THE UNIVERSITY of ADELAIDE

ENG 4001 Research Project
Progress Report

COVID or flu? That's the question!

Group: Abbott-2022s1-EEE-UG-13148

Prepared by:
Andreas Kotsanis (a1720576)
Nivin Jose Kovukunnel (a1749677)
Mohammad Shafaie (a1748929)

Supervisors:
Professor Derek Abbott
Dr Mohsen Dorraki

Executive Summary

With the rampant spread of COVID-19, many testing centres have become overwhelmed, suffering from long patient queue times as well as a delayed turnaround time between testing and receiving the result. To reduce the impact on testing and healthcare centres, research must be undertaken to investigate novel techniques for COVID-19 diagnosis. One technique, which will be the focus of this report, is to employ deep learning models for automated detection of COVID-19 using a patient's chest X-ray image. In particular, the aim of this project is to construct various deep learning models to determine whether patients with COVID-19 can be differentiated from those with viral pneumonia.

As part of the project's scope, each model will be analysed for its efficiency, relevancy to image classification, interpretability, and framework support availability. This project will utilise two chest X-ray image datasets to train and test these models. The first dataset comes from the University of Montreal, which contains 330 chest X-ray images. The second dataset is from the COVID-19 Radiography Database which contains 15153 chest X-ray images, collated from various studies across several Asian Universities. In both datasets, each image is attributed to one of the following classes: COVID-19, viral pneumonia and normal.

This project consists of three objectives. The first objective was to perform image pre-processing and data augmentation to prepare the dataset for the deep learning models. Currently, the following pre-processing techniques have been implemented: image compression (converting each image to one standardised size) and pixel normalisation. Several data augmentation techniques have also been implemented such as rotating, shifting and flipping the images. The second objective is to design and implement various deep learning models to perform three-way classification of the chest X-ray images from both datasets. At this stage, four deep learning convolutional neural network architectures have been implemented for chest X-ray image classification. These models consist of three pre-trained architectures (VGG-16, Inception-V3 and DenseNet-121) and one custom made convolutional neural network architecture that was specifically designed for this project. The final objective was to evaluate each model using a series of performance metrics such as accuracy, precision, recall and F1-score. From the architectures that have currently been implemented, the best performing model was found to be DenseNet-121 which obtained the highest testing accuracy (91%) for both datasets.

Nonetheless, the project was subject to some limitations as the used dataset did not include key patient descriptors such as age, gender, contamination period, medical history. Furthermore, the COVID-19 strain type could not be retrieved from the metadata. The future work for this project includes applying more advanced pre-processing techniques such as image contrasting and region of interest extraction. In addition to this, other pre-trained CNN architectures such as ResNet, MobileNet and EfficientNet will be evaluated for chest X-ray classification. Alternatively, this project could also consider other novel image classification techniques such as Vision Transformers.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
AUC	Area Under the Curve
CLAHE	Contrast Limited Adaptive Histogram Equalisation
COVID-19	Coronavirus 2019
CIR	Contrast Improvement Ratio
CNN	Convolutional Neural Network
CT	Computer Tomography
CXR	Chest X-Ray
DenseNet	Dense Connected Convolutional Networks
GGO	Ground Glass Opacity
HOG	Histogram-Oriented Gradient
ML	Machine Learning
MobileNet	Mobile Network
RAT	Rapid Antigen Testing
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
RT-PCR	Reverse transcription-polymerase chain reaction
SVM	Support Vector Machines
VGG	Visual Geometry Group
WHO	World Health Organisation

1 Introduction

1.1 Project Overview

To increase the testing accessibility of Coronavirus Disease 2019 (COVID-19), it is critical to develop new methods that are capable of detecting this disease. The focus of this project, 'COVID or flu? That's the question!' is to explore advanced image classification techniques to correctly classify chest X-ray (CXR) images into one of the following categories: COVID-19, normal and viral pneumonia.

1.2 Background

According to World Health Organisation (WHO), there has been a total of approximately 452.2 million COVID-19 cases and 6.3 million deaths, as of March 14, 2022 [1]. The COVID-19 virus has caused an ongoing global pandemic and with newer variants being developed rapidly, the world is struggling to adapt. The first case of the COVID-19 virus was reported in December 2019, in Wuhan, Hubei Province, China, from where it began to transmit rapidly to the rest of the countries around the world [2].

For the diagnosis of COVID-19, various methods are used with the most common method being the Reverse transcription-polymerase chain reaction (RT-PCR) testing [3]. Although RT-PCR tests can be cost-effective, patients can expect a delay in testing and receiving results, especially during an outbreak. Numerous studies also concluded that RT-PCR testing has low sensitivity during the early stages of the infection, contributing to false-negative results [4], [5], [6]. Chest imaging using X-rays and computer tomography (CT) scans are protocols currently practiced by healthcare centres to patients, that show strong respiratory symptoms [7]. Contrast to other popular methods such as RT-PCR testing and Rapid Antigen Testing (RAT), the process of using CXR imaging is very simple, fast and provides greater accuracy due to its high sensitivity during the early stages of the infection [8]. Viral pneumonia is still one of the leading causes of death [9]. According to WHO, chest imaging using X-rays is the best method for diagnosing pneumonia [10]. For more information on the techniques and/or methods involved in CXR imaging refer to Appendix A1.

Over the recent years, there has been a significant development in the areas of Artificial Intelligence (AI) and Machine Learning (ML). With increasing computational power and growing amount of quality available data, various ML methods have already shown good performance for medical imaging diagnosis [11]. However, there are still areas of improvement in the analysis, as it requires proficiency and incorporates a diverse range of techniques to improve, accelerate and generate an accurate diagnosis. Several studies have showed that deep learning methods, more specifically, Convolutional Neural Networks (CNNs), have achieved better performance on image classification problems in comparison to other traditional ML models [12], [13], [14]. In this project, existing studies will be used as a guide to verify the work being conducted. Furthermore, this project will also focus on exploring deep learning models such as CNNs to perform accurate CXR image classification.

1.3 Motivation

The main motivation behind this project is to develop a new method for diagnosing COVID-19 which can serve as a viable alternative to patients who require high accuracy testing with a quick turnaround time. In this project, advanced image classification techniques will be implemented to accurately differentiate normal patients from those with COVID-19. In addition to this, it is also important to be able to distinguish patients with COVID-19 and those with viral pneumonia. In comparison to viral pneumonia, COVID-19 is high transmissible and can display little to no symptoms, especially during the incubation period. Therefore, it is also important to differentiate patients with COVID-19 and viral pneumonia, to help contain the spread of the COVID-19 virus and to assign appropriate medical treatments and measures.

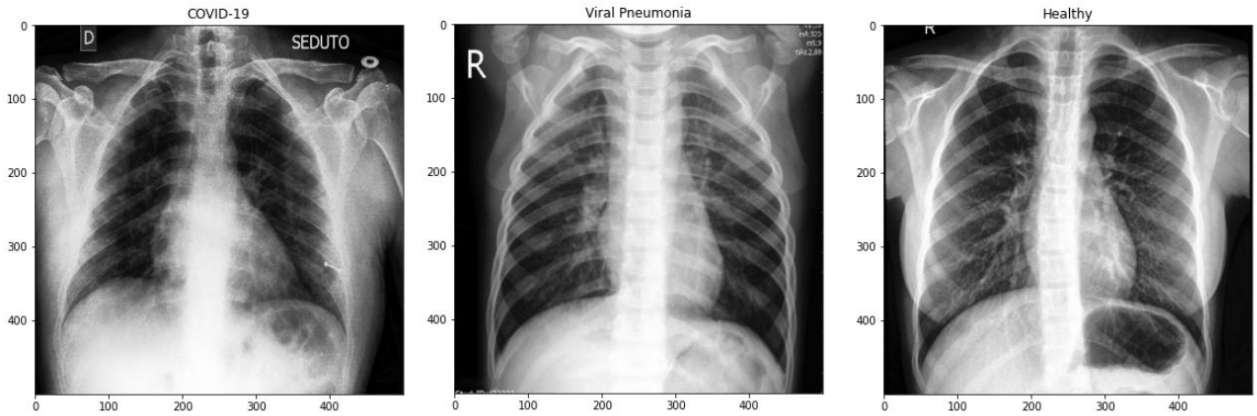

Figure 1 Chest X-ray image of COVID-19, Viral Pneumonia and Healthy patients from the University of Montreal dataset

Both COVID-19 and viral pneumonia display similar symptoms such as coughing, fever and shortness of breath. To an untrained eye, both diseases show similar characteristics in CXR images, as illustrated in Figure 1. As such, differentiating patients with COVID-19 from those with viral pneumonia using CXR images can be tedious, even for expert radiologists. Therefore, this project provides healthcare centres with an alternative method using automated ML models to detect COVID-19 and viral pneumonia diseases using CXR images. This process is also automated which has the additional benefit of lessening the burden and stress on healthcare workers, especially during outbreaks when there is a significant influx of patients that need to undergo testing.

1.4 Aims and Scope

The aim of this project is to explore advanced techniques for image classification to determine whether CXR images from patients with COVID-19 can be differentiated from those with normal lungs or viral pneumonia. This project will involve designing and constructing ML models, that can extract specific features from CXR images and then learn from it, to perform accurate classification.

The scope of this project involves identifying and selecting numerous effective ML methods, that can be used to perform classification on CXR images. The ML model's efficiency, relevancy to image classification problems, interpretability, and framework support availability, will all be analysed for selection. This project will require an analysis of CXR images from normal patients as well as those that have been diagnosed with COVID-19 and viral pneumonia. There is no need to perform experiments or contact healthcare centres for CXR data as there exists numerous online datasets containing CXR images of COVID-19, viral pneumonia and normal patients, that are free and publicly available.

This project will make use of two CXR image datasets. The first dataset (which will be referred to as Dataset A in this document) is from the University of Montreal which consists of 130 COVID-19, 100 viral pneumonia and 100 normal CXR images. The journal article, 'COVID-19 Image Data Collection: Prospective Predictions Are the Future' by Cohen *et al.* [15] also provided a 'Data Availability Statement,' allowing public use of their data. The second dataset (referred to as Dataset B) is the COVID-19 Radiography Database, which was created from various research across several Asian universities. The journal articles, 'Can AI Help in Screening Viral and COVID-19 Pneumonia?' by Chowdhury *et al.* [16] and 'Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images' by Rahman *et al.* [17] have provided a 'Data Availability Statement,' allowing public use of their data. The dataset constitutes of 3616 COVID-19, 1,345 viral pneumonia and 10,192 normal CXR images.

1.5 Objectives

The technical objectives for this project are summarised in Table 1. This table also includes the specific tasks that will need to be undertaken for each objective as well as the expected deliverables/outcomes.

Table 1 Project objectives and key specifications/outcomes

#	Objective description	Specifications	Deliverables / outcomes
1.	Perform image pre-processing and data augmentation on the dataset provided to prepare the dataset for training.	• Resize images to 224 × 224 pixel size. • Pixel value range will be normalised to the range 0 to 1. • Additional data will be generated via data augmentation.	• All CXR images have been pre-processed and are ready for training the ML models. • Additional CXR images generated via data augmentation to increase robustness of ML models.
2.	Design and construct ML models to differentiate between COVID-19 and viral pneumonia in CXR images.	• 60% of the dataset provided will be used to train the ML models. Another 20% will be used to validate and fine tune the ML models. The remaining 20% will be used to testing and evaluation. • Various CNN architectures will be trained to perform classification on CXR images.	• Multiple ML models capable of classifying between COVID-19, viral pneumonia and normal CXR images.
3.	Evaluate ML models to determine the best performing model.	• Hold-out set was used to evaluate the ML models. • Several evaluation metrics will be used to evaluate models.	• Accuracy, precision, recall and F1-score quantities determined for each ML model. • Selection of best performing ML model that can accurately classify COVID-19, viral pneumonia and normal CXR images.

1.6 Document Overview

The next chapter in this document is the literature review. This chapter will concisely summarise the previous research that has been undertaken regarding CXR image classification for COVID-19 detection. Particular emphasis will be placed on the different methods and techniques that these studies implemented as well as the overall performance of the ML models. This chapter will then be proceeded by the method that was used to execute this project. This will include a summary of the data preparation techniques that were used as well as a description of the various CNN architectures that have been implemented. The Results chapter consists of an evaluation of the CNN architectures that have been implemented using performance metrics such as accuracy, precision, recall and F1-score. The subsequent chapter will then address the limitations of this project with reference to the methodology and the results obtained. This report will conclude with a Completion Plan chapter which will outline the remaining tasks that have yet to be completed.

2 Literature Review

2.1 Introduction

Ever since the emergence of the global COVID-19 pandemic, numerous studies have attempted to detect traces of COVID-19 in CXR images using a wide range of image classification techniques. Although a large number of these studies have already demonstrated a high classification accuracy for CXR images, there is limited research comparing these different deep learning techniques for the same dataset. As such, the aim of this literature review is to collate the results from these various studies and identify the techniques and/or models that yielded the best results for CXR image classification.

2.2 Findings

This review consists of two sections. In the first section, an analysis is performed looking into the data pre-processing techniques that have been used to prepare CXR images for the deep learning models. The second section will then draw upon literature regarding various image classification techniques that have previously been used to classify CXR images.

2.2.1 Data Preparation

The main techniques that have been used by earlier studies to prepare the CXR data can be allocated into one of the following categories: data pre-processing and data augmentation. In this case, data pre-processing refers to a group of techniques that are used to improve the quality and/or contrast of a CXR image. On the other hand, data augmentation refers to a group of techniques that are used to artificially increase the size of a dataset by creating modified copies of existing images.

2.2.1.1 Data Pre-processing

In this case, data pre-processing refers to several techniques that are used to enhance the quality of the dataset that is used to train the ML model. One of these techniques involves adjusting the contrast on the CXR images to make features encompassing the lungs more prominent. In a study conducted by Reynaldi *et al.* [18], the contrast in CXR images was adjusted by applying a Contrast Limited Adaptive Histogram Equalisation (CLAHE) filter. This filter partitions the image into similarly-sized non-overlapping regions and then performs adaptive histogram equalisation on each section [19]. As part of their results, the researchers of this study [18] compared the accuracy of an ML model with and without using the CLAHE filter. The results from this experiment showed increased accuracy, sensitivity and specificity when the CLAHE filter was applied. These results are supported by a separate study conducted by Umri *et al.* [20], in which the accuracy of the CNN model was increased by 1% after the CLAHE filter was used to pre-process the dataset. The CLAHE filter can also be used in conjunction with other pre-processing techniques such as intensity normalisation to remove low or high frequencies from the CXR images [21].

Contrast enhancement can also be performed by means of mathematical morphology [22]. Mathematical morphology involves creating a structural element from the original image. This structural element consists of binary numbers (1s and 0s), which is then overlayed with the

original image to suppress some of the complex background tissue commonly found in medical CXR images [23]. An example of contrast enhancement using mathematical morphology was used successfully implemented in a study by Sarki *et al.* [22] to prepare a CXR dataset for an ML model. The effectiveness of contrast enhancement via mathematical morphology was demonstrated in a study by Kimori [23], in which this technique was compared against several other popular contrast enhancement methods, including CLAHE. The results showed a higher contrast improvement ratio (CIR) for the mathematical morphology technique over the other methods that were considered in this paper [23].

Data segmentation is another technique that can be applied to the CXR images during the data pre-processing stage. This technique is used to crop down the CXR images to eliminate irrelevant information such as the patient's arms, neck, stomach etc., which may otherwise affect how the ML model classifies the images. This technique was applied by Tabik *et al.* [24] during their investigation into how deep learning models can be used to detect COVID-19 from a set of CXR images. To accomplish this, without manually cropping the images, a U-Net segmentation model [25] was employed to isolate the lungs in the CXR images.

A slightly less common data pre-processing technique is wavelet decomposition. This technique was used by Singh *et al.* [26] to convert the CXR images from the spatial domain to the frequency domain, from which multiresolution analysis can then performed. This technique offers several advantages such as reducing the background noise that plagues CXR images while also improving the image contrast in the process [27].

2.2.1.2 Data Augmentation

Data augmentation is a technique that is used to expand the number of images contained in a dataset by making copies of the existing data with some slight alterations. This technique has been used in various studies [24], [28], [29] with the aim of increasing the ML model performance by subjecting it to a wider and richer dataset. Data augmentation was used by Li *et al.* [29], when attempting to classify COVID-19 CXR images using a novel neural network model. In this study, the authors noted that the number of COVID-19 CXR images (400) were far fewer than those in the viral pneumonia and normal classes (7000 and 10000 respectively). To avoid the introduction of biases in the ML model, the COVID-19 dataset was expanded twofold via data augmentation [29]. A similar process was used by Mishra *et al.* [30] to expand a limited dataset of 1800 CXR to over 15000 images. The data augmentation techniques that were implemented in these studies include sheering, zooming, shifting, rotating and (horizontally) flipping the CXR images [29, 30]. One way to employ these techniques is to use python libraries, such as OpenCV, as described in the study conducted by Hernandez *et al.* [31]. Unlike previous studies [29, 30], the study conducted by Hernandez *et al.* [31] also included Gaussian noise as a data augmentation technique. This was done with the aim of increasing the robustness of the ML model to poor CXR image quality. When altering these images, the authors in studies [29], [30] noted that they had to be careful to not create images that were drastically different than those in the original dataset. Thus, the authors opted to only used one data augmentation technique at a time for each particular CXR image. In addition to this, the degree that these images were rotated/shifted was limited to a range of +/- 10% [29], [30].

2.2.2 Image Classification Techniques

The two main image classification techniques that will be discussed in this section are CNNs and tradition computer vision techniques. CNNs are deep learning algorithms specific to image classification. It employs end-to-end learning which requires the algorithm to determine an underlying pattern within a group of images so that they may be correctly classified. There are various architectures for CNNs. The main architectures that will be discussed in this literature review are: ResNet, VGG, DenseNet and MobileNet. On the other hand, in traditional computer vision techniques, the user is responsible for extracting the features of interest from these images manually. These extracted features are then used to train an ML model so that these images may be accurately classified.

2.2.2.1 Convolutional Neural Networks

ResNet is one of the most common CNN architectures for image classification. This architecture is renowned for its quick training despite consisting of network layers that are considerably deeper than many other typical CNN architectures [32]. A variation of this architecture, ResNet-101, was used by Reynaldi *et al.* [18] to identify CXR images for normal patients and those with COVID-19. This model was able to accurately classify these images with a reported accuracy of 99.61%, a recall of 99.62% and a specificity of 99.6% [18]. These results are supported by a separate, yet similar, study which used support vector machines (SVM) alongside the ResNet-50 architecture to categorise CXR images as either COVID-19, pneumonia or normal [33]. In this study, Narin [33] reported an overall accuracy, recall and specificity of 94.86%, 96.04% and 96.62% respectively using a SVM Cubic model.

The VGG architecture is another popular CNN model that has been used to classify CXR images. Umri *et al.* [20] used the VGG-16 architecture (consisting of 16 convolutional layers) to achieve up to a 99% training accuracy and a 97% validation accuracy when classifying CXR images into COVID-19 and normal classes. In a similar study, Sarki *et al.* [22] used a pre-trained VGG-16 model to achieve a 100% accuracy for binary classification (COVID-19/normal) and an 87.5% accuracy for tertiary classification (COVID-19/pneumonia/normal). Alam *et al.* [34] used VGG-19 architecture (with 19 convolutional layers) to a 99.49% accuracy. This study also implemented individual feature extraction techniques such as histogram-oriented gradient (HOG) which is believed to have increased the accuracy of the model [34].

DenseNet is a CNN architecture which consists of many convolutional layers but makes use of concatenation to reduce the number of hyperparameters required [35]. Chaudhary *et al.* [36] implemented a pre-trained DenseNet architecture to detect the presence of COVID-19 in CT scans. This was done in a two-stage classification framework which yielded an overall accuracy of 89.3% for a three-class classification: COVID-19, pneumonia and normal. Similar results were also obtained by Albahli, Ayub and Shiraz [37] who used a pre-trained DenseNet architecture to obtain an accuracy of 92% using the same three-class classification identified earlier. Montalbo [35] was able to boost the performance of the DenseNet architecture by using a lightweight model with partial layer freezing and feature extraction to classify CXR images. With these techniques, Montalbo [35] achieved an accuracy of 99.84%, a precision of 99%, a recall of 100% and a F1-score of 99%. Bohmrah and Kaur [38] used CXR images to trained several variations of DenseNet architectures including DenseNet-121, DenseNet-169 and DenseNet-201. The best performing

model was DenseNet-201, which achieved an accuracy of 95.2% [38]. The authors also experimentally determined that DenseNet architectures produced better results using the RMSprop optimiser over the more commonly used Adam and Adamax optimisers [38].

MobileNet is a CNN architecture designed for mobile and vision-based embedded applications [39]. Mohammadi *et al.* [40] achieved an accuracy of 99.1% when using the MobileNet architecture to perform binary classification of CXR images (COVID-19 or normal). A modified version of this model was used by Tangudu, Kakarla and Venkateswarlu [41] to perform the same binary classification as above. This model used a residual separable convolutional block in conjunction with a pre-trained MobileNet model [41]. Overall, this model achieved a 99% accuracy for two publicly available datasets (COVID5K and COVIDRD) [41]. Another modified MobileNet model was also used by Jia, Lam and Xu [42] to perform a 5-way classification for CXR images (COVID-19, tuberculosis, viral pneumonia, bacterial pneumonia and normal). The modified MobileNet model was based on the architecture of MobileNet-V3_Small and aimed to overcome some of the issues that were present in the original model such as vanishing gradients and overfitting [42]. This modified MobileNet architecture produced similar results to that which was used in [41], with an overall accuracy of 99.6% for the 5-way classification of CXR images [42].

Some comparative studies have also been conducted to determine the best performing CNN architecture for COVID-19 diagnosis via CXR images. El Asnaoui and Chawki [21] measured the accuracy of several CNN architectures such as VGG-16, VGG-19, DensNet-101, Inception_ResNet-V2, Inception-V3, ResNet-50 and MobileNet-V2. The models with the highest accuracies were reported to be Inception_ResNet-V2 (92.18%) and DensNet101 (88.09%) [21]. On the other hand, studies performed by Jabber *et al.* [43] and Akter *et al.* [44] have demonstrated that a modified version of the MobileNet-V2 architecture can outperform other CNN models including DenseNet-121 and the ResNet-50V2. In a separate study, Shorfuzzaman and Masud [45] conducted a comparative study with several major CNN architectures including VGG-16, ResNet-50V2, MobileNet and DenseNet. The authors concluded that ResNet-50V2 was the best performing CNN model in terms of accuracy, precision, sensitivity, specificity and F1-score [45]. This was closely followed by the MobileNet architecture which managed to outscore ResNet50-V2 in the Area Under Curve (AUC) metric [45]. These results are supported by Hernandez *et al.* [31] who performed a similar study where they compared their own custom CNN model to the likes of ResNet-50, VGG-16 and DenseNet-121. Once again, the ResNet-50 architecture outscored all the other models (including the custom CNN model) for all the performance metrics including accuracy, precision, recall and the F1-score [31]. Santoso and Pernomo [46] also created their own CNN model which they used to compare against the ResNet-50, Inception-V3 and Xception architectures. In this case, the custom CNN model outperformed the other models in this study [46]. However, the researchers noted that the computational time for their custom model exceeded that of the other models. Another comparative study was performed by Rahaman *et al.* [47] using 15 pre-trained CNN models to determine which one could best classify CXR images into normal, pneumonia and COVID-19 categories. In this case, the VGG-19 architecture was found to outperform the other models in this study [47].

2.2.2.2 CNNs vs Traditional Computer Vision Techniques

Ever since deep learning techniques became mainstream in digital image processing, traditional computer vision techniques have become obsolete [12]. This can be seen in a study performed by Hedjazi, Kourbane and Genc [14], in which a comparison is drawn between CNN and classical ML methods. In this case, the CNN model (AlexNet) was shown to have a validation accuracy of at least 86.75%, thus outperforming the traditional ML method with a maximum validation accuracy of 83% [14]. This result was consistent for the two datasets that were considered in this study, the smaller of which contained approximately 6000 images and the larger containing nearly 300,000 [14]. On the other hand, O'Mahoney *et al.* [12] notes that deep learning methods (such as CNNs) can sometimes be unnecessary as traditional computer vision techniques can sometimes classify images with fewer lines of code and greater efficiency. Another advantage for traditional computer vision techniques is that the algorithms are not class-specific and can thus be used to detect features from any image in the training dataset [12]. This differs from CNNs which are required to learn the features of each class of images separately. As such, a poorly constructed training dataset will likely result in a subpar performing CNN model. That being said, the major advantage of CNNs is the elimination of having to manually perform feature extraction, which is perhaps the most time-consuming process in traditional ML methods [14]. In another study, López-Cabrera *et al.* [48] investigated the limitations of deep learning methods for CXR image classification and compared those to traditional ML methods. In this paper the authors suspected that the patterns extracted via deep learning approaches can be subtle and often overlap with other viral pneumonias [48]. Furthermore, results from their study showed that many CNN models suffered from shortcut learning, in which the model would use simple characteristics to classify the image as opposed to learning and capturing the true essence of the underlying data. This was observed when the authors noted that a large portion of the region outside the lungs was being used to classify the CXR image [48]. This result is highly irrational and can be partially attributed to improper image pre-processing. In this case, the authors recommend using traditional ML methods which they believe are better at generalising CXR image classification for new/unseen datasets [48].

2.2.3 Comparison of CNN models

A summary of the main literature discussed in Section 2.2.2.1 can be found in Table 2 below. It should be noted here that the data size column corresponds to the size of the original data, i.e., prior to performing any data augmentation. Most the literature considered here implemented CNN architectures in conjunction with Transfer Learning (TL). Furthermore, for comparison-based studies which considered more than one CNN architecture, the results for the best performing model were selected for display. These studies are indicated by an asterisk (*).

Table 2 Summary of key literature that performed X-ray classification using CNN models

Study	Classes	Data size	Data Preparation Techniques	Model(s)	Performance Metrics	Average Performance (%)
Reynaldi *et al.* [18]	COVID-19 Normal	1281 1281	CLAHE	ResNet-101 + TL	Accuracy Recall Specificity	99.61 99.62 99.60
Narin [33]	COVID-19 Pneumonia Normal	219 1345 1341	N/A	ResNet-50 + SVM Cubic + TL	Accuracy Recall Specificity	94.86 96.04 96.62
Umri *et al.* [20]	COVID-19 Normal	100 100	CLAHE	VGG-16 + TL	Accuracy	97
Sarki *et al.* [22]	COVID-19 Pneumonia Normal	296 3875 1341	Math. Morphology	VGG-16 + TL	Accuracy Recall Specificity	87.50 96.43 100
Alam *et al.* [34]	COVID-19 Normal	1979 3111	Region of Interest	VGG-19 + TL	Accuracy Recall Specificity	99.49 93.65 95.7
Chaudhary *et al.* [36]	COVID-19 Pneumonia Normal	171 60 76	N/A	DenseNet-121 + TL	Accuracy	89.3
Albahli, Ayub and Shiraz [37]	COVID-19 Pneumonia Normal	590 6057 8851	Data Augmen.	DenseNet-512 + TL	Accuracy Recall Specificity	92 85 99
Montalbo [35]	COVID Pneumonia Normal	1281 4657 3270	N/A	Custom DenseNet + TL	Accuracy Precision Recall F1-score	97.99 98.38 98.15 98.26
Bohmrah and Kaur [38]	COVID-19 Pneumonia Normal	111 70 70	N/A	DenseNet-201 + TL	Accuracy Precision Recall F1-score	86 92 91 91
Mohammadi *et al.* [40]	COVID-19 Normal	181 364	Data Augmen.	MobileNet + TL	Accuracy Precision Recall F1-score	99.1 100 98.0 99.0
Tangudu, Kakarla and Venkateswarlu [41]	COVID-19 Normal	1341 1200	N/A	MobileNet + TL	Accuracy Recall Specificity	99.65 99.65 100.0
Jia, Lam and Xu [42]	COVID-19 Tuberculosis B. Pneumonia V. Pneumonia Normal	1770 1436 1700 1345 1341	N/A	MobileNet	Accuracy Recall Precision	95.0 95.0 95.3
El Asnaoui and Chawki [21]*	Coronavirus B. Pneumonia Normal	1724 2780 1583	CLAHE	Inception_ ResNet-V2	Accuracy Recall Specificity Precision	92.18 92.11 96.06 92.38

					F1-score	92.07
Jabber *et al.* [43]*	COVID-19 Non-COVID	500 6000	N/A	MobileNet	Accuracy Recall Specificity Precision F1-score	98.6 87.8 99.3 87.8 87.8
Akter *et al.* [44]*	COVID-19 Normal	3616 10192	Data Augmen. CLAHE	MobileNet + TL	Accuracy Recall Specificity Precision F1-score	98 98 97 97 97
Shorfuzzaman and Masud [45]*	COVID-19 Pneumonia Normal	226 226 226	Data Augmen.	ResNet-50V2	Accuracy Recall Specificity Precision F1-score	98.15 98.26 98.89 97.87 98.06
Hernandez *et al.* [31]*	COVID-19 Pneumonia Normal	1234 4576 16627	N/A	ResNet-50 + TL	Accuracy Recall Precision F1-score	90.63 91.67 90.00 90.72
Santoso and Pernomo [46]*	COVID-19 Pneumonia Normal	206 206 206	N/A	FCovNet	Accuracy Recall	99.55 96.67
Rahaman *et al.* [47]*	COVID-19 Pneumonia Normal	260 300 300	Data Augmen.	VGG-19	Accuracy Recall Precision F1-score	89.3 89.7 90.8 89.6

2.3 Review Conclusions

Looking through the literature in Table 2, a large amount of research has already been conducted on the classification of CXR images using CNN models. Although various performance metrics were used, the accuracy metric is the only one that is common to all of these studies. As such, this will be the major performance metric that will be used to compare the results from these studies.

First, by examining the Classes column, it is evident that the studies that performed binary classification (i.e., studies that only consisted of two classes) had on average a higher accuracy metric. This result is expected as it is relatively easy to differentiate a normal patient from one that contains COVID-19. However, when additional classes are added, such as viral pneumonia, the model may struggle to differentiate between the various illnesses (especially if these diseases display similar CXR characteristics). Out of the seven studies that performed binary classification, the minimum testing accuracy (97%) corresponded to the study conducted by Umri *et al.* [20]. Despite this impressive result, this accuracy is slightly lower compared to the rest of the binary classification studies, which contained an average performance accuracy of 99.08%. This is likely attributed to the low data size that was used by Umri *et al.* [20], which only contained 100 CXR images for each class. Due to this small data size, the VGG-16 model was likely inadequately trained compared to the models implemented in some of these other studies which used

thousands of CXR images for training. In this case, the researchers of this study could have benefitted from implementing data augmentation to increase the training data sample size. This technique was used by Mohammadi *et al.* [40] and who was able to achieve an accuracy of 99.1%, despite the fact that the original data size only contained approximately 500 CXR images. However, it should be noted that these two studies used different CNN architectures and thus their results cannot be fully compared.

On the other hand, the average accuracy for the studies that performed a three-way classification is approximately 92.5%. Overall, the ResNet architecture consistently displayed a high performance accuracy with respect to the other studies. This can be seen in studies [21], [31], [33] and [45] which used the ResNet architecture to obtain classification accuracies of 92.18%, 90.63%, 94.86% and 98.15% respectively. In one particular study [35], the custom DenseNet architecture was shown to perform reasonably well, with an overall accuracy of 97.99%. However, when considering the other studies that used DenseNet, this architecture had a considerably lower average accuracy of approximately 89%. In addition to this, the DenseNet architecture is infamous for its long training times and larger number of hyperparameters needed [35]. As such this model is likely less suited to performing image classification compared to some of the other architectures considered here. Similarly, the VGG architecture also had a lower than average performance compared to ResNet with an average classification accuracy of 88.4% [22], [47].

The CLAHE filter was implemented in several studies in this literature review [18], [20, 21] and [44]. These studies used this filter to improve the contrast of the CXR and to highlight small features that might otherwise be missed by the CNN model when performing image classification. The studies that implemented this filter achieved an average accuracy of over 95% despite the various CNN architectures and data sizes used. An interesting data pre-processing technique that was used by Alam *et al.* [34] is region of interest (ROI) extraction. ROI extraction was used to remove any unnecessary labels (such as numbers and letters) that commonly appear on the edges of CXR images. This process encourages the CNN model to focus in on the lung region of the CXR image and thus lessens the likelihood of the model using shortcut learning when classifying the image [48]. Unfortunately, ROI extraction appears to be overlooked in most of the studies that were considered in this literature review. This pre-process technique can be quite tedious to implement (especially for datasets with thousands of images), which is likely why many researchers have elected to skip this process entirely. Finally, it is worth noting that a large number of these studies implemented CNNs via TL. TL was used to reduce the training time required by reusing the model hyperparameters from a different (albeit similar) task.

In summary, upon reviewing the previous literature, the following techniques will be considered for implementation in our project:

1. Data augmentation to increase/balance out the data size between the different classes.
2. Image contrast enhancement via a CHALE filter or mathematical morphology.
3. Region of interest extraction to remove unwanted labels and to focus in on the lung region.
4. Implementation of various CNN architectures including but not limited to ResNet, VGG and DenseNet.
5. An evaluation of various models using performance metrics such as accuracy, recall, F1-score.

3 Methods

3.1 Introduction

This chapter recounts the methodology that was used to classify the CXR images using ML models. The first section describes the techniques that were used to prepare the data. These techniques are categorised into the following subsections: data distribution, pre-processing and augmentation. This chapter will then discuss the various CNN architectures that were implemented to classify the CXR images. This includes a custom CNN model that was designed from scratch and several pre-trained CNN architectures such as VGG-16, Inception-V3 and DenseNet-121.

3.2 Data Preparation

The data preparation techniques that were used to prepare the CXR for the ML models are data distributions, data pre-processing and data augmentation. Data distribution was performed to evenly distribute the dataset to ensure that each class (COVID-19, viral pneumonia and normal) contained the same amount of CXR images. Data pre-processing was then applied to normalise the CXR images for the ML models. Finally, data augmentation was used to improve the ML model performance by subjecting it to a larger and richer dataset.

3.2.1 Data Distribution

The experiments were performed on the two different datasets separately. For dataset A, all the CXR images were used during the training process. For dataset B, a data imbalance exists, as shown in Figure 2, which may cause false classification performances [49]. Furthermore, to minimise the training time required for the second dataset, a random sample of 1345 CXR images were selected for each class. Since viral pneumonia class contained the least amount of CXR images (1345), this sample size was used to randomly select the CXR images from the COVID-19 and normal classes to help even out the distribution.

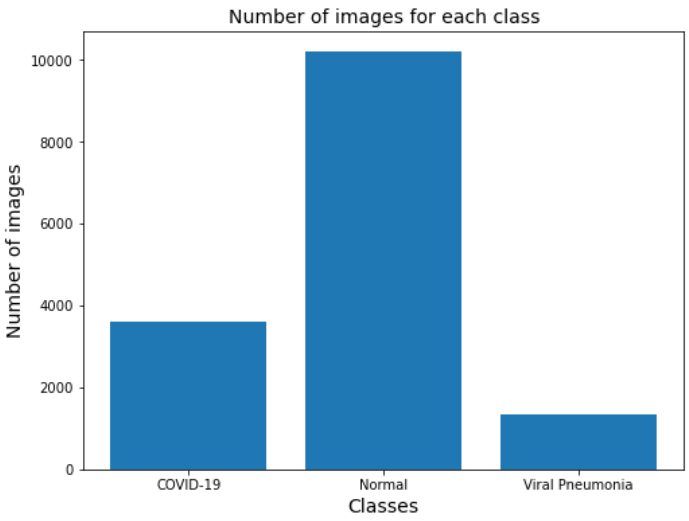

Figure 2 Image distribution between classes for dataset B

For both datasets, 60% of the data were used to train the ML models, 20% were used for validation purposes and the remaining 20% were used for testing. The distribution of the train, validation, and test sets for the datasets A and B are summarised in Tables 3 and 4, respectively.

Table 3 Distribution of the training, validation and testing sets for dataset A

Set	**Count**			**Percentage**
	COVID-19	**Normal**	**Viral Pneumonia**	
Train	78	60	60	60%
Validation	26	20	20	20%
Test	26	20	20	20%
Total	130	100	100	100%

Table 4 Distribution of the training, validation and testing sets for dataset B

Set	**Count**			**Percentage**
	COVID-19	**Normal**	**Viral Pneumonia**	
Train	807	807	807	60%
Validation	269	269	269	20%
Test	269	269	269	20%
Total	1345	1345	1345	100%

3.2.2 Data Pre-processing

The CXR images present in each dataset were in JPEG and PNG formats, with different sizes varying from 299 × 299 to 4248 × 3480 pixels. For performing the experiments, the image size of 224 × 224 was selected, as this was the required input size for all of the TL models used in this project. The nearest neighbour interpolation method was used to compress (i.e., down sample) the images to the target size. Pixel normalisation is very important, as it aids in improving the computational time of the deep learning models [50]. The learning process of the deep learning models can be significantly slower, when the inputs hold large integer values. The CXR images in the datasets were normalised, such that the pixel value ranged from 0 to 1. This was performed by multiplying every pixel value with a multiplication factor. Both datasets contain 8-bit RGB colour images, and the pixel values ranged from 0 to 255. Therefore, the multiplication factor can be computed using the following expression:

$$\text{Multiplication factor} = \frac{\text{pixel value} - \min(\text{pixel value})}{\max(\text{pixel value}) - \min(\text{pixel value})} = \frac{\text{pixel value}}{255}.$$

3.2.3 Data Augmentation

Several studies have shown that data augmentation have been effective in increasing the performances in medical image classification problems [51], [52]. Data augmentation will generate additional images with slight variations using existing data. Data augmentation was performed on the training set to increase the training set size and to reduce overfitting. Overfitting occurs when the ML model performs too well on the training set but, is unable to generalise to the testing set which contains unseen data. As the ML models are exposed to a larger dataset during training process, the models are being forced to generalise.

Table 5 provides a summary of the data augmentation methods applied on the training set. For this project, the following data augmentation techniques were used: rotation, horizontal shifting, vertical shifting and horizontal flipping. For rotation augmentation, each image was randomly rotated within a maximum range of ±20°. Horizontal and vertical shift augmentation were responsible for shifting the CXR image pixels horizontally or vertically by a certain fraction. In horizontal flip augmentation, the pixel columns in the CXR images were reversed (i.e., the image was flipped horizontally).

Table 5 Data augmentation techniques applied on the training datasets

Type of Augmentation	Value
Rotation range	20°
Width shift range	20%
Height shift range	20%
Horizontal flip	True

3.3 CNN Architectures

According to a study conducted by Litjens *et al.* [53], [54], CNNs are one of the most common techniques for medical image analysis. The is likely due to the CNNs distinctive ability to extract and retain the complex features of the input images. A CNN architecture comprises of multiple sequential convolutional and pooling layers. A CNN's main aim is to learn and retain patterns of images associated with each class, while reducing the dimensionality of the image. For more information on the configuration of the CNN layers, refer to Appendix A. During this project, a custom CNN model was designed and implemented from scratch to perform image classification. The CNN model was implemented using the Keras library in the Python programming language. Several pre-trained CNN models were also implemented via TL using the Keras library.

3.3.1 Custom CNN Model

A diagram of the custom CNN architecture that was designed and implemented for CXR image classification is shown in Figure 3. The custom CNN architecture consists of 3 convolutional layers. This architecture takes in an input image of size 224 × 224. For each convolutional layer, the rectified linear unit (ReLU) activation function is used. After each convolutional layer, a 2 × 2 max pooling layer is applied to reduce the dimensionality of the image. The first convolutional layer uses 16 3 × 3 kernel filters. The second convolutional layer uses 32 3 × 3 kernel filters. The final

convolutional layer uses 64 3 × 3 kernel filters. To perform classification on the output obtained from the convolutional layers, multiple dense layers are used. The first dense layer consists of 512 neurons with the activation function being ReLU. The second dense layer consist of 3 neurons with a softmax activation function.

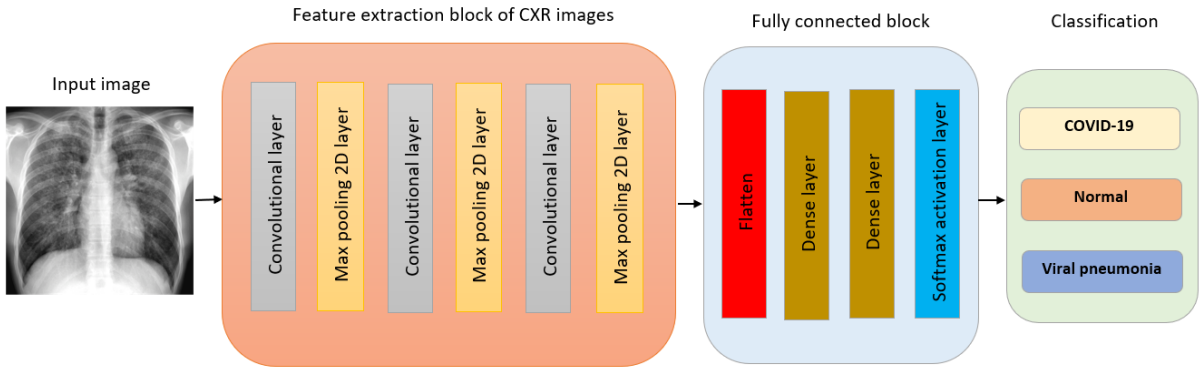

Figure 3 Architecture of the custom CNN model implemented

3.3.2 CNN models with Transfer Learning

TL is a popular ML technique that uses pre-trained models (often with the same weights) to solve a new problem. TL is a well-researched technique in ML and is generally shown to result in increased model performance while minimising training time, overfitting and the number of parameters required [55].

This section provides a brief overview of the TL methods used in this project. The pre-trained models along with the pre-trained weights, made available in Keras Applications have been utilised for this project. For this project, the models, VGG-16, Inception-V3, and ResNet-50 with pre-trained weights on the ImageNet database have been used. The ImageNet database is a large-scale hierarchical image database, consisting of 3.2 million images with over 5247 classes [56]. A study by Alexander Ke *et al.* [57], found that pre-trained models on ImageNet dataset yielded a significant boost in performance compared to other architectures for CXR interpretation.

3.3.2.1 VGG-16

The VGG-16 architecture consists of 16 convolutional layers. Like the proposed CNN architecture above, VGG-16 is also designed for input images of size 224 × 224 × 3. The first two layers start off by making use of 64 kernel filters of size 3 × 3 [58]. The subsequent convolutional layers make use of progressively more filters from 128 (for layers 3 and 4) to 256 (for layers 5 to 7) and finally to 512 (for layers 8 to 13) [58]. Convolutional layers 2, 4, 7, 10 and 13 also make use of max pooling which is used to select the most prominent features contained in those layers [59]. Dense layers consisting of 4096 neurons are then used for the 14th and 15th convolutional layers. These layers are then fed into a dense softmax layer which also utilises ReLU activation to classify these images. In total, this model consists of approximately 140 million hyperparameters [59].

3.3.2.2 Inception-V3

The inception-V3 architecture is made up of 42 convolutional layers. Like VGG-16, Inception-V3 also makes use of max pooling layers throughout the model and terminates with a Softmax layer for image classification [60]. Where these architectures differ however, is that Inception-V3 employs three inception-based modules which serve as a "multi-level function extractor" [22]. These modules compute convolutions of sizes 1 × 1, 3 × 3 and 5 × 5 for the same network layer, which are all then concatenated via a filter before being passed onto the next network layer [22].

3.3.2.3 DenseNet-121

In a DenseNet architecture, each convolutional layer is connected to each other layer via a Densely Connected Convolutional Network [61]. In this study, DenseNet-121 was used which consists of four DenseBlocks consisting of 6, 12, 24 and 16 dense layers respectively [62]. Each dense layer is made up of convolutions of 1×1 and 3×3. Between each DenseBlock is a transition layer which consists of 1×1 convolution followed by a 2×2 average pool [63]. The Softmax function is once again used to perform image classification in the final layer of this architecture [61] .

3.4 Summary

In summary, two experiments were performed using datasets A and B. Data pre-processing techniques such as image compression and pixel normalisation were applied. The data augmentation techniques, rotation augmentation, horizontal augmentation, vertical augmentation and horizontal flip augmentation were performed to generate additional training data. To perform classification on CXR images, four different CNN architectures, custom CNN, VGG-16, Inception-V3 and DenseNet-121, were implemented.

4 Results

4.1 Introduction

This chapter will summarise the major results obtained from this project. In this study, the CNN architectures will be evaluated according to the following performance metrics: accuracy, precision, recall and F1-score. Definitions for these performance metrics can be found in the Appendix B. To help visualise the performance of these ML models, a series of confusion matrices were also created.

4.2 Evaluation of CNN Architectures

Tables 6 and 7 show the performance evaluation on the test data for the ML models on the two datasets experimented. The tables include the precision, recall and F1-score values obtained for each class and the overall accuracy of the ML models when evaluated on the testing set.

Table 6 Performance evaluation of ML models for dataset A

Model	**Classes**	**Precision**	**Recall**	**F1-score**	**Testing accuracy (%)**
Custom CNN	COVID-19	1.00	0.96	0.98	86%
	Normal	0.73	0.95	0.83	
	Viral Pneumonia	0.87	0.65	0.74	
VGG-16	COVID-19	1.00	1.00	1.00	89%
	Normal	0.84	0.80	0.82	
	Viral Pneumonia	0.81	0.85	0.83	
Inception-V3	COVID-19	0.96	0.96	0.96	86%
	Normal	0.79	0.75	0.77	
	Viral Pneumonia	0.81	0.85	0.83	
DenseNet-121	COVID-19	1.00	1.00	1.00	**91%**
	Normal	0.79	0.95	0.86	
	Viral Pneumonia	0.94	0.75	0.83	

Table 7 Performance evaluation of ML models for dataset B

Model	**Classes**	**Precision**	**Recall**	**F1-score**	**Testing accuracy (%)**
Custom CNN	COVID-19	0.87	0.69	0.77	82%
	Normal	0.72	0.84	0.77	
	Viral Pneumonia	0.90	0.94	0.92	
VGG-16	COVID-19	0.90	0.87	0.88	89%
	Normal	0.81	0.91	0.86	
	Viral Pneumonia	0.99	0.91	0.95	
Inception-V3	COVID-19	0.73	0.93	0.82	84%
	Normal	0.89	0.61	0.72	
	Viral Pneumonia	0.94	0.98	0.96	
DenseNet-121	COVID-19	0.93	0.85	0.89	**91%**
	Normal	0.83	0.92	0.87	
	Viral Pneumonia	0.98	0.96	0.97	

According to Tables 6 and 7, the best performing model for COVID-19 detection in CXR images is DenseNet-121. This model produced the highest values for precision, recall and F1-score for both datasets A and B. This was closely followed by the VGG-16 architecture which obtained the second highest scores for precision, recall and F1-score. Both models, DenseNet-121 and VGG-16, obtained a perfect score of 1 for precision, recall and F1-score on dataset A for the COVID-19 class. The DenseNet-121 and VGG-16 models on dataset B obtained high F1-score values of 0.89 and 0.88, respectively. Therefore, the DenseNet-121 model outperformed the VGG-16 model on dataset B for the COVID-19 class. The high F1-score values indicate that, the models DenseNet-121 and VGG-16 produced low false positives and low false negatives for the COVID-19 class. Overall, both models were successful in detecting the CXR images of COVID-19 patients.

The DenseNet-121 and VGG-16 models also produced high values of precision, recall and F1-score for the normal class on both datasets. Therefore, both models are capable of identifying CXR images of normal patients and differentiating them from patients with COVID-19 and viral pneumonia. The custom CNN model did produce a high recall value on dataset A. The recall value, however, was very low on dataset B. This meant that, the CNN model produced a larger number of false negatives on dataset B. The Inception-V3 model produced low recall values of 0.75 and 0.61 on dataset A and B, respectively. Thus, the Inception-V3 model produced a large number of false negatives for the normal class on both datasets. A large number of false negatives is detrimental, as it means that the model is not capable of correctly detecting normal CXR images. Therefore, the custom CNN and Inception-V3 models cannot successfully differentiate normal CXR images from COVID-19 and viral pneumonia CXR images.

All three pre-trained models, DenseNet-121, VGG-16, and Inception-V3 outperformed the custom CNN model on the viral pneumonia class, as higher values of precision, recall and F1-score were

obtained on both datasets. The custom CNN model produced a low recall value of 0.65 on dataset A, indicating a large number of false negatives. Therefore, the custom CNN model cannot successfully detect CXR images of patients with viral pneumonia on dataset A. All the models produced high values of precision, recall and F1-score on dataset B. Therefore, on dataset B, the models were able to successfully detect viral pneumonia CXR images and differentiate them from COVID-19 and normal CXR images.

From Tables 6 and 7, it can be observed that the DenseNet-121 model outperformed all other ML models, as it produced the highest testing accuracy (91%) on both datasets. For each class, the DenseNet-121 model consistently produced high values of precision, recall and F1-score, on both datasets. The VGG-16 model also performed well on both datasets, as it obtained a testing accuracy of 89% on datasets A and B. In comparison with the pre-trained models (VGG-16, Inception-V3 and DenseNet-121), the custom CNN architecture designed by the team produced the worst performance, as it obtained the lowest testing accuracy of 86% and 82% on dataset A and B respectively.

4.3 Confusion Matrices

Figure 4 illustrates the confusion matrices obtained from the ML models when evaluated on the testing set of dataset A. Figure 5 shows the confusion matrices obtained from the ML models when evaluated on the testing set of dataset B.

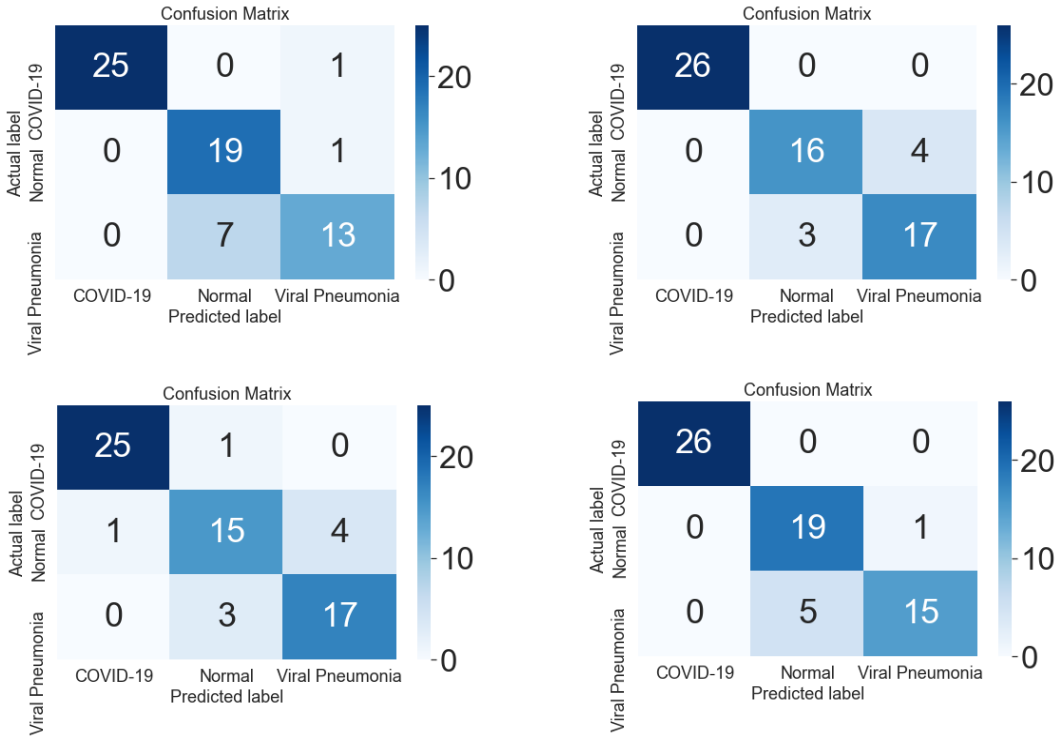

Figure 4 Confusion matrices for custom CNN (top left), VGG-16 (top right), Inception-V3 (bottom left) and DenseNet-121 (bottom right) for dataset A

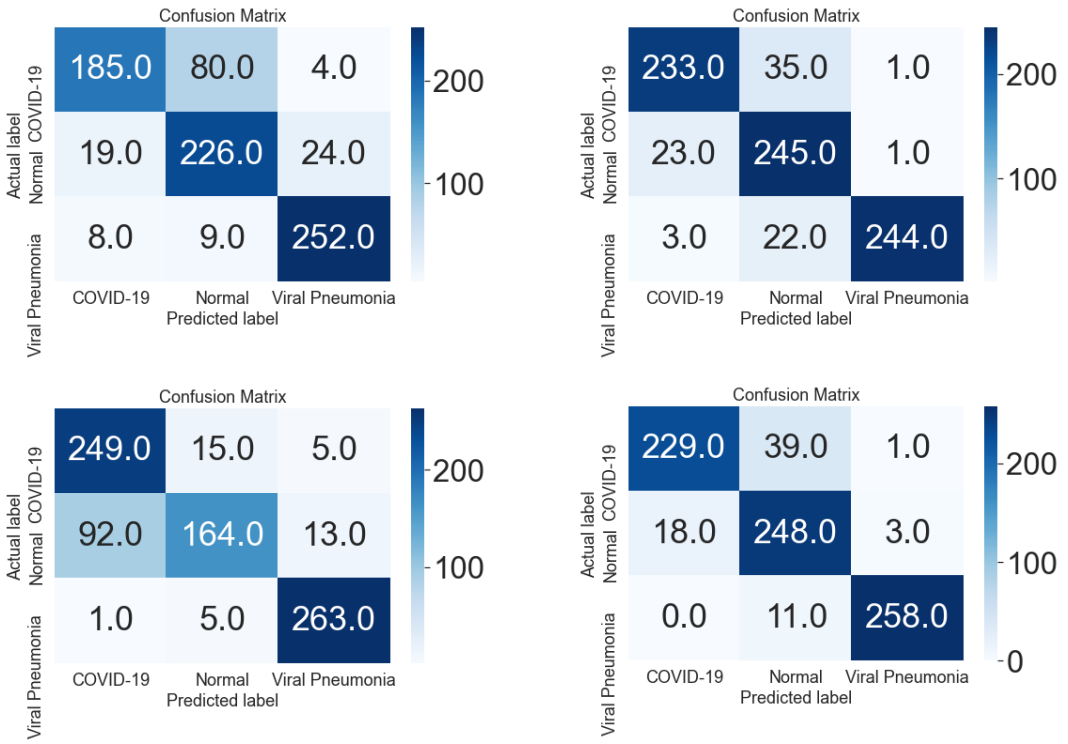

Figure 5 Confusion matrices for custom CNN (top left), VGG-16 (top right), Inception-V3 (bottom left) and DenseNet-121 (bottom right) for dataset B

All the ML models were highly successful in correctly predicting the CXR images of the COVID-19 patients on dataset A. This can be observed in Figure 4, as a low number of misclassifications were produced for each ML model on the COVID-19 class. The custom CNN model and the DenseNet-121 model produced the least number of misclassifications for the normal class on dataset A. However, the custom CNN model produced the worst performance on classifying the viral pneumonia class for dataset A, as a high number of misclassifications was observed.

The custom CNN model also produced the worst performance in classifying CXR images of COVID-19 patients on dataset B. As shown in Figure 5, the custom CNN model was able to correctly classify 185 CXR images of COVID-19 patients. However, the model incorrectly classified 80 COVID-19 CXR images as normal CXR images. Therefore, the custom CNN model is incapable of successfully differentiating CXR images of COVID-19 and normal patients. The Inception-V3 model had the highest success in classifying CXR images of COVID-19 patients. As shown in Figure 5, the Inception-V3 model correctly classified 249 CXR images of COVID-19 patients and only incorrectly classified 20 COVID-19 CXR images.

The DenseNet-121 model produced the best performance in classifying CXR images of normal patients on dataset B. The model correctly classified 248 CXR images of normal patients and only misclassified 21 normal CXR images, as shown in Figure 5. The Inception-V3 model produced the worst performance in classifying normal CXR images. As illustrated in Figure 5, the model incorrectly classified 92 normal CXR images as COVID-19 CXR images. Therefore, the Inception-V3 model is not able to successfully differentiate between COVID-19 CXR images and normal CXR images on dataset B.

Referring to Figure 5, all the ML models were highly successful in detecting viral pneumonia in CXR images and differentiating them from the other classes, as evident from the low number of misclassifications observed. The Inception-V3 model outperformed all the other models in classifying viral pneumonia CXR images on dataset B. The Inception-V3 model correctly classified 263 viral pneumonia CXR images and only incorrectly classified 6 viral pneumonia CXR images.

The Custom CNN implemented produced the worst performance in classifying CXR images on both datasets. All three pre-trained models produced high values of precision, recall and f1-score on both datasets. Moreover, all the pre-trained models scored higher values of accuracy, when evaluated on the testing set. Therefore, this is indicating that the pre-trained models are more capable of generalising to unseen data, compared to the custom CNN model. The DenseNet121 model generally produced the best performance in classifying CXR images of COVID-19 and normal patients. The Inception-V3 produced the best performance in classifying viral pneumonia CXR images. However, the Inception-V3 model outputted a large number of incorrect classifications on the normal CXR images on both datasets, which is detrimental. Therefore, based on the experiments considered, the DensetNet-121 model can be considered the best performing model.

4.4 Summary

The results demonstrate that both the designed and pre-trained CNN models can successfully extract important features from the CXR images to perform classification. The Custom CNN model implemented produced the worst performance in classifying CXR images on both datasets. All three pre-trained models produced high values of precision, recall and F1-score on both datasets. Moreover, all the pre-trained models scored higher values of accuracy, when evaluated on the testing set. Therefore, this indicates that the pre-trained models are more capable of generalising to unseen data, compared to the custom CNN model. The DenseNet-121 model had a high success rate of classifying COVID-19, normal and viral pneumonia CXR images. This model also produced the highest testing accuracy (91%) for both datasets. Therefore, DenseNet-121 was considered to be the best performing model for CXR image classification.

5 Limitations

5.1 Introduction

This project was subjected to several limitations that impacted the quality and generalisability of the results. The major limitations of this study are mainly the small dataset size and missing metadata. These limitations and their impact on the project's results are discussed in the following subsections.

5.2 Sample Size

The first limitation was the low number of CXR images provided in dataset A. This dataset only contained a total of 330 samples which only included 130 CXR images for COVID-19 class. This sample size is generally considered to be very small, compared to some other ML studies which use a recommended sample size of at least 1000 images [64]. A small sample size is generally insufficient for the models to adequately learn the characteristics of the image which often leads to biased results and poor image classification.

To combat this limitation, the COVID-19 Radiography Database (dataset B) was used. This dataset consists of 3616 COVID-19, 1,345 viral pneumonia and 10,192 normal CXR images. Although a larger dataset often leads to greater generalisation, the effect of biasing may still be prevalent due to a data imbalance between the classes of the dataset. To obtain an even distribution of data, a random sample of 1,345 CXR images were selected from each class.

5.3 Missing Metadata

Both datasets were missing key metadata such as COVID-19 strain type (Delta, Omicron etc.), as well as some patient descriptors such as a patient's age, gender, contamination period and medical history. This metadata could help to further generalise the results obtained in this study.

For example, different COVID-19 strains are known to affect the lungs differently. One of these variants (Omicron) causes less damage to the lungs compared to other variants such as Delta [65]. Since the variant type was not included in the CXR metadata, the performance of these ML models could not be generalised for different COVID-19 strains. Another limitation could be the patient's sex, as the male and female immune systems differ in the way that they treat infectious diseases [66]. A patient's age could also influence the interpretation of the results in this study. Government health officials reported that the severity of COVID-19 increases with the patient's age [67]. For example, older patients are more likely to suffer from other respiratory diseases which may also be visible in CXR images. These images can cause the ML models to produce misleading classification performances. Similarly, a patient's medical history can also limit the interpretability of our results. For example, CXR images from patients with a history of lung disease can affect the learning and classification of the ML models. Another important limitation is contamination period of COVID-19. For example, during the incubation period the patients may be asymptomatic and thus there may be no distinguishable COVID-19 features in the CXR image [68]. In this case, the ML model may misclassify a COVID-19 CXR image as normal.

5.4 Summary

In summary, several limitations were imposed on the datasets used. One of these limitations was a low sample size which can lead to biased results and poor image classification. To mitigate the effect of this limitation, a second dataset containing a larger number of CXR images was used in this project. Another limitation that was identified was the lack of key metadata. This includes the strain of the COVID-19 virus as well as several other patient descriptors that could provide insight into the patients' health at the time when the CXR images were taken. With these important factors missing, the resulting obscure dataset may in turn lead to subpar and/or inaccurate ML model performance

6 Completion Plan

The ML models implemented produced good performances in classifying the CXR images into the following three classes: COVID-19, normal and viral pneumonia. However, the results achieved by the models were not as great as some of the models from the literature review. Therefore, improvements can be made to the methods used, to obtain a more accurate classification of the CXR images.

For the remainder of this project, the following tasks are to be considered:

1. More advanced pre-processing techniques, such as ROI extraction and image contrasting needs to be applied.

2. Implementation of several other pre-trained CNN architectures mentioned in the literature review, such as ResNet, MobileNet and EfficientNet.

3. Investigation into novel image classification techniques, such as Vision Transformers.

4. Selection of the best performing ML model for CXR image classification.

5. Comparing the results of the best performing model to the previously developed models in literature review.

7 Conclusion

The project thus far has seen progression towards answering the question “Covid or Flu? That’s the question!”. Despite having limitations, the ML models implemented were able to achieve satisfactory results on both datasets. The pre-trained models, VGG-16, Inception-V3 and DenseNet-121 produced a higher classification accuracy compared to the custom CNN model that was implemented. The DenseNet-121 model outperformed all the other models that were implemented, as it produced the highest classification accuracy on the testing set. Nonetheless, there are still areas of improvement in the methodology that was used in this project. At the pre-processing stage, ROI extraction needs to be performed on the CXR images, so that only the relevant area of the chest is kept. Ideally, this will lead to an improvement in the classification accuracy of the ML models and reduce the likelihood of shortcut learning. Furthermore, other pre-trained CNN architectures, including but not limited to ResNet, MobileNet and EfficientNet, could be explored for CXR image classification. Vision transformers are a viable alternative to CNNs for image classification, which will also be considered for implementation during the second half of this project.

References

[1] WHO, "WHO Coronavirus (COVID-19) Dashboard," 13 May 2022. [Online]. Available: https://covid19.who.int/. [Accessed 16 May 2022].

[2] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang *, et al.*, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature,* vol. 579, no. 7798, pp. 270-273, 2020.

[3] Z. C. Brooks and S. Das, "COVID-19 Testing: Impact of Prevalence, Sensitivity, and Specificity on Patient Risk and Cost," *American journal of clinical pathology,* vol. 154, no. 5, pp. 575-584, 2020.

[4] J. Hellewell, T. W. Russell, R. Beale, G. Kelly, C. Houlihan, E. Nastouli *, et al.*, "Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections," *BMC medicine,* vol. 19, no. 1, p. 106, 2021.

[5] S. Mallett, A. J. Allen, S. Graziadio, S. A. Taylor, N. S. Sakai, K. Green *, et al.*, "At what times during infection is SARS-CoV-2 detectable and no longer detectable using RT-PCR-based tests? A systematic review of individual participant data," *BMC medicine,* vol. 18, no. 1, p. 346, 2020.

[6] D. Jarrom, L. Elston, J. Washington, M. Prettyjohns, K. Cann, S. Myles *, et al.*, "Effectiveness of tests to detect the presence of SARS-CoV-2 virus, and antibodies to SARS-CoV-2, to inform COVID-19 diagnosis: a rapid systematic review," *BMJ evidence-based medicine,* vol. 27, no. 1, pp. 33-45, 2020.

[7] M. Elsharkawy, A. Sharafeldeen, F. Taher, A. Shalaby, A. Soliman, A. Mahmoud *, et al.*, "Early assessment of lung function in coronavirus patients using invariant markers from chest X-rays images," *Scientific reports,* vol. 11, no. 1, p. 12095, 2021.

[8] X. Meng and Y. Liu, "Chest Imaging Tests versus RT-PCR Testing for COVID-19 Pneumonia: There Is No Best, Only a Better Fit," *Radiology,* vol. 297, no. 3, p. 345, 2020.

[9] "Leading Causes of Death," Centres for Disease Control and Prevention, 13 January 2022. [Online]. Available: https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm. [Accessed 16 May 2022].

[10] WHO, Standardization of Interpretation of Chest Radiographs for the Diagnosis of Pneumonia in Children/World Health Organization Pneumonia Vaccine Trial Investigators' Group, Switzerland: WHO, 2001.

[11] A. Anaya-Isaza, L. Mera-Jiménez and M. Zequera-Diaz, "An overview of deep learning in medical imaging," *Informatics in medicine unlocked,* vol. 26, p. 100723, 2021.

[12] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova *, et al.*, "Deep Learning vs. Traditional Computer Vision," *Advances in Computer Vision,* pp. 128-144, 24 April 2019.

[13] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image," *Journal of big data,* vol. 6, no. 1, pp. 1-18, 2019.

[14] M. A. Hedjazi, I. Kourbane and Y. Genc, "On identifying leaves: A comparison of CNN with classical ML methods," in *2017 25th Signal Processing and Communications Applications Conference (SIU)*, Antalya, Turkey, 2017.

[15] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong and M. Ghassemi, "COVID-19 Image Data Collection: Prospective Predictions are the Future," *Journal of Machine Learning for Biomedical Imaging,* vol. 2, pp. 1-38, 2020.

[16] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub *, et al.*, "Can AI Help in Screening Viral and COVID-19 Pneumonia?," *IEEE Access,* vol. 8, pp. 132665-132676, 2020.

[17] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. Abul Kashem *, et al.*, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Computers in biology and medicine,* vol. 132, p. 104319, 2021.

[18] D. Reynaldi D.S., B. S. Negara, S. Sanjaya and E. Satria, "COVID-19 Classification for Chest X-Ray Images using Deep Learning and Resnet-101," in *2021 International Congress of Advanced Technology and Engineering*, Taiz, Yemen, 2021.

[19] B. Soni and P. Mathur, "An Improved Image Dehazing Technique using CLAHE and Guided Filter," in *2020 7th International Conference on Signal Processing and Integrated Networks*, Noida, India, 2020.

[20] B. K. Umri, M. W. Akhyari and K. Kusrini, "Detection of Covid-19 in Chest X-ray Image using CLAHE and Convolutional Neural Network," in *2020 2nd International Conference on Cybernetics and Intelligent System*, Manado, Indonesia, 2020.

[21] K. El Asnaoui and Y. Chawki, "Using X-ray iamges and deep learning for automated detection of coronavirus disease," *Journal of biomolecular structure & dynamics,* vol. 39, no. 10, pp. 3615-3626, 2021.

[22] R. Sarki, K. Ahmed, H. Wang, Y. Zhang and K. Wang, "Automated detection of COVID-19 through convolutional neural network using chest x-ray images," *PloS one,* vol. 17, no. 1, p. 262052, 2022.

[23] Y. Kimori, "Mathematical morphology-based approach to the enhancement of morphological features in medical images," *Journal of clinical bioinformatic,* vol. 1, no. 1, p. 33, 2011.

[24] S. Tabik, A. Gomez-Rios, J. L. Martin-Rodriguez, I. Sevillano-Garcia, M. Rey-Area, D. Charte *, et al.*, "COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images," *IEEE journal of biomedical and health informatics,* vol. 24, no. 12, pp. 3595-3605, 2020.

[25] E. Mineo, "Kaggle," 2018. [Online]. Available: https://www.kaggle.com/code/eduardomineo/u-net-lung-segmentation-montgomery-shenzhen/notebook. [Accessed 16 5 2022].

[26] K. K. Singh and A. Singh, "Diagnosis of COVID-19 from chest X-ray images using wavelets-based depthwise convolution network," *Big Data Mining and Analytics,* vol. 4, no. 2, pp. 84-93, 2021.

[27] Q. Hou and W. Li, "An improved enhancement self-daptive algorithm for X-ray digital image based on wavelet decomposition," *2010 3rd International Congress on Image and Signal Processing,* vol. 3, pp. 1041-1044, 2010.

[28] D. Haritha, M. K. Pranathi and M. Reethika, "COVID Detection from Chest X-rays with DeepLearning: CheXNet," in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, Patna, India, 2020.

[29] J. Li, D. Zhang, Q. Liu, R. Bu and Q. Wei, "COVID-GATNet: A Deep Learning Framework for Screening of COVID-19 from Chest X-Ray Images," in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2020.

[30] M. Mishra, V. Parashar and R. Shimpi, "Development and evaluation of an AI System for early detection of Covid-19 pneumonia using X-ray (Student Consortium)," in *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, New Delhi, India, 2020.

[31] D. Hernandez, R. Pereira and P. Georgevia, "COVID-19 detection through X-Ray chest images," in *2020 International Conference Automatics and Informatics (ICAI)*, Varna, Bulgaria, 2020.

[32] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016.

[33] A. Narin, "Detection of Covid-19 Patients with Convolutional Neural Network Based Features on Multi-class X-ray Chest Images," in *2020 Medical Technologies Congress*, Antalya, Turkey, 2020.

[34] N.-A. Alam, M. Ahsan, M. A. Based, J. Haider and M. Kowalski, "COVID-19 Detection from Chest X-ray Using Feature Fusion and Deep Learning," *Sensors,* vol. 21, no. 4, p. 1480, 2021.

[35] F. J. P. Montalbo, "Diagnosing Covid-19 chest x-rays with a lightweight truncated DenseNet with partial layer freezing and feature fusion," *Biomedical signal processing and control,* vol. 68, p. 102583, 2021.

[36] S. Chaudhary, S. Sadbhawna, V. Jakhetiya, B. N. Subudhi, U. Baid and S. C. Guntuku, "Detecting Covid-19 and Community Acquired Pneumonia Using Chest CT Scan Images With Deep Learning," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, ON, Canada, 2021.

[37] S. Albahli, N. Ayub and M. Shiraz, "Coronavirus disease (COVID-19) detection using X-ray images and enhanced DenseNet," *Applied soft computing,* vol. 110, p. 107645, 2021.

[38] M. K. Bohmrah and H. Kaur, "Classification of Covid-19 patients using efficient fine-tuned deep learning DenseNet model," *Global Transitions Proceedings,* vol. 2, no. 2, pp. 476-483, 2021.

[39] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang and T. Weyand, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *Computer Vision and Pattern Recognition,* vol. 1, pp. 1-9, 2017.

[40] R. Mohammadi, M. Salehi, H. Ghaffari, A. A. Rohani and R. Reiazi, "Transfer Learning-Based Automatic Detection of Coronavirus Disease 2019 (COVID-19) from Chest X-ray Images," *Journal of biomedical physics and engineering,* vol. 10, no. 5, pp. 559-568, 2020.

[41] V. S. K. Tangudu, J. Kakarla and I. B. Venkateswarlu, "COVID-19 detection from chest X-ray using MobileNet and residual seperable convolutional block," *Soft Computing,* vol. 26, no. 5, pp. 2197-2208, 2022.

[42] G. Jia, H.-K. Lam and Y. Xu, "Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method," *Computers in biology and medicine,* vol. 134, p. 104425, 2021.

[43] B. Jabber, J. Lingampalli, C. Z. Basha and A. Krishna, "Detection of Covid-19 Patients using Chest X-ray images with Convolution Neural Network and Mobile Net," in *2020 3rd International Conference on Intelligent Sustainable Systems*, Thoothukudi, India, 2021.

[44] S. Akter, F. M. J. M. Shamrat, S. Chakraborty, A. Karim and S. Azam, "COVID-19 Detection Using Deep Learning Algorithm on Chest X-ray Images," *Biology,* vol. 10, no. 11, p. 1174, 2021.

[45] M. Shorfuzzaman and M. Masud, "On the Detection of COVID-19 from Chest X-Ray Images Using CNN-Based Transfer Learning," *Computers, materials & continua,* vol. 64, no. 3, pp. 1359-1381, 2020.

[46] F. Y. Santoso and H. D. Purnomo, "A Modified Deep Convolutional Network for COVID-19 detection based on chest X-ray images," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems*, Yogyakarta, Indonesia, 2020.

[47] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, M. A. Rahman, Q. Wang *, et al.*, "Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches," *Journal of X-ray science and technology,* vol. 28, no. 5, pp. 821-839, 2020.

[48] J. D. López-Cabrera, R. Orozco-Morales, J. A. Portal-Díaz, O. Lovelle-Enríquez and M. Pérez-Díaz, "Current limitations to identify covid-19 using artificial intelligence with chest x-ray imaging (part ii). The shortcut learning problem," *Health and technology,* vol. 11, no. 6, pp. 1331-1345, 2021.

[49] A. Fernandez, S. C. N. V. del Rio and F. Herrera, "An insight into imbalanced Big Data classification: outcomes and challenges," *Complex & Intelligent Systems,* vol. 3, no. 2, pp. 105-120, 2017.

[50] P. Sane and R. Agrawal, "Pixel normalization from numeric data as input to neural networks: For machine learning and image processing," in *2017 International Conference on Wireless Communications, Signal Processing and Networking*, Chennai, India, 2017.

[51] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy *, et al.*, "Classification of breast cancer histology using Convolutional Neural Networks," *PloS one,* vol. 12, no. 6, p. 177544, 2017.

[52] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger and H. Greenspan, "Synthetic data augmentation using GAN for improved liver lesion classification," in *2018 IEEE 15th International Symposium on Biomedical Imaging*, DC, USA, 2018.

[53] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian *, et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis,* vol. 42, pp. 60-88, 2017.

[54] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian *, et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis,* vol. 42, pp. 60-88, 2017.

[55] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE transactions on knowledge and data engineering,* vol. 22, no. 10, pp. 1345-1359, 2010.

[56] J. Deng, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, FL, USA, 2009.

[57] A. Ke, W. Ellsworth, O. Banerjee, A. Y. Ng and P. Rajpurkar, "CheXtransfer: performance and parameter efficiency of ImageNet models for chest X-Ray interpretation," in *2021 Proceedings of the Conference of Health, Inference, and Learning*, NY, USA, 2021.

[58] A. Thite, "Introduction to VGG16 | What is VGG16?," Great Learning Team, 1 October 2021. [Online]. Available: https://www.mygreatlearning.com/blog/introduction-to-vgg16/. [Accessed 22 May 2022].

[59] M. Hassan, "VGG16 - Convolutional Network for Classification and Detection," Neurohive, 20 November 2018. [Online]. Available: https://neurohive.io/en/popular-networks/vgg16/. [Accessed 22 May 2022].

[60] X. Xia, C. Xu and B. Nan, "Inception-v3 for flower classification," in *2017 2nd International Conference on Image, Vision and Computing*, Chengdu, 2017.

[61] A. Arora, "DenseNet Architecture Explained with PyTorch Implementation from TorchVision," GitHub, 2 August 2020. [Online]. Available: https://amaarora.github.io/2020/08/02/densenets.html. [Accessed 22 May 2022].

[62] A. Ahmed, "Architecture of DenseNet-121," OpenGenus, [Online]. Available: https://iq.opengenus.org/architecture-of-densenet121/#:~:text=In%20short%2C%20DenseNet%2D121%20has%20120%20Convolutions%20and%204%20AvgPool,use%20features%20extracted%20early%20on.. [Accessed 22 May 2022].

[63] E. T. Hastui, A. Bustamam, P. Anki, R. Amalia and A. Salma, "Performance of True Transfer Learning using CNN DenseNet121 for COVID-19 Detection from Chest X-Ray Images," in *2021 IEEE International Conference on Health, Instrumentation & Measurement, and Natural Sciences*, Medan, Indonesia, 2021.

[64] A. Vabalas, E. Gowen, E. Poliakoff, A. J. Casson and E. Hernandez-Lemus, "Machine learning algorithm validation with a limited sample size," *PloS One,* vol. 14, no. 11, p. 224365, 2019.

[65] CDC, "What you need to know about variants," Centers for Disease Control and Prevention, 26 April 2022. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/variants/about-variants.html#:~:text=The%20Omicron%20variant%20causes%20more,of%20all%20variants%2C%20including%20Omicron. [Accessed 25 May 2022].

[66] B. G. M., "Coronavirus: Why Men are More Vulnerable to Covid-19 Than Women?," *SN Compr Clin Med,* vol. 2, no. 7, pp. 874-876, 2020.

[67] D. o. Health, "Coronavirus (COVID-19) advice for older people and carers," Australian Government, 30 May 2022. [Online]. Available: https://www.health.gov.au/node/18602/coronavirus-covid-19-advice-for-older-people-and-carers#:~:text=More%20information-,Older%20people%20are%20at%20risk,or%20a%20weakened%20immune%20system. [Accessed 25 May 2022].

[68] L. Maragakis, "Coronavirus Diagnosis: What should I expect?," Health, 24 January 2022. [Online]. Available: https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/diagnosed-with-covid-19-what-to-expect . [Accessed 25 May 2022].

[69] RadiologyInfo, "Chest X-ray," RadiologyInfo, 15 June 2020. [Online]. Available: https://www.radiologyinfo.org/en/info/chestrad . [Accessed 15 May 2022].

[70] D. Weatherspoon, "Chest X-Ray," Healthline, 18 November 2018. [Online]. Available: https://www.healthline.com/health/chest-x-ray. [Accessed 15 May 2022].

[71] "X-Ray Examinations," Better Health, 6 February 2019. [Online]. Available: https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/x-ray-examinations. [Accessed 15 May 2022].

[72] "X-ray," NHS, 20 April 2022. [Online]. Available: https://www.nhs.uk/conditions/x-ray/ . [Accessed 15 May 2022].

[73] L. A. Rousan, E. Elobeid, M. Karrar and Y. Khader, "Chest X-ray findings and temporal lung changes in patients with COVID-19 pneumonia," *BMC Pulmonary Medicine,* vol. 20, no. 1, p. 245, 2020.

[74] R. Yasin and W. Gouda, "Chest X-ray findings monitoring COVID-19 disease course and severity," *Egyptian Journal of Radiology and Nuclear Medicine,* vol. 51, no. 1, pp. 1-18, 2020.

[75] A. Cattamanchi, "What is ground glass opacity," Medical News Today, 29 March 2021. [Online]. Available: https://www.medicalnewstoday.com/articles/ground-glass-opacity . [Accessed 15 May 2022].

[76] C. Bao, X. Liu, H. Zhang, Y. Li and J. Liu, "Coronavirus Disease 2019 (COVID-19) CT Findings: A Systematic Review and Meta-analysis," *Journal of the American College of Radiology,* vol. 17, no. 6, pp. 701-709, 2020.

[77] "Machine Learning," IBM Cloud Education, 15 July 2020. [Online]. Available: https://www.ibm.com/cloud/learn/machine-learning . [Accessed 16 May 2022].

[78] S. Brown, "Machine Learning, Explained," MIT Management, 21 April 2021. [Online]. Available: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained. [Accessed 16 May 2022].

[79] "What is Machine Learning (ML)?," Berkeley, 26 June 2020. [Online]. Available: https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/. [Accessed 16 May 2022].

[80] A. Bajaj, "Performance Metrics in Machine Learning," Neptune Blog, 18 March 2022. [Online]. Available: https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide . [Accessed 16 May 2022].

[81] "Convolutional Neural Network (CNN) in Machine Learning," GeeksforGeeks, 28 Decemeber 2020. [Online]. Available: https://www.geeksforgeeks.org/convolutional-neural-network-cnn-in-machine-learning/ . [Accessed 16 May 2022].

[82] S. Saha, "A comprehensive guide to convolutional neural networks - the ELI5 way," Towards Data Science, 16 December 2018. [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 . [Accessed 16 May 2022].

[83] J. Brownlee, "How do Convolutional Layers Work in Deep Learning Neural Networks?," Machine Learning Mastery, 17 April 2020. [Online]. Available: https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/ . [Accessed 16 May 2022].

[84] "What is pooling in a convolutional neural network (CNN): Pooling layers explained," Programmathically, 5 December 2021. [Online]. Available: https://programmathically.com/what-is-pooling-in-a-convolutional-neural-network-cnn-pooling-layers-explained/ . [Accessed 16 May 2022].

[85] S. Saxena, "Introduction to Softmax for Neural Network," Analytics Data, 5 April 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/04/introduction-to-softmax-for-neural-network/ . [Accessed 16 May 2022].

Appendices

Appendix A: Theory

In this part of the report, a number of technical information related to the project topic will be discussed.

A1 CXR Radiography

CXR is one the imaging techniques that is used to look inside the body, specifically the chest area as the ROI. Hence, this technique is useful for analysing the heart, lungs, and chest area for medical diagnosis and treatment purposes [69]. This imaging techniques uses small amount of ionising radiation to produce these images [70]. In small dosages, this radiation is considered to be non-lethal, and is thus not likely to cause any serious health complications for the patient [71].

A CXR image is typically undertaken in a special room purposefully equipped with a portable X-ray camera supported by a metallic frame. Prior to proceeding with the examination, the patient will be asked by the technician to remain in front of X-ray film so when the radiation beam passes through the body, it can be captured and recorded [71], as shown in Figure 6.

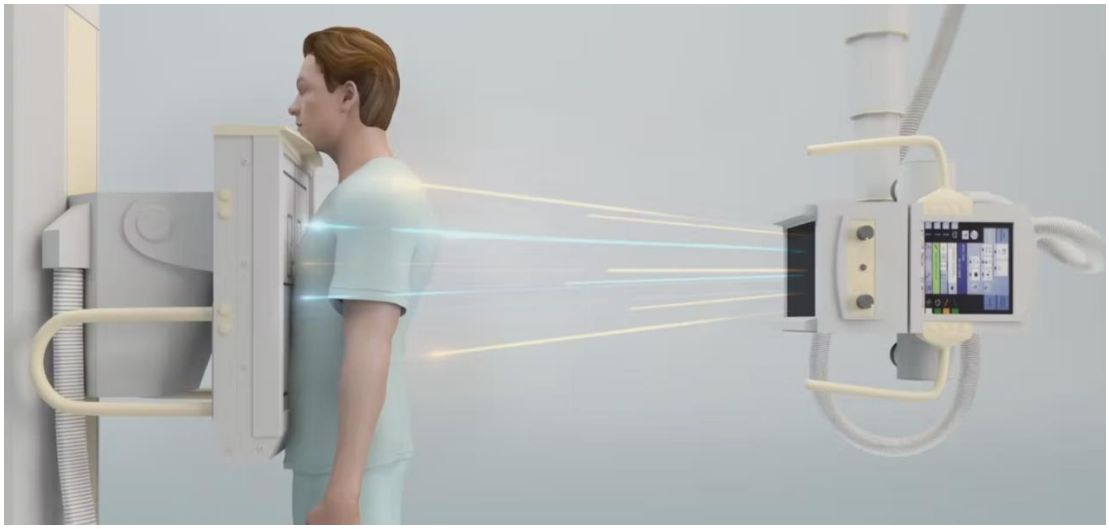

Figure 6 Depiction of patient undergoing CXR imaging

If the patient does not remain still during this procedure, the film will appear to be blurry and the resulting CXR will not be sufficient for the radiologists to correctly report on the patient's condition [70].

A CXR is a 2D image where body parts in the chest area appear in different colours, depending on the intensity of the radiation passing through them [69]. For instance, more radiation will pass through soft body tissue compared to dense structures such as bones. Thus, the X-ray will show the dense structures as a bright white colour, soft tissues as a grey colour and any air inside the lungs as a dark colour [72]. After the X-ray examination is completed, the patient may

go home, and the X-ray image would be sent to a radiologist so a technical analysis report would be prepared for the patient's general practician [72].

CXR imaging has several advantages when compared to other medical imaging techniques. The use of X-ray imaging is prevalent inside emergency rooms as it is significantly faster and more convenient to use [69]. More medical centres and clinics intend to equip their place with X-ray machines as it is considered to be a cost-effective machine for medical diagnosis [69].

For patients with COVID-19 symptoms, their CXR images can show signs of abnormalities with the prominent ones being the presence of Ground Glass Opacity (GGO) in the lower region of the lungs called, "Lower Lobes" [73] and consolidation of the lungs [74].

GGO refers to the grey area that can be present on CXR images and in the case of lung organs, it can be described as a sign of an unusual condition for the lungs since healthy lungs should appear as a black colour on the CXR images [75]. GGO can be caused due to many reasons that can compromise the normal function of the lungs [75]. One of the main causes of GGO is due to an increased density in the lungs which can be a result of fluid build-up, inflammation or tissue damage in the lung region.

However, the scale of the size of GGO depends on various important factors such as the severity of disease and at what period the image is taken [20], [76]. In the case of a common COVID-19 disease, GGO may start to develop from the first few days when the symptoms are shown and reaches the peak level from day 5 to day 10 since the first symptoms [73], [76]. In the case of a severe COVID-19 patient, GGO would compromise the whole lobe and lungs in general.

As discussed, one of the other common features in the CXR images of patients with COVID-19, is the consolidation of the lungs. On a CXR, this can be seen as a white area with similar features to GGO. However, in this case, the blood vessels are concealed in the X-ray image [73].

A2 Machine Learning Techniques

ML is a field of computer and data science which uses advanced algorithms and a large dataset, to develop and train a model, so that a machine can learn, think and understand like a human without the need of any pre-programming [77].

ML runs in different arrangements called supervised, unsupervised, semi-supervised and reinforcement learning. For this project, the focus is placed on supervised ML methods, where the ML model is trained on a dataset and learns the patterns associated with the data [78]. To achieve this, the dataset should be separated into two sets: training set (labelled) and testing set (unlabelled). From the training set, the model would be able to learn the pattern present in the data [78]. The model can then be evaluated using the testing dataset to measure the model's effectiveness in learning and classifying images [79]. Several evaluation metrics, such as precision, recall, accuracy and F1-score, can be used to measure the model's performance [80].

In this project, emphasis is placed on deep learning image classification techniques, such as CNNs. CNN is a unique technique where convolutional layers are used to classify data such as image portrays, as seen in Figure 7. The first convolutional layer is responsible for extracting the high-level features associated with the images, such as major shape structures. A pooling layer is then used to further reduce the size of the image [81]. The purpose of this reduction in size is to reduce computation power to process the data. These two layers together subsequently make up a single layer of a CNN. There can typically be multiple copies of this single layer in the case where image data has significantly more complex patterns. However, as more layers are added, there is an increased demand on the computational power required to process the image [82].

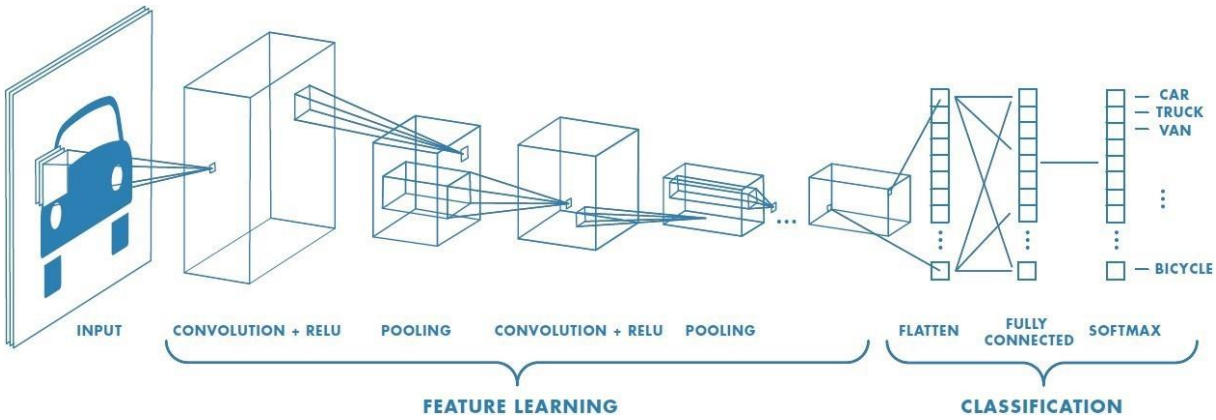

Figure 7 Diagram depicting the layers of a CNN model

The learning process for a CNN model starts with a convolutional layer that can be described as the main feature of the CNN modelling technique. This simply runs a linear operation of multiplication using the input function (I) and a set of weight function called kernel or filter (f) to produce a function map or feature map $F(t)$ from the operation [22] [83].

$$F(t) = (I * f)(t)$$

Given the input value is one dimensional, and t only takes inputs that are integers, the following equation can be constructed [22]:

$$F(t) = \sum_{a} \mathrm{I}(a).f(t-a)$$

If the given input value is two-dimensional, the variables would be $I(m,n)$ and $f(a,b)$ and the equation can be written as the following [22]:

$$F(t) = \sum_{a}\sum_{b} \mathrm{I}(a,b).f(m-a,n-b)$$

This equation can be re-expressed by switching the filter using the commutative law [22]:

$$F(t) = \sum_{a}\sum_{b} \mathrm{I}(m-a,n-b).f(a,b)$$

However, in neural networks, the cross-correlation function is applied instead, which has the same form as the convolution equation above but without flipping the filter variable [22].

$$F(t) = \sum_{a}\sum_{b} \mathrm{I}(m+a,n+b).f(a,b)$$

Once the feature map is produced, each feature is passed through the next layer called ReLU which is defined as an activation function. This function aims to convert the negative input values to zero so that the training of the ML model on the input data can be more efficient and faster [22], [83].

The final layer in this technique is called max pooling, whose function is to extract the maximum pixel value or element in the feature map input from the previous layer [84]. Once the learning process is finished, classification of the input data is initiated by having a connected layer which takes the flattened output from the max-pooling layer to calculate the probability values for classification purposes [22], [82].

The classification stage also consists of an activation function called softmax. This function is typically used for classification problems that deal with multiple classes (i.e., 2 or more classes). This function takes the input from the output of the previous layer to perform the final probability calculation [22], [82]. The equation for this function can be expressed as [85]:

$$\text{Softmax Function } (Z_i) = \frac{e\ (Z_i)}{\sum_j e\ (Z_j)}$$

In this equation, Z represents the neuron values from the output of the previous layer which passes through an exponential function and is normalised by dividing it by the sum of the neuron values (Z).

Appendix B: Evaluation Metrics

Accuracy can be defined as the number of correct predictions over the total number of predictions. The accuracy score can be defined using the following expression:

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

Where TP is the true positives, TN is the true negatives, FP is the false positives and FN is the false negatives. The accuracy score should not be used alone to evaluate a model's performance as it can produce misleading results on an imbalanced dataset. Therefore, it is important to consider other evaluation metrics.

Recall is another evaluation metric that allows to measure the ML model's ability to determine positive cases. Recall can be defined as the number of positive class predictions over the total number of positives. Recall be defined using the following expression:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Precision is an evaluation metric that measures the number of true positives over the total number of positive predictions. Mathematically, it can be expressed using the definition below:

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

An F1 score is an evaluation metric that considers both recall and precision. Therefore, F1 score can be defined to be the harmonic mean of recall and precision. It can be defined using the expression shown below,

$$\text{F1 score} = 2\,\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

A confusion matrix was used to visualise the performance of the ML model on a test dataset. Figure 8 shows the format of a confusion matrix. For each class, a confusion matrix provides information about the True Positives, True Negatives, False Negatives and False Positives. Using the confusion matrix, the accuracy, recall, precision and F1-scores can be computed.

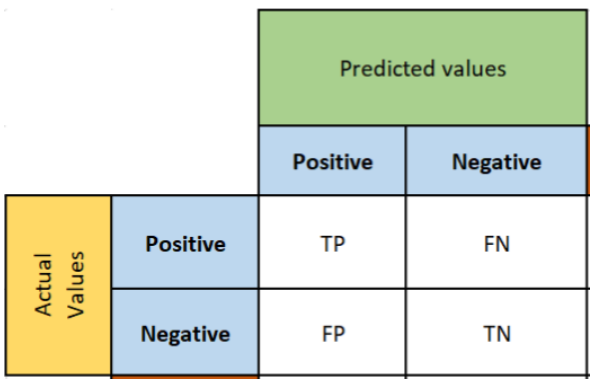

Figure 8 Example of a confusion matrix