing the residual connections in CNNs and designing an ultra-deep learning model (50 to 152 layers) (He et al., 2016). This model achieved an incredible performance (3.6% top five error), which means, for the first time, a computer model could beat human brains (with 5% to 10% error) in image classification. Contrary to the extremely deep representation of ResNet, it can handle the vanishing gradients as well as the degradation problem (saturated accuracy) in deep networks by utilizing residual blocks (Glorot and Bengio, 2010).

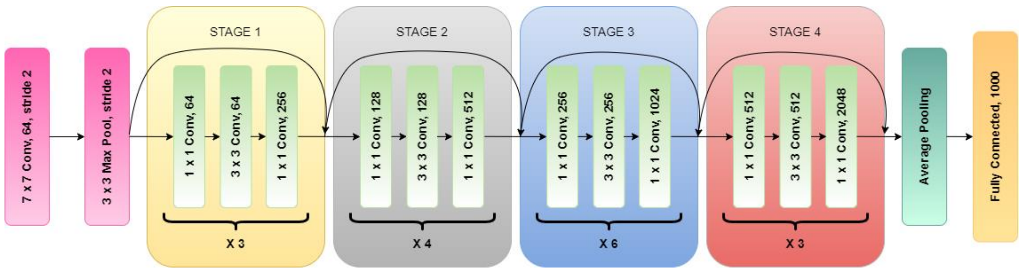

Figure 2.5: Architecture of ResNet-50.

In the last few years, several variations of ResNet have been proposed. The first group of methods tried to increase the number of layers more and more. Current CNN models may include more than 1,000 layers (Huang et al., 2016). Finally, in 2017, ResNeXT was proposed as an extension of ResNet and VGGNet (Xie et al., 2017). This simple model includes several branches in a residual block, each performing a transformation that is finally aggregated by a summation operation. This general model can be further reshaped by other techniques such as AlexNet. ResNeXT outperforms its original version (ResNet) using half of the layers and also improves the Inception-v3 as well as Inception-ResNet networks on the ImageNet dataset. Figure 2.6 demonstrates the revolution of depth and performance in image classification (e.g., ImageNet) over time. The problem of supervised image classification is regarded as "solved" and the ImageNet classification challenge concluded in 2017.

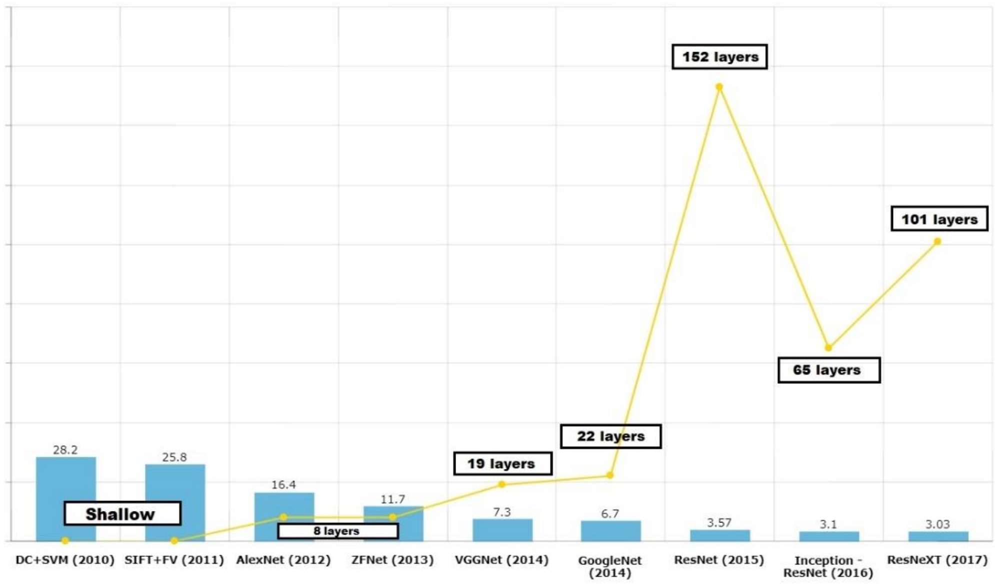

Figure 2.6: The network top five errors (%) and layers in the ImageNet classification over time.

2.2.3 Training

Forward and backward propagation are two important processes in training the neural network. In general, feedforward means moving forward with provided input and weights (assumed in 1st run) till the output. And, backward propagation, as the name suggests, is moving from output to input. In backward propagation, I reassign weights based on the loss and then forward propagation runs. These two processes are independent for the training of the model.

2.2.3.1 Forward Propagation

In terms of Neural Networks, forward propagation is important, and it will help to decide whether assigned weights are good to learn for the given problem statement. There are two major steps performed in forward propagation technically:

1. Sum the product
   It means multiplying weight vectors with the given input vector (x * w). And then, it keeps going on till the final layer, where the decision is made.
2. Pass the sum through an activation function
   The sum of the product of weight and input vector is passed in each layer and applies an activation function to produce the output. The output of one layer becomes the input of the next layer to be multiplied with weight vectors in that layer. This process goes on till the output layer activation function.

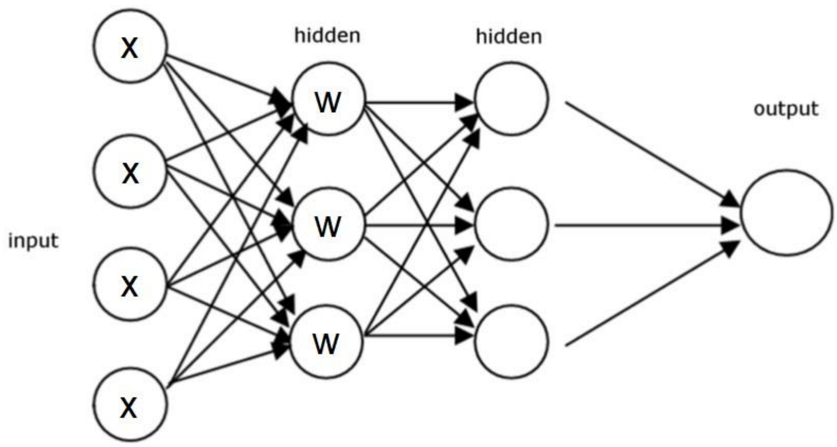

Figure 2.7: Feedforward Neural Network.

2.2.3.2 Backward Propagation

In neural networks, backward propagation is one of the important algorithms for training the feed-forward network. Once passing through the forward network, the predicted output is obtained to compare with the target output. Based on this, the total loss can be calculated and concluded whether the model is good to go or not. If not, the loss value is used to recalculate weights again for the forward pass. This weight re-calculation process is made simple and efficient using back-propagation.

In other words, backward propagation is the practice of fine-tuning the weights of a neural net based on the error rate (i.e., loss) obtained in the previous epoch (i.e., iteration). Proper tuning of the weights helps us to ensure lower error rates and make the model reliable by increasing its generalization.

As discussed above, there are two important processes involved in the training of any neural network: (1) Forward Propagation: Receive input data, process the information, and generate output. (2) Backward Propagation: calculate errors and update the parameters of the network. The following diagram explains how the forward and backpropagation algorithm works.

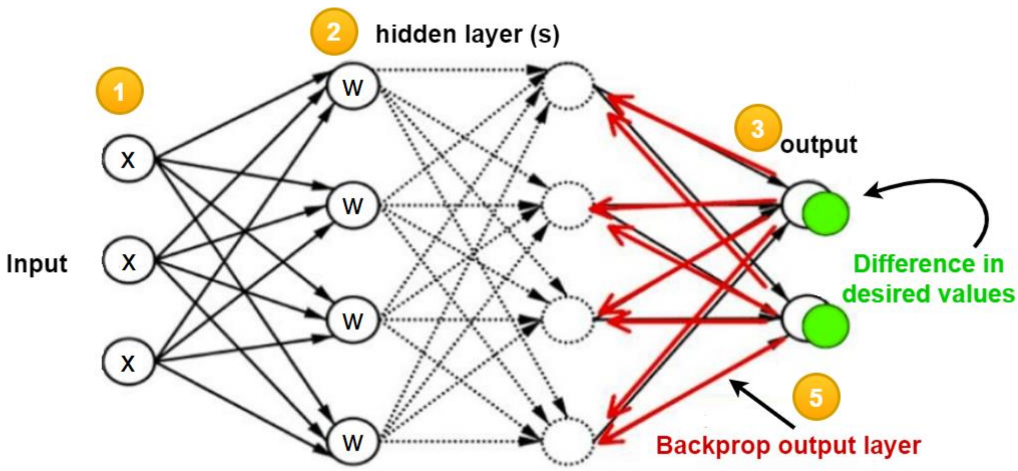

Figure 2.8: Forward and Backward Propagation in Neural Network.

1. Inputs X, arrive through the preconnected path.
2. Input is modelled using real weights W. The weights are usually randomly selected.

3. Calculate the output for every neuron from the input layer to the hidden layers, to the output layer.
4. Calculate the error in the outputs:
   ErrorB = Actual Output – Desired Output
5. Travel back from the output layer to the hidden layer to adjust the weights such that the error is decreased.
6. Keep repeating the process until the desired output is achieved.

2.2.4 Hyper-parameter Tuning

Deep learning algorithms are widely applied to various areas, like computer vision, natural language processing, and machine translation since they have had great success solving many types of problems. Common types of deep learning architectures include deep neural networks (DNNs), feedforward neural networks (FFNNs), deep belief networks (DBNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs) and many more (Yin et al., 2017). All these deep learning models have similar hyper-parameters since they have similar underlying neural network architecture. Compared with other machine learning models, deep learning models benefit more from hyperparameter tuning (HPO) since they often have many hyper-parameters that require tuning.

The first set of hyper-parameters is related to the construction of a deep learning model. Hence, named model design hyper-parameters. Since all neural network models have an input layer and an output layer, the complexity of a deep learning model mainly depends on the number of hidden layers and the number of neurons in each layer, which are two main hyper-parameters to build deep learning models (Koutsoukas et al., 2017). These two hyper-parameters are set and tuned according to the complexity of the datasets or the problems. Deep learning models need to have enough capacity to model objective functions (or prediction tasks) while avoiding over-fitting.

On the other hand, some other hyper-parameters are related to the optimization and training process of deep learning models; hence, they are categorized as optimizer hyper-parameters. The learning rate is one of the most important hyper-parameters in deep learning models (Ozaki, Yano and

Onishi, 2017). It determines the step size at each iteration, which enables the objective function to converge. A large learning rate speeds up the learning process, but the gradient may oscillate around a local minimum value or even cannot converge. On the other hand, a small learning rate converges smoothly, but will largely increase model training time by requiring more training epochs. An appropriate learning rate should enable the objective function to converge to a global minimum in a reasonable amount of time. Another common hyper-parameter is the dropout rate. Dropout is a standard regularization method for deep learning models proposed to reduce over-fitting. In dropout, a proportion of neurons are randomly removed, and the percentage of neurons to be removed should be tuned.

Batch size and the number of epochs are the other two deep learning hyper-parameters that represent the number of processed samples before updating the model, and the number of complete passes through the entire training set, respectively (Soon et al., 2017). Batch size is affected by the resource requirements of the training process and the number of iterations. The number of epochs depends on the size of the training set and should be tuned by slowly increasing its value until validation accuracy starts to decrease, which indicates over-fitting. On the other hand, deep learning models often converge within a few epochs, and the following epochs may lead to unnecessary additional execution time and over-fitting, which can be avoided by the early stopping method. Early stopping is a form of regularization whereby model training stops in advance when validation accuracy does not increase after a certain number of consecutive epochs. The number of waiting epochs, called early stop patience, can also be tuned to reduce model training time.

Apart from traditional deep learning models, transfer learning is a technology that obtains a pre-trained model on the data in a related domain and transfers it to other target tasks (Han, Liu and Fan, 2018). To transfer a deep learning model from one problem to another problem, a certain number of top layers are frozen, and only the remaining layers are retrained to fit the new problem. Therefore, the number of frozen layers is a vital hyper-parameter to tune if transfer learning is used.

2.2.5 Validation and Testing

In the context of neural networks, the ultimate goal of training a neural network is to find a set of neural network weights and bias values so that the input data generates output values that best match the target values. A simplistic approach would be to use all the available data items to train the neural network. However, this approach would likely find weights and bias values that match the data extremely well. In fact, probably with 100 per cent accuracy. But when presented with a new, previously unseen set of input data, the neural network would likely predict very poorly. This phenomenon is called over-fitting. To avoid over-fitting, the idea is to separate the available data into a training data set (typically 80 per cent to 90 per cent of the data) that is used to find a set of good weights and bias values, and a test set (the remaining 10 per cent to 20 per cent of the data) that is used to evaluate the quality of resulting neural network.

The simplest form of cross-validation randomly separates the available data into a single training set and a single test set. This is called hold-out validation. But the hold-out approach is somewhat risky because an unlucky split of the available data could lead to an ineffective neural network. One possibility is to repeat hold-out validation several times. This is called repeated sub-sampling validation. But this approach also entails some risk because, although unlikely, some data items could be used only for training and never for testing, or vice versa.

For this reason, some suggest using the K-Fold cross-validation scheme to accurately describe the predictive performance of neural networks. K-Fold is a validation technique in which the data is split into K-subsets and the holdout method is repeated K-times where each K subsets is used as the test set and the other K-1 subsets are used for the training purpose. Then the average error from all these K trials is computed, which is more reliable as compared to the standard handout method. So, with this technique, there is no need to be concerned about how the data is actually divided. The images below, i.e., Figure 2.9 and Figure 2.10, give better insights into how it works.

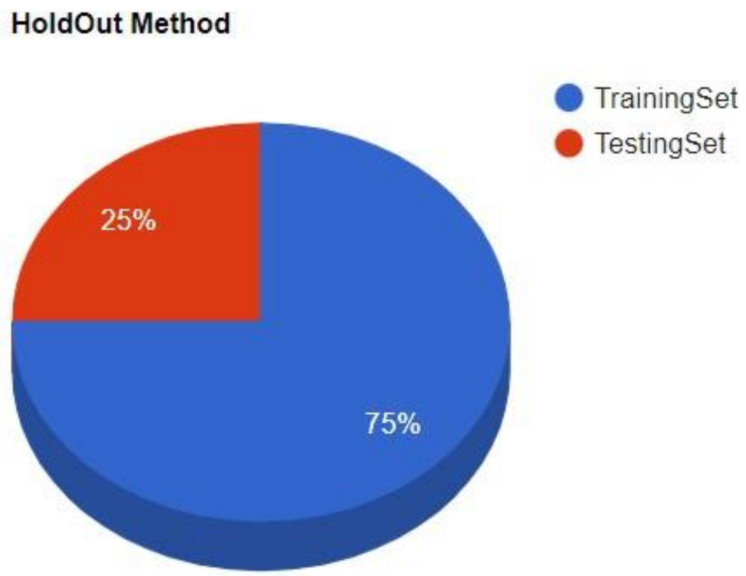

Figure 2.9: Pie chart represents how data is split in the holdout method.

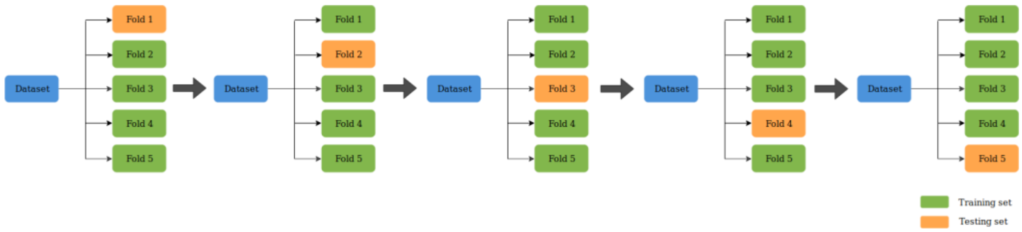

Figure 2.10: K-Fold cross-validation.

In K-Fold cross-validation, i.e., Picture 10, the dataset is divided into five subsets, i.e., K = 5. Each time, one of the subsets or folds is selected as the testing set, while the remaining folds are used as the training set. Each iteration represented above is nothing but a holdout method with different training and testing data. As a result, the advantage of K-fold cross-validation is that all observations or patterns in the available sample are used for testing and most of them are also used for training the model. The cross-validation analysis will yield valuable insights into the reliability of the neural networks with respect to sampling variation.

2.2.6 Summary

This section has highlighted how deep learning is getting more popular owing to its powerful feature extraction and learning abilities that traditional machine learning methods do not have. It also discussed the deep CNN architectures with the lowest error rates, beginning from the LeNet model to the ResNeXT model. The two common methods to learn during training neural networks, forward and backward propagation, were also covered. Following that, the section discussed several hyper-parameters that needed to be tuned, including model design hyper-parameters, function type hyper-parameters, optimizer hyper-parameters, batch size, number of epochs, and the number of frozen layers when using transfer learning. Finally, the section introduced the K-Fold cross-validation method for validation and testing in neural networks.

2.3 Deep Learning Models for COVID-19 Related X-ray

2.3.1 InstaCovNet-19

InstaCovNet-19 is a deep convolutional architecture (DCNN) used for the detection of patients with COVID-19 using chest Xray images (Gupta, Gupta and Katarya, 2021). Transfer Learning and multiple pre-trained DCNNs are used, like Inceptionv3, MobileNetV2, ResNet101, NASNet and Xception. These models were first imported with their pre-trained weights matrix (on ImageNet). Then these models were fine-tuned for the dataset. The fine-tuned models were then combined using the Integrated Stacking technique, making the stacked model a larger and more robust model. Two image pre-processing techniques are used i) fuzzy colour image enhancement ii) stacking.

DATASET: i) COVID19 Radiography database by Kaggle ii) ChestXray dataset

ACCURACY: 99.08%, SENSITIVITY: 99.00%

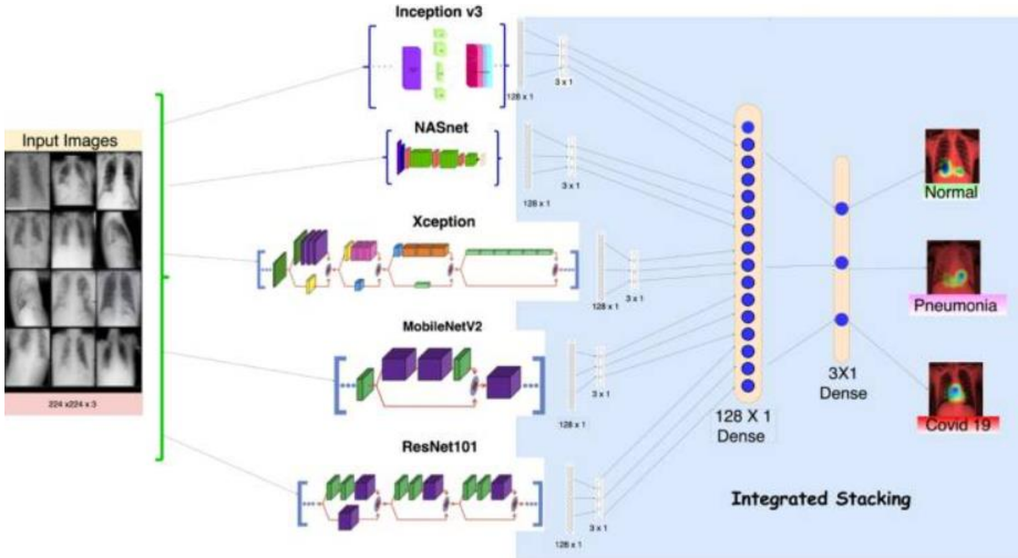

Figure 2.11: Architecture of InstaCovNet-19 Integrated stacked model.

2.3.2 COVID-AID

COVID-AID: COVID-19 AI detector, a novel deep neural network-based model to triage patients for appropriate testing (Mangal et al., 2020). This model contains pretrained CheXnet with 121-layer Densenet. DenseNet is quite similar to ResNet with some differences. ResNet uses an additive method which merges the previous layer (identity) with the future layer, whereas DenseNet concatenates the output of the previous layer with the future layer (Agarwal et al., 2022). An output of the previous layer acts as an input of the second layer by using composite function operation. This composite operation consists of the convolution layer, pooling layer, batch normalization, and non-linear activation layer. These connections mean that the network has L(L+1)/2 direct connections. L is the number of layers in the architecture. Deep CNN backbone followed by fully connected layer.

A two-stage training is used:

1. Densenet's backbone weights are frozen and only the final connected layer is trained. Batch size=16, Number of epochs=30 and the lowest validation loss is selected for the next stage.
2. In the second step, network weights are initialized from above, but the whole network is trained end to end.

DATASET: 1) Covid chest Xray images 2) Chest Xray pneumonia.

ACCURACY: 90.50%, SENSITIVITY: 100%

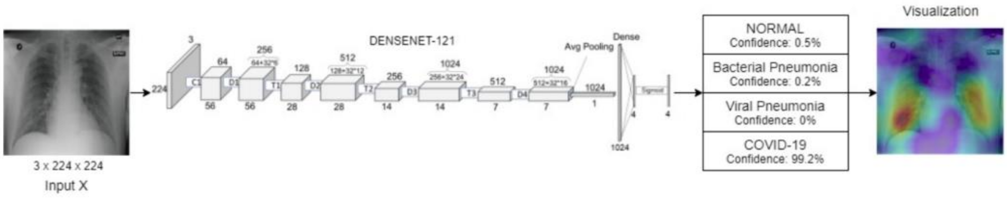

Figure 2.12: Architecture of CovidAID model.

2.3.3 DetraC Deep Model

DeTraC method of image classification is used, which consists of three phases (Abbas, Abdelsamea and Gaber, 2021). First phase: Train the pretrained CNN model (AlexNet, VGG19, ResNet, GoogLeNet, SqueezeNet) to extract deep local features from images. Second phase: Training is accomplished using a sophisticated gradient descent optimization method. Third phase: Composition layer to refine the final classification layer of images. This method can detect irregularities in the dataset by investigating class boundaries using class decomposition. For the decomposition of classes K-mean clustering method is used (Wu et al., 2008).

DATASET: i) 80 samples from the Japanese Society of Radiological Technology ii) COVID-19 Image data collection

MODELS	ACCURACY	SENSITIVITY
AlexNet	95.66%	97.53%
VGG19	**97.35%**	**98.23%**
ResNet	95.12%	97.91%
GoogLeNet	94.71%	97.88%
SqueezeNet	94.90%	95.70%

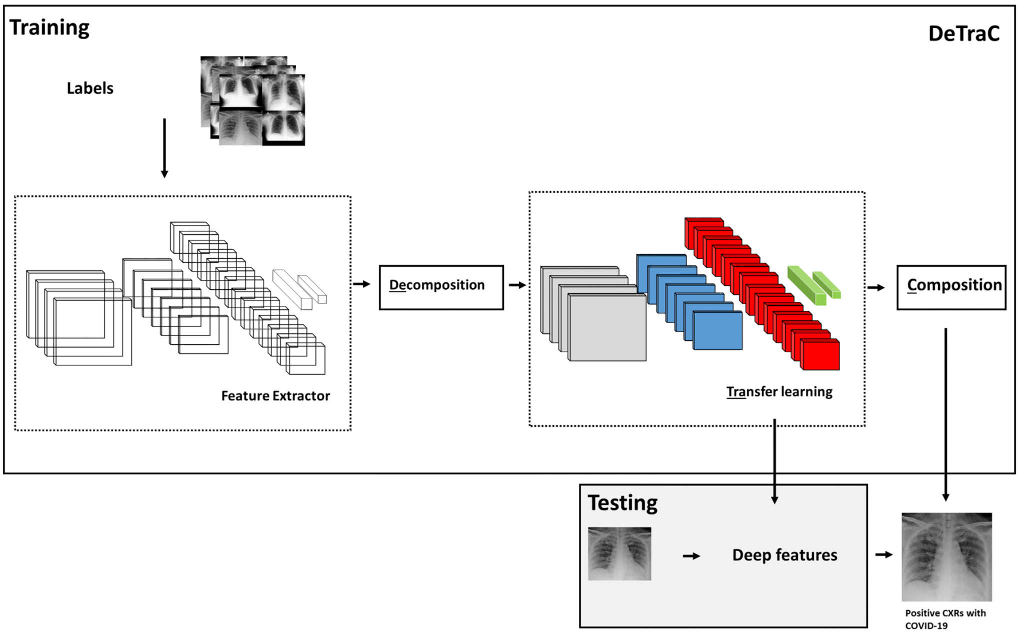

Figure 2.13: Architecture of DeTraC model.

2.3.4 CoroDet

CoroDet is a Novel CNN model which uses chest X-ray images to detect COVID-19. It is a new 22-layer CNN model (Hussain et al., 2021). It consists of 9-layer Conv2d followed by Maxpooling2D, 9-layer Conv2d followed by Max Pooling and at last Flatten layer followed by a Dense layer. Adam optimizer has been used. Training is carried out for 50 epochs with a learning rate=0.0001.

DATASET: COVID-R dataset has been used, which consists of 2843 COVID-19 images, 3108 Normal images, 1439 Pneumonia images

ACCURACY: 94.20%

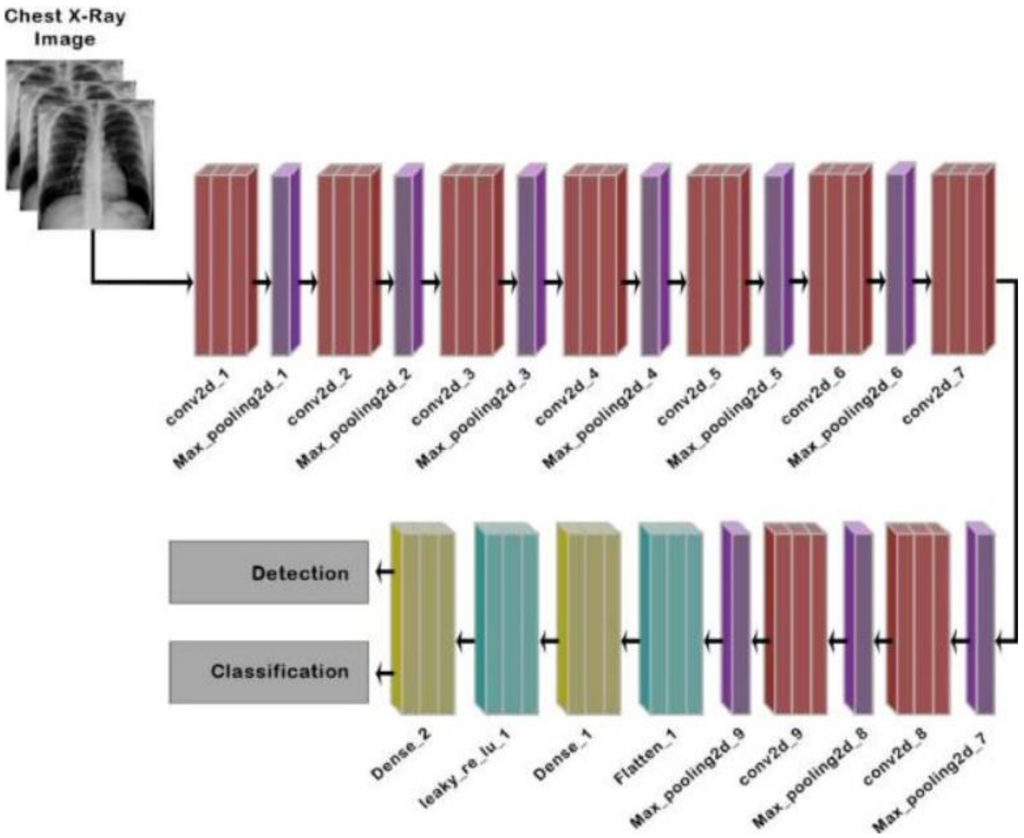

Figure 2.14: Architecture of the proposed 22-layer CNN model.

2.3.5 DeepCoroNet

Deep coronet is a new approach based on the deep Long Short-Term Memory (LSTM) model (Demir, 2021). Instead of a transfer-learning approach, the deep LSTM model is designed from scratch. Pre-processing of images is done using the sobel gradient and Marker-Controlled Watershed Segmentation (MCWS) are applied to raw images followed by a deep LSTM model, which increases classification performance (Huang, Li and Chen, 2018). Deep LSTM model consists of sequence data creating block and LSTM network sequence data creating block consists of convolution operation, Batch Normalization, Relulayer. The LSTM model is a modified version of recurrent neural networks (Yu et al., 2019). This layer is followed by a fully connected layer, Relu and dropout fully connected, which gives output to the SoftMax layer, which give probable scores of classes.

DATASETS: COVID-19 and Normal CXR images are taken from the Kaggle repository

ACCURACY: 100%, SENSITIVITY: 100%

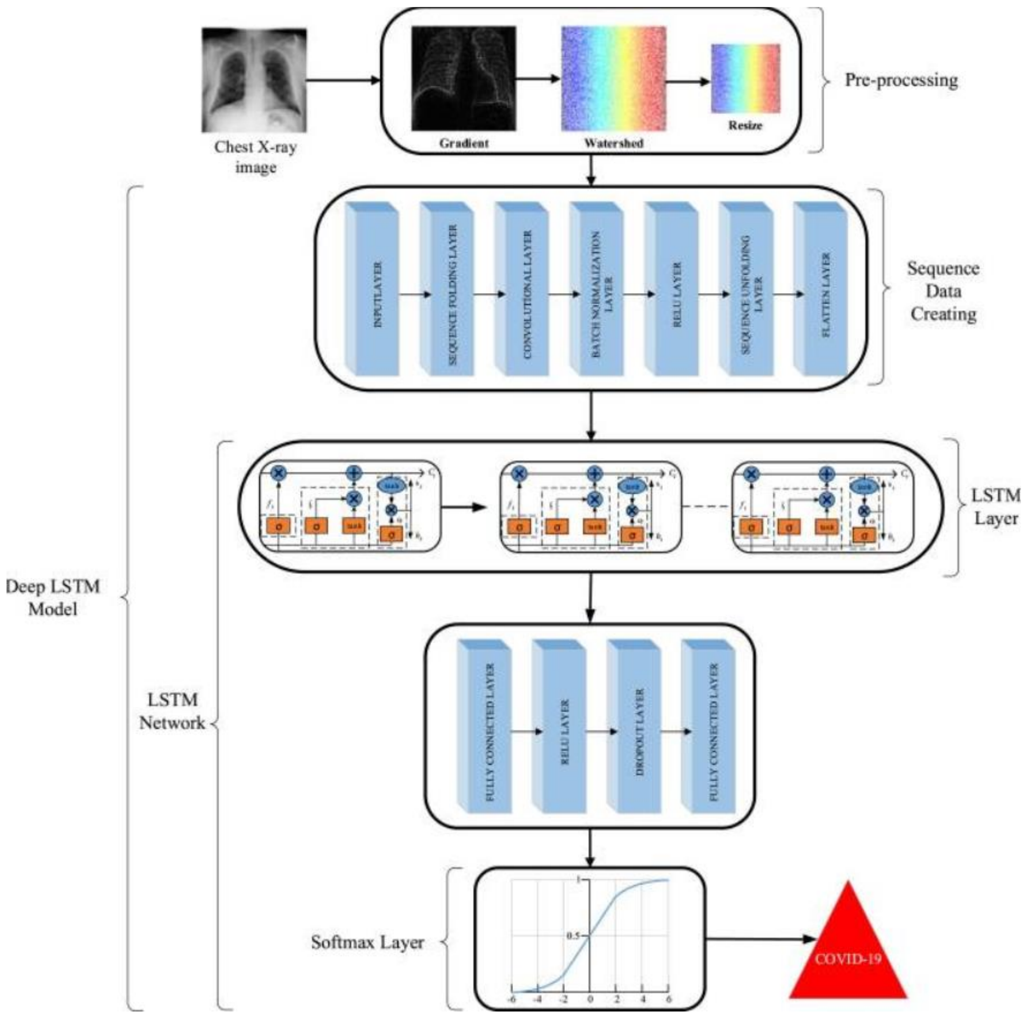

Figure 2.15: Architecture of DeepCoroNet.

2.3.6 CVDNet

CVDNET, a deep CNN model used to classify COVID-19 images from normal and other pneumonia cases using Chest X-ray images (Ouchicha, Ammor and Meknassi, 2020). It is based on a residual neural network and uses two parallel levels with different kernel sizes to capture the local and global features of the input. This architecture was trained on a small dataset but achieved promising results. For the convolution, it employs the concept of residual technique, which enhances the performance of this model.

DATASET: Kaggle Covid 19 Radiography database. The total number of images are 2,905. Out of which Normal (1341 images), COVID-19 (219 images), and Viral Pneumonia (1345 images) are used.

ACCURACY: 96.69%, SENSITIVITY: 96.84%

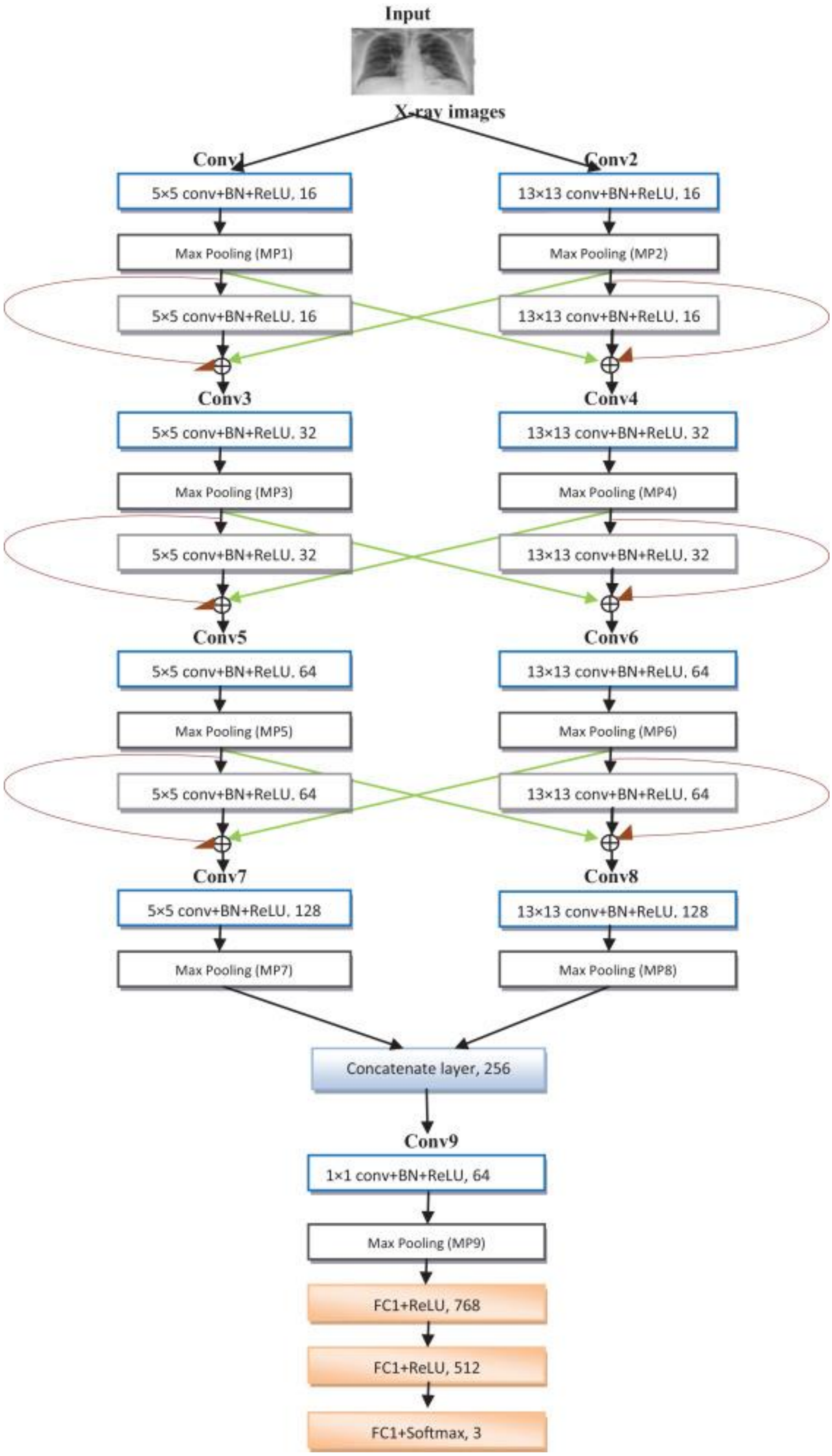

Figure 2.16: Architecture of the proposed CVDNet model.

2.3.7 EDL-COVID

Ensemble deep learning is a hybrid learning paradigm that can produce effective results by combining various machine learning models intelligently (Polikar, 2012). The combined strength of models offsets individual model variances and biases and provides a composite prediction where the final accuracy is better than the accuracy of the individual model. In EDL-COVID, instead of taking multiple models for ensembling, multiple model snapshots of a deep learning network, COVIDNet have been taken. COVIDNet network is used with its multiple snapshots and the cosine annealing learning rate is used to change the learning rate aggressively but systematically to generate different model weights over training epochs by allowing the learning rate to start high and decrease to a minimum value of zero at the relatively rapid speed (Tang et al., 2021).

COVIDX DATASET: It is a combination of five different datasets: Actual med covid 19 dataset, Covid19 Image data collection, Covid19 radiography database collection, Covid19 CXR dataset Initiative, RSNA pneumonia detection challenge.

ACCURACY: 95.00%, SENSITIVITY: 95.23%

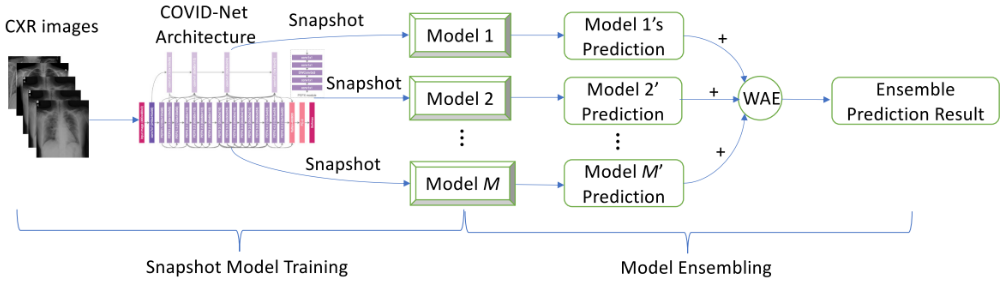

Figure 2.17:Overall flow for EDL-COVID ensemble model training. It consists of two phases, namely, snapshot model training, and model ensembling.

2.3.8 Summary

The comparison of the above-described deep learning models is shown in Table 2.3. Seven different deep learning models for COVID-19 related X-rays are surveyed, and their performance is analyzed on the basis of two parameters: Accuracy and Sensitivity.

Table 2.3: Comparison of the above-described deep learning models.

MODELS	ACCURACY	SENSITIVITY
InstaCovNet-19	99.08%	99.00%
COVID-AID	90.50%	100%
DeTraC	97.35%	98.23%
CoroDet	94.20%	NA
DeepCoroNet	100%	100%
CVDNet	96.69%	96.84%
EDL-COVID	95.00%	95.23%

2.4 Training Techniques

With the advent of deep learning techniques, feature extraction can be done automatically rather manually and thus achieves recognition accuracy at a higher level than ever before. Deep learning employs a convolutional neural network (CNN) which performs feature extraction. A CNN convolves learned features with input data and uses a 2D convolutional layer, making this architecture suitable for processing 2D images. CNN learned to detect different features of an image by using tens or hundreds of hidden layers (Ouchicha, Ammor and Meknassi, 2020). The relevant features are not pretrained; they are learned while the network trains on a collection of images (Mangal et al., 2020). The two most common techniques researchers used to train a network architecture are: training from scratch and transfer learning.

2.4.1 Training from Scratch

Training a deep architecture from scratch, requires a very large, labelled data set. A newly designed network architecture will learn the features from the dataset and is tested for its performance. This proves beneficial when a new algorithm is used for designing the layered architecture (Jain et al., 2021). This is a less common approach because with the large amount of data and rate of learning, these networks typically take days or weeks to train, and it is not economical.

2.4.2 Transfer Learning

Most researchers prefer to use the transfer learning approach in training a deep architecture. It is a process that involves fine-tuning a pretrained model (Pambudi, Widayanti and Edastama, 2021). In this, a model developed for a particular task is reused as the starting point for a model on a second task, such as the AlexNet CNN model is trained on the Image Net database but by applying the transfer learning approach, it can be used for other classification problems. Transfer learning performs best in situations where the training examples are insufficient for training a model from scratch. Tajbakhsh et al. demonstrated that a pre-trained CNN with adequate fine tuning might outperform or perform as well as a CNN trained from scratch.

After reviewing several related articles, the reviewed works that utilized transfer learning can be categorized into four groups. In the first group, a pre-trained CNN on a large-scale natural image dataset was used to initialize the weights of a new network that will be trained on the target data. When performing transfer learning, the last layer of the pretrained model architecture is replaced with a fully connected layer with the same number of classes as the new dataset. The architecture is retrained to use the model for the new dataset (Chakraborty et al, 2021). This method is based on the fact that the early layer features are more generic (e.g., edges), whereas the later-layer features are more specific to a particular task or dataset (Yamashita et al., 2018).

The second group is similar to the first in that the last layer of the architecture is replaced and redefined. The only difference is that in the first group, only the last layer is retrained, whereas in this group, some layers can be redefined and retrained according to the context (Cai and Peng, 2021). One major disadvantage of these methods is that the size of the input image cannot be changed. Therefore, if the pretrained model uses a smaller image dimension and transfer learning has to be conducted on a dataset with a larger image dimension, resizing the image is compulsory. Resizing a large image to a smaller image can affect the performance of the model in some cases.

In the third group of studies, a pretrained model is used to extract the deep features of the images of a prepared custom dataset. Then, the extracted deep features are input into a linear machine learning classifier such as support vector machine (SVM) for classification. For example, Sethy and Behera used eleven established model architectures that are pretrained on the ImageNet dataset to extract the deep features: AlexNet, DenseNet201, GoogleNet, InceptionV3, ResNet18, ResNet50, ResNet101, VGG16, VGG19, XceptionNet, and InceptionResNetV2. A slightly different approach is applied by Ozkaya et al. for the classification of X-ray images. Similarly, features are extracted from three networks, namely, VGG-16, GoogleNet and ResNet-50, for the classification of CT images. The features are fused, and to reduce the redundancy of the features, the t-test method is used to rank the features based on frequency. The final constructed feature vector is input into a binary SVM classifier for classification.

In the last group, transfer learning was implemented using a model pre-trained on a similar target domain. For example, Afshar et al. trained a model on a radiography dataset of patients with and without pneumonia. They then trained the model further on COVID-19 CXR images. The studies in this group claimed that the use of models trained on ImageNet is not the best option for medical applications because the source (natural images) and target domains (e.g., CXR images) are different (Basu, Mitra and Saha, 2020; Afshar et al., 2020). However, the results of a comparative study by Cheplygina did not fully support this assumption; the study examined 12 articles that compared the use of medical images to natural images in transfer learning in medical imaging research. The goal of the study was to determine which source images are better in medical transfer learning tasks: natural images such as ImageNet or medical images. Among the 12 articles examined, the study found that six articles supported each claim, i.e., each claim is supported equally; therefore, the study concluded that the selection of the model and source data depends on the task at hand among other factors.

2.4.3 Summary

In short, there are two main methods for training a deep architecture model: training from scratch and transfer learning. However, due to the duration of development and the amount of data required, training from scratch is not considered in our project. Instead, the project will use the transfer learning technique.

There are mainly four strategies that gradually evolved when using transfer learning to train models. Initially, the first group of studies used different state-of-the-art pretrained CNN models to initialize the weights of a new network that will be trained on the target data. The early layers of the network model were frozen, and their weights were kept unchanged while the final layer was fine-tuned. The final layer of the pretrained model architecture is replaced with a fully connected layer with the same number of classes as the new dataset. The second group of studies then began to redefine and retrain some layers based on the context in order to increase model accuracy and extract more features to supply additional information to the fully connected layer in a CNN. Different hyper-parameters are fine-tuned, and some or all CNN layers are unfrozen to be retrained during the training process.

Nevertheless, while the transfer learning approach helps achieve better results with a smaller data set than training from scratch, it still needs a rather large, labelled dataset. One of the main issues in the first and second groups of studies is that they do not consider the limited dataset of COVID-19 cases when using CNN for training and classification. It leads to a question mark about the robustness of the classification model because deep learning models trained on limited datasets are not generalized, and thus, such models are not reliable. Moreover, when CNN is used for classification, it takes a lot of time for training. To get good enough results, it is necessary to fine-tune the CNN parameters during training. As a result, the computational complexity grows, as does the execution time. So, rather than using a pre-trained network as a classifier in the transfer learning strategy to detect COVID-19, the third group of studies used a machine learning algorithm as the classifier. The machine learning algorithm uses deep features extracted from the fully connected layer of the pre-trained network to classify X-ray images of COVID-19 patients, pneumonia patients, and healthy people.

Finally, the fourth group of studies claimed that using ImageNet-trained models for medical applications is not the ideal solution because the source (natural images) and target domains (e.g., CXR images) are different. As a result, transfer learning was implemented using a model pre-trained on a similar target domain. However, a research study did not fully support this claim and concluded that the selection of model and source data is dependent on the task at hand, among other considerations.

Table 2.4: Comparison of Transfer Learning Techniques.

	Pros	Cons	In-Text Citation
First group – fine tune last layer	• Require lesser training time and computational costs as compared to second group, because it retains the useful feature extractors trained during the initial stage.	• Limited state-of-the-art CNN learning ability, thus poor performance than second group. • Performance can be limited because CNN models' pre-training is performed based on natural images (ImageNet dataset). • The internal logic of CNN is not explicitly known and require other techniques for	(Chakraborty et al, 2021; Wang et al., 2020; Khan et al., 2020)

		visual interpretation of CNN decision-making.	
Second group – fine tune some or all layers	• Tailored design of CNN model can extract unique feature entirely. Thus, performance is better than first group if the model is tuned properly.	• Image resizing is compulsory. This can affect the performance of the model in some cases. • Training time and computational costs are the highest among the discussed techniques because specific modifications such as architecture adjustments and parameter tuning need to be applied to the pre-trained model. • Require a relatively large amount of data to be advantageous as the new trainable parameters are inserted into the network. • Performance can	(Apostolopoulos and Mpesiana, 2020; Cai and Peng, 2021; Khan et al., 2020)

		be limited because CNN models' pre-training is performed based on natural images (ImageNet dataset). • The internal logic of CNN is not explicitly known and require other techniques for visual interpretation of CNN decision-making.	
Third group – CNN with machine learning	• Require lesser amount of data as compared to other techniques as traditional machine learning classifier is not as data hungry as CNN classifier. • Require lesser training time and computational costs as compared to second group, because it retains the useful feature extractors trained during the initial	• Performance can be limited because CNN models' pre-training is performed based on natural images (ImageNet dataset).	(Sethy and Behera, 2020)

	stage. • The decision-making of traditional machine learning classifiers is explainable and interpretable.		
Fourth group – TL with similar target domain	CNN models' pre-training is performed based on same target domain. Thus, the model is able to extract intricate features specific to the target.	• Lack of publicly available pretrained CNN model on same domain. • Robustness of the publicly available pre-trained CNN model on same domain is not proven. • The internal logic of CNN is not explicitly known and requires other techniques for visual interpretation of CNN decision-making.	(Basu, Mitra and Saha, 2020; Afshar et al., 2020; Khobahi et al., 2020)

2.5 Evaluation Metrics

Evaluation metrics adopted within deep learning tasks play a crucial role in achieving the optimized classifier (Hossin and Sulaiman, 2015). They are utilized to optimize the classification algorithm during the training stage. This means that the evaluation metric is utilized to discriminate and select the optimized solution. For the time being, the evaluation metric is also utilized to measure the efficiency of the created classifier, e.g., as an evaluator, within the model testing stage using hidden data. As given in Eq. 1, true negative (TN) and true positive (TP) are defined as the number of negative and positive instances, respectively, which are successfully classified. In addition, false negative (FN) and false positive (FP) are defined as the number of misclassified positive and negative instances, respectively. Next, some of the most well-known evaluation metrics are listed below.

1. Accuracy: Calculates the ratio of correct predicted classes to the total number of samples evaluated (Eq. 1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Sensitivity or Recall: Utilized to calculate the fraction of positive patterns that are correctly classified (Eq. 2).

$$Sensitivity = \frac{TP}{TP + FN}$$

3. Specificity: Utilized to calculate the fraction of negative patterns that are correctly classified (Eq. 3).

$$Specificity = \frac{TN}{FP + TN}$$

4. Precision: Utilized to calculate the positive patterns that are correctly predicted by all predicted patterns in a positive class (Eq. 4).

$$Precision = \frac{TP}{TP + FP}$$

5. F1-Score: Calculates the harmonic average between recall and precision rates (Eq. 5).

$$F1_{score} = 2 \text{ x } \frac{Precision\ x\ Recall}{Precision + Recall}$$

6. False Positive Rate (FPR): This metric refers to the possibility of a false alarm ratio as calculated in (Eq. 6).

$$FPR = 1 - \text{Specificity}$$

7. Area Under the ROC Curve (AUC): AUC is a common ranking type of metric. It is utilized to conduct comparisons between learning algorithms, as well as to construct an optimal learning model. In contrast to probability and threshold metrics, the AUC value exposes the entire classifier ranking performance. The following formula is used to calculate the AUC value for a two-class problem (Eq. 7)

$$AUC = \frac{S_p - n_p(n_n + 1)\ /\ 2}{n_p n_n}$$

8. Mean Absolute Error (MAE): MAE is a simple way to measure error magnitude. It consists of the average of the absolute differences between the predictions and the observed values. (Eq. 8).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

9. Root Mean Squared Error (RMSE): RMSE measures the quadratic mean of the differences between the predictions made by a model and the actual values (residuals) (Eq. 9).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)\ ^2}{N}}$$

In conclusion, this project will quantify the model's classification performance using nine evaluation metrics outlined above. These metrics include accuracy, sensitivity, specificity, precision, F1-Score, False Positive Rate, AUC, MAE, and RMSE.

2.6 Prior Works

This project assessed the significance and uniqueness of several publications and their associated datasets using a range of methods, including developing models and frameworks from scratch, as well as leveraging transfer learning in combination with specialized feature extraction techniques. The table below summarized the key findings from the assessment.

Table 2.5: Articles Proposing Novel Methods for COVID-19 Detection via CXR Images

Author	**Title**	**Techniques**	**Dataset**	**Result**		**Remarks**
				No. of class	**Accuracy**	
Asif Iqbal Khan, Junaid Latief Shah, and Mohammad Mudasir Bhat	CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images	The authors proposed a model called CoroNet, which uses Xception CNN architecture as base model, with a dropout layer and two fully connected layers added at the end.	COVID-19: 284 Pneumonia Bacterial: 330 Pneumonia Viral: 327 Normal: 310	2 3 4	99.00% 95.00% 89.60%	https://github.com/ieee8023/covid-chestxray-dataset https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

Shahin Khobahi, Chirag Agarwal, and Mojtaba Soltanalian	CoroNet: A Deep Network Architecture for Semi-Supervised Task-Based Identification of COVID-19 from Chest X-ray Images	Autoencoders were used for feature extraction. A fine-tuned CNN pre-trained model (ResNet-18) was used to perform the task of classification.	COVID-19: 99 Pneumonia: 9579 Normal: 8851	3	93.50%	https://github.com/ieee8023/covid-chestxray-dataset https://www.cc.nih.gov/drd/summers.html

Ioannis D. Apostolopoulos and Tzani A. Mpesiana	Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks	A pre-trained model (VGG19, MobileNet v2, Inception, Xception, Inception ResNet v2) is used only as a feature extractor. The pre-trained model retains both its initial architecture and all the learned weights. Only the fine-tuning layers that are closer to the output features are adjusted.	Dataset I COVID-19: 224 Bacterial pneumonia: 700 Normal: 504 Dataset II COVID-19: 224 Bacterial and Viral Pneumonia: 714 Normal: 504	2 3	96.78% 94.72%	https://github.com/ieee8023/covid-chestxray-dataset https://www.kaggle.com/andrewmvd/convid19-X-rays https://data.mendeley.com/datasets/rscbjbr9sj/2
Y. Pathak, P.K. Shukla, A. Tiwari, S. Stalin, S. Singh, and P.K. Shukla	Deep Transfer Learning Based Classification Model for COVID-19 Disease	The authors utilized the Resnet-50 model as a feature extractor, and transfer learning was used to tune the initial parameters of deep layers.	COVID-19: 413 Non COVID-19: 439	2	93.02%	https://ieeexplore.ieee.org/ielx7/6287639/8948470/09144185.pdf https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7183816/

Muhammad Faroo & Abdul Hafeez	COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs	The authors proposed a model called COVID-ResNet, which uses ResNet50 CNN architecture as a base model, and applies different training techniques, including progressive resizing, cyclical learning rate finding, and discriminative learning rates to gain fast and accurate training.	Total: 5941 COVID-19: 68	4	96.23%	https://github.com/lindawangg/COVID-Net
Linda Wang, Zhong Qiu Lin, and Alexander Wong	COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images	The authors used generative synthesis (GenSynth) to generate a deep CNN network architecture named COVID-Net by specifying the design requirements.	COVID-19: 358 Pneumonia: 5538 Normal: 8066	3	93.33%	https://github.com/lindawangg/COVID-Net

Mohammad Rahimzadeh & Abolfazl Attar	A MODIFIED DEEP CONVOLUTIONAL NEURAL NETWORK FOR DETECTING COVID-19 AND PNEUMONIA FROM CHEST X-RAY IMAGES BASED ON THE CONCATENATION OF XCEPTION AND RESNET50V2	The authors proposed a concatenated neural network that is designed by concatenating the extracted features of Xception and ResNet50V2, and then connecting the concatenated features to a convolutional layer that is connected to the classifier.	COVID-19: 180 Pneumonia: 6054 Normal: 8851	3	91.40%	https://github.com/ieee8023/covid-chestxray-dataset https://www.kaggle.com/c/rsna-pneumonia-detection-challenge

Kishore Medhi, Md. Jamil, and Md. Iftekhar Hussain	Automatic Detection of COVID-19 Infection from Chest X-ray using Deep Learning	The authors proposed a multi-layer CNN architecture consisting of four convolutional layers with ReLU activation function, four maxpooling layers, one flatten layer, two fully connected layers, and one softmax activation layer. A 2D Gaussian filter and segmentation techniques were used in data pre-processing.	COVID-19: >150 Normal: Unknown	2	93.00%	https://www.kaggle.com/bachrr/covid-chest-xray
Saban Ozturk, Umut Ozkaya, and Mucahid Barstugan	Classification of Coronavirus Images using Shrunken Features	The authors ensembled several feature extraction algorithms and used a stacked autoencoder with principal component analysis to make decisions. Finally, a support vector machine (SVM) was used for classification.	Ards: 4 COVID-19: 101 Normal: 2 Pneumonia: 2 SARS: 11 Streptococci: 6	6	90.00%	https://github.com/ieee8023/covid-chestxray-dataset

Arpan Mangal, Surya Kalia, and Harish Rajgopal	CovidAID: COVID-19 Detection Using ChestX-Ray	This model contains pre-trained CheXNet, a 121-layer Dense Convolutional Network (DenseNet) backbone, followed by a fully connected layer. The CheXNet's final classifier of 14 classes was replaced with a classification layer of 4 classes, each with a sigmoid activation to produce the final output. Densenet's backbone weights are frozen, and only the final connected layer is trained.	COVID-19: 155 Bacterial pneumonia: 2780 Viral pneumonia: 1493 Normal: 1583	3 4	90.50% 87.20%	https://github.com/ieee8023/covid-chestxray-dataset https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

ChaimaeOuchicha, OuafaeAmmor, and MohammedMeknassi	CVDNet: A novel deep learning architecture for detection of coronavirus (Covid-19) from chest x-ray images	The authors proposed CVDNet, a Deep Convolutional Neural Network (CNN) model to classify COVID-19 infection from normal and other pneumonia cases using chest X-ray images. The proposed architecture is based on the residual neural network and is constructed by using two parallel levels with different kernel sizes to capture local and global features of the inputs.	COVID-19: 219 Viral pneumonia: 1345 Normal: 1341	3	96.69%	https://www.kaggle.com/tawsifurrahman/covid19-radiography-database

Nour Eldeen M. Khalifa, Mohamed Hamed N. Taha, Aboul Ella Hassanien, and Sally Elghamrawy	Detection of coronavirus (covid-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest x-ray dataset	Transfer learning on a pre-trained model (AlexNet, SqueezeNet, GoogleNet, and ResNet18) in combination with GAN was performed. Only the last fully connected layer was fine-tuned. The GAN network helped in overcoming the overfitting problem by generating new images.	Total: 624	2	99.00%	https://data.mendeley.com/datasets/rscbjbr9sj/2

Prabira Kumar Sethy, Santi Kumari Behera, Pradyumna Kumar Ratha, and Preesat Biswas	Detection of coronavirus Disease (COVID-19) based on Deep Features and Support Vector Machine	The authors utilized pre-trained CNN models (AlexNet, VGG16, ResNet50) for deep feature extraction. The deep features obtained from these networks are individually fed to SVM for classification.	COVID-19: 127 Pneumonia: 127 Normal: 127	3	95.33%	https ://github.com/ieee8023/covid -chest Xray-dataset https ://www.Kaggle.com/andrewmvd/convid19-X-rays https://data.mendeley.com/datasets/rscbjbr9sj/2

Sanhita Basu, Sushmita Mitra, and Nilanjan Saha	Deep Learning for Screening COVID-19 using Chest X-Ray Images	The authors built a model from scratch trained on more than 100,000 CXR images and fine-tuned on COVID-19 samples by using transfer learning to overcome the problem of a limited number of COVID-19 samples.	COVID-19: 305 Pneumonia: 322 Other diseases: 51119 Normal: 57910	4	90.13%	https://www.sirm.org/category/senza-categoria/covid-19/ https://radiopaedia.org/search?utf8=%E2%9C%93&q=covid&scope=all&lang=us https://github.com/ieee8023/covid-chestxray-dataset https://www.kaggle.com/nih-chest-xrays/data

Lawrence O. Hall, Rahul Paul, Dmitry B. Goldgof, and Gregory M. Goldgof	Finding covid-19 from chest x-rays using deep learning on a small dataset	Transfer learning on pre-trained models from ImageNet (VGG16 and ResNet50) is performed by replacing the last three layers of both VGG16 and ResNet50 with the trainable part, followed by a 64-unit connected layer with dropout, 10-fold validation, and a classification layer with a sigmoid output.	COVID-19: 135 Pneumonia: 320	2	89.20%	https://github.com/ieee8023/covid-chestxray-dataset https://radiopaedia.org/search?utf8=%E2%9C%93&q=covid&scope=all&lang=us http://www.sirm.org/en/ https://www.kaggle.com/nih-chestxrays/data

M.E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al Emadi, et al.	Can AI Help in Screening Viral and COVID-19 Pneumonia?	Transfer learning on pre-trained models (AlexNet, ResNet18, DenseNet201 and SqueezeNet) is performed, and the Softmax and classification layers of the pre-trained networks are modified.	COVID-19: 423 Pneumonia: 1485 Normal: 1579	2 3	99.70% 99.50%	https://www.kaggle.com/tawsifurrahman/ https://www.sirm.org/category/senza-categoria/ https://github.com/ieee8023/covid-chestxray-dataset https://radiopaedia.org/search?lang=us&page=4&q=covid+19&scope=all&utf8=%E2%9C%93 https://threadreaderapp.com/thread/1243928581983670272.html https://nihcc.app.box.com/v/ChestXray-NIHCC

Enzo Tartaglione, Carlo Alberto Barbano, Claudio Berzovini, Marco Calandri, and Marco Grangetto	Unveiling COVID-19 from CHEST X-Ray with Deep Learning: A Hurdles Race with Small Data	Pretrained CNN models (ResNet-18, Resnet-50, COVID-Net, and DenseNet-121) are utilized as feature extractors. The feature extractor is pre-trained on combinations of CXR datasets, and then fine-tuned on COVID data.	COVID-19: 405 Non COVID-19: 473	2	91.00%	https://github.com/ieee8023/covid-chestxray-dataset Locally collected from Hospital in Piedmont (CDSS) https://data.mendeley.com/datasets/rscbjbr9sj/2/files/f12eaf6d-6023-432f-acc9-80c9d7393433 https://www.kaggle.com/c/rsna-pneumonia-detection-challenge http://openi.nlm.nih.gov/imgs/collections/NLM-MontgomeryCXRSet.zip

Dailin Lv , Wuteng Qi, Yunxiang Li, Lingling Sun, and Yaqi Wang	A cascade network for detecting COVID-19 using chest X-rays	The authors proposed a Cascade-SEMEnet consisting of a SEME-ResNet50 for detecting the type of lung infection and a DenseNet169 for the subdivision of viral pneumonia. This model is used to diagnose lung disease and COVID-19.	Dataset I Normal: 1591 Bacteria pneumonia: 2772 Viral pneumonia: 1493 Dataset II COVID-19: 125 Pneumonia: 316	2 3	97.10% 85.60%	https://data.mendeley.com/datasets/rscbjbr9sj/2 https://github.com/ieee8023/covid-chestxray-dataset
Tianyang Li, Zhongyi Han, Benzheng Wei, Yuanjie Zheng, Yanfei Hong, Jinyu Cong, Wei Zhang, Xue Zhu,	Robust Screening of COVID-19 from Chest X-ray via Discriminative Cost-Sensitive Learning	The authors built a framework called discriminative cost-sensitive learning, which offers two features, i.e., fine-grained classification and cost-sensitive learning.	COVID-19: 239 Pneumonia: 1000 Normal: 1000	3	97.01%	https://github.com/ieee8023/covid-chestxray-dataset https://www.kaggle.com/andrewmvd/convid19-x-rays https://www.cc.nih.gov/drd/summers.html

Rahul Kumar, Ridhi Arora, Vipul Bansal, Vinodh J Sahayasheela, Himanshu Buckchash, Javed Imran, Narayanan Narayanan, Ganesh N Pandian, and Balasubramanian Raman	Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers	Pre-trained CNN model (ResNet152) is utilized for feature extraction. The SMOTE technique is used to balance the dataset. Machine learning classifiers (Random Forest, XGB, etc.) use the concerned features to classify the chest X-ray images.	COVID-19: 62 Pneumonia: 5610 Normal: 2916	3	97.70%	https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia https://towardsdatascience.com/covid19-public-dataset-on-gcp-nlp-knowledge-graph-193e628fa5cb

Jianpeng Zhang, Yutong Xie, Guansong Pang, Zhibin Liao, Johan Verjans, Wenxing Li, Zongji Sun, Jian He, Yi Li, Chunhua Shen, and Yong Xia	Viral Pneumonia Screening on Chest X-rays Using Confidence-Aware Anomaly Detection	Anomaly detection is used in the last layer to classify COVID-19 CXR images. This layer generates a scalar anomaly score that assigns statistically significantly large classification scores and anomaly scores to CXR images with COVID-19.	X-VIRAL dataset viral pneumonia: 5977 Normal: 37393 X-COVID dataset COVID-19: 106 Normal: 107 Open-COVID dataset COVID-19: 493 SARS: 16 MERS: 10	2	80.65%	https://github.com/ieee8023/covid-chestxray-dataset https://www.cc.nih.gov/drd/summers.html

Emtiaz Hussain, Mahmudul Hasan, Md Anisur Rahman, Ickjai Lee, Tasmi Tamanna, and Mohammad Zavid Parvez	CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images	The authors proposed a novel 22-layer CNN model for COVID detection using chest X-ray and CT images. The proposed CNN model consists of convolutional, max pooling, dense layer, flatten layer, and three activation functions, namely Sigmoid, ReLU, and Leaky ReLU.	COVID-19: 2843 Bacteria and viral pneumonia: 1439 Normal: 3108	2 3 4	99.10% 94.20% 91.20%	https://github.com/ieee8023/covid-chestxray-dataset https://github.com/UCSD-AI4H/COVID-CT https://github.com/agchung/Figure1-COVID-chestxray-dataset https://github.com/agchung/Actualmed-COVID-chestxray-dataset https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset https://www.kaggle.com/khoongweihao/covid19-xray-dataset-train-test-sets

Fatih Demir	DeepCoroNet: A deep LSTM approach for automated detection of COVID-19 cases from chest X-ray images	The author proposed a new approach based on the deep Long Short-Term Memory (LSTM) model. The proposed LSTM network consists of 5 layers, which includes the LSTM, the fully connected, the Rectified Linear Unit (ReLU), the dropout, and the softmax layers. The classification process is performed with the activation function in the softmax layer.	COVID-19: 361 Pneumonia: 500 Normal: 200	3	100%	https://www.kaggle.com/bachrr/covid-chest-xray https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia https://nihcc.app.box.com/v/ChestXray-NIHCC/file/220660789610

Shanjiang Tang, Chunjiang Wang, Jiangtian Nie, Neeraj Kumar, Senior Member, IEEE, Yang Zhang, Member, IEEE, Zehui Xiong, Member, IEEE, and Ahmed Barnawi	EDL-COVID: Ensemble Deep Learning for COVID-19 Case Detection From Chest X-Ray Images	The authors proposed EDL-COVID, an ensemble deep learning model employing deep learning and ensemble learning. The EDL-COVID model is generated by combining multiple snapshot models of COVID-Net, which has pioneered an open-sourced COVID-19 case detection method with deep neural network processed chest X-ray images, by employing a proposed weighted averaging ensembling method that is aware of different sensitivities of deep learning models on different class types.	COVID-19: 573 Pneumonia: 6053 Normal: 8851	3	95.00%	https://github.com/agchung/Actualmed-COVID-chestxray-dataset https://github.com/ieee8023/covid-chestxray-dataset https://kaggle.com/tawsifurrahman/covid19-radiography-database https://github.com/agchung/Figure1-COVID-chestxray-dataset https://kaggle.com/c/rsna-pneumonia-detection-challenge

Anunay Gupta, Anjum, Shreyansh Gupta, and Rahul Katarya	InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using Chest X-ray	The authors proposed an ensemble model called InstaCovNet-19, which combines 5 pretrained CNN models (Inception v3, MobileNetV2, ResNet101, NASNet, and Xception) using the Integrated Stacking technique. Each pretrained model is fine-tuned on COVID-19 X-ray images by replacing the classification part of the model with two dense layers of 128x1 and 2x1 or 3x1. The outputs of every one of the models are then combined and passed through a dense layer with 128 nodes having Relu function as an activation function.	COVID-19: 361 Pneumonia: 1341 Normal: 1345	2 3	99.52% 99.08%	https://www.kaggle.com/tawsifurrahman/covid19-radiography-database?rvi=1 https://github.com/ieee8023/covid-chestxray-dataset

Ali Narin, Ceren Kaya, and Ziynet Pamuk	Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks	Pre-trained models (ResNet50, ResNet101, ResNet152, InceptionV3, and Inception-ResNetV2) are utilized as a base model with a global spatial average pooling layer, a fully connected layer, and a logistic layer added at the end.	COVID-19: 341 Bacterial Pneumonia: 2772 Viral pneumonia: 1493 Normal: 2800	2 2 2	96.10% 99.50% 99.70%	https://github.com/ieee8023/covid-chestxray-dataset https://nihcc.app.box.com/v/ChestXray-NIHCC https://www.kaggle.com/paultimothymooney/chest-xray-pnemonia

Narinder Singh Punn & Sonali Agarwal	Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks	A pre-trained CNN model (NASNet-Large, DenseNet169, InceptionV3, ResNet18, and Inception ResNet V2) was utilized as a base model. Fine-tuning was performed by keeping nontrainable layers as the base model and adding four trainable convolutional layers, one fully connected layer and one classification layer.	COVID-19: 108 Pneumonia: 515 Tuberculosis: 58 Normal: 533	2 3 4	97.00% 94.00% 95.00%	https://github.com/nspunn1993/COVID-19-PA-CXR-fused-dataset

Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng & Bo Xu	A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19)	Instead of using the whole image, region of interests (RoIs) or image patches are provided as input. These image patches are inputted into a CNN pre-trained network (Inception v3) for feature extraction, followed by a fully connected classification layer for classification.	COVID-19: 325 Normal: 740	2	79.30%	CT images are locally collected from Xi'an Jiaotong University First Affiliated Hospital (center 1), Nanchang University First Hospital (center 2), and Xi'an No.8 Hospital of Xi'an Medical College (center 3)

Asmaa Abbas, Mohammed M. Abdelsamea, and Mohamed Medhat Gaber	Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network	Pre-trained CNN models (ALexNet, VGG, ResNet, GoogleNet, and SqueezeNet) were utilized for feature extraction. Training was accomplished using sophisticated gradient descent optimization method. A composition layer was added to refine the final classification layer of images. This method can detect irregularities in the dataset by investigating class boundaries using class decomposition. For the decomposition of classes, the K-mean clustering method is used.	COVID-19: 1053 Normal: 80 SARS: 11		93.10%	https://github.com/ieee8023/covid-chestxray-dataset www.macnet.or.jp/jsrt2/cdrom_nodules.html

Pedro R. A. S. Bassi & Romis Attux	A Deep Convolutional Neural Network for COVID-19 Detection Using Chest X-rays	ChexNet was trained with transfer learning twice. A pre-trained CheXNet (already trained on ImageNet and then on ChestX-ray14) was utilized and trained on the COVID-19 database.	COVID-19: 439 Pneumonia: 1255 Normal: 370	3	97.80%	https://github.com/ieee8023/covid-chestxray-dataset https://sirm.org/en/category/articles/covid-19-database/ https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

CHAPTER 3

METHODOLOGY

This section discusses the proposed method for accurately predicting COVID-19 using CXR images, which consists of a deep feature learning model for feature extraction and a machine learning classifier for classification.

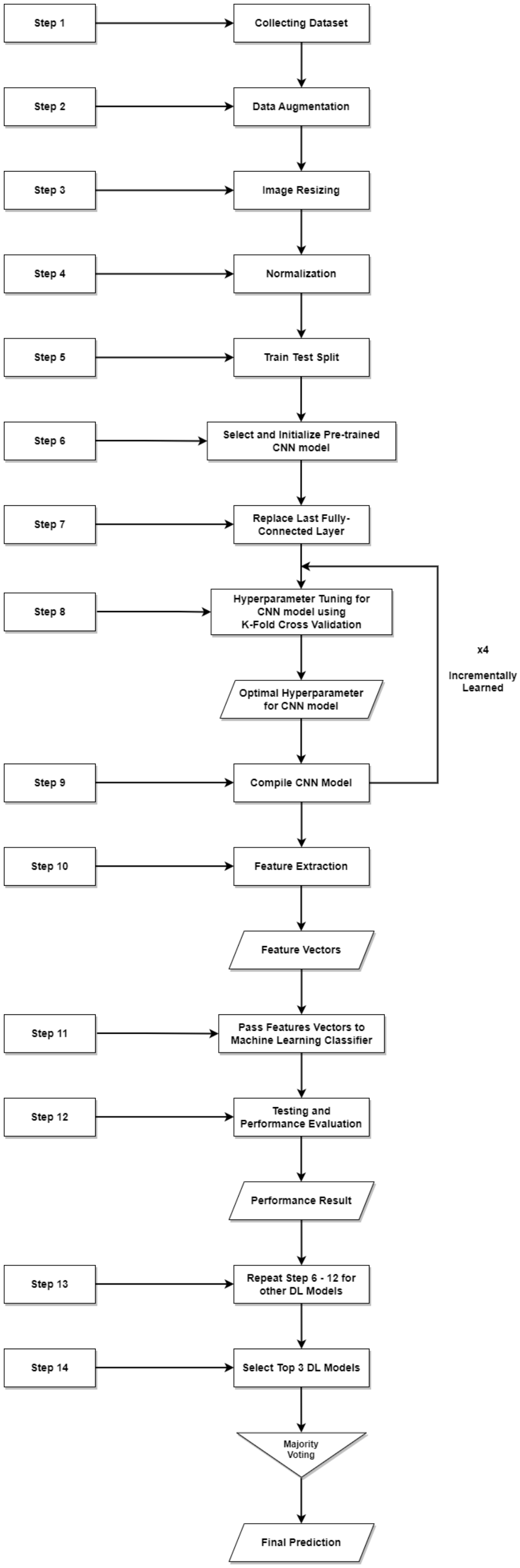

Figure 3.1:
The workflow of the proposed method to demonstrate the process from beginning to end. Details on each step are given below.

Step 1: Collecting Dataset

During this step, data for some predefined categories were collected from various publicly available online sources. In particular, in the case of detecting COVID-19 from chest radiography imaging, a dataset consisting of not only COVID-19 cases but also healthy cases, as well as cases of viral pneumonia-like cases, was prepared. Such a comprehensive dataset enables the model to distinguish properly between cases, resulting in a more accurate classification.

Step 2: Data Augmentation

One possible solution to increase the amount of available data and avoid overfitting issues is data augmentation techniques. Data augmentation incorporates a collection of methods that improve the attributes and size of training datasets. Thus, deep learning models can perform better when these techniques are employed. There are a number of image augmentation techniques.

1. Flipping: Flipping the horizontal axis is more commonly used than flipping the vertical axis. Flipping has been verified as valuable on datasets like ImageNet and CIFAR-10 (Krizhevsky and Hinton, 2009; Deng et al., 2009). Moreover, it is highly simple to implement. Flipping is label-preserving except for text.

2. Rotation: Rotation augmentations can be obtained by rotating an image left or right within 0 to 360 degrees around the axis. The rotation degree parameter greatly determines the suitability of the rotation augmentations. However, the data label cannot be preserved post-transformation when the rotation degree increases.

3. Translation: To avoid positional bias within the image data, a very useful transformation is to shift the image up, down, left, or right. For instance, it is common that the whole dataset images are centred. Moreover, the tested dataset should be entirely made up of centred images to test the model. The

spatial dimensions of the image post-augmentation are preserved using this padding.

Step 3, 4, 5: Image resizing, Normalization, Train test split

The acquired CXR images have variable shapes and sizes, which makes effective classification difficult. Image pre-processing was performed to ensure effective classification. The CXR images were resized to meet the input requirements of different CNNs. For instance, SqueezeNet requires images resized to 227 × 227 pixels while MobileNetV2, ResNet18, ResNet101, VGG19, and DenseNet201 require images resized to 224 × 224 pixels. InceptionV3 requires images resized to 299×299 pixels. All images were resized according to the pre-trained model standards.

After that, the dataset was normalized within a range of 0 and 1. Every pixel of images present in the dataset was multiplied by a factor of 1/255. This has been done to make the dataset consistent in terms of pixel intensity. Before proceeding to the next phase, the dataset was split into three parts: the training set (66.67%), the validation set (16.67%), and the testing set (16.67%).

Steps 6, 7: Select and Initialize Pre-trained CNN model, Replace Last Fully Connected Layer

This project evaluated eleven pre-trained CNN models, including five comparatively shallow networks (ResNet50V2, MobileNetV2, VGG16, Xception, and DenseNet121) and six deep networks (ResNet152, InceptionV3, Inception ResNetV2, VGG19, DenseNet201, and NasNetLarge). After pre-processing the data, a pre-trained CNN model was selected to instantiate the model's convolutional base, retaining both its initial architecture and all learned weights. The model's hyper-parameters were initialized, including the optimizer, learning rate, batch size, epoch, and dropout rate. The last fully connected layer of the pre-trained CNN was replaced with a new fully connected layer with three prediction classes, and only the fully connected layer was trained while the remaining layers' weights were frozen.

Step 8: Hyperparameter Tuning for CNN model using K-Fold Cross Validation

To overcome the limitations of computing resources, the training set of 14,400 images was divided into four subsets, each containing 3,600 images. Then, the deep learning model was trained and fine-tuned on the first subset of 3,600 images. To optimize the model's hyperparameters, a 5-fold cross-validation approach was used. The first subset was randomly divided into 5 folds, where 4 out of the 5 folds were used for training the CNN model, and the other fold was used for validation. This approach of validation was repeated 5 times by shifting the validation and training folds. The average result was calculated based on the result of each individual fold, and the configuration with the highest validation accuracy was considered the optimum set of hyperparameters.

Step 9: Compile CNN Model

After fine-tuning and optimizing the hyperparameters of the deep learning model on the first subset, the model was compiled with the optimal set of hyperparameters and was saved for the next round of training with the second subset of 3,600 images. Steps 8 and 9 were repeated for the remaining subsets of the training set, each compiled with their respective optimal set of hyperparameters.

Steps 10, 11: Feature Extraction, Pass Feature Vectors to Machine Learning Classifier

After finishing the training of the deep learning model on all four subsets of the training set, the model was used to extract features in the CXR images. The final feature representation obtained was interpreted as a one-dimensional vector. These acquired feature vectors were then fed into a machine learning predictive classifier to perform the classification task. For this purpose, the XGBoost classifier was utilized to classify the CXR images into three categories: COVID-19, Normal, and Pneumonia.

Step 12: Training and Performance Evaluation

During this step, the XGBoost classifier was tested on CXR images that had not been shown to the model during previous training steps. The performance of the model in predicting new cases was examined, and the generalization ability was investigated. To evaluate the efficacy of the model, the confusion matrix, along with Area under Curve (AUC), were estimated to gain an understanding of the proposed methodology and its potential for detailed classification. Different metrics, such as accuracy, sensitivity, specificity, precision, F1-Score, False Positive Rate, AUC, MAE, and RMSE, were used to measure the usefulness and productivity of the classification model.

Step 13: Repeat Step 6 – 12 for other DL Models

Repeated the steps outlined in Step 6 through 12 for all the remaining deep learning models.

Step 14: Select Top 3 DL Model and Apply Majority Voting Approach

Selected the top three deep learning models based on their accuracy and applied a majority voting approach to their predictions to obtain the final prediction.

CHAPTER 4

EXPERIMENTAL RESULTS

4.1 Performance Evaluation of the Best Performing Model for Each Approach

The performance metrics of the best performing model from each approach were evaluated in the following aspects.

4.1.1 Accuracy Evaluation

Eleven deep learning models were trained using three approaches, i.e., single deep learning model approach, incrementally learned single deep learning model approach, and incrementally learned multiple deep learning models with majority voting approach. Table 4.1 presents the prediction accuracy for the best-performing models for each approach. It shows that the incrementally learned multiple deep learning models with majority voting approach using ResNet152V2, DenseNet201, and VGG16 outperformed the single deep learning model approach by about 3.22% and performed relatively better than the incrementally learned single deep learning model approach by about 0.02%.

Table 4.1: Overall Prediction Accuracy.

Best Performing Model	**Accuracy (%)**
Single model: ResNet152V2 + XGBoost	91.36
Incremental learned model: ResNet152V2 + XGBoost	94.56
Voting: i) ResNet152V2 + XGBoost ii) DenseNet201 + XGBoost iii) VGG16 + XGBoost	**94.58**

However, for a multiclass problem, judging a model's effectiveness solely on higher accuracy is insufficient. It is necessary to consider the other two important class-level metrics, namely, sensitivity and positive predictive value (PPV) as well.

4.1.2 Sensitivity Evaluation

In medical analysis, the sensitivity of a disease can be interpreted as the proportion of people with a certain disease that have been successfully identified. Taking COVID-19, for example, achieving a high sensitivity is quite important since no affected people should be omitted during COVID-19 testing; otherwise, the affected people who have been omitted cannot receive immediate treatment, and they can affect others. Table 4.2 gives a sensitivity analysis of the best-performing model for each approach concerning each class type. It can be observed that it is seldom to have a model that works best for all three classes. For example, the incrementally learned single deep learning model has the highest sensitivity in the Normal class but not for two other class types. In comparison, the incrementally learned multiple deep learning models with majority voting obtained the highest sensitivities for both the Pneumonia and COVID-19 classes, although its sensitivity for the Normal class is not the best across all models. From a practical point of view, there is no doubt to consider the voting approach since highly sensitive screening for infectious diseases such as COVID-19 is very important.

Table 4.2: Sensitivities of the Best Performing Model for Each Approach in Each Class.

Sensitivity (%)			
Best Performing Model	Normal	Pneumonia	COVID-19
Single model: ResNet152V2 + XGBoost	91.00	93.00	86.00
Incremental learned model: ResNet152V2 + XGBoost	**97.00**	92.00	88.00

Voting: i) ResNet152V2 + XGBoost ii) DenseNet201 + XGBoost iii) VGG16 + XGBoost	96.00	**97.00**	**89.00**

4.1.3 PPV Evaluation

Positive predictive value (PPV) denotes the probability of positive results that are true positive results in diagnostic tests. If this value is low, there are many false positives, and follow-up testing is required for a more reliable result. For COVID-19 screening, if a model's PPV is low, it cannot be judged or confirmed that a person with a positive test result is a true COVID-19 case, and additional accurate testing is necessary. Table 4.3 presents the PPV analysis for the best-performing model of each approach on each class type. Still, no model performs the best for all class types. The incrementally learned single deep learning model has the highest PPV for the pneumonia class, while the majority voting approach achieves the highest PPVs for the Normal and COVID-19 classes.

Table 4.3: PPV of the Best Performing Model for Each Approach in Each Class.

Positive Predictive Value (%)			
Best Performing Model	Normal	Pneumonia	COVID-19
Single model: ResNet152V2 + XGBoost	86.00	92.00	93.00
Incremental learned model: ResNet152V2 + XGBoost	94.00	**97.00**	86.00
Voting: i) ResNet152V2 + XGBoost ii) DenseNet201 + XGBoost iii) VGG16 + XGBoost	**97.00**	89.00	**96.00**

In summary, although no model outperformed others on all metrics for all class types, the incrementally learned multiple deep learning models with majority voting approach is the best choice for COVID-19 case detection since it performs relatively better than others on accuracy, sensitivity, and PPV for the COVID-19 class type.

4.2 Further Performance Evaluation of Incrementally Learned Multiple Deep Learning Models with Majority Voting Approach

In this section, the incrementally learned multiple deep learning models with majority voting approach was further evaluated from various perspectives, including confusion matrix, ROC curves, and training and validation loss.

4.2.1 Confusion Matrix

Figure 4.1 presents the confusion matrix for the proposed approach analysing the test dataset, which consists of CXR images of 1200 COVID-19 cases, 1200 pneumonia cases, and 1200 normal cases. For COVID-19 testing, only 124 out of 1200 CXR images of COVID-19 were not detected correctly, and 15 out of 3600 CXR images were mistakenly identified as COVID-19. This indicates that the error ratio is relatively small compared to the total number of CXR images.

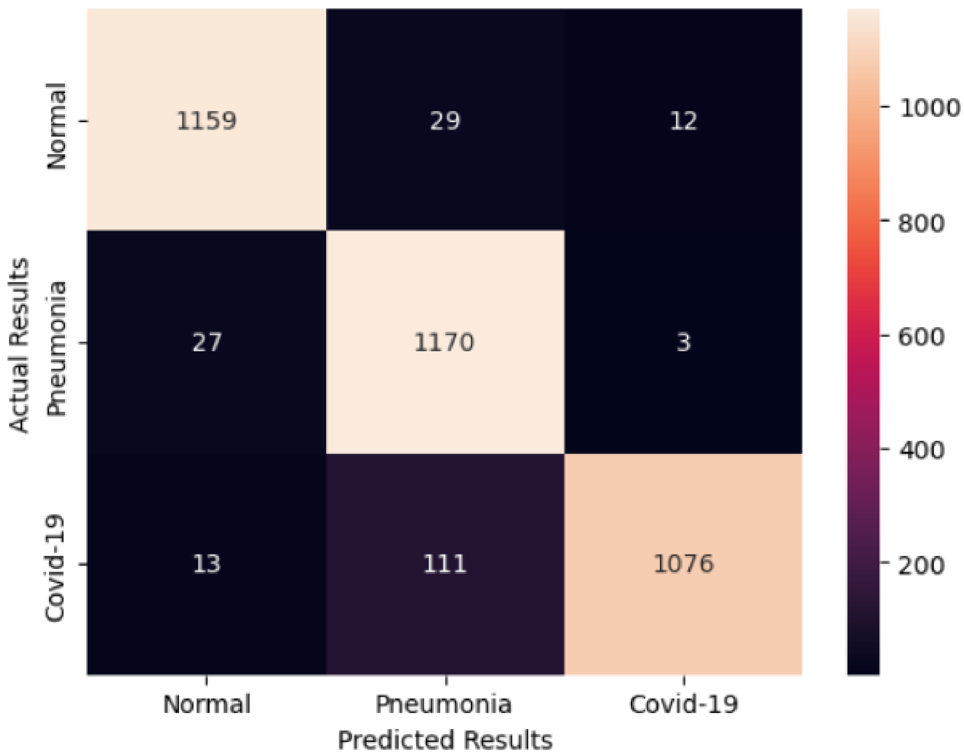

Figure 4.1: The Confusion Matrix for the Incrementally Learned Multiple Deep Learning Models with Majority Voting on the test dataset containing 1200 normal cases, 1200 pneumonia cases, and 1200 COVID-19 cases.

4.2.2 ROC Curves

To show the detection capability of the incrementally learned multiple deep learning models with majority voting approach, ROC curves were generated to depict its prediction on the test dataset with respect to each class type, as shown in Figure 4.2. A larger area under the ROC curve indicates a better prediction ability. It can be observed that the ROC area for each class under COVID-19 is much closer to the maximum value of one, indicating that the proposed method has a good prediction capability for COVID-19 in practice.

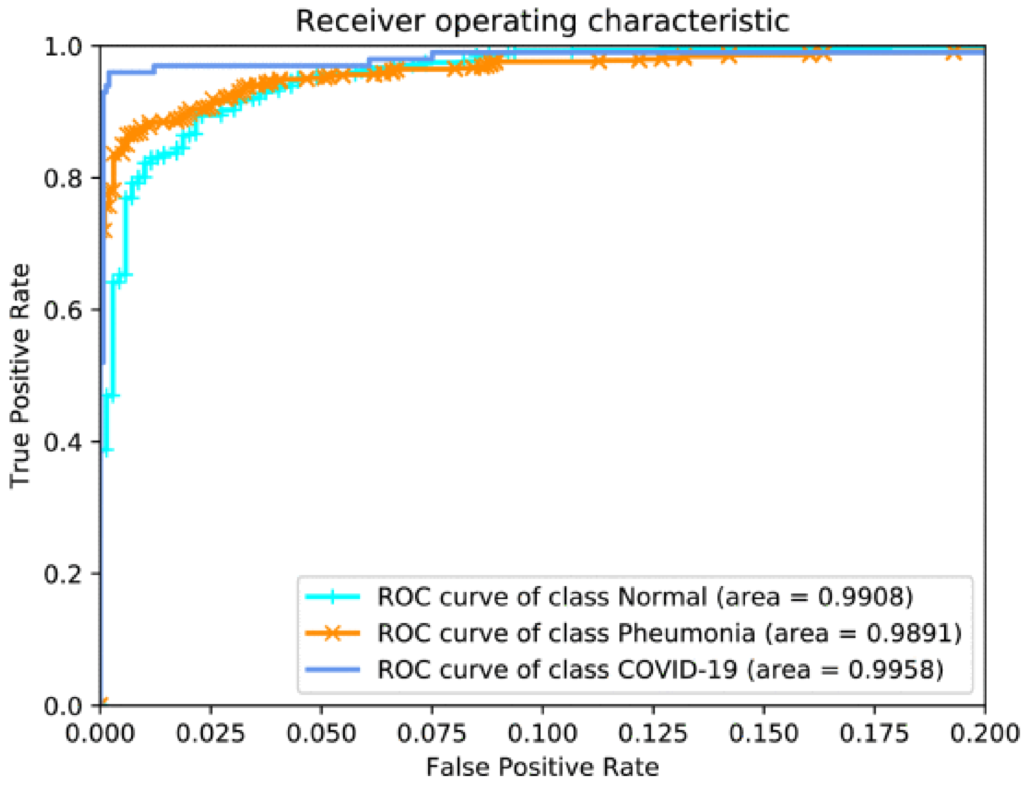

Figure 4.2:ROC curves of the Incrementally Learned Multiple Deep Learning Models with Majority Voting approach for Prediction on the Test Dataset with Respect to Each Class Type.

4.2.3 Training and Validation Loss

The proposed approach with ResNet152, DenseNet201, and VGG16 (the combination of the best performer) has shown good learning curves, as can be observed from Figure 4.3, 4.4, and 4.5, which depict the training and validation loss for each individual model. The learning curves provide insight into how the learning performance changes over the number of epochs and help diagnose any problems that can lead to an underfit or an overfit model. The training and validation loss gradually decrease over the number of epochs and reach a point of stability, indicating good fits. Moreover, the generalization gap between the training and validation loss learning curves is minimal (nearly zero in an ideal situation), indicating that the model is not overfitting the data and can generalize well to new, unseen data. These findings suggest that the proposed approach is highly promising and could perform exceptionally well in real-world scenarios. As such, this model can be confidently tested in actual environments, and it can be expected to produce favourable outcomes.

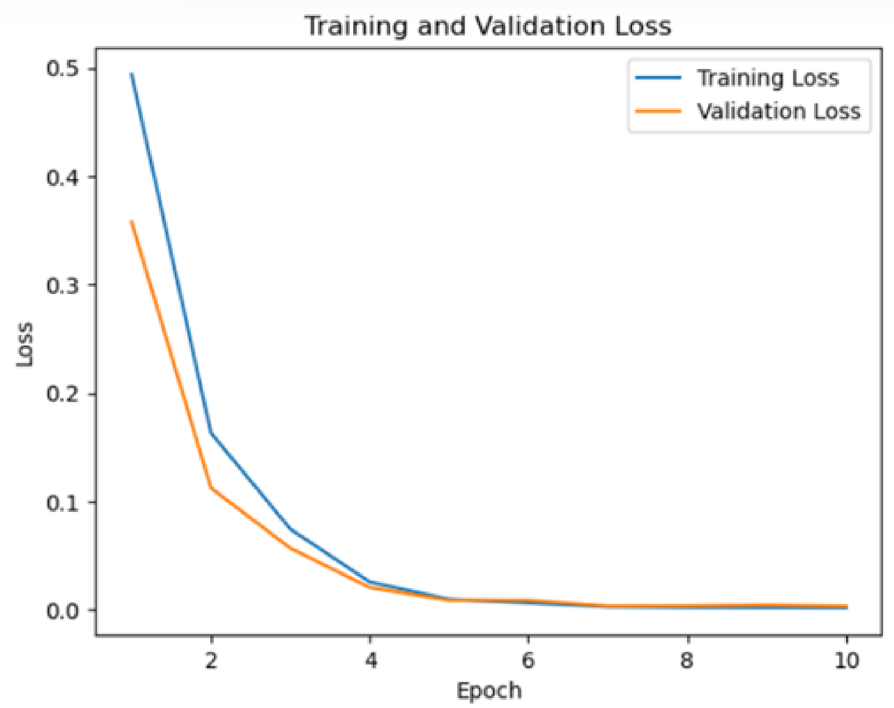

Figure 4.3: Training and Validation Loss for ResNet152.

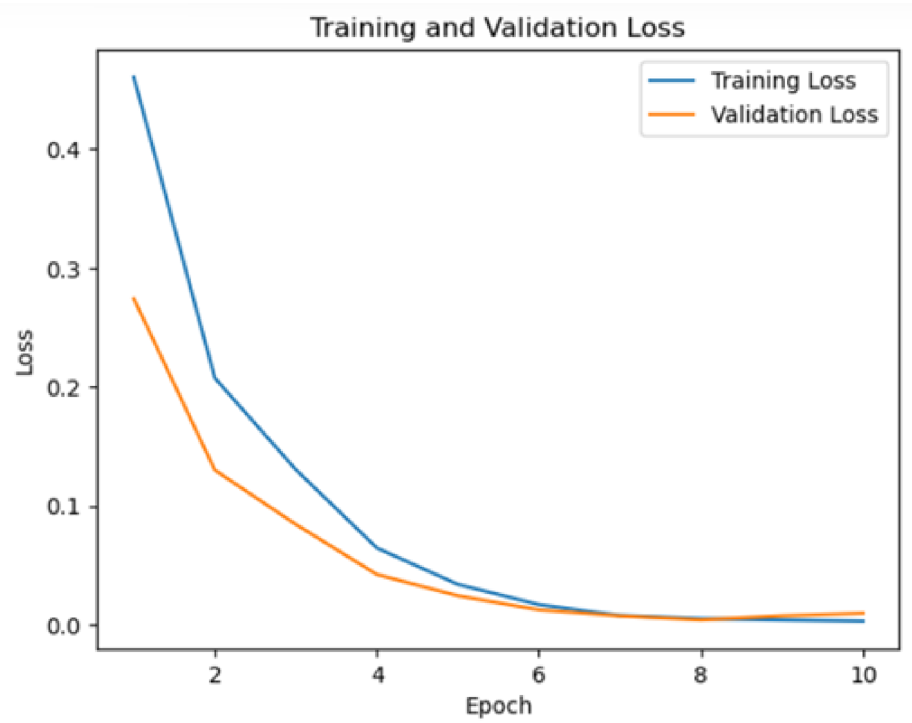

Figure 4.4: Training and Validation Loss for DenseNet201.

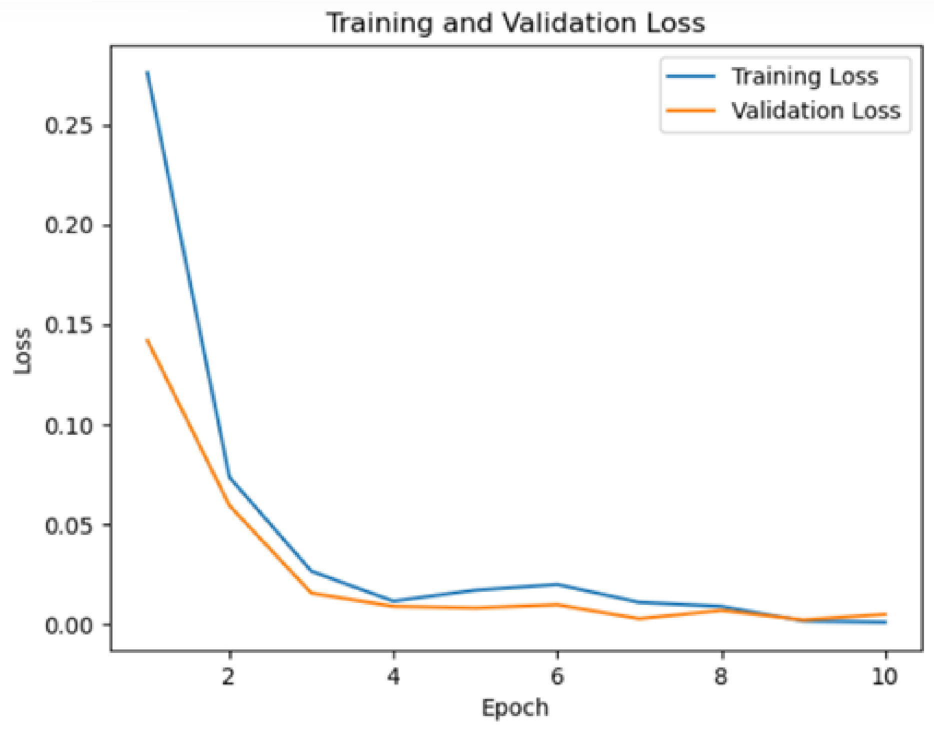

Figure 4.5: Training and Validation Loss for VGG16.

CHAPTER 5

CONCLUSIONS

The COVID-19 coronavirus is a recent virus that leads to pneumonia, which can be detected using CXR images. This paper investigated the use of (i) single CNNs, (ii) incrementally learned single CNNs, and (iii) incrementally learned multiple CNNs with majority voting to extract features from CXR images. Then, an XGBoost classifier was used with each of these CNNs to detect COVID-19. The proposed model addressed the limitations of computing resources by using an incrementally learned approach and provided a robust solution for detecting COVID-19 from CXR images. Additionally, the use of majority voting approach slightly improved the detection accuracy. The dataset used in this research consisted of 22,900 CXR images with three categories: Normal, Pneumonia, and COVID-19. The dataset is split into 66.67% for training, 16.67% for validation, and 16.67% for testing. Through the paper, eleven pretrained CNNs (ResNet50V2, ResNet152, DenseNet121, DenseNet201, VGG16, VGG19, MobileNetV2, Inception ResNetV2, InceptionV3, Xception, NasNetLarge) were selected as deep transfer learning models. The results show that using the XGBoost classifier with incrementally learned single CNN and incrementally learned multiple CNNs gave good and comparable detection accuracy (94.56% and 94.58%). The best performer was the incrementally learned multiple CNNs with majority voting, which used ResNet152, DenseNet201, and VGG16. These results demonstrate the effectiveness of our proposed method in detecting COVID-19 from CXR images and its potential for clinical applications.

REFERENCES

Abbas, A., Abdelsamea, M.M. and Gaber, M.M., 2021. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Applied Intelligence, 51(2), pp.854-864.

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

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

Agarwal, V., Lohani, M.C., Bist, A.S., Harahap, E.P. and Khoirunisa, A., 2022. Analysis Of Deep Learning Techniques For Chest X-Ray Classification In Context Of Covid-19. ADI Journal on Recent Innovation, 3(2), pp.208-216.

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z. and Xia, L., 2020. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology.

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z. and Xia, L., 2020. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. and Farhan, L., 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8(1), pp.1-74.

Apostolopoulos, I.D. and Mpesiana, T.A., 2020. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Physical and engineering sciences in medicine, 43(2), pp.635-640.

Basu, S., Mitra, S. and Saha, N., 2020, December. Deep learning for screening covid-19 using chest x-ray images. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 2521-2527). IEEE.

Bejani, M.M. and Ghatee, M., 2021. A systematic review on overfitting control in shallow and deep neural networks. Artificial Intelligence Review, 54(8), pp.6391-6438.

Bernheim, A., Mei, X., Huang, M., Yang, Y., Fayad, Z.A., Zhang, N., Diao, K., Lin, B., Zhu, X., Li, K. and Li, S., 2020. Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. Radiology.

Bodenreider, O., 2005. The Unified Medical Language System Integrating Biomedical Terminology.

Brenner, D.J. and Hall, E.J., 2007. Computed tomography—an increasing source of radiation exposure. New England journal of medicine, 357(22), pp.2277-2284.

Bustos, A., Pertusa, A., Salinas, J.M. and de la Iglesia-Vayá, M., 2020. Padchest: A large chest x-ray image dataset with multi-label annotated reports. Medical image analysis, 66, p.101797.

Cai, Z. and Peng, C., 2021, January. A study on training fine-tuning of convolutional neural networks. In 2021 13th International Conference on Knowledge and Smart Technology (KST) (pp. 84-89). IEEE.

Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K.G. and Murphy, K., 2021. Deep learning for chest X-ray analysis: A survey. Medical Image Analysis, 72, p.102125.

Candemir, S., Jaeger, S., Palaniappan, K., Musco, J.P., Singh, R.K., Xue, Z., Karargyris, A., Antani, S., Thoma, G. and McDonald, C.J., 2013. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE transactions on medical imaging, 33(2), pp.577-590.

Chakraborty, S., Mondal, R., Singh, P.K., Sarkar, R. and Bhattacharjee, D., 2021. Transfer learning with fine tuning for human action recognition from still images. Multimedia Tools and Applications, 80(13), pp.20547-20578.

Chan, J.F.W., To, K.K.W., Tse, H., Jin, D.Y. and Yuen, K.Y., 2013. Interspecies transmission and emergence of novel viruses: lessons from bats and birds. Trends in microbiology, 21(10), pp.544-555.

Chen, Y., Liu, Q. and Guo, D., 2020. Emerging coronaviruses: genome structure, replication, and pathogenesis. Journal of medical virology, 92(4), pp.418-423.

Cheplygina, V., 2019. Cats or CAT scans: Transfer learning from natural or medical image source data sets?. Current Opinion in Biomedical Engineering, 9, pp.21-27.

Chouhan, V., Singh, S.K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R. and De Albuquerque, V.H.C., 2020. A novel transfer learning based approach for pneumonia detection in chest X-ray images. Applied Sciences, 10(2), p.559.

Chowdhury, M.E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Al Emadi, N. and Reaz, M.B.I., 2020. Can AI help in screening viral and COVID-19 pneumonia?. IEEE Access, 8, pp.132665-132676.

Chowdhury, N.K., Rahman, M. and Kabir, M.A., 2020. PDCOVIDNet: a parallel-dilated convolutional neural network architecture for detecting COVID-19 from chest X-ray images. Health information science and systems, 8(1), pp.1-14.

Cohen, J.P., Dao, L., Roth, K., Morrison, P., Bengio, Y., Abbasi, A.F., Shen, B., Mahsa, H.K., Ghassemi, M., Li, H. and Duong, T.Q., 2020. Predicting covid-19 pneumonia severity on chest x-ray with deep learning. Cureus, 12(7).

Cohen, J.P., Dao, L., Roth, K., Morrison, P., Bengio, Y., Abbasi, A.F., Shen, B., Mahsa, H.K., Ghassemi, M., Li, H. and Duong, T.Q., 2020. Predicting covid-19 pneumonia severity on chest x-ray with deep learning. Cureus, 12(7).

Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q. and Ghassemi, M., 2020. Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988.

Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). Ieee.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

Deng, L., 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA transactions on Signal and Information Processing, 3.

Domhan, T., Springenberg, J.T. and Hutter, F., 2015, June. Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves. In Twenty-fourth international joint conference on artificial intelligence.

European Centre for Disease Prevention and Control. An agency of the European Union. Link: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases

Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P. and Ji, W., 2020. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology.

Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P. and Ji, W., 2020. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology.

Glorot, X. and Bengio, Y., 2010, March. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (pp. 249-256). JMLR Workshop and Conference Proceedings.

Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Ou, C.Q., He, J.X., Liu, L., Shan, H., Lei, C.L., Hui, D.S. and Du, B., 2019. China medical treatment expert group for Covid-19. Clinical characteristics of coronavirus disease, 382(18), pp.1708-1720.

Gupta, A., Gupta, S. and Katarya, R., 2021. InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using Chest X-ray. Applied Soft Computing, 99, p.106859.

Ha, H.Y., Yang, Y., Pouyanfar, S., Tian, H. and Chen, S.C., 2015, November. Correlation-based deep learning for multimedia semantic concept detection. In International Conference on Web Information Systems Engineering (pp. 473-487). Springer, Cham.

Hall, L.O., Paul, R., Goldgof, D.B. and Goldgof, G.M., 2020. Finding covid-19 from chest x-rays using deep learning on a small dataset. arXiv preprint arXiv:2004.02060.

Han, D., Liu, Q. and Fan, W., 2018. A new image classification method using CNN transfer learning and web data augmentation. Expert Systems with Applications, 95, pp.43-56.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Hossin, M. and Sulaiman, M.N., 2015. A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process, 5(2), p.1.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X. and Cheng, Z., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet, 395(10223), pp.497-506.

Huang, G., Sun, Y., Liu, Z., Sedra, D. and Weinberger, K.Q., 2016, October. Deep networks with stochastic depth. In European conference on computer vision (pp. 646-661). Springer, Cham.

Huang, H., Li, X. and Chen, C., 2018. Individual tree crown detection and delineation from very-high-resolution UAV images based on bias field and marker-controlled watershed segmentation algorithms. IEEE Journal of selected topics in applied earth observations and remote sensing, 11(7), pp.2253-2262.

Huh, M., Agrawal, P. and Efros, A.A., 2016. What makes ImageNet good for transfer learning?. arXiv preprint arXiv:1608.08614.

Hussain, E., Hasan, M., Rahman, M.A., Lee, I., Tamanna, T. and Parvez, M.Z., 2021. CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. Chaos, Solitons & Fractals, 142, p.110495.

Jacofsky, D., Jacofsky, E.M. and Jacofsky, M., 2020. Understanding antibody testing for COVID-19. The Journal of arthroplasty, 35(7), pp.S74-S81.

Jaeger, S., Candemir, S., Antani, S., Wáng, Y.X.J., Lu, P.X. and Thoma, G., 2014. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quantitative imaging in medicine and surgery, 4(6), p.475.

Jain, R., Gupta, M., Taneja, S. and Hemanth, D.J., 2021. Deep learning based detection and analysis of COVID-19 on chest X-ray images. Applied Intelligence, 51(3), pp.1690-1700.

Kakodkar, P., Kaka, N. and Baig, M.N., 2020. A comprehensive literature review on the clinical presentation, and management of the pandemic coronavirus disease 2019 (COVID-19). Cureus, 12(4).

Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F. and Dong, J., 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5), pp.1122-1131.

Khalifa, N.E.M., Loey, M., Taha, M.H.N. and Mohamed, H.N.E.T., 2019. Deep transfer learning models for medical diabetic retinopathy detection. Acta Informatica Medica, 27(5), p.327.

Khalifa, N.E.M., Loey, M.O.H.A.M.E.D. and Taha, M.H.N., 2020. Insect pests recognition based on deep transfer learning models. J. Theor. Appl. Inf. Technol, 98(1), pp.60-68.

Koutsoukas, A., Monaghan, K.J., Li, X. and Huan, J., 2017. Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. Journal of cheminformatics, 9(1), pp.1-13.

Krizhevsky, A. and Hinton, G., 2009. Learning multiple layers of features from tiny images.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Kumar, S., Maurya, V.K., Prasad, A.K., Bhatt, M.L. and Saxena, S.K., 2020. Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). Virusdisease, 31(1), pp.13-21.

Lakhani, P. and Sundaram, B., 2017. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology, 284(2), pp.574-582.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), pp.2278-2324.

Li, Y. and Xia, L., 2020. Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management. Ajr Am J Roentgenol, 214(6), pp.1280-1286.

Li, Y., Yao, L., Li, J., Chen, L., Song, Y., Cai, Z. and Yang, C., 2020. Stability issues of RT‐PCR testing of SARS‐CoV‐2 for hospitalized patients clinically diagnosed with COVID‐19. Journal of medical virology, 92(7), pp.903-908.

Liu, C., Cao, Y., Alcantara, M., Liu, B., Brunette, M., Peinado, J. and Curioso, W., 2017, September. TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network. In 2017 IEEE international conference on image processing (ICIP) (pp. 2314-2318). IEEE.

Lowe, D.G., 1999, September. Object recognition from local scale-invariant features. In Proceedings of the seventh IEEE international conference on computer vision (Vol. 2, pp. 1150-1157). Ieee.

Mahase, E., 2020. Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate.

Mangal, A., Kalia, S., Rajgopal, H., Rangarajan, K., Namboodiri, V., Banerjee, S. and Arora, C., 2020. CovidAID: COVID-19 detection using chest X-ray. arXiv preprint arXiv:2004.09803.

Mangal, A., Kalia, S., Rajgopal, H., Rangarajan, K., Namboodiri, V., Banerjee, S. and Arora, C., 2020. CovidAID: COVID-19 detection using chest X-ray. arXiv preprint arXiv:2004.09803.

Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R. and Muharemagic, E., 2015. Deep learning applications and challenges in big data analytics. Journal of big data, 2(1), pp.1-21.

Narin, A., Kaya, C. and Pamuk, Z., 2021. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. Pattern Analysis and Applications, 24(3), pp.1207-1220.

Ng, M.Y., Lee, E.Y., Yang, J., Yang, F., Li, X., Wang, H., Lui, M.M.S., Lo, C.S.Y., Leung, B., Khong, P.L. and Hui, C.K.M., 2020. Imaging profile of the COVID-19 infection: radiologic findings and literature review. Radiology: Cardiothoracic Imaging, 2(1).

Nishiura, H., Linton, N.M. and Akhmetzhanov, A.R., 2020. Serial interval of novel coronavirus (COVID-19) infections. International journal of infectious diseases, 93, pp.284-286.

Ouchicha, C., Ammor, O. and Meknassi, M., 2020. CVDNet: A novel deep learning architecture for detection of coronavirus (Covid-19) from chest x-ray images. Chaos, Solitons & Fractals, 140, p.110245.

Ouchicha, C., Ammor, O. and Meknassi, M., 2020. CVDNet: A novel deep learning architecture for detection of coronavirus (Covid-19) from chest x-ray images. Chaos, Solitons & Fractals, 140, p.110245.

Ozaki, Y., Yano, M. and Onishi, M., 2017. Effective hyperparameter optimization using Nelder-Mead method in deep learning. IPSJ Transactions on Computer Vision and Applications, 9(1), pp.1-12.

Özkaya, U., Öztürk, Ş. and Barstugan, M., 2020. Coronavirus (COVID-19) classification using deep features fusion and ranking technique. In Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach (pp. 281-295). Springer, Cham.

Pambudi, A., Widayanti, R. and Edastama, P., 2021. Trust and Acceptance of E-Banking Technology Effect of Mediation on Customer Relationship Management Performance. ADI Journal on Recent Innovation, 3(1), pp.87-96.

Perlman, S. and Netland, J., 2009. Coronaviruses post-SARS: update on replication and pathogenesis. Nature reviews microbiology, 7(6), pp.439-450.

Pham, T.D., 2021. Classification of COVID-19 chest X-rays with deep learning: new models or fine tuning?. Health Information Science and Systems, 9(1), pp.1-11.

Polikar, R., 2012. Ensemble learning. In Ensemble machine learning (pp. 1-34). Springer, Boston, MA.

Pormohammad, A., Ghorbani, S., Khatami, A., Farzi, R., Baradaran, B., Turner, D.L., Turner, R.J., Bahr, N.C. and Idrovo, J.P., 2020. Comparison of confirmed COVID‐19 with SARS and MERS cases ‐ Clinical characteristics, laboratory findings, radiographic signs and outcomes: a systematic review and meta‐analysis. Reviews in Medical Virology, 30(4), p.e2112.

Prompetchara, E., Ketloy, C. and Palaga, T., 2020. Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. Asian Pacific journal of allergy and immunology, 38(1), pp.1-9.

Robinson, C., Trivedi, A., Blazes, M., Ortiz, A., Desbiens, J., Gupta, S., Dodhia, R., Bhatraju, P.K., Liles, W.C., Lee, A. and Kalpathy-Cramer, J., 2021. Deep learning models for COVID-19 chest x-ray classification: Preventing shortcut learning using feature disentanglement. medRxiv.

Rubin, G.D., Ryerson, C.J., Haramati, L.B., Sverzellati, N., Kanne, J.P., Raoof, S., Schluger, N.W., Volpi, A., Yim, J.J., Martin, I.B. and Anderson, D.J., 2020. The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society. Chest, 158(1), pp.106-116.

Rubin, G.D., Ryerson, C.J., Haramati, L.B., Sverzellati, N., Kanne, J.P., Raoof, S., Schluger, N.W., Volpi, A., Yim, J.J., Martin, I.B. and Anderson, D.J., 2020. The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society. Chest, 158(1), pp.106-116.

Salehi, S., Abedi, A., Balakrishnan, S. and Gholamrezanezhad, A., 2020. Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients. Ajr Am J Roentgenol, 215(1), pp.87-93.

Sethy, P.K. and Behera, S.K., 2020. Detection of coronavirus disease (covid-19) based on deep features.

Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y. and Shen, D., 2020. Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. IEEE reviews in biomedical engineering, 14, pp.4-15.

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

Smith, L.N., 2018. A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay. arXiv preprint arXiv:1803.09820.

Song, Y., Zheng, S., Li, L., Zhang, X., Zhang, X., Huang, Z., Chen, J., Wang, R., Zhao, H., Chong, Y. and Shen, J., 2021. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. IEEE/ACM transactions on computational biology and bioinformatics, 18(6), pp.2775-2780.

Soon, F.C., Khaw, H.Y., Chuah, J.H. and Kanesan, J., 2018. Hyper‐parameters optimisation of deep CNN architecture for vehicle logo recognition. IET Intelligent Transport Systems, 12(8), pp.939-946.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

Tahamtan, A. and Ardebili, A., 2020. Real-time RT-PCR in COVID-19 detection: issues affecting the results. Expert review of molecular diagnostics, 20(5), pp.453-454.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B. and Liang, J., 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning?. IEEE transactions on medical imaging, 35(5), pp.1299-1312.

Tang, S., Wang, C., Nie, J., Kumar, N., Zhang, Y., Xiong, Z. and Barnawi, A., 2021. EDL-COVID: Ensemble deep learning for COVID-19 case detection from chest X-ray images. IEEE Transactions on Industrial Informatics, 17(9), pp.6539-6549.

Vayá, M.D.L.I., Saborit, J.M., Montell, J.A., Pertusa, A., Bustos, A., Cazorla, M., Galant, J., Barber, X., Orozco-Beltrán, D., García-García, F. and Caparrós, M., 2020. Bimcv covid-19+: a large annotated dataset of rx and ct images from covid-19 patients. arXiv preprint arXiv:2006.01174.

Wang, L., Lin, Z.Q. and Wong, A., 2020. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Scientific Reports, 10(1), pp.1-12.

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G. and Tan, W., 2020. Detection of SARS-CoV-2 in different types of clinical specimens. Jama, 323(18), pp.1843-1844.

West, C.P., Montori, V.M. and Sampathkumar, P., 2020, June. COVID-19 testing: the threat of false-negative results. In Mayo clinic proceedings (Vol. 95, No. 6, pp. 1127-1129). Elsevier.

WHO Coronavirus (COVID-19) Dashboard [Internet]. World Health Organization. 2021 [cited 26 October 2021]. Available from: https://covid19.who.int/

Wikramaratna, P., Paton, R.S., Ghafari, M. and Lourenco, J., 2020. Estimating false-negative detection rate of SARS-CoV-2 by RT-PCR. MedRxiv, 2020.

Winther, H.B., Laser, H., Gerbel, S., Maschke, S.K., Hinrichs, J.B., Vogel-Claussen, J., Wacker, F.K., Höper, M.M. and Meyer, B.C., 2020. COVID-19 image repository. Figshare (Dataset).

World Health Organization, 2020. Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance, 2 March 2020 (No. WHO/COVID-19/laboratory/2020.4). World Health Organization.

World Health Organization, 2020. WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S. and Zhou, Z.H., 2008. Top 10 algorithms in data mining. Knowledge and information systems, 14(1), pp.1-37.

Xia, J., Tong, J., Liu, M., Shen, Y. and Guo, D., 2020. Evaluation of coronavirus in tears and conjunctival secretions of patients with SARS‐CoV‐2 infection. Journal of medical virology, 92(6), pp.589-594.

Xiao, A.T., Tong, Y.X. and Zhang, S., 2020. False‐negative of RT‐PCR and prolonged nucleic acid conversion in COVID‐19: rather than recurrence. Journal of medical virology.

Xie, S., Girshick, R., Dollár, P., Tu, Z. and He, K., 2017. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500).

Yamashita, R., Nishio, M., Do, R.K.G. and Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. Insights into imaging, 9(4), pp.611-629.

Yan, Y., Chen, M., Sadiq, S. and Shyu, M.L., 2017. Efficient imbalanced multimedia concept retrieval by deep learning on spark clusters. International Journal of Multimedia Data Engineering and Management (IJMDEM), 8(1), pp.1-20.

Yan, Y., Chen, M., Shyu, M.L. and Chen, S.C., 2015, December. Deep learning for imbalanced multimedia data classification. In 2015 IEEE international symposium on multimedia (ISM) (pp. 483-488). IEEE.

Yang, Y., Yang, M., Shen, C., Wang, F., Yuan, J., Li, J., Zhang, M., Wang, Z., Xing, L., Wei, J. and Peng, L., 2020. Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. MedRxiv.

Yang, Y., Yang, M., Yuan, J., Wang, F., Wang, Z., Li, J., Zhang, M., Xing, L., Wei, J., Peng, L. and Wong, G., 2020. Laboratory diagnosis and monitoring the viral shedding of SARS-CoV-2 infection. The innovation, 1(3), p.100061.

Yin, W., Kann, K., Yu, M. and Schütze, H., 2017. Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv:1702.01923.

Yu, Y., Si, X., Hu, C. and Zhang, J., 2019. A review of recurrent neural networks: LSTM cells and network architectures. Neural computation, 31(7), pp.1235-1270.

Zhang, J., Xie, Y., Li, Y., Shen, C. and Xia, Y., 2020. Covid-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv preprint arXiv:2003.12338, 27.

Zhao, X., Liu, L., Qi, S., Teng, Y., Li, J. and Qian, W., 2018. Agile convolutional neural network for pulmonary nodule classification using CT images. International journal of computer assisted radiology and surgery, 13(4), pp.585-595.

Zou, L., Zheng, J., Miao, C., Mckeown, M.J. and Wang, Z.J., 2017. 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI. Ieee Access, 5, pp.23626-23636.