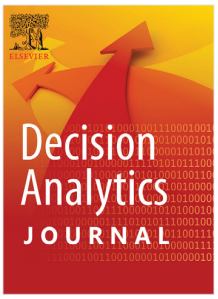
Journal Pre-proof



A comparative analysis of machine learning algorithms for detecting COVID-19 using lung X-ray images

Susmita Hamal, Bhupesh Kumar Mishra, Robert Baldock, William Sayers, Tek Narayan Adhikari, Ryan M. Gibson

PII: S2772-6622(24)00064-X
DOI: https://doi.org/10.1016/j.dajour.2024.100460
Reference: DAJOUR 100460

To appear in: *Decision Analytics Journal*

Received date : 25 June 2023
Revised date : 25 February 2024
Accepted date : 8 April 2024

Please cite this article as: S. Hamal, B.K. Mishra, R. Baldock et al., A comparative analysis of machine learning algorithms for detecting COVID-19 using lung X-ray images, *Decision Analytics Journal* (2024), doi: https://doi.org/10.1016/j.dajour.2024.100460.

Title Page

Manuscript Title: "A Comparative Analysis of Machine Learning Algorithms for Detecting COVID-19 Using Lung X-Ray Images"

Authors' Details:

Susmita Hamal, School of Engineering and Computing, University of Gloucestershire, UK, Email: susmita.hamal@gmail.com

Bhupesh Kumar Mishra, Data Science, AI & Modelling Centre (DAIM), University of Hull, UK, Email: bhupesh.mishra@hull.ac.uk, Corresponding Author

Robert Baldock, School of Pharmacy & Biomedical Sciences, University of Portsmouth, UK, Email: robert.baldock@port.ac.uk

William Sayers, School of Engineering and Computing, University of Gloucestershire, UK, Email: wsayers@glos.ac.uk

Tek Narayan Adhikari, School of Engineering and Computing, University of Gloucestershire, UK, Email: tadhikari@glos.ac.uk

Ryan M. Gibson, Department of Computing, Glasgow Caledonian University, UK, Email: Ryan.Gibson@gcu.ac.uk

A Comparative Analysis of Machine Learning Algorithms for Detecting COVID-19 Using Lung X-Ray Images

Abstract: Machine intelligence has the potential to play a significant role in diagnosing, managing, and guiding the treatment of disease, which supports the rising demands on healthcare to provide rapid and accurate interpretation of clinical data. The global pandemic caused by the Severe Acute Respiratory Syndrome Coronavirus (SARSCoV-2) exposed a need for rapid clinical data interpretation in response to an unprecedented burden on the healthcare system. A new healthcare challenge has arisen – post-COVID syndrome or 'long COVID'. Symptoms of the post-COVID syndrome can persist for months following infection with SARS-CoV-2, often characterised by fatigue, breathlessness, dizziness, and pain. Despite this additional healthcare burden, no tests can diagnose, monitor, or determine the efficacy of treatments/interventions to support recovery. In this paper, an array of machine-learning algorithms is trained to evaluate and detect COVID-19-associated changes to lung tissue from X-ray images. X-ray images are classified from open sources into three categories: COVID-19 patients, patients with pneumonia, and unaffected otherwise healthy individuals using existing Machine Learning (ML) and pre-trained deep learning models. Prioritising models with the fewest false positives and false negatives assessed the performance of different models in detecting COVID-19-associated lung tissue. In addition, image pre-processing, data augmentation, and hyperparameter tuning are used to achieve the best accuracy in the models. Different ML models, including K Nearest Neighbour (KNN), and decision trees (DT), as well as transfer learning models such as Convolutional Neural Network (CNN), Visual Geometry Group (VGG-16, VGG-19), ResNet50, DenseNet201, Xception, and InceptionV3, were tested to evaluate the performance of these models for X-ray images classification. The comparative analysis indicates that VGG-19 with augmentation performed best among the ten algorithms with a training accuracy of 99%, testing accuracy of 98%, and precision of 90% for COVID-19, 90% for normal, and 100% for pneumonia. This higher accuracy for detecting COVID-19-associated lung changes on X-ray may be further developed to stratify patients suffering from post-COVID syndrome. This may enable future intervention studies to determine the efficacy of treatments or better track patients' prognoses to be optimised.

Keywords: Machine Learning**,** Transfer Learning, Convolutional Neural Network, Visual Geometry Group, ResNet50, DenseNet201, Xception, InceptionV3, post-COVID syndrome, long COVID, Image Classification.

1. Introduction

In 2019, an outbreak of pneumonia initially spread in Wuhan city in China and the infectious agent was later named SARS-CoV-2 by the World Health Organization (WHO). SARS-CoV-2 is an enveloped virus with a positive sense, single-stranded ribonucleic acid (RNA) genome [1]. The virus predominantly affects the human respiratory system and is transferred readily from one person to another through coughing, sneezing and physical contact [2]. It can spread through touching or exposure to contaminated surfaces as the virus can survive for several days depending upon the surface and environment [3]. Primary symptoms of this disease include fever, dry cough, loss of taste and smell, sore throat, and muscle pain [4]. Despite significant research efforts to identify, characterise and limit mortality associated with COVID-19 infections, a new challenge has arisen regarding the longer-term impact of viral infections. Post-COVID syndrome (PCS) also known as 'long-COVID' has been estimated to affect 45% of COVID-19 infection survivors [5]. According to government statistics an estimated 2 million people self-reported experiencing ongoing symptoms of COVID or 'long-COVID'

with 48% of these individuals reporting shortness of breath as a symptom [6]. In their meta-analysis, they identified fatigue, breathlessness, and myalgia as the most experienced symptoms by examining pooled prevalence from 194 studies of patients with PCS. Furthermore, the authors identified that of all clinical investigations undertaken within the studies included, abnormalities were most identified by X-ray and Computerised Tomography (CT). The full potential of chest X-rays to identify and characterise changes in PCS has yet to be fully explored. However, the ability to rapidly detect, identify and stratify COVID-19-associated lung changes in PCS has the potential to aid diagnosis and treatment monitoring.

ML plays an essential role in the healthcare sector, especially in identifying diseases, and outbreaks and preventing diseases [7, 8]. It can support rapidly evolving information to assist public health experts in making complex decisions. ML has been used extensively in the health sector to identify diseases and diagnose illnesses such as lung cancer, cervical cancer, breast cancer, and many more through medical images [9, 10]. Chen et al. [11] consider ML a helpful tool, especially with the growing pressure on limited time and health resources, by improving the efficiency and effectiveness of human efforts to combat this pandemic. Wootton et al. [12] consider X-ray images sensitive and essential tools for diagnosing lung diseases. However, the underlying limitations such as difficulty in data collection, lack of multi-model assessment, delays in realising benefits, poor internal validations, and data privacy make X-ray images complex for diagnosis [13]. Also, in recent years, data science has been used to analyse health-related data to identify diseases using various deep learning, data mining, and ML approaches. ML is considered one of the required fields in ML. It is the process of selecting the right algorithms that can learn from previous datasets and help to predict unseen data is always challenging [14]. Furthermore, with the use of deep learning, ML enables computers to learn from experiences by allowing computational models consisting of multiple processing layers to comprehend a representation of data with various abstraction layers such as facial recognition, drug discovery, and speech recognition [15]. In other words, deep learning models can classify abnormalities and have been reported to aid doctors or radiologists in achieving expert-level diagnostic expertise and predictive analytics [16, 17]. In this context, chest X-rays have become one of the most accessible and cost-effective tools for triage patients [18].

As early detection of COVID-19 is vital during the pandemic, access to healthcare is very necessary. However, the lack of test kits, hospital beds, and required resources in hospitals hinders healthcare access [19]. Subsequently, it is difficult to track and test the infection in developing countries due to insufficient test kits, lack of skilled human resources, and other prejudices [20]. This raises the chances of transmission leading to the volume of infected people spike. Thus, it is necessary to diagnose the infection at an early stage to prevent further transmission. This paper aims to determine how effectively COVID-19-associated lung changes can be detected from chest X-ray images using ML algorithms as well as identifying the best algorithm for detecting these changes. Comparative analysis among 10 different algorithms: Decision Tree, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Convolutional Neural Network (CNN), Visual Geometry Group (VGG-16, VGG-19), DenseNet201, Resnet50, InceptionV3 and Xception are applied to build to explore image classifier for X-ray image with higher accuracy. In the comparative analysis, 10 ML models classified the X-ray image into COVID-19 patients, patients with pneumonia and unaffected otherwise healthy individuals. In addition, data augmentation is implemented to train these models. By doing so, the research aims to pinpoint effective models that can alleviate the strain on healthcare systems and ultimately save lives, particularly in the context of diseases like COVID-19. As a result, there has been improved performance of certain algorithms that previously were suboptimal in other tests and hence the enhancement in model accuracy. This comparative analysis will help enable future interventions to explore the effectiveness of

treatments or better track the prognoses of the long COVID patients to be optimised. The comparative results have shown that VGG-19 with data augmentation achieved the best performance with a training accuracy of 99% and testing accuracy of 98%. The model achieved a precision of 90 % for COVID-19, 90 % for normal and 100% for pneumonia. Importantly, establishing the most appropriate model for detecting COVID-19-associated changes is a fundamental first step in establishing a decision aid to help detect and monitor the treatment of patients with PCS. Also, the development of an algorithm capable of accurately distinguishing between healthy pneumonia or COVID-19-associated lung damage is of significant value in detecting or monitoring the prevalence or treatment. In addition, this comparative analysis also reveals some of the challenges such as data collection methods and quality along with the consistency challenges which can be part of consideration in similar directions.

The rest of the paper is organised as follows. A thorough review of the use of different algorithms is discussed in section 2. The computational model-building methodology is presented in section 3. The results of different ML algorithms are presented in section 4. Finally, the discussion and conclusions are presented in section 5.

2. Detection of COVID-19 from X-ray images using different ML models

This section reviews different earlier approaches that have been applied for COVID-19 detection from X-ray images. These approaches are divided into three different categories: multi-model, CNN and Deep Learning approaches.

2.1 Multi-Model approaches.

ML models have been used extensively in the health sector for identifying diseases and diagnosis of illnesses such as lung cancer, cervical cancer, breast cancer, and others using medical images [9]. After the COVID-19 pandemic, various machine learning, deep learning and hybrid learning models were used to help identify diseases. For example, Racic et al. [10] use a model to perceive whether chest X-ray changes can be detected and classify images based on pneumonia. Similarly, Ahmed et al. [21] use five ML models such as decision tree, random forest, neural network (NN), naive Bayes, logistic regression, and k-nearest neighbour learning (KNN) where NN classifier yielding the most effective results when detecting COVID-19-associated pneumonia with a score of 97%. Similarly, Khan et al. [22] use pre-processing, feature extraction and histogram of oriented gradients in their ML model to detect COVID-19-infected patients. In the study, out of five ML algorithms, SVM provides promising results with an accuracy of 96 %.

2.2 Convolution Neural Network Approaches.

CNN appears as an alternative approach to traditional ML algorithms for image-based classification problems. Ahmed et al. [21], used CNN models with rich filter families and the weight-sharing feature extractor SqueezeNet. This study indicated the utility of CNN features and NN classifiers that can obtain the highest detection of COVID-19 cases from X-ray images. Furthermore, highly accurate models were found to help interpret the screening and detection of COVID-19 quickly and accurately [23]. Similarly, the result of the study by Karar et al. [24] shows that CNN model accuracy was up to 90 % in identifying COVID-19 cases using deep learning for X-ray image detection. CNN has been used in various medical diagnoses with promising results [25]. For example, Mohammad-Rahimi et al. [26] identify CNN, long short-term memory (LSTM), generative adversarial networks (GAN), and residual neural network, out of which the CNN method achieved 99 % accuracy while classifying COVID-19 patients from other causes of pneumonia used on chest X-ray images. Likewise, Hafeez et al. [27] work for detecting COVID-19-associated changes from X-ray images. The study proposes a CNN model for binary

(Normal and COVID), and multiclass images (Normal, COVID, and Virus Bacteria), the model performed an accuracy of 97 % for detecting normal, 89% for COVID and 84% for pneumonia respectively.

To identify an alternative approach for RT-PCR (Reverse Transcription Polymerase Chain Reaction) tests, Redie et al. [28] proposed a modified CNN, that can detect COVID-19 using over 10,000 X-ray images, including both binary and multiclass X-ray images. With image pre-processing, the model performs at 99.53% accuracy for binary class and 94.18 % for multiclass. Furthermore, Chaddad et al. [29] applied different CNN-based techniques with pre-trained models to identify COVID-19 in X-ray images where the model stood out with an impressive accuracy of 99.09% in accurately classifying COVID-19. In a similar study conducted by Haritha et al. [30], a transfer learning from pre-trained GoogleNet, one of the CNN architectures and named InceptionV1. In the study, they used open sources of X-ray images and achieved a testing accuracy of 98.5%.

2.3 Deep Learning Approaches.

In recent years, Deep learning (DL) approaches have appeared as widely used approaches for image-based classification tasks. Kumar et al. [31] classified COVID-19 using chest X-rays, VGG-16, VGG-19, ResNet50, and InceptionV3 models with a modified deep-learning architecture. The accuracy achieved was 98.61% for modified-VGG-16, 97.22% for modified-VGG-19, 95.13% for modified-ResNet50, and 99.31% for InceptionV3. These modifications involved incorporating a new-age architecture by introducing features such as average pooling and dense layers, dropout, and activation functions. Consequently, InceptionV3 demonstrated better image classification compared to any other model. Another study by Asif et al. [32] introduced a model based on Deep Convolutional Neural Network (DCNN), specifically InceptionV3, using transfer learning to enhance the accuracy of detection. The study involved analysing three types of images, including 864 COVID-19 cases, 1345 instances of viral pneumonia, and 1341 normal chest X-ray images. The DCNN-based InceptionV3 model demonstrated a classification accuracy exceeding 97%. Another study by Chen et al. [33] uses Transfer Learning and assessed various models, including VGG-16, VGG-19, Inception-V3, Inception-ResNet, Xception, ResNet152-V2, and DenseNet201, using pre-processing and image segmentation. Among these models, VGG-16 stood out by achieving superior performance, with an accuracy of 98% after 10 epochs. Similarly, Taresh et al. [34] introduced pre-trained CNN models, including InceptionV3, Xception, InceptionResNetV2, MobileNet, VGG-16, DenseNet169, NasNetLarge, and DenseNet121. These models had varying selections of frozen layers and a different number of trainable convolution blocks. VGG-16 outperformed all other models in achieving accuracy, F1 score, precision, specificity, and sensitivity of 98.72%, 97.59%, 96.43%, 98.70%, and 98.78%, respectively.

Nonetheless, with a similar approach to the imbalance dataset, a study by Elagili et al. [35] conducted a study utilising CNN-based transfer learning from chest X-rays with three classes: normal, pneumonia, and COVID-19. They employed pre-trained VGG-16 and DenseNet121, and the latter demonstrated higher accuracy. The DenseNet121 pre-trained model achieved an accuracy of 94%, surpassing the 89% achieved by VGG-16. Performance metrics indicated that DenseNet121 achieved an overall accuracy of 94%, with precision, recall, and F-score values of 91%, 95%, and 93%, respectively. Another study by Monga et al. [36] employed transfer learning with multiclass chest X-ray images to classify three categories: COVID-19, pneumonia, and normal. They utilised deep learning classifiers including DenseNet201, Xception, ResNet50V2, VGG-16, VGG-19, and InceptionResNetV2. Model comparisons involved image pre-processing, including rescaling and normalisation. The evaluation

based on accuracy, precision, and recall as performance parameters revealed that DenseNet201 emerged as the most effective deep learning model, achieving an accuracy of 82.2%.

A comparative study by Alshehri et al. [37] which incorporated both X-ray and computed tomography (CT) scan images and various models such as CoreNet, InceptionV3, InceptionResNetV2, MobileNetV2, NASNetMobile, VGG-16, VGG-19, and Xception. The CT scan images yielded superior results compared to X-ray images, and among the diverse models tested, Xception demonstrated superior performance. Similarly, an experiment conducted by Mane et al. [38] aimed to classify CNN designs from scratch using four transfer learning models (Inception, Xception, ResNet, and VGG-19), and the study found that the Xception model achieved a higher accuracy of 94.8% compared to the others. Sakib et al. [39] proposed a deep transfer learning-based framework utilising a pre-trained network (ResNet-50). The model exhibited a performance of 96% accuracy, with precision, recall, F1-score, and specificity all reaching 1.00. Another study by Sethy et al. [40] aimed to identify infected individuals using X-ray images by employing ResNet50 along with the SVM model. This approach showed outstanding performance, achieving an accuracy of 95.38%, outperforming other models such as AlexNet, GoogleNet, ResNet18, VGG-16, VGG-19, InceptionV3, XceptionNet, and InceptionResNetV2.

Additionally, Gouda et al. [41] proposed two distinct approaches. The first involves image pre-processing, incorporating normalisation and resizing for chest X-ray images, while the second utilises augmentation for constructing ML models. Their study revealed that the modified ResNet50 exhibited superior performance, achieving an overall accuracy of 99.63%. Seeking more precise and reliable outcomes. Alam et al. [42] advocated for VGG-19 with histogram-oriented gradient in comparison to traditional ML models such as ANN (Artificial Neural Network), KNN, and SVM. VGG-19 demonstrated a high accuracy of 98.36% compared to other models. Awan et al. [43] utilised the Apache Spark approach in a comprehensive data framework, applying it to deep learning models like InceptionV3, ResNet50, and VGG-19. This method demonstrated notable performance, with the ResNet50 model achieving an accuracy of 98.55%, and the VGG-19 model also achieving a comparable accuracy of 98.55%.

2.4 Overview of previous studies

Overall, different study findings are summarised in Table 1 highlighting the prevalent use of deep learning models, with only a few incorporating traditional ML models. The overarching goal across most studies is the accurate and rapid detection of COVID-19 in an accessible manner. A common trend emerges, indicating a preference for DL models, which are widely acknowledged for their superior performance. Transfer learning models such as VGG-16, VGG-19, ResNet50, DenseNet, Inception, Xception, Alex, and Google Net are frequently employed in X-ray image datasets, as indicated in the table. Many studies emphasise the reliability of using chest X-ray images as a tool for COVID-19 detection, underscoring its potential utility for doctors and radiographers.

Table 1: Performance summary of different ML models for COVID-19 X-ray Classification

Author	Model Employed	Train Dataset	Test Dataset	Accuracy	Precision	Recall	F1-Score	AUC
Ahmed et al. [21]	NN	-	-	97.24%	97%	98%	97%	100%
Khan et al. [22]	SVM	1760	440	96%	95%	98%	96%	-

Hemdan et al. [24]	VGG-19	40	10	90%	83%	100%	91%	70%
Hafeez et al. [27]	CNN	149	37	89%	91%	97%	95%	55%
Redie et al. [28]	CNN	8000	2000	94.13%	82%	83%	89%	-
Kumar et al. [31]	Modified VGG-16	576	144	98.61%	97%	97%	97%	-
Asif et al. [32]	InceptionV3	2840	710	97%	-	-	-	-
Chen et al. [33]	VGG-16			95%	95.48%	95.41%	95.41%	-
Taresh et al. [34]	VGG-16	3575	311	98.28%	97.59%	98.78%	98.72%	98%
Elagili et al. [35]	DenseNet121	1789	449	97%	-	-	-	-
Monga et al. [36]	DenseNet201	612	90	82.20%	81%	100%	90%	-
Alshehri et al. [37]	Xception	1278	320	84%	-	91%	-	-
Mane et al. [38]	Xception	13639	1514	94.80%	-	-	-	-
Sethy et al. [40]	ResNet50+SVM	-	-	95.38%	93%	97%	95%	-
Alam et al. [42]	HOG and CNN	1979	3111	98.36%	-	-	-	100%
Awan et al. [43]	Resnet50, VGG-19Inception	856	207	98.50%	99%	99%	99%	99%

3. Methodology: Machine Learning Algorithms

ML aims to enable computers to learn without being programmed and is designed to understand complex problems as ML can provide high performance in detecting COVID-19-associated changes [14]. In this study, at first, images related to COVID-19 and viral pneumonia and normal X-rays are collected and hence different ML algorithms are applied to build classifiers to identify COVID-19, Pneumonia and normal X-ray images and compare with their performance.

3.1 Data Collection and Description:

The dataset has been used from the following two open-source Kaggle chest X-ray images, including X-ray images of COVID-19 and viral pneumonia and normal X-rays. The dataset from the Kaggle COVID-19 radiography database (https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database) contains 3616 COVID-19 images, 10,192 normal and 6,012 lung opacity (non-COVID) infected lungs, and 1345 viral pneumonia images. Similarly, Chest X-ray images (https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia) contain 5,863 images dataset with pneumonia, normal and bacterial pneumonia. Out of which, 2561 images depict bacterial pneumonia, 1345 images show viral pneumonia and 1341 are normal images.

3.2 K-Nearest Neighbour (KNN)

KNN is used to train and classify the dataset based on similarity and distance measures. This classifier is a nonparametric learning algorithm that requires an integer k to determine the number of neighbours. When k is equal to 1, the nearest neighbours determine the sample class [44]. The main aim of KNN is to calculate the distance and find the closest neighbour. Furthermore, each object votes for the class

and the most voted for the class will be the outcome. As depicted in equation 1, KNN computes the distance between the target sample and the features of other samples. Eventually, the target samples are classified based on the highest frequency within the KNN-derived neighbourhood. KNN increases neighbourhood size selection via the introduction of the generalised mean distance-based k Nearest Neighbour [89].

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad \text{----------(1)}$$

3.3 Support Vector Machine (SVM)

SVM is a supervised learning method that creates a hyperplane in a multidimensional space separated by different classes that minimise any error. The SVM finds the maximum marginal hyperplane which divides the datasets into categories of positive and negative [45, 46]. SVM shows proficiency in non-linear classification through the deployment of kernel tricks. This involves implicitly changing input data into a higher-dimensional feature space, accomplished by presenting margins between the two classes. The advantages of using the SVM model come with speed, efficiency, and accuracy [47].

3.4 Decision Tree (DT)

A Decision Tree (DT) is a supervised machine-learning method to address classification problems. The structure of a decision tree includes nodes, with each branch indicating the test outcome and each node containing a class label. Nodes within the decision tree play a pivotal role in decision-making, encompassing root nodes, internal nodes, and child nodes. Branches emanate from nodes, representing potential outcomes and involving the splitting of input variables linked to the target. Moreover, DT enable the prediction of an object by considering collective observations, including the potential consequences of a series of decisions [48]. DT excel in discerning meaningful information from extensive raw text datasets and effectively handling unwanted noise.

3.5 Deep learning (DL)

DL models have been popular with excellent performance, especially in the medical image data field such as retina images, chest X-ray images or brain MRI images [49]. There are multiple benefits of using deep learning models such as maximum unstructured data utilisation, elimination of the need for feature engineering, high-quality results, cost-effectiveness, and the need for data labelling [50]. However, it is important to note that applications in the medical sector are sensitive, so high accuracy with an efficient model is needed. Therefore, DL has a high potential to be used in X-ray image applications as it can learn robust and accurate features. In the DL model, CNN is one popular approach for analysing images among various deep learning methods. It has made remarkable achievements in the medical field because it automatically extracts translationally invariant features using the convolution of the input images and filter [51]. Different algorithms and architectures of DL are used in this study as follows.

3.5.1 Convolutional Neural Network (CNN)

According to Racic et al. [10], CNN helps learn essential features like recognizing the shape and edges of the Image, and pre-trained models benefit from the knowledge acquired in learning primary elements of the images from the database of the existing image. A detailed description of the architecture is provided below.

1. Convolutional Layer: The layer is responsible for extracting features and patterns by recognizing input images. Images are passed through a filter consisting of feature maps and kernels, and the size of the kernel filter layers is either 3x3 or 5x5.

2. Pooling Layer: The layer extracts combinations of features from CNN which are invariant to translational shifts and minor distortions. It helps to reduce overfitting and regulate the complexity of the model by feature map with uniform features. There are different dimensions of pooling layers, such as 2X2 or global average pooling.

3. Fully connected layer: This is a crucial layer of CNN where the input features are extracted from different stages of the network and are compared and analysed with the output of all preceding layers. Fully connected layers are used to connect the rectified linear unit (ReLU) with the SoftMax activation function to predict the output images in the last layer of CNN.

4. Activation function: The activation function transforms the weighted sum to one node for a layer and uses it for the activation node for the input. The primary purpose of this function is to help in learning the feature patterns. There are various activation functions such as sigmoid, SoftMax, max out, SWISH, and ReLU.

5. Batch normalisation: This helps every layer of the network to learn independently and is used to normalise the output of previous layers, which helps in the internal covariance shift in feature maps. It helps prevent overfitting in the model and makes the model more efficient.

6. Dropout: It is used to regularise the CNN network by skipping some connection with a particular random probability that generates several thinned network architectures. These thin layers are taken as an approximation of all proposed networks.

7. Flatten Layers: It converts the pooled feature map to a one-dimensional array as an input to the next layer in the form of a single-long feature vector. Then it relates to a deep neural network to create fully connected layers. The architecture of the CNN is presented below in Figure 4.

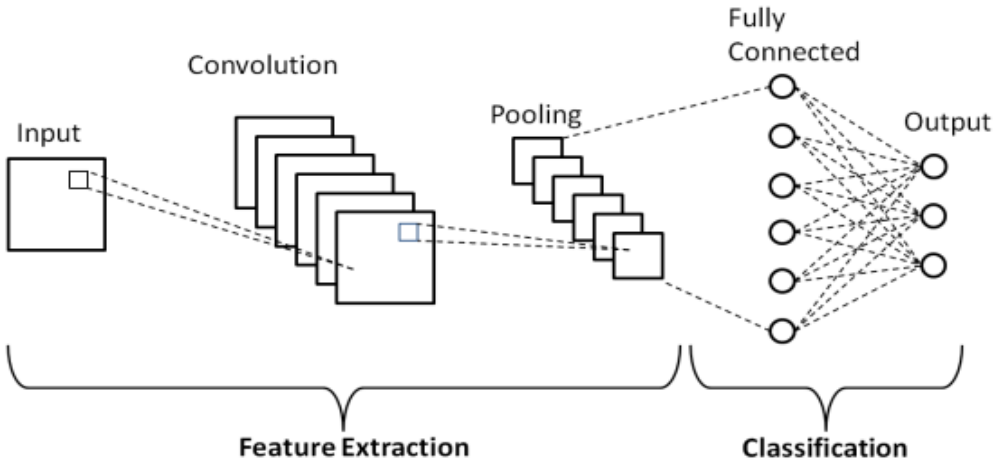


Figure 4. The basic architecture for CNN [52]

3.5.2 Visual Geometry Group (VGG)

VGG explores the effect of increasing the depth of the convolutional network on its accuracy [53]. Due to their small convolution filters of size 3X3 show significant improvement. However, due to their vast network, they consume more time.

a) VGG-16

VGG-16 is deep with 16 weight layers, including 13 convolutional layers with a filter size of 3x3 and fully connected layers. The configurations of fully connected layers in VGG-16 are equivalent to stride, and the padding of all CNN layers is fixed to 1 pixel [30]. The architecture of VGG-16 is shown in Figure 5. The main advantage of using VGG-16 architecture is that it generalises well to other datasets. However, it is slow in training and needs more space for running the models, which makes this model ineffective [54].

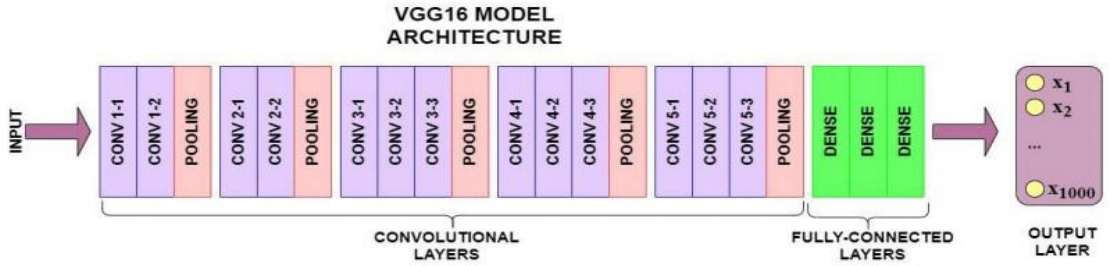


Figure 5. The basic architecture of VGG-16 [53]

b) VGG-19

VGG-19 has filters of 3×3 and a stride of 1 designed to achieve high accuracy in large-scale image recognition applications [55]. VGG-19 has a depth of convolution/max-pooling and fully connected layers with 19 layers in the base model using the SoftMax activation function. The architecture of VGG-19 is shown in Figure 6. VGG-19 is a viral method for image classification due to the use of multiple 3 × 3 filters in each convolutional layer [56].

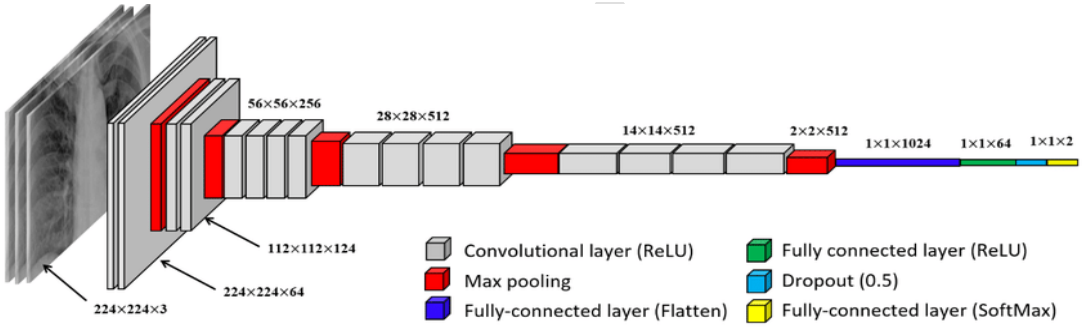


Figure 6. The basic architecture of VGG19 [56]

3.5.3 Resnet 50

One well-known model that performs better in medical image detection is ResNet [57]. ResNet identifies shortcut connections and skips one or more layers in the network, which helps the network provide a direct path to the early layers and make gradient updates for those layers much easier, as shown in Figure 7. The architecture of the ResNet50 model consists of 50 layers and uses ImageNet, which consists of more than 20 thousand categories created for image recognition competitions [58, 59]. The benefit of this model is that it adds more layers that solve the problem of vanishing or exploding gradient degradation problem by skipping the connections that act as gradients allowing the gradient to flow undisturbed [60].

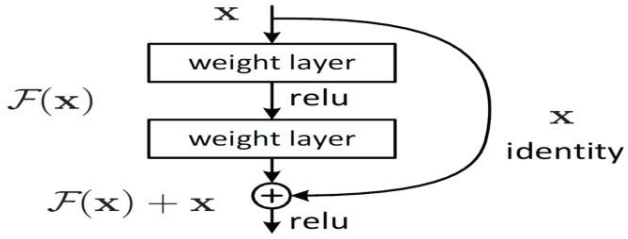


Figure 7. Residual learning: a building block [61]

3.5.4 DenseNet201

Dense Convolutional Network (DenseNet) architecture offers 201-layer densely connected convolutional concrete of all feature maps from previous layers, as shown in Figure 8. All the feature maps propagate to the last layers and are connected to the newly generated feature maps [62]. So, the DenseNet inputs the feature and helps to reuse the feature and prevent exploding or gradient vanishing to a certain extent [63]. However, the limitation of using this model is that excessive connections decrease the networks' computation and parameter efficiency and make them vulnerable to overfitting [64].

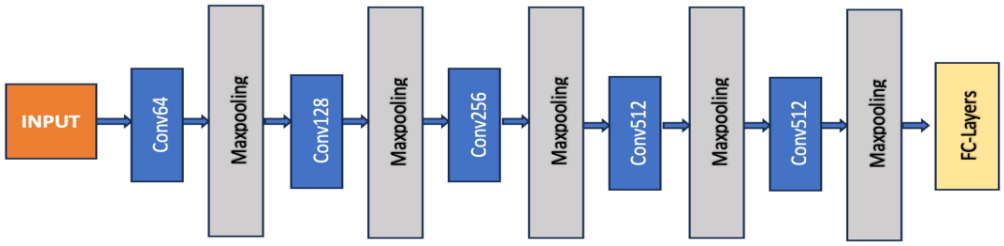


Figure 8. Structure diagram of the DenseNet201 model [64]

3.5.5 InceptionV3

The InceptionV3 model is an improved version of inception such as label smoothing [60]. As shown in Figure 9, InceptionV3 7x7 convolution and propagating label information on the network allows for propagating label information lower down the network. The main benefit of using InceptionV3 is its 24 million parameters, which give better accuracy to the ImageNet dataset [65].

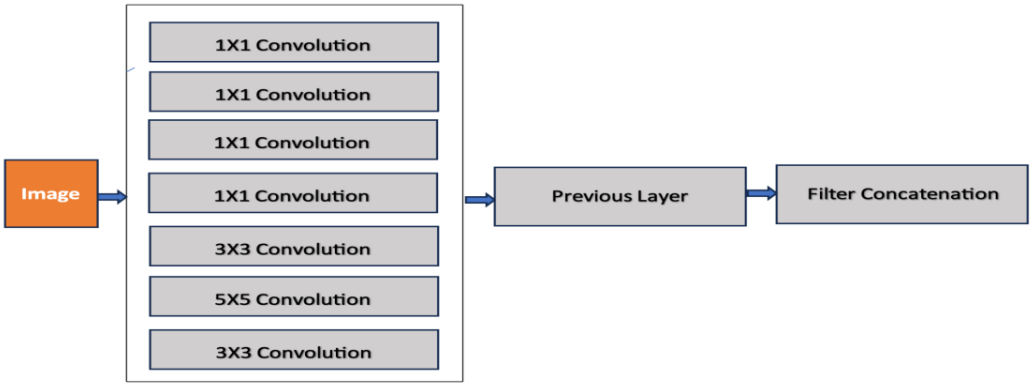


Figure 9. Structure diagram of the InceptionV3 model [66]

3.5.6 Xception

The Xception network is the modified version of the inception with 22.9 million parameters [67]. With more parameters and deep architecture, Xception models help to solve many image classification and object detection problems [24]. The layout of the Xception architecture used in this study is shown in Figure 10. There are three blocks in the model, the first block is the entry flow where the data goes, then the data is passed to the middle and is repeated eight times, and lastly, the last is the exit flow. As the model is lighter and has a smaller number of connections, it is robust and more vital compared to the Inception model [54].

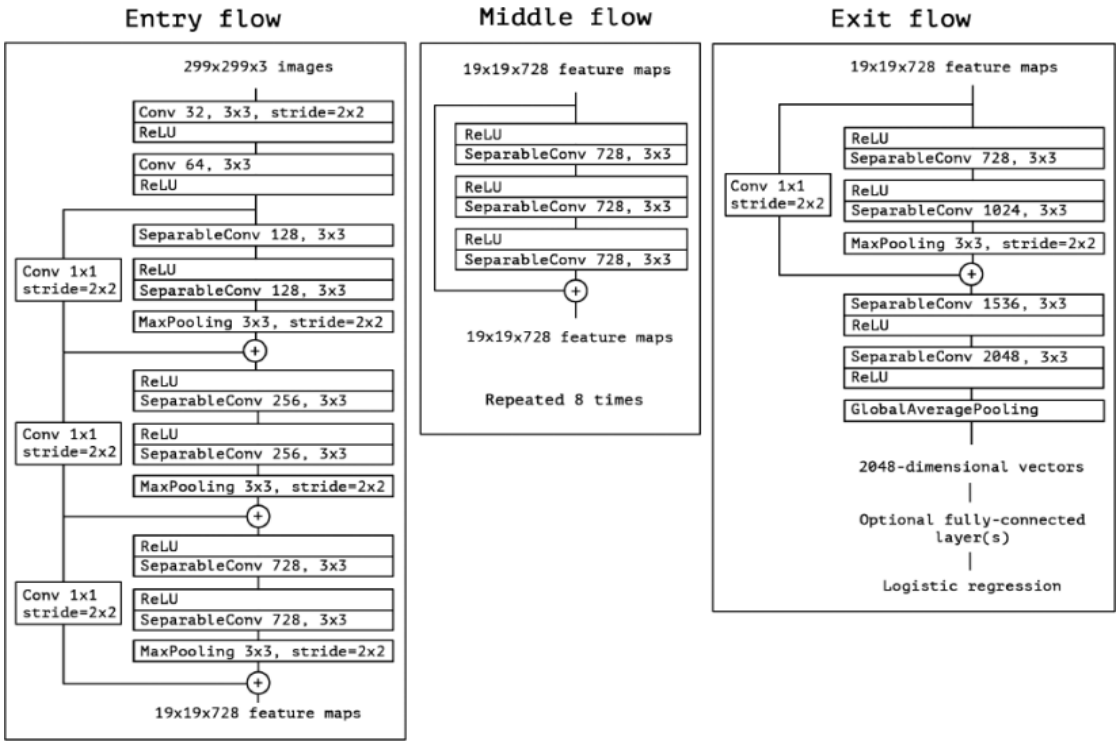


Figure 10. Structure diagram of the InceptionV3 model [68]

3.6 Transfer Learning with convolutional neural network

According to Lytras et al. [69], Transfer Learning (TL) refers to a process where a model trained on one problem is used in some way related to a second problem. TL is mainly used for image recognition and helps build a new model with fewer image datasets [43, 57]. One of the significant drawbacks of training the medical images is the lack of sufficient data, so using transfer learning allows for processing less data [34, 59]. This study uses TL, comparative testing of several models namely, CNN, VGG-16/VGG-19, Resnet 50, DensNet201, InceptionV3 and Xception with and without augmentation. The main aim of this comparison is to identify whether each model tested can detect COVID-19 from the multiclass comprising of COVID-19, pneumonia, and normal image datasets.

3.7 Data augmentation

Deep learning models have deep network structures with many parameters, which requires a dataset with a large sample. However, the number of images in this study has been limited. To overcome this limitation, data augmentation has been applied which is a technique that eliminates the limited dataset insufficiency by increasing image data with the use of existing images [70]. Data Augmentation gives the data size enlightenment in training data making the dataset compatible for better prediction of the deep learning model [71,72]. Besides, data augmentation can also reduce the overfitting of the model

by increasing its accuracy [73,74]. Furthermore, it helps to manage the data by enhancing the augmentation quality and producing an updated version of images [75]. In this study, data augmentation has been applied using an image data generator including rotation, cropping, zooming, and flipping. The Image was rotated with a range of 10, with a width shift range of 0.1, height shift range of 0.1, zoom range of 0.1, horizontal flip True, and fill mode as nearest during the augmentation of the data.

3.8 Additional model considerations: Overfitting and Hyperparameter tuning.

Overfitting in ML is a situation when ML models memorise the training dataset without learning important features, trends, or boundaries [70]. Overfitting can be seen when the error in the validation dataset is higher than in the training dataset, and as a result, the performance is poor for unseen data. For example, it is difficult to acquire medical image data due to concerns about patient confidentiality and the cost of getting high-quality images. Another problem in deep learning is that its architecture requires a large amount of training data. Due to data shortages, lack of quality images and uneven datasets, especially in the medical sector, cause either low-performing models or overfitting [76]. In this study, for the experiment's multi-classification of COVID-19, pneumonia, normal X-ray images with batch sizes from 10 and 32 were used. Using Adam optimizer with weight as ImageNet, cross-entropy learning rates=0.0001 with epochs 10 up to 100 epochs were used to identify the best-performing models. Data augmentation is also with rotation range =10, width shift range=0.1, height shift range=0.1, zoom range=0.1, horizontal flip=True and fill mode=nearest. A total of 6000 training images were generated using augmentation.

As performance measures for the classification of the models, the model accuracy curve, model loss curve and confusion matrix are also used. A confusion matrix has been used in this study to visualise the prediction. The confusion matrix consists of four terms: True Positive (TP) when the model classifies the diseased person as a disease; True Negative (TN), when a model classifies a non-diseased person as non-diseased; False Positive (FP) as positive even if the person does not have any disease, also known as type 1 error; False-negative (FN) person does not have the disease but is predicted as a person having a disease, also known as Type II error. The confusion matrix helps visualise the models' class predicted [77]. Besides that, statistical measure precision, recall, F1 score, and area under the curve (AUC) accuracy are also used. This study will identify the most widely used metrics for evaluation for multi-classification in each class COVID-19, normal and viral pneumonia. This study used to assess performance using four metrics used here to evaluate the model's accuracy, precision, recall, and F1-score. as explained in the paragraph below.

3.9 Receiver operating characteristics (ROC) and Area under the curve (AUC)

The ROC and AUC curves plot the true positive rate with the false-positive rate. The Area's high value under the curve demonstrates that the model performs well in the test data set. ROC is a probability curve and AUC represents the degree or measure of separability and helps identify how much the model can distinguish between classes. The higher the AUC, the better the model predicts [78].

4. Results Analysis:

4.1 Data Pre-processing:

In this study, datasets consisted of COVID-19 (3016), pneumonia (3016) and normal (3016), with a total of 9048 X-ray images have been used. All three datasets were combined and created in a CSV followed by random shuffling to ensure equal and random representation. As the data were added from multiple sources, the quality of images was widely varied, blurred, low intensity and less clear in the border and edges. As a result, this could lead to inaccurate diagnosis of the disease [79]. Therefore, pre-

processing helps reduce and remove the effects on the performance of the models caused by data inconsistency by increasing the model's accuracy [80]. Also, resizing and enhancement are used to classify the disease in X-rays, and CT scans in COVID-19 diagnosis to enhance the image quality. The original shape of the image is in the shape 299,299,3. However different models' architecture needs different shapes such as Xception and InceptionV3 architecture expects an image size of 299 × 299 × 3, whereas DenseNet201 architecture expects an input size of 224 × 224 × 3 (Asif et al., 2022). So, to address the architecture used in this model, the image data shape has been converted to a 224 × 224 × 3 shape. Histogram equalisation (HE) provides contrast enhancement in consumer electronics, medical image processing, image matching and searching, speech recognition, and texture synthesis. This approach is popular and widely used due to its simplicity and effectiveness [81], In this study, the set-up parameters of the filter were used for pre-processing. The Image was normalised by a factor of 1/255 and converted into pre-processed images. After that, the dataset was divided into training, validation, and test datasets to train the models, check model performance and overfitting and test to determine how the dataset classifies the images from trained models. Finally, unseen data has been used to test further to identify whether models predict accurately or not. Data augmentation has been used to mitigate the image insufficiency problem and to prevent the overfitting of the model. In addition, a total of 6000 images have been augmented through rotating, cropping, zooming, and flipping.

4.2 Model Building: X-ray Image Classifier

This study used transfer learning models such as CNN VGG-16, VGG-19, Resnet50, DenseNet201, InceptionV3, and Xception. In addition, traditional ML models such as KNN, SVM and DT were used. In the implementation of these models, feature extraction of the image was performed in terms of pixel values in the image matrix, which has been set as 224 x 224, and flattens the RGB (Red, Green and Blue) pixel intensities into a single list of numbers. All the experiments were conducted with and without augmentation. Among the KNN, SVM and DT, KNN has shown the highest accuracy for both without augmentation and with augmentation with accuracy of 88% and 89% respectively. Whereas the SVM and DT have accuracies of 81% and 76% without data augmentation and 80% and 78% with augmentation respectively. Among these three models, data augmentation increased the accuracy level for the KNN and SVN models but decreased for the DT model. For each of the transfer learning models. The model has been trained from 10 to 100 epochs using a categorical cross-entropy loss function. ImageNet with Adam optimizer with learning rates 0.0001, batch size 10 to 32 and SoftMax as an activation function were used. The image generator from Keres was used for data augmentation. Dropout 0.5 was applied in the fully connected layers to avoid overfitting in the model. Among the datasets, 70% of images were used as the training data, 15% of images were used as the validation data and the rest 15% of the images were used as test datasets to train all models, comprising KNN, SVM, DT, CNN, VGG-16, VGG-19, DenseNet201, InceptionV3, Xception and Resnet50. In addition, images were resized to 224 x 224. The performance measures after the experiment results from all these models are summarised in Table 2 with the corresponding scores.

Table 2. Classification of Accuracy for Training, Testing and Validation

Classifier	**Without Augmentation**					**With Augmentation**				
	Training		Validation		Test	Training		Validation		Test
	Loss	Accuracy	Loss	Accuracy	Accuracy	Loss	Accuracy	Loss	Accuracy	Accuracy
CNN	0.11	96%	0.18	93%	92%	0.34	0.90	0.53	90	90%
VGG-16	**4.91**	**99%**	**0.06**	**98%**	**98%**	1.98	99%	0.12	97%	97%
VGG-19	1.54	99%	0.19	97%	97%	**4.63**	**99%**	**0.11**	**98%**	**98%**
DenseNet201,	5.4	98%	0.32	96%	95%	0.005	96%	0.3	95%	94%
Resnet50,	0.28	90%	0.32	86%	86%	0.36	91%	0.29	90%	90%
InceptionV3	0.13	91%	4.00	93%	90%	0.07	99%	0.48	93%	92%

Classifier	Without Augmentation					With Augmentation				
	Training		Validation		Test	Training		Validation		Test
	Loss	Accuracy	Loss	Accuracy	Accuracy	Loss	Accuracy	Loss	Accuracy	Accuracy
Xception	2.08	92%	0.54	93%	91%	0.41	98%	4.8	93%	92%

From the Table 2 data, it can be observed that the VGG-16 appeared as the highly accurate model with training 99 % accuracy with 4.91 loss in the training dataset, 98 % validation accuracy with 0.06 loss and 98 % test accuracy when the models were trained, validated, and tested without augmentation. Another model that performed better without an augmented dataset is VGG-19 with 99% training accuracy, 97% validation accuracy and 97% test accuracy. The DT algorithm appeared as the model with the lowest training accuracy of 76%. When the dataset is enriched with augmented data, the VGG-19 model has the best model with 99 % accuracy with a 4.63 loss in the training dataset and 98 % validation accuracy with a 0.11 loss and 98% test accuracy. On the other hand, the training accuracy of VGG-16 remains 99% but validation accuracy and test accuracies dropped to 97%. With augmented data, the performance of DenseNet201 decreases by 2 % in training, and by 1 % in validation and test accuracy. The performance of Resnet50 is less than other deep learning models, but with the augmented data the accuracy has increased by 1 % in training accuracy and by 4% in validation and test accuracy. When comparing traditional ML with deep learning, the deep learning models performed better than conventional ML models. However, among the traditional machine learning algorithms, KNN performed better compared to the other two models. With the augmentation of ML models, only the KNN models' accuracy increased by 1 %, whereas the DT's accuracy increased by 2 %, and the SVM models' accuracy decreased by 1 %.

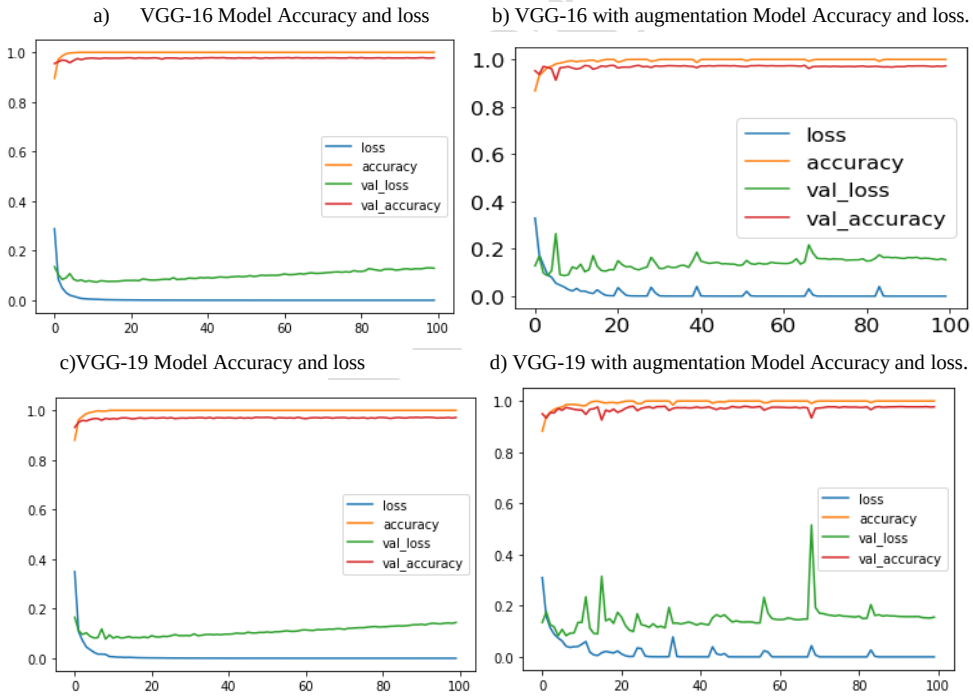
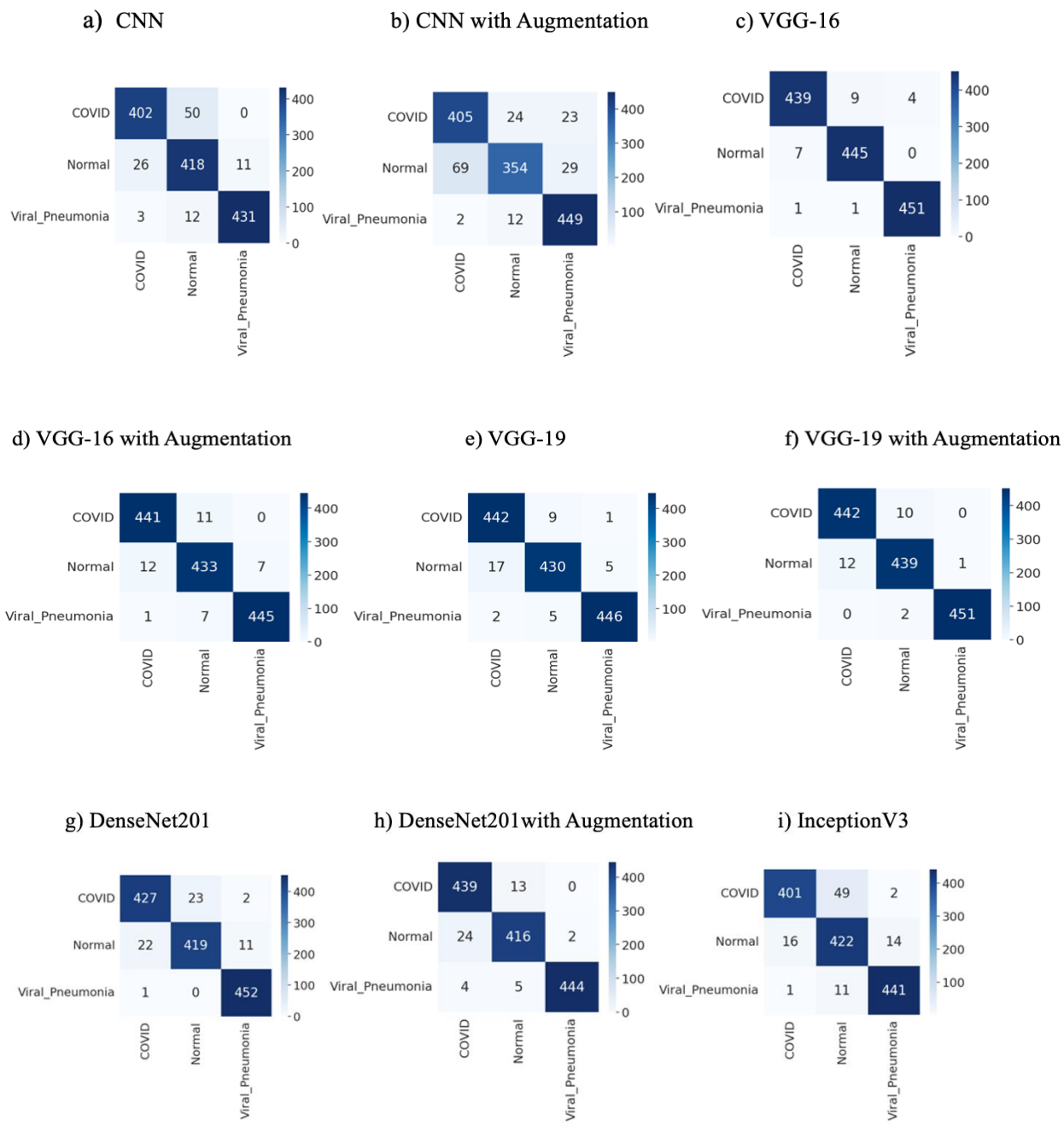


Figure 11. Visualising loss and accuracy best-performed model

VGG-16 without augmentation and VGG-19 with augmentation are visualised in Figure 8. VGG-16 possesses a loss of 4.91 for 100 epochs with 98% accuracy. In the case of VGG-19, the model performs with 99 % accuracy and a loss of 4.63 for 100 epochs. Accuracy for non-augmented data does not fluctuate and remains as it is after 20 epochs, and loss decreases from 10 epochs and remains stable.

However, the validation loss starts increasing slightly after 40 epochs. There is a high fluctuation in accuracy and loss function for augmented data. Overall, while comparing the loss function between two highly performed models, it seems VGG-19 with augmentation has a slightly lower loss compared to VGG-16.

In a further comparative analysis, confusion matrices were drawn to measure the model's performance using test data as shown in Figure 9. The confusion matrix from CNN, VGG-16, VGG-19, DenseNet201, InceptionV3, Xception and Resnet50 with and without augmentation is presented. Out of a total of 1357 test images, 452 included COVID-19, 452 were normal, and 453 were pneumonia.



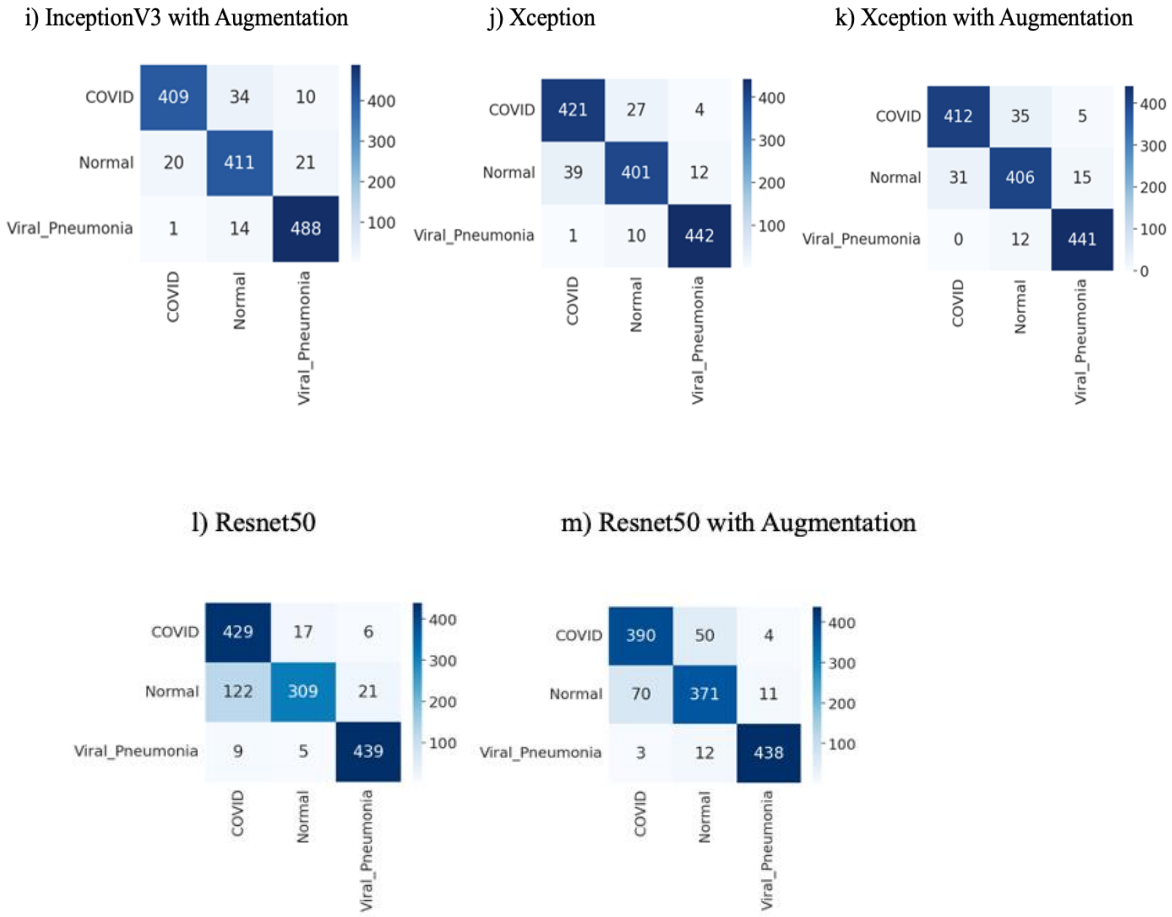


Figure 12 (a-m). Visualising Confusion Matrix for predicting each class.

VGG-19 with the augmentation model classifies 442 COVID-19 cases correctly and misclassifies 10 cases as normal whereas this model classifies has the highest accuracy for pneumonia cases with only 2 wrong classification. However, in the case of VGG-16, out of 452 COVID-19 images, it only classified 439 as COVID-19 and misclassified 9 as normal and 4 as pneumonia. While classifying pneumonia X-ray images, DenseNet201, without an augmentation model, classified 452 images correctly and misclassified 1 image as COVID-19. Similarly, VGG-16 without augmentation performed better in predicting normal images. Out of 452 X-ray images, it predicted 445 with the correct class and misclassified 7 as normal. From the above confusion matrix, it can also be observed that VGG-19 with augmentation performs better for COVID-19, DenseNet201 without augmentation classified pneumonia better and VGG-16 with augmentation classified as normal without misclassifying any images. Overall, VGG-19 performs better in classifying the images with inaccuracies compared to others.

A further comparative analysis was conducted to identify how the models would classify three categories of X-ray images using sensitivity, F1 score and accuracy value. The corresponding results of each model are shown in Table 3. In Table 3, C states COVID-19, N normal and P pneumonia. It shows that VGG-19 with augmentation performed better with a 98 % AUC score, with the precision for COVID-19 being 97 %, normal being 97 % and pneumonia at 99 %. Recall for COVID-19 is 98%, normal with 97 % and 99% for pneumonia and the F1 score for COVID-19 images is 98%, normal image with 97 % and 100% for pneumonia images. Similarly, VGG-16 without augmentation had 97% accuracy, the precision for COVID-19 97%, normal with 97 % and 99% for pneumonia. Recall for

COVID-19 achieved 98 %, normal 97 % and 99 % for pneumonia. Likewise, the F1 score for COVID-19 is 97 %, normal 97 %, and for pneumonia, 99 %. Also, with a 96 % AUC score, DenseNet201 performs better than all CNN, InceptionV3, Resnet50 and Xception. The precision for COVID-19 is 96 %, for normal, 94 % and 99 % for pneumonia. Recall for COVID-19 is 95 %, normal 95 %, and for pneumonia, 99 %. F1 score for COVID-19 is 95 %, normal 94 % and pneumonia 99 %. Overall, with high recall from VGG-16 and VGG-19 without augmentation performs better with a higher accuracy of 98 %.

Table 3. Performance metrics for test data

Classifier	**Without Augmentation**										**With Augmentation**									
	Precision (%)			Recall (%)			F1-score (%)			AUC	Precision (%)			Recall (%)			F1-score (%)			AUC
	C	N	P	C	N	P	C	N	P	(%)	C	N	P	C	N	P	C	N	P	(%)
CNN	93	86	98	89	92	95	91	89	97	92	91	93	90	90	91	99	87	91	94	90
VGG-16	**97**	**97**	**99**	**98**	**97**	**99**	**97**	**97**	**100**	**98**	97	96	98	98	96	98	97	96	98	97
VGG-19	96	97	99	98	95	98	97	96	99	97	**97**	**97**	**99**	**98**	**97**	**99**	**98**	**97**	**99**	**98**
Denset201	96	94	99	95	95	99	95	94	99	96	92	96	99	97	92	98	95	94	99	96
Resnet50	77	93	94	95	88	97	85	90	96	92	90	90	97	89	91	97	91	91	97	91
InceptionV3	96	88	96	89	93	97	92	90	97	93	95	92	94	90	91	99	93	91	96	93
Xception	91	92	97	93	89	98	92	90	97	93	93	90	96	91	90	97	92	90	96	93

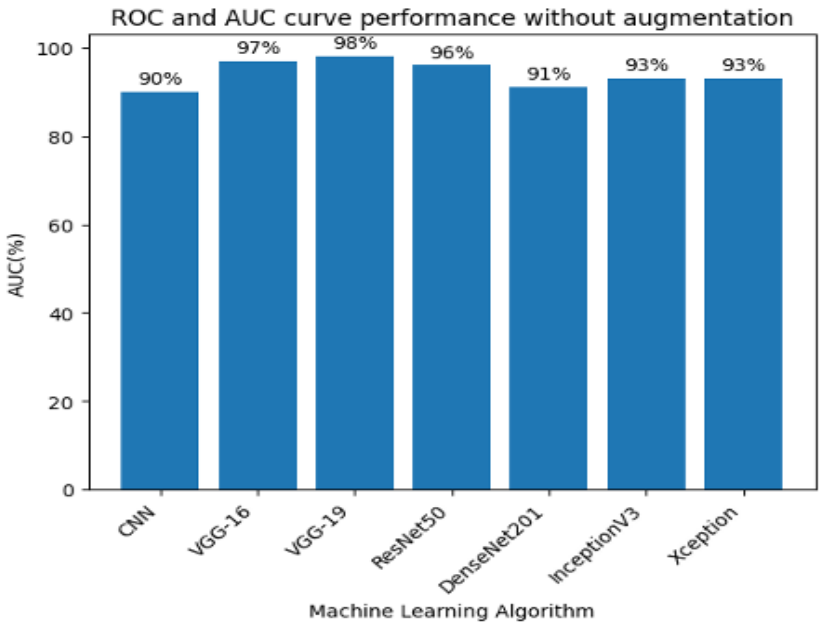


Figure 13. ROC and AUC curve performance without augmentation (%)

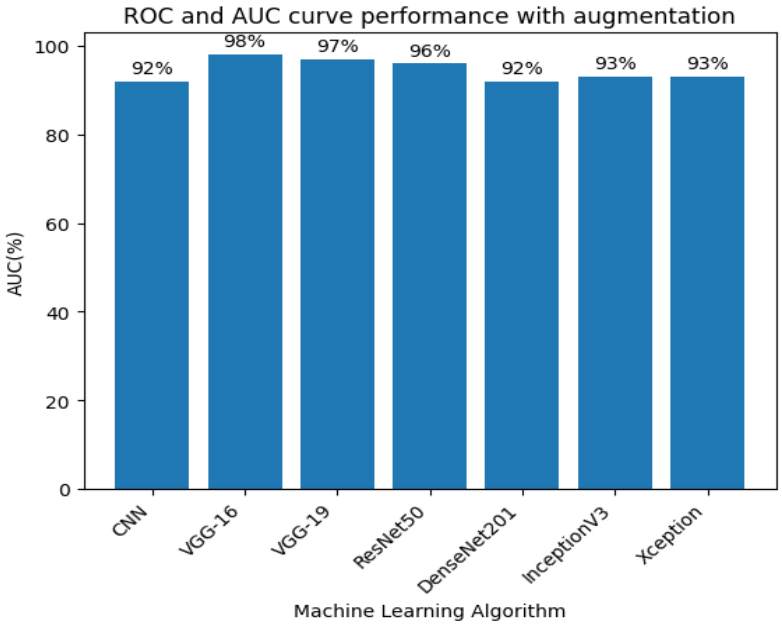
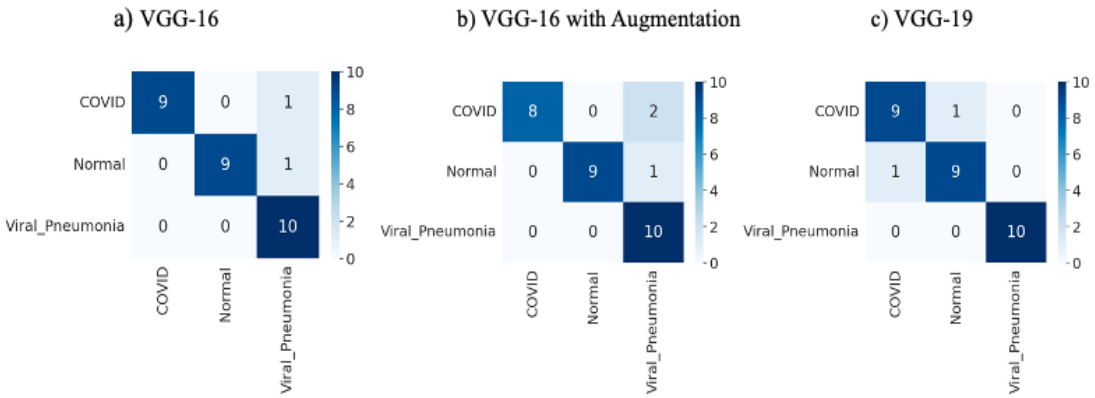


Figure 14. ROC and AUC curve performance with augmentation (%)

In addition, the Receiver Operating Characteristics Curve (ROC) and AUC) for all the models used in this study were analysed as shown in Figures 13 and 14. Figure 13 shows ROC and AUC curve performance without augmentation, and Figure 14 illustrates the graph of the augmented images dataset. VGG-16 performs better in non-augmented images with an accuracy of 98 %, and VGG-19 performs better in an augmented dataset with an accuracy of 98 %. Further to test the model, using CNN, VGG-16, VGG-19, DenseNet201, Resnet50, InceptionV3 and Xception, 30 new unseen image datasets without pre-processing were tested to identify how the models perform with unseen data. Both augmented and non-augmented models were used. After the training to visualise how the model performed, a confusion matrix was created as shown in Figure 15.



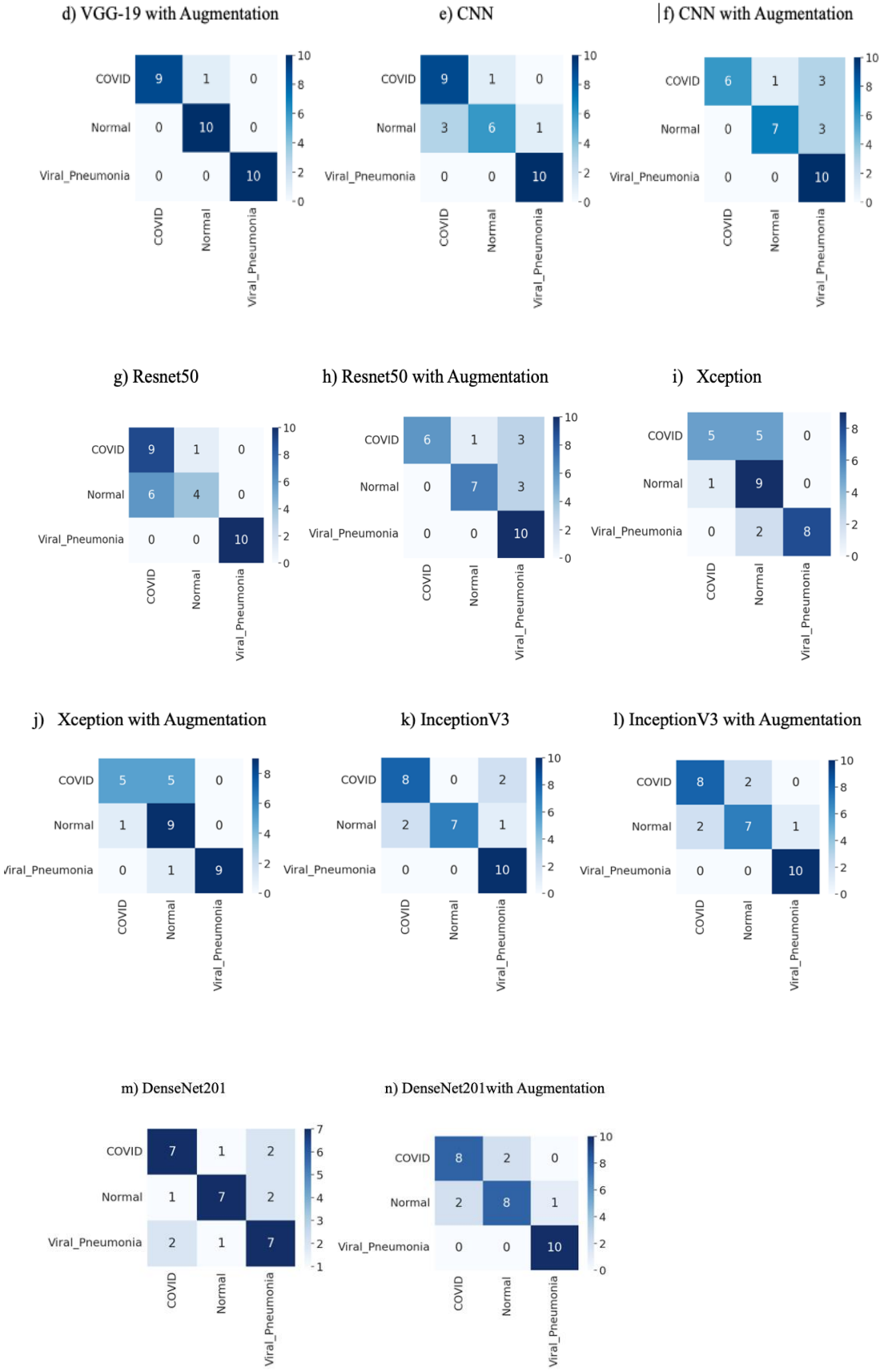


Figure 15 (a-m). Visualising Confusion Matrix for predicting each class.

Analysing the confusion matrix from, Figure 15, shows that VGG-19 with augmentation predicted most of the classes correctly. Out of 10 COVID-19 images, the model predicts 9 X-ray images to its correct class and 1 misclassified as unaffected or otherwise healthy. However, in the case of pneumonia, all the X-ray images are predicted accurately. Out of all models, VGG-19 with augmentation was able to predict COVID-19 along with other classes with the least false positives and false negatives.

Table 4. Performance Metrics for Unseen Data

Classifier	Without Augmentation										With Augmentation									
	Precision (%)			Recall (%)			F1-score (%)			AUC	Precision (%)			Recall (%)			F1-score (%)			AUC
	C	N	P	C	N	P	C	N	P	(%)	C	N	P	C	N	P	C	N	P	(%)
CNN	75	85	90	75	85	90	90	60	100	83	100	80	60	60	70	100	75	77	76	76
VGG-16	100	90	94	83	100	90	100	90	94	93	75	85	90	90	60	100	81	70	95	90
VGG-19	**81**	**90**	**85**	**88**	**80**	**84**	**100**	**100**	**100**	**93**	**90**	**90**	**100**	**90**	**90**	**100**	**90**	**90**	**100**	**96**
DenseNet 201	70	77	63	70	77	63	70	70	70	70	80	80	100	80	80	100	80	80	100	86
Resnet50	60	80	100	90	40	100	72	50	70	76	100	87	62	60	70	100	70	70	70	76
Inception V3	80	100	76	80	70	100	80	80	80	83	80	77	90	80	70	100	80	70	95	83
Xception	83	50	62	56	90	69	100	80	88	73	80	60	100	50	90	90	60	70	90	76

From Table 4, result analyses show that VGG-19 with data augmentation achieved the best performance with a 96 % AUC score. The model achieved a precision of 90 % for COVID-19, 90 % for normal and 100% for pneumonia. Recall for COVID-19 and unaffected individuals is 90 % and 100 % for pneumonia. Similarly, the F1 score for COVID-19 and unaffected individuals is 90 %, and 100 % for pneumonia. The second-highest performing models are VGG-16 and VGG-19, with 93 %.

5. Discussion and Conclusion

This paper evaluated a range of ML models to determine whether COVID-19-associated lung changes could be classified from X-ray images and a distinction made between COVID-19, Pneumonia, and unaffected otherwise healthy individuals. The experimental results identified VGG-19 with data augmentation as the best-performing algorithm among ten tested ML models. The precision achieved 98 % for COVID-19 and unaffected otherwise healthy individuals and 100% for pneumonia. Similarly, the recall (sensitivity) for COVID-19 is 98%, unaffected 97% and 99% for pneumonia. Lastly, the F1 score is 98% for COVID-19 unaffected 97% and 99% for pneumonia. The overall AUC score of the model was 96%. Hence using VGG-19, presented in this paper with image pre-processing, tuning and augmentation performs the best model when compared to any deep learning models. In the performance evaluation metric, recall identifies the actual positive cases of whether the person has COVID-19 or not, and high precision indicates a low false-positive rate and high recall indicates a low false negative. Thus, VGG-19 with augmentation exhibits the best precision, F1 score and accuracy. When compared with other similar studies such as Makris et al.'s [82] deep learning models, VGG-16 and VGG-19 where they have the highest accuracy of 95%, the presented model in this study appeared as a highly accurate model with a test accuracy of 98%. This comparative study also reveals that deep learning models have a powerful learning ability with feature extraction, which can improve image classification. Although deep learning theory has achieved higher accuracy for detecting image

classification, it has problems such as excessive gradient propagation path and overfitting. These limitations had been overcome using data augmentation through rotating, flipping or injecting noise into image data or by improving the architecture in the deep network. Similarly, using various image data pre-processing techniques has improved the quality of the image and prevented bias in the data.

5.1 Limitations, assumptions and future directions

Despite promising results, the proposed methods need to be tested clinically to identify the robustness of the model. However, more importantly, the presented work helps in establishing the most appropriate model for detecting COVID-19-associated changes and is a fundamental first step in establishing a decision aid to help detect and monitor the treatment of patients with PCS. Besides, this study also reveals some of the challenges and strategies with data collection methods, quality and consistency which need to be considered in similar research works. As future work, to refine the model and improve performance, multiple feature extraction, segmentation, and cross-validation can be utilised. Besides that, this study is solely based on publicly available data, which has a direct impact on the quality of consistency as imaging techniques between hospitals/counties/countries vary and can influence the performance of the algorithm. Finally, to check the reliability of the model, it needs to be tested with a large dataset with a diverse group and on a larger scale. In addition, in the future, a clinical decision aid designed to classify COVID-19-associated lung changes on X-rays for patients with PCS could be developed. It may be appropriate to adopt or adapt scoring systems used for other medical conditions affecting lung tissues with changes visible by X-ray to provide an indicator of the extent of lung damage. For example, the Northern Score or the Chrispin-Norman score to quantitatively assess the radiological features of lung changes in Cystic Fibrosis patients – a genetic condition for which routine chest X-ray imaging is undertaken to monitor the disease [83]. Importantly, the ability to score the extent of lung tissue changes in PCS would permit better stratification of patient groups and allow more accurate monitoring of response to treatments. Moreover, this would prompt a switch from a classification problem to one requiring regression to calculate scores reflecting the extent of lung tissue change. Significant challenges remain with this approach as it would require either expert annotation of the existing dataset with the scoring system or collection and collation of a new dataset with assigned scores. One key advantage of the latter is that the time post-infection could also be captured with repeated measures (X-ray images) being taken of patients through the course of PCS. Ultimately, we believe that ML will streamline medical care and aid the development of future treatments for patients suffering from long-term chronic conditions. Besides, in future, it could also explore to classify the extent of damage and changes over time of the long COVID. Exploration of these challenges will be helpful in future research directions to the long COVID-19 diagnosis system.

References

1. Pan, F., et al., *Time course of lung changes at chest CT during recovery from coronavirus disease 2019 (COVID-19).* Radiology, 2020. **295**(3): p. 715-721.
2. Debata, B., P. Patnaik, and A. Mishra, *COVID- 19 pandemic! It's impact on people, economy, and environment.* Journal of Public Affairs, 2020. **20**(4): p. e2372.
3. Baker, C.A. and K.E. Gibson, *Persistence of SARS-CoV-2 on surfaces and relevance to the food industry.* Current Opinion in Food Science, 2022. **47**: p. 100875.
4. Yu, Y. and P. Chen, *Coronavirus disease 2019 (COVID-19) in neonates and children from China: a review.* Frontiers in pediatrics, 2020. **8**: p. 287.
5. O'Mahoney, L.L., et al., *The prevalence and long-term health effects of Long Covid among hospitalised and non-hospitalised populations: A systematic review and meta-analysis.* EClinicalMedicine, 2023. **55**: p. 101762.
6. *Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK: 1 April 2021*. 2021 [cited 2021 03/04/2023]; Prevalence of ongoing symptoms following coronavirus

(COVID-19) infection in the UK: 1 April 2021]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/1april2021.

7. Hashiguchi, T.C.O., J. Oderkirk, and L. Slawomirski, *Fulfilling the Promise of Artificial Intelligence in the Health Sector: Let's Get Real.* Value in Health, 2022. **25**(3): p. 368-373.
8. Khanna, V.V., et al., *A machine learning and explainable artificial intelligence triage-prediction system for COVID-19.* Decision Analytics Journal, 2023. **7**: p. 100246.
9. Kourou, K., et al., *Machine learning applications in cancer prognosis and prediction.* Computational and structural biotechnology journal, 2015. **13**: p. 8-17.
10. Račić, L., T. Popović, and S. Šandi. *Pneumonia detection using deep learning based on convolutional neural network*. in *2021 25th International Conference on Information Technology (IT)*. 2021. IEEE.
11. Chen, J. and K.C. See, *Artificial intelligence for COVID-19: rapid review.* Journal of medical Internet research, 2020. **22**(10): p. e21476.
12. Wootton, D. and C. Feldman, *The diagnosis of pneumonia requires a chest radiograph (x-ray)—yes, no or sometimes?* Pneumonia, 2014. **5**(1): p. 1-7.
13. Castiglioni, I., et al., *AI applications to medical images: From machine learning to deep learning.* Physica Medica, 2021. **83**: p. 9-24.
14. Erdaw, Y. and E. Tachbele, *Machine learning model applied on chest X-ray images enables automatic detection of COVID-19 cases with high accuracy.* International Journal of General Medicine, 2021. **14**: p. 4923.
15. Aral, R.A., et al. *Classification of trashnet dataset based on deep learning models*. in *2018 IEEE International Conference on Big Data (Big Data)*. 2018. IEEE.
16. Rajpurkar, P., et al., *Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists.* PLoS medicine, 2018. **15**(11): p. e1002686.
17. Kuvvetli, Y., et al., *A predictive analytics model for COVID-19 pandemic using artificial neural networks.* Decision Analytics Journal, 2021. **1**: p. 100007.
18. Dalglish, S.L., *COVID-19 gives the lie to global health expertise.* The lancet, 2020. **395**(10231): p. 1189.
19. Nzeribe, E., et al., *COVID-19 and its impacts: The situation in Niger republic.* Clinical epidemiology and global health, 2021. **11**: p. 100797.
20. Giri, A.K. and D.R. Rana, *Charting the challenges behind the testing of COVID-19 in developing countries: Nepal as a case study*. 2020, Chinese Medical Journals Publishing House Co., Ltd. 42 Dongsi Xidajie …. p. 53-56.
21. Ahmed, M., et al. *Automatic COVID-19 pneumonia diagnosis from x-ray lung image: A Deep Feature and Machine Learning Solution*. in *Journal of Physics: Conference Series*. 2021. IOP Publishing.
22. Khan, N., et al., *COVID-19 classification based on Chest X-Ray images using machine learning techniques.* Journal of Computer Science and Technology Studies, 2020. **2**(2): p. 01-11.
23. Santa Cruz, B.G., et al., *Public covid-19 x-ray datasets and their impact on model bias–a systematic review of a significant problem.* Medical image analysis, 2021. **74**: p. 102225.
24. Hemdan, E.E.-D., M.A. Shouman, and M.E. Karar, *Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images.* arXiv preprint arXiv:2003.11055, 2020.
25. Kayalibay, B., G. Jensen, and P. van der Smagt, *CNN-based segmentation of medical imaging data.* arXiv preprint arXiv:1701.03056, 2017.
26. Mohammad-Rahimi, H., et al., *Application of machine learning in diagnosis of COVID-19 through X-ray and CT images: a scoping review.* Frontiers in cardiovascular medicine, 2021. **8**: p. 638011.
27. Hafeez, U., et al., *A CNN based coronavirus disease prediction system for chest X-rays.* Journal of Ambient Intelligence and Humanized Computing, 2022: p. 1-15.
28. Redie, D.K., et al., *Diagnosis of COVID-19 using chest X-ray images based on modified DarkCovidNet model.* Evolutionary Intelligence, 2022: p. 1-10.

29. Chaddad, A., L. Hassan, and C. Desrosiers, *Deep CNN models for predicting COVID-19 in CT and x-ray images.* Journal of medical imaging, 2021. **8**(S1): p. 014502.
30. Haritha, D., N. Swaroop, and M. Mounika. *Prediction of COVID-19 Cases Using CNN with X-rays*. in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*. 2020. IEEE.
31. Kumar, V., et al., *COV-DLS: Prediction of COVID-19 from X-Rays Using Enhanced Deep Transfer Learning Techniques.* Journal of Healthcare Engineering, 2022. **2022**.
32. Asif, S., et al. *Classification of COVID-19 from chest X-ray images using deep convolutional neural network*. in *2020 IEEE 6th international conference on computer and communications (ICCC)*. 2020. IEEE.
33. Chen, A., et al. *Detecting Covid-19 in chest X-rays using transfer learning with VGG16*. in *CSBio'20: proceedings of the eleventh international conference on computational systems-biology and bioinformatics*. 2020.
34. Taresh, M.M., et al., *Transfer learning to detect covid-19 automatically from x-ray images using convolutional neural networks.* International Journal of Biomedical Imaging, 2021. **2021**.
35. Elagili Elagili, A., I. Ighneiwa Ighneiwa, and Z. Rajab Rajab. *Deep Learning to Improve COVID-19 Detection: Using CNN-Based Transfer Learning from Chest X-Ray Images*. in *The 7th International Conference on Engineering & MIS 2021*. 2021.
36. Monga, J.S., Y.S. Champawat, and S. Kharb, *Transfer Learning Based Multiclass Classification For Covid-19 Detection Using Chest X-Rays*, in *Computer Vision and Robotics*. 2022, Springer. p. 213-227.
37. Alshehri, E., et al., *COVID-19 Diagnosis from Medical Images Using Transfer Learning.* Saudi Journal of Health Systems Research, 2022: p. 1-8.
38. Mane, P.R., R. Shenoy, and G. Prabhu. *Comparison of Classification Models Based on Deep learning on COVID-19 Chest X-Rays*. in *Journal of Physics: Conference Series*. 2022. IOP Publishing.
39. Sakib, S., et al., *Detection of COVID-19 disease from chest X-ray images: a deep transfer learning framework.* MedRxiv, 2020.
40. Sethy, P.K. and S.K. Behera, *Detection of coronavirus disease (covid-19) based on deep features.* 2020.
41. Gouda, W., et al. *Detection of COVID-19 Based on Chest X-rays Using Deep Learning*. in *Healthcare*. 2022. MDPI.
42. Alam, N.-A.-., et al., *COVID-19 Detection from Chest X-ray Images Using Feature Fusion and Deep Learning.* Sensors, 2021. **21**(4): p. 1480.
43. Awan, M.J., et al., *Detection of COVID-19 in chest X-ray images: A big data enabled deep learning approach.* International journal of environmental research and public health, 2021. 18(19): p. 10147.
44. Mathew, C. and P. Asha. *Detection of Covid-19 from chest X-ray scans using machine learning*. in *AIP Conference Proceedings*. 2022. AIP Publishing LLC.
45. Barstugan, M., U. Ozkaya, and S. Ozturk, *Coronavirus (covid-19) classification using ct images by machine learning methods.* arXiv preprint arXiv:2003.09424, 2020.
46. Shabbir, A., et al., *Exploratory data analysis, classification, comparative analysis, case severity detection, and internet of things in COVID-19 telemonitoring for smart hospitals.* Journal of Experimental & Theoretical Artificial Intelligence, 2022: p. 1-28.
47. Eljamassi, D.F. and A.Y. Maghari. *COVID-19 detection from chest X-ray scans using machine learning*. in *2020 International Conference on Promising Electronic Technologies (ICPET)*. 2020. IEEE.
48. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry. 2015 Apr 25;27(2):130-5. doi: 10.11919/j.issn.1002-0829.215044. PMID: 26120265; PMCID: PMC4466856
49. Harsono, M., et al., *A Newborn Infant with Congenital Central Hypoventilation Syndrome and Pupillary Abnormalities: A Literature Review.* American Journal of Perinatology Reports, 2022. **12**(03): p. e139-e143.
50. Jain, G., et al., *A deep learning approach to detect Covid-19 coronavirus with X-Ray images.* Biocybernetics and biomedical engineering, 2020. **40**(4): p. 1391-1405.

51. LeCun, Y. *Deep learning & convolutional networks*. in *Hot Chips Symposium*. 2015.
52. L. Hertel, E. Barth, T. Käster, and T. Martinetz, Deep Convolutional Neural Networks as Generic Feature Extractors,2017: p. 1-4
53. S. Gholami, Erfan Zarafshan, A. Sheikh, and Shib Sankar Sana, Using deep learning to enhance business intelligence in organizational management, 2023, pp. 337–35354. Mahanty, C., et al., COVID-19 detection with X-ray images by using transfer learning. Journal of Intelligent & Fuzzy Systems, 2022(Preprint): p. 1-10.
55. Horry, M.J., et al., COVID-19 detection through transfer learning using multimodal imaging data. Ieee Access, 2020. 8: p. 149808-149824.
56. M. Y. Kamil, A deep learning framework to detect Covid-19 disease via chest X-ray and CT scan images, 2021.p. 844.
57. Minaee, S., et al., Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. Medical image analysis, 2020. 65: p. 101794.
58. Hossain, M.B., et al., Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images. Informatics in Medicine Unlocked, 2022. 30: p. 100916.
59. Narin, A., C. Kaya, and Z. Pamuk, Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. Pattern Analysis and Applications, 2021. 24: p. 1207-1220.
60. Mukti, I.Z. and D. Biswas. Transfer learning based plant diseases detection using ResNet50. in 2019 4th International conference on electrical information and communication technology (EICT). 2019. IEEE.
61. Reddy, A.S.B. and D.S. Juliet. Transfer learning with ResNet-50 for malaria cell-image classification. in 2019 International Conference on Communication and Signal Processing (ICCSP). 2019. IEEE.
62. Wang, S.-H. and Y.-D. Zhang, DenseNet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2020. 16(2s): p. 1-19.
63. Huang, G., et al. Densely connected convolutional networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
64. Liu, W. and K. Zeng, SparseNet: A sparse DenseNet for image classification. arXiv preprint arXiv:1804.05340, 2018.
65. Kong, L. and J. Cheng, Classification and detection of COVID-19 X-Ray images based on DenseNet and VGG16 feature fusion. Biomedical Signal Processing and Control, 2022. 77: p. 103772.
66. Bhardwaj, P. and A. Kaur, A novel and efficient deep learning approach for COVID- 19 detection using X- ray imaging modality. International Journal of Imaging Systems and Technology, 2021. 31(4): p. 1775-1791.
67. Fei-Fei, L., J. Deng, and K. Li, ImageNet: Constructing a large-scale image database. Journal of vision, 2009. 9(8): p. 1037-1037.
68. Endah, S.N. and I.N. Shiddiq. Xception architecture transfer learning for garbage classification. in 2020 4th International Conference on Informatics and Computational Sciences (ICICoS). 2020. IEEE.
69. Lytras, M., et al., Artificial intelligence and big data analytics for smart healthcare. 2021: Academic Press.
70. Fawzi, A., Samulowitz, H., Turaga, D. and Frossard, P. (2016). Adaptive data augmentation for image classification. [online] IEEE Xplore. doi:10.1109/ICIP.2016.7533048.
71. Shorten, C. and Khoshgoftaar, T.M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1). doi:10.1186/s40537-019-0197-0.
72. Chen, N., Xu, Z., Liu, Z., Chen, Y., Miao, Y., Li, Q., Hou, Y. and Wang, L. (2022). Data Augmentation and Intelligent Recognition in Pavement Texture Using a Deep Learning. IEEE Transactions on Intelligent Transportation Systems, pp.1–10. doi:10.1109/tits.2022.3140586.
73. Perez, L. and Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621

74. Sevi, M. and AYDIN, İ. (2020). COVID-19 Detection Using Deep Learning Methods. [online] IEEE Xplore. doi:10.1109/ICDABI51230.2020.9325626.
75. Ayalew, A.M., Salau, A.O., Abeje, B.T. and Enyew, B. (2022). Detection and classification of COVID-19 disease from X-ray images using convolutional neural networks and histogram of oriented gradients. Biomedical Signal Processing and Control, 74, p.103530. doi:10.1016/j.bspc.2022.103530.
76. Rahaman, M.M., 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, 2020. 28(5): p. 821-839.
77. Mikołajczyk, A. and M. Grochowski. Data augmentation for improving deep learning in image classification problem. in 2018 international interdisciplinary PhD workshop (IIPhDW). 2018. IEEE.
78. Yang, D., et al., Detection and analysis of COVID-19 in medical images using deep learning techniques. Scientific Reports, 2021. 11(1): p. 19638.
79. Bhattacharyya, A., Bhaik, D., Kumar, S., Thakur, P., Sharma, R. and Pachori, R.B. (2022). A deep learning-based approach to automatically detect COVID-19 cases using chest X-ray images. Biomedical Signal Processing and Control, 71, p.103182.
80. Senan, E.M., Alzahrani, A., Alzahrani, M.Y., Alsharif, N. and Aldhyani, T.H.H. (2021). Automated Diagnosis of Chest X-Ray for Early Detection of COVID-19 Disease. Computational and Mathematical Methods in Medicine, 2021, pp.1–10. doi:10.1155/2021/6919483.
81. Lim, S.H., Mat Isa, N.A., Ooi, C.H. and Toh, K.K.V. (2013). A new histogram equalization method for digital image enhancement and brightness preservation. Signal, Image and Video Processing, 9(3), pp.675–689. doi:10.1007/s11760-013-0500-z.
82. Dilshad, S., et al., *Automated image classification of chest X-rays of COVID-19 using deep transfer learning.* Results in physics, 2021. **28**: p. 104529.
83. Makris, A., I. Kontopoulos, and K. Tserpes. *COVID-19 detection from chest X-Ray images using Deep Learning and Convolutional Neural Networks*. in *11th hellenic conference on artificial intelligence*. 2020.

Highlights of the paper

The following are the major highlights of the paper.

i. Use machine learning algorithms to detect COVID-19-associated lung changes from chest X-ray images.
ii. Compare ten different machine learning algorithms for performance.
iii. Use machine learning to classify the X-ray image into COVID-19 patients, patients with pneumonia, and healthy individuals.
iv. Enhance model accuracy and improve its accuracy with data augmentation.
v. Discuss future interventions for the effectiveness of machine learning in diagnosing and treating long COVID-19 patients.

Sincerely

Bhupesh Kumar Mishra

07/01/2024

Authors' Details:
Susmita Hamal, School of Engineering and Computing, University of Gloucestershire, UK, Email: susmita.hamal@gmail.com
Bhupesh Kumar Mishra, Data Science, AI & Modelling Centre (DAIM), University of Hull, UK, Email: bhupesh.mishra@hull.ac.uk, Corresponding author
Robert Baldock, School of Pharmacy & Biomedical Sciences, University of Portsmouth, UK, Email: robert.baldock@port.ac.uk
William Sayers, School of Engineering and Computing, University of Gloucestershire, UK, Email: wsayers@glos.ac.uk
Tek Narayan Adhikari, School of Engineering and Computing, University of Gloucestershire, UK, Email: tadhikari@glos.ac.uk
Ryan M. Gibson, Department of Computing, Glasgow Caledonian University, UK, Email: Ryan.Gibson@gcu.ac.uk

Declaration of Conflict of Interest

On behalf of all the co-authors, I attest this article has not been published in whole elsewhere and is prepared following the instructions to authors. All authors have contributed to this manuscript, reviewed and approved the current form of the manuscript.

Sincerely

Bhupesh Kumar Mishra

25/06/2023

Authors' Details:
Susmita Hamal, School of Engineerinf and Computing, University of Gloucestershire, UK, Email: susmita.hamal@gmail.com
Bhupesh Kumar Mishra, Data Science, AI & Modelling Centre (DAIM), University of Hull, UK, Email: bhupesh.mishra@hull.ac.uk
Robert Baldock, School of Pharmacy & Biomedical Sciences, University of Portsmouth, UK, Email: robert.baldock@port.ac.uk
William Sayers, School of Engineerinf and Computing, University of Gloucestershire, UK, Email: wsayers@glos.ac.uk
Tek Narayan Adhikari, School of Engineerinf and Computing, University of Gloucestershire, UK, Email: tadhikari@glos.ac.uk
Ryan M. Gibson, Department of Computing, Glasgow Caledonian University, UK, Email: Ryan.Gibson@gcu.ac.uk