MSc in Data Analytics
Research Project Report

Akshen Doke
Student ID: x18191592

School of Computing
National College of Ireland

Supervisor : Prof. Rashmi Gupta

National College of Ireland

MSc Project Report

School of Computing

Student Name:	Akshen Doke
Student ID:	x18191592
Programme:	MSc in Data Analytics **Year:** 2020
Module:	Research Project
Supervisor:	Prof. Rashmi Gupta
Submission Due Date:	1st February 2021.
Project Title:	Image Classification: Detection of covid19, normal and pneumonia from chest x-ray image dataset using ensemble methods.

Word Count:7434 **Page Count** 22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Akshen Doke.

Date: 1st February 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Image Classification: Detection of covid-19, normal and pneumonia from chest x-ray image dataset using ensemble methods.

Abstract

Covid-19 is a lung disease which originated from Wuhan, China and has spread across the globe resulting in loss of human lives and also causing an economic loss in various countries. This paper aims to talk about how a method was developed using deep neural networks which will help us detect this disease using just chest x-ray image. Since Covid-19 chest x-ray images are somewhat similar to that of pneumonia, they are also being taken as a set for training the models so that they can help us detect and distinguish covid-19 from pneumonia accurately without any confusion. In total, 7 models were trained some created from scratch while most adopted using transfer learning methodology on the dataset available publicly and ensemble techniques of two kinds i.e., voting based and weighted were used to get the final output for the given input. Overall, the output obtained was close to 91% based on popular evaluation metrics used for multi-class classification which is very good but still it can raise some concerns as even a single inaccurate diagnosis might result in loss.

Keywords: **Transfer Learning, Ensemble Methods, Deep Neural Networks, Chest X-ray**

1 Introduction

1.1 Background

X-ray technology, which was invented in the 19th century, changed the nature of medical diagnosis of the human body and created what we know as radiology today which helps gain insights in disease detection. Treatments of diseases in lung region, accidental damage to bones etc its detection became easy and precise. Later in the early 20th century computers were invented which made way to calculate and process complex statistical formulas on large datasets quickly which in-turn gave birth to a field called data science and machine learning. Using machine learning principles many researchers especially in the field of computer vision worked on developing computer vision-based models which classified or detected any tumour or object it was required to.

1.2 Motivation

As industrialization happened, it led to air pollution also cigarette consumption gave rise to a number of patients related to lung diseases. In late 2019 the world saw a new type of virus emerging called Covid-19 which is spreading rapidly across the globe, this increased the demand for radiologist but in reality, the supply and reach of radiologist and doctors is low because this pandemic was unexpected as well as radiology being a hard skill to master in less time not many people can be trained for it quickly. This problem made computer vision researchers in many countries work on a model which will help detect this virus automatically and in less time as compared to human examination (Krizhevsky, et al., 2012; Maguolo & Nanni, 2020; Tawsifur Rahman, 2020).

1.3 Research Question and Objective

"Can a reliable and efficient model be created with use of ensemble method for detection of covid-19?"

To achieve and solve the above-mentioned question, following objectives were defined:

- Data pre-processing and data visualisation.
- Model creation, training via transfer learning.
- Model creation from scratch and it's training.
- Visualizing output and parameter optimization for improvement.
- Unifying the trained models as one black box to create an ensemble network.
- Testing of the ensemble networks and evaluating its performance.
- Identifying short comings and future scope.

1.4 Research Overview

In this paper using principles of computer vision for multiclass classification of x-ray images we have tried to get a highly precise ensemble model for identification of lung related diseases like covid-19 and pneumonia (Wang, et al., n.d.). For this project we have used deep neural network which is a technique in computer vision which helps us to pass our data i.e., image via many layers which does the work of identifying features which in turn helps to detect or classify our input image in desired classes. The process continues in loops called epochs were in the network updates its weights and biases to improve its performance of identification or detection (Krizhevsky, et al., 2012).

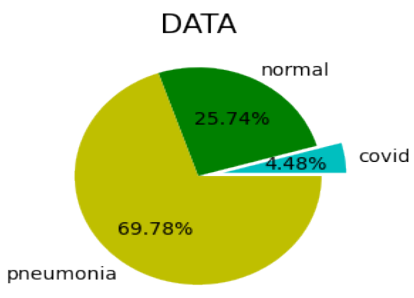


Figure 1: Pie Chart of Data Distribution

We also have performed some basic exploratory data analyses in which we visualize and pre-process the dataset to get a visual representation and ratio of data spread which we have used on our deep neural network-based computer vision models since we have 3 types of chest x-ray images as shown in Figure 1 above, later on we had split that dataset into training and testing set which was further passed to our computer vision models for training, validating and evaluating their performances. For models we have created some from the beginning while others were inherited using principles of transfer learning. After the process of training was completed, we have also visualised to see based on what regions the model thinks and has made the prediction, this is done using class activation method called GRAD-CAM as shown in Figure 2 below.

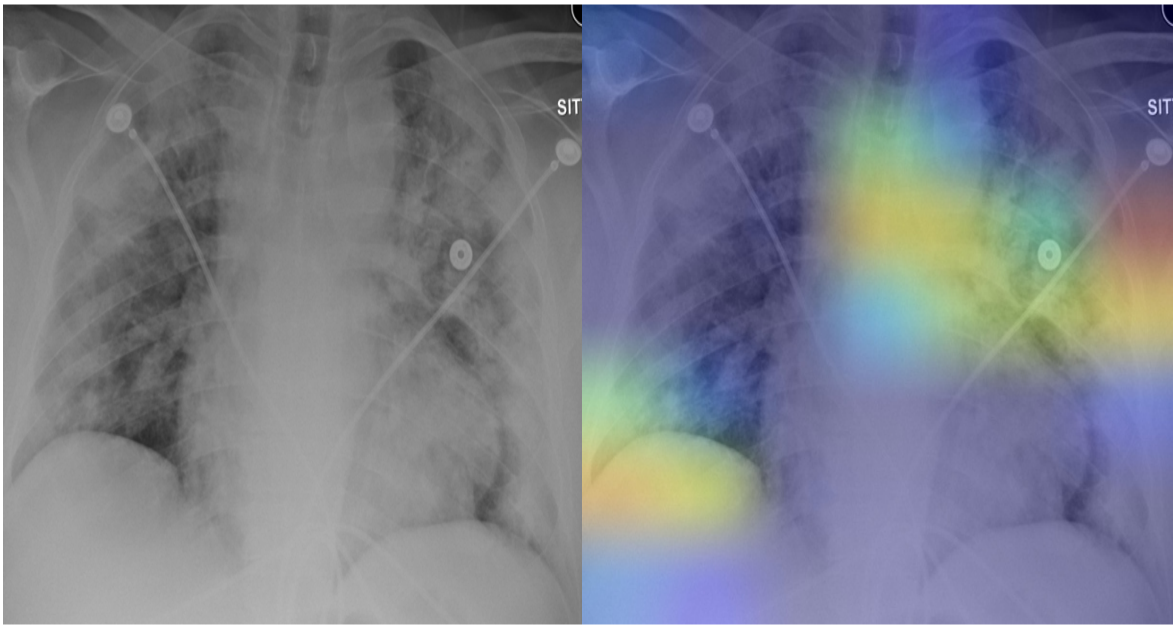


Figure 2: Regular Covid X-Ray, Covid X-ray Image with Grad-CAM

After the models were built, we then combine their outputs, and an ensemble network was formed which acted as a single black box virtually to each input given to it. Inside this ensemble each model we had trained acted independently and gave the prediction based on its own interpretation, which was then counted for voting, in another ensemble network which was weight based, models with high performance and more reliability were weighted more i.e., they were given a chance to vote twice and whatever was the majority was considered as our final output. Such ensembles which have more than two models increases the confidence in the output and reliability.

2 Related Works

In this section, we've gone through works of some researchers related to computer vision, pneumonia and chest x-ray related disease detection in general. This helped us to understand by learning and identifying the gaps of technology as to what can be done and what cannot.

- Transfer Learning Techniques for Chest X-ray classification

  Pneumonia is a disease which not only affects senior citizens but children alike, it was documented that more than a million children below age 15 died and children aged less than 5 were 18% of the total deaths while around more than a billion people worldwide have suffered from some or other form of pneumonia or chest related disease in their lifetime (Rahman, et al., 2020). It was observed that Chest X-ray is one of the best ways to detect pneumonia, but the models do sometimes lead to misclassification to other types of diseases. Here 4 types of CNN based models like ResNet18, DenseNet201, AlexNet (Krizhevsky, et al., 2012) and SqueezeNet were used to classify two known types of pneumonia called viral and bacterial and the dataset used had around 5247 images of patient chest x ray along with some normal ones as well. Considering the dataset to be small for a deep neural network, data augmentation techniques were also used to increase the training data for the models. For this dataset it was found that DenseNet201 was performing better compare to other models.

- Work on Covid-19 dataset of chest x-ray images

  The major chest disease addressed in this paper is covid-19 on which even other researchers have tried and created models, (Maguolo & Nanni, 2020) came up with a very unique method to compare and test protocols used to detect covid-19 from chest x-ray image dataset and came out with a conclusion that most models show similar outcomes. The authors gathered data from 4 unique data providers and also did an out of box method over them wherein they covered the centre of the chest x-ray images with a mask of rectangular black box and ran it over a model based on AlexNet which was trained to recognise the source of the dataset. The model performed well and proved that deep neural networks can give biased output, this was to prove the point of overfitting wherein the model trained on a dataset works well on that data but when taken to the real world it failed to perform as expected on an unknown data.
  Also, it was trained to classify based on the source rather than medical diagnosis. This method was very unique and raised serious concerns over use of machine learning, computer vision models in the field of medical diagnostics.

- Building of the most popular model AlexNet Architecture

  British-Canadian Computer Scientist Geffrey Hilton who is regarded as the Father of Neural Networks known for his pioneering work in the field of research related to deep neural networks developed a deep convolutional neural network model called AlexNet (Krizhevsky, et al., 2012) which was a breakthrough in the domain of computer vision[1]. The model they built was able to classify up to one thousand different classes, they achieved this feat by training models with around 600,000 plus neurons

[1] https://en.wikipedia.org/wiki/Geoffrey_Hinton

consisting of 5 layers of convolution. The model was trained on a sub-set of ImageNet database, which consisted around 1.2million pictures and around 50,000 to 150,000 for validation and testing respectively. Considering the sheer size of the dataset and length of the network they spread it out across two Giga Texel Shader eXtreme 580 Graphics Processing Units each of 3GB RAM using the principles of parallelism. As the architecture was dense problems occurring due to overfitting were observed which was later dealt using Data Augmentation and Dropout techniques where in each neuron in the hidden layer with probability of 0.5 was set to 0 this way they didn't contribute to backpropagation. The result was AlexNet which is still one of the most neural network architectural model which we've used in this project.

- ### One of the best and recognised work on pneumonia detection using chest x-ray images.

More than a million people suffer from pneumonia just in the United States of America every year out of which nearly half of them die because of their illness. CheXNet a 121-layer dense convolutional neural network model was built by (Rajpurkar, et al., 2017) to detect pneumonia. Authors of this paper claim their model performs on par with radiologists and in some cases even exceeds their predictions and at the same time it was able to outperform all the major known research results for 14 pathologies of the ChestX-ray14 dataset. Some limitations claimed and observed which are worth looking over is that they made use of only frontal chest x-ray images to test their trained model and the same was given to radiologist who were volunteering for this experiment, also the radiologist nor the model was allowed to access current or previous patient information of any sort in any way which had an impact on the performance of human radiologist. This is worth noting since in order to maintain patient privacy, no other patient information is taken in the dataset.

- ### X-ray Image Segmentation

Identifying and detecting objects, region of interest or highlights it improves the quality of image classification in computer vision and is known as image segmentation. In X-ray Image Segmentation, methods such as Laplacian, Sobel and Prewitt were used but the issue with it was that it failed in cases where noise was present. (Saad, et al., 2014) talks about an algorithm to deal with early edge detection that helps to manage lower and upper threshold units for image input noise. Although detecting of covid-19 or pneumonia in our case using this method is not a good approach since the virus is spreading unevenly for each patient, this can only be used for detection of boundary reaches for lung or ribs.

- ### Multi-Label Chest X-Ray Classification

As per (Baltruschat, et al., 2018) over 23,000 chest x-ray images didn't go through certified radiologist at Queen Alexandra Medical Institution alone and a handful of patients suffered because their x-ray images weren't being properly assessed. This is not just the only medical institution which has these issues as aging population is growing due to an increased life span, the requirement for healthcare workers and clinicians have also driven up. This observation made them to build a computer vision model which makes use of transfer learning and classifies images based on the features learned by their model. The model they created is loosely based on ResNet-50, they fined tuned some layers of that model to suit their needs at the same time they made use of Grad-CAM which helped them get insights of their model. The dataset they used was of 14 different chest related diseases, which we've also seen in other researches cited above. Their optimized model ResNet-38 achieved better results in 5 out of 14 classes compared to the works they've cited also they observed substantial variability when different splits of the same dataset were considered.

- Class Visualization

(Selvaraju, et al., 2019) devised a method to identify neurons which were important through the Grad-CAM technique as even though computer vision models enable us to classify things, detect objects and do semantic segmentations, it's still hard to interpret individual components. In order to get a reliable model, we need to know how and why the model came to the predicted conclusion. This interpretability is not just useful when it comes to understanding the conclusion part of the model but also while training it as it can help us make the training process smoother by helping us debug and fine tune the model efficiently, diagnosing classification models for considerable biased errors. As shown in the example given that the model to classify doctors and nurses was trained on a dataset which had 78% images of doctors who were men and 93% images of nurses who were women so the model instead of checking for props like stethoscope went for features like facial dimensions, hair length etc and classified all women as nurses while all men as doctors. This issue was identified using grad-cam technique which highlights the region of interest based on which the model made it's classification. Further the model was retrained with more images this time balancing it out with more female doctors and more male nurses to lower the bias results. This helped to demonstrate that grad-cam technique can help not just visualize the output of the trained model but also fix issues if any.

- Detection of Abnormality in Digital Chest Radiography with the aid of Computer Detection Systems

(Carrillo-de-Gea, et al., 2016) proposed a method to perform automatic normality classification of posteroanterior digital chest radiographs which is able to detect anything which can be classified as different from normality. Initially, images of 3000 by 3000 pixels with depth of 12 bits per pixel were taken with an average age of 55. These images were reduced in pixel depth by 4 bits i.e. from 12 to 8 after which decimation is applied to the image using super sampling interpolation which reduced the size by 2000 i.e. to 1000 x 1000 which is considered the standard resolution for further steps.

In the next step of segmentation, image is segmented to locate the position of both lungs which then helps them to determine the region of interest. Samples of both left and right lungs were extracted and the location with maximum correlation is selected as the expected position of each lung, after which a grid of 3 by 4 region is generated. For feature extraction, they made use of LBP histogram for each reach obtained, later these features were classified based on distances between histograms i.e. using Bhattacharyya distance two histograms are computed. Later on the experimental results obtained from the classification were 90% for the best classifier speaking about the disadvantage the method implied by the researchers relies mainly on texture information which in case of some diseases which affect only the intensity of the images would be hard to detect.

- ## Detection of Abnormality and it's Localization using DNN in Chest X-ray region.

(Islam, et al., 2017) made use of ensemble models to improve the classification accuracy for abnormality detection in chest region for x-rays compared to a single model. They made use of some open datasets like JSRT dataset which contained around 250 images of which 154 had lung nodules i.e. malignant and benign cases and 93 didn't have any nodules each of size 2048 x 2048 pixels and a grey-scale colour depth of 12 bits, Shenzhen Dataset from China which had two classes i.e. normal and tuberculosis and Indiana chest X-ray dataset the largest amongst the 3 with 7284 chest x-ray images of both frontal and lateral with diseases like cardiomegaly, pulmonary and pleural effusion on which the first studied performances of some already built deep convolutional network over different abnormalities. For their experiment, they trained many models via transfer learning like AlexNet, VGG based and ResNet and for each they found some model performing better than other on certain diseases while some giving high sensitivity and specificity etc. For the ensemble of models, they trained the variants of the same mentioned model and used a simple linear averaging of probability on individual model as bagging and boosting are implied it might result in a biased model as the number of dataset is low for such huge model with multiple layers. They also tried to increase the number of models in the ensemble and found that gradually a consistent performance was seen after 9 models. For better understanding, like other researchers listed above, they too went to visual depiction of model prediction to actually understand how and why the models were making the classification they were giving rather than just blindly treating it as a black box. Speaking in simple words they wanted to identify the features which contributed more to the output of the model. They took the localization approach for cardiomegaly abnormality and highlighted the 20% area which was more sensitive to the region where the heart is larger than normal heart. They performed the same experiment on around 50 samples of cardiomegaly and normal images and found the result to be mostly consistent. Based on the localization observation approach seeing sensitive region they came to a conclusion that characteristic features in the shape of heart and its surrounding regions is alone to detect cardiomegaly, the lungs are less important when it comes to detecting it. However, while applying similar methods on pointed features for that of nature like bone fracture and lung nodule, the localization method failed.

Table 1: Summary table for above mention related works

Author(s)	Research Question/Purpose	Title	Search Terms	Data	Finding
(Maguolo & Nanni, 2020)	To prove that several testing protocols for recognition are always right for neural network to identify covid-19	A Critic Evaluation of Methods for COVID-19 Automatic Detection from X-Ray Images (2020)	Covid-19; Covid-19 Diagnosis; Convolutional Neural Networks; X-Ray Images	108,948 images of 32,717 different patients, classified into 8 different sectors	Training of models can be biased.
(Rahman, et al., 2020)	Report on advances in accurate detection of pneumonia using transfer learning methods	Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray. (2020)	pneumonia; bacterial and viral pneumonia; chest X-ray; deep learning; transfer learning; image processing	5247 chest X-ray images consisting of bacterial, viral, and normal chest x-rays images were pre-processed and trained for the transfer learning-based classification	Authors used 4 models out of which DenseNet201 outperformed rest of the models in classification.
(Selvaraju, et al., 2019)	Proposed a Novel way to make any CNN-based model more transparent in visual explanation	Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization (2019)	CNN, VGG16, ResNet	Variants of ImageNet dataset	Results show that the authors were able to achieve what they intended to do.
(Baltruschat, et al., 2018)	Building a model using transfer learning and testing with and without fine tuning it.	Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification (2019)	X-Ray, Deep Learning, Convolutional Neural Networks	ChestX-ray14 consists of 112,120 frontal chest X-rays from 30,805 patients	
(Wang, et al., n.d.)	To demonstrate that thoracic diseases can be detected and even spatially-located via a unified weakly-supervised multi-label image classification and disease localization framework	ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases (2017)	Thoracic Disease Detection	108,948 frontal view X-ray images of 32,717 unique patients with the text mined eight disease image labels	Authors conducted quantitative performance on the ChestX-ray8 db using transfer learning techniques.
(Islam, et al., 2017)	DCNN based classification and localization on the publicly available datasets	Abnormality Detection and Localization in Chest X-Rays using Deep Convolutional Neural Networks(2017)	Abnormality detection	7284 CXRs, both frontal and lateral images. Another one with 247 chest X-rays, among which 154 have lung nodules and 93 have no nodules.	The DCNN architecture they used didn't perform well on all abnormalities.
(Rajpurkar, et al., 2017)	Develop an Algorithm which can detect	CheXNet: Radiologist-Level Pneumonia Detection	CheXNet	Chest Xray dataset of 100,000 frontal	Authors created a 121 Dense Convolutional Network called

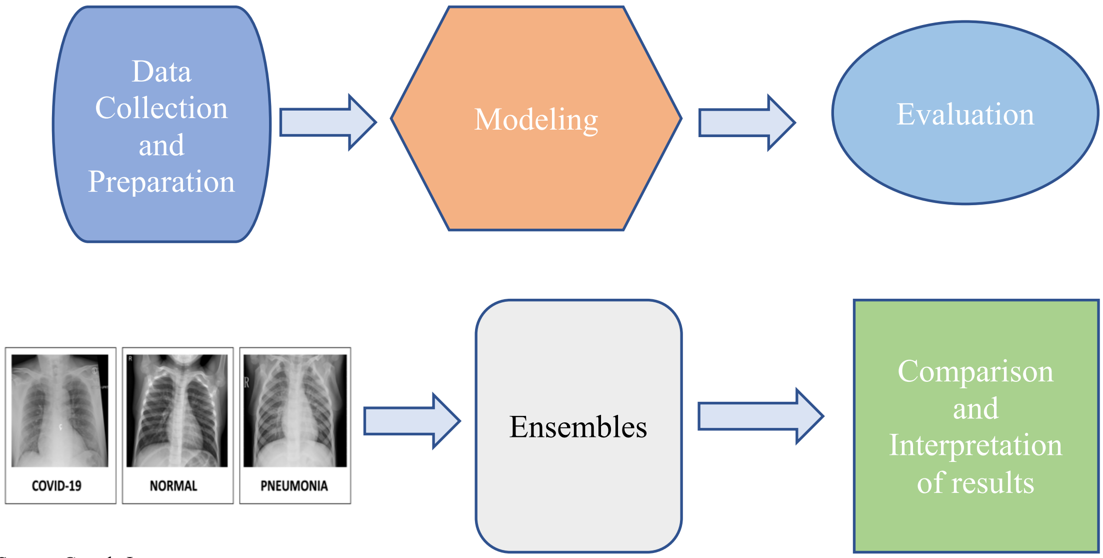
	pneumonia from chest x-ray.	on Chest X-Rays with Deep Learning (2017)		view X-ray images with 14 diseases.	CheXNet which was able to perform better than radiologist given the data under certain circumstances.
(Carrillo-de-Gea, et al., 2016)	Objective of the paper is to perform an automatic normality classification of posteroanterior chest radiographs.	A Computer-Aided Detection System for Digital Chest Radiographs (2016)	Medical Imaging	DICOM images of chest radiographs (23 women and 25 men) were provided by HGURSM, Spain to perform the test.	A new approach for detection of normality in chest radiographies was given which is based on LBP.
(K. He, 2016)	Proposed a learning framework to ease the training of deep networks	Deep Residual Learning for Image Recognition (2016)	Deep neural network, ILSVRC 2015, Residual Learning, COCO object detection.	ImageNet dataset, CIFAR-10	
(Saad, et al., 2014)	Segmentation of Lungs based on object detection techniques.	Image Segmentation for Lung region in Chest Xray Images using Edge Detection and Morphology(2014)	Lungs, Image edge detection, shape, noise, biomedical imaging.	247 CXR images with standard size of 2048 x 2048 pixel	Researcher did build a method for lung segmentation however there are some changes and improvements needed as pointed out by the authors themselves.
(Krizhevsky, et al., 2012)	Create a Model which works efficiently over huge datasets	ImageNet Classification with Deep Convolutional Neural Networks (2012)	AlexNet, ILSVRC-2012	1.2 million high-resolution images were trained for ImageNet LSVRC-2010 contest into the 1000 classes	Large networks tend to achieve good results on huge datasets with supervised learning. Authors want to try the same on video sequence.

- ## Conclusion

After going through the papers in Table 1, we can see that computer vision has evolved tremendously and a lot of research has been carried out to use this technology uplift medical image diagnostics. To the best of my knowledge, combination of transfer learning models into ensembles has not been performed on covid-19 dataset which leaves area for exploration. In further section we can see how using this novel approach a computer vision model is trained and evaluated.

3 Research Methodology

In most cases, whenever an x-ray image of patient is taken, first it is examined by a radiologist who writes a report of his/her interpretation and then the x-ray image and report are forwarded to the doctor who has asked for it. The research in this paper focuses on reducing this time by automating detection of covid-19 and helping doctors to take decisions quickly in order to save patients life. The work for this research is carried out by following simple process illustrated in Figure 3 below.



Source: Google Images

Figure 3: Process

The Novel approach proposed in this paper is the use of ensemble technique to make the final prediction more reliable and efficient.

3.1 Data Collection, Exploration and Preparation

Since the beginning of January 2020, the virus has been spreading across the globe like wildfire, some medical institutions and researchers also published various data on this virus which ranges from its symptoms, measures to take for prevention and if highly suspected, how to contact the medical authorities for its treatment. For collecting data over Covid-19 we made use of the available dataset COVID-19 RADIOGRAPHY DATABASE, also while trying to detect covid-19 from chest x-rays manually it was observed that radiologist got confused between covid-19 and pneumonia (Joanne Cleverley, 2020) hence we also take pneumonia images from here so that our models learn how to distinguish and don't fall for the similarity trap.

Images downloaded from the provider were collected from various sources also the provider made sure to hide vital information of patients while giving the x-ray images considering privacy hence in order to make them consistent, data was renamed, and their extensions were set to ".png" using a simple bash script. The results stored in the folder by the name of type they belonged to i.e., covid-19, normal and pneumonia. Later they were split into two categories one called training and the other for testing, the ratio for which is 70-30 respectively. The split was carried out using random shuffling of images. Also, when the split was made, each class was automatically assigned a label which in this case would be the index of the folders occurrence i.e., 0, 1 and 2 by the keras package when it was loaded.

This newly created data folders were then uploaded on Google Drive as we need them to connect with our cloud infrastructure called Google Colab[2]. This cloud environment allows us to be flexible in our use of resources on demand and also has all the packages required to achieve the objective of our project.

3.2 Data Augmentation

Overall, our data count is low so in order to make most of our dataset we used a technique called data augmentation wherein we performed operations like rotation, transformation by flipping, zooming and scaling so that our model to be trained gets various inputs and also it doesn't see the same image twice as this is helpful to avoid a training problem called over-fitting of data, another plus point of data augmentation is that the transformed images which are generated aren't stored in the secondary memory instead they're generated on fly while training and dumped once the process is over therefore saving storage space. Also, before setting the model to train on our dataset, we had to set the number of batch size which made sure that only that many numbers of images were given in training at the particular moment by updating and validating the weights of our model. This feature setup can also be called as hyper-parameter optimization. The test set was augmented as well and used for validation in order to deal with low data points.

3.3 Modelling

The pre-processed data was then sent as input to our 7 models, and output from each one was later combined to get our polling results.

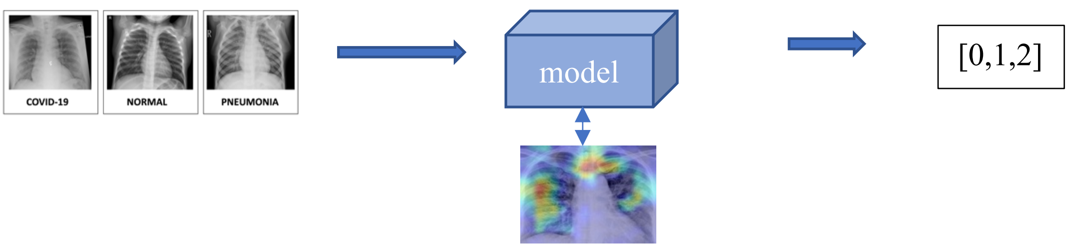


Figure 4: Model training process

For first 5 models which were going to use in our ensembles, we make use of transfer learning methodology which enabled us to adopt a pre-trained model for our current dataset.

[2] https://research.google.com/colaboratory/faq.html

The models used in the ensemble network were as follows: -

- <u>DenseNet</u>

As the name suggests, it has many layers hence the name dense, the speciality about this neural network architectural model is that output from each layer is given as input to every layer coming after it.

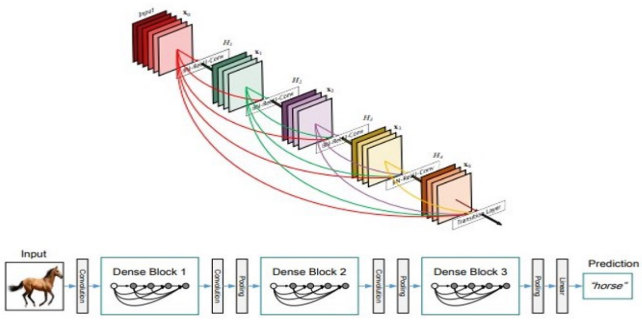


Figure 5: DensetNet Architecture Source: (Gao Huang, 2018)

- <u>Resnet</u>

This is one of the most successful neural network architectures ever built for computer vision problems as it showed that the more layers you add to a network, it is not the case that the performance will increase contrary to the popular belief reason being is that after each layer you have feature extracted which extract and update the weights of the model and downsize further but this can’t be done after a certain period as the input gets very small (K. He, 2016).

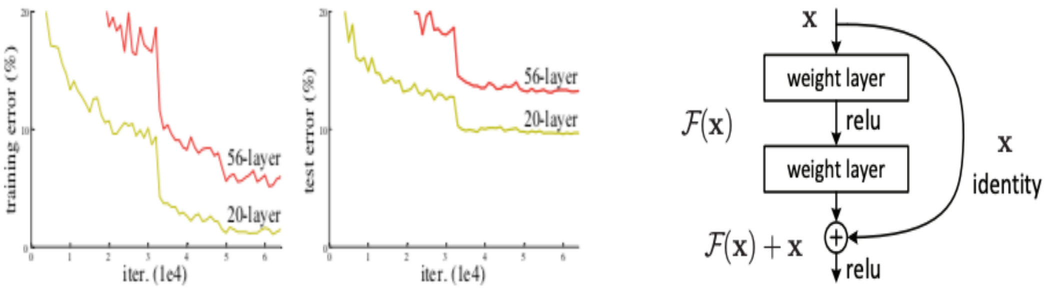


Figure 6: RenseNet Source: (K. He, 2016)

They also introduced a technique called skip layer where in a layer can jump directly by skipping its very next layer and pass on the output to the one after that, this helps to make more complex connections via just a few layers.

- Alexnet

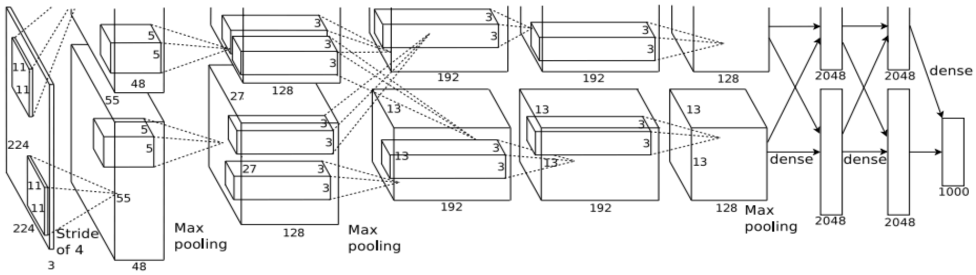


Figure 7: AlexNet Architecture Source: (Krizhevsky, et al., 2012)

As discussed earlier in the literature review section, this network was built by Geffrey Hilton along with his team, it is bigger and deeper than LeNet and includes ReLu activation, Dropout layers etc to avoid overfitting issues. It was also the winner of an Image Classification problem called ILSVRC-2012 (Krizhevsky, et al., 2012).

- VGG16

VGG stands for Visual Geometric Group and 16 is the number of layers present in the model it is also present in a 19-layer variant but for this project we use the 16 layer one. It is bigger and deeper in comparison to Alexnet. This model is also available in keras package and it takes 224 * 224 size of image as input (Simonyan & Zisserman, 2015).

- Xception

This model was inspired by Inception model where in the modules of inception are replaced by depth wise separable convolutions, it has the same number of parameters (Chollet, 2017) like that of Inception V3 but it still slightly outperforms V3 on the ImageNet dataset.

- NASnet

"NAS" stands for Neural Architecture Search, which makes use of reinforcement learning search methodology in order to optimize configuration of the architecture. It is mostly used for object detection. Basically, it is kind of an automated way to create the most efficient model for the dataset we have, each model has a controller which has certain blocks and cells, each block comprises of filters, pooling layers called operators and each cell is a collection of 5 blocks. When the data is feed, the controller generates a child model which is trained on our input data, it tried all permutation and combinations of cells to get the highest validation accuracy and at the end the controller is rewarded. NASnet often gives unimaginable neural architecture and works efficiently for the particular dataset which it was provided (Zoph, 2018).

- <u>MyModel</u>

This model was built from scratch by combining some layers from AlexNet and DenseNet, it was an experimental model, yet it performed as good as the previous ones when trained on the given dataset. Input for this model was standard 224 by 224 with 6 convolutional layers and 2 fully connected layers along with dropouts at each end of the fully connected layers finally with a dense layer having SoftMax as activation function. Also, occasionally pooling layers were added to reduce the spatial dimensionality in order to bring down the parameters for computation in the network.

Class Visualization Feature

As presented in (Selvaraju, et al., 2019), keras the package used to implement the project also provided us an option to visualize the last layers and see which is the region of interest or area in the image based on which the model is making the given prediction. This technique helped us to understand the behaviour of our model and if required, further fine-tune it by optimizing and updating the hyperparameters of that model.

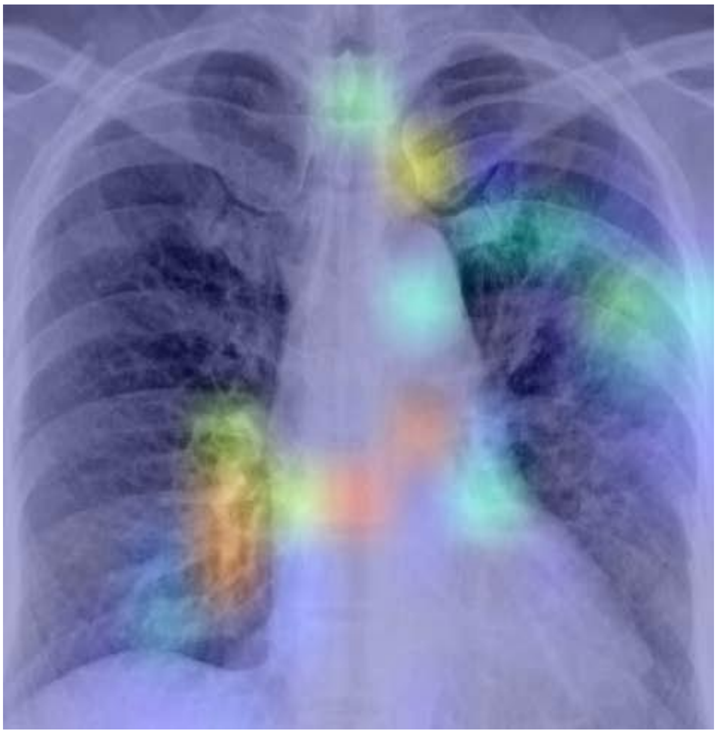


Figure 8: Class Visualization Source: NASNet output

The Figure 8 is the interpretation of NASNet for a covid-19 input image, the model classified it correctly as a covid-19 image and we can also see the regions based on which the model has made such prediction. However, note that ensemble being a collection of models, the layer configuration of each model present in the ensemble network is different hence getting such a mapping for an output from the ensemble would be impossible. For this project since all the models came out to be with good accuracy, we couldn't fully utilize layer visualization to its full power.

Model	Epochs	Implementation Type	Input Size	Accuracy
AlexNet	15	Coded from scratch	[150, 150]	0.9268
Vgg16	25	Transfer Learning	[299, 299]	0.9790
NASNet	15	Transfer Learning	[331, 331]	0.9705
Xception	15	Transfer Learning	[299, 299]	0.9732
DenseNet	25	Transfer Learning	[224, 224]	0.9677
ResNet	25	Transfer Learning	[224, 224]	0.8999
MyModel	15	Coded from scratch	[224, 224]	0.9502

Table 2: Summary table for model training outcome

As given in Table 2 above, the accuracy of each of our 7 models trained on the training set, they were run for the mentioned epochs i.e., number of training time and the size of input image for each model depended on the size of input layer which is the first layer it had. The reason why epochs are different for each model is because it was observed that for some models after a certain number of epochs, the resultant model was overfitting, or it had no accuracy improvement whatsoever and was just consuming GPU time. After the training, each model was saved with **".h5"** extension and downloaded.

Once our models were trained, we passed our test image sets to them individually and output from each model was stored in a list since the test image set had multiple images then they were clubbed to form as a single ensemble output as stated in Figure 9 below.

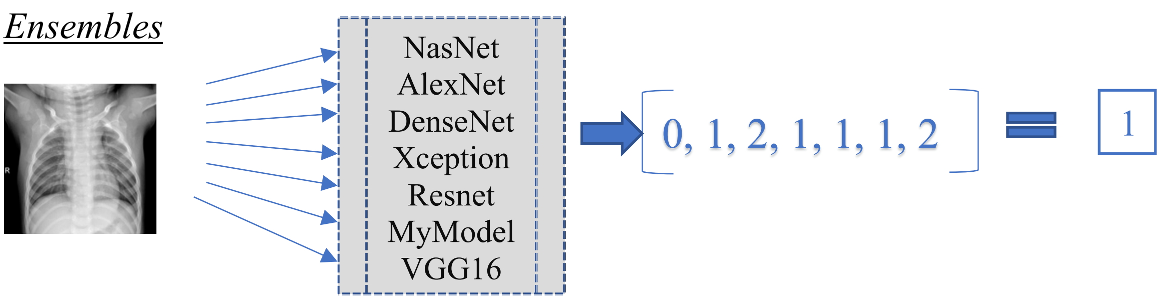


Figure 9: Ensemble Model Layout and poll-based working

As stated in the data preparation stage that each class would be assigned a label as per the index of their occurrences, we will also get the output as class indexes from the ensembles based on the two ways which are polling and weighted polling[3].

[3] https://machinelearningmastery.com/ensemble-machine-learning-algorithms-python-scikit-learn/

Polling

One of the easiest methods to understand and implement is polling/voting based ensembles, here each model in the ensemble gives its output as seen in Figure 9 which is then counted to check which class has been voted the most and the answer to that is considered the final answer. Since each model present in the ensemble can only vote once so in this case, the number of models in the ensemble network plays a very crucial role in order to increase or decrease the overall accuracy of the ensemble as a whole. In some cases, to maintain reliability only the models with high accuracy are appended in the ensemble network, since in our case all models are trained well as in Table 2, we include all the models in the polling-based ensemble network, performance of which is discussed in the evaluation section.

Weighted Polling

In this method of ensemble network, certain models in the ensemble are given multiple chances to vote or check the input image and their output is considered 2x times in count. In our case since models with complex architectures, different input image size and good efficiency like NasNet, Xception and DenseNet were taken into consideration for these perks and hence their predictions were counted twice. This method of weighted polling also known as weighted averaging helps to increase the overall reliability and tends to perform better for very large and complex datasets.

3.4 Evaluation Metrics

Later we evaluated both the methods with popular evaluation metrics for classification problems which are confusion matrix, F1 score and Matthews Correlation Coefficient. Also, the library which was used to implement this called **'sklearn'** had a metrics function to create a report for classification problems which gave the precision, recall, macro average and weighted average.

Confusion Matrix for Polling based Ensemble is plotted in Figure 10 below.

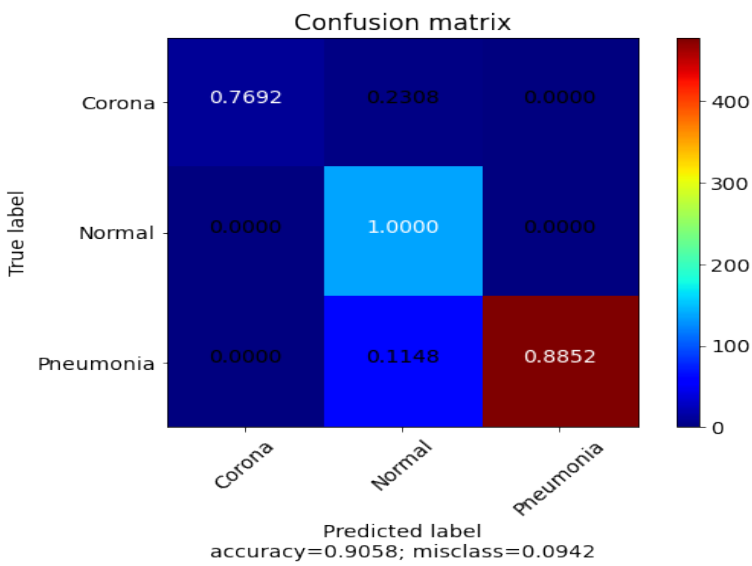


Figure 10: Polling based Ensemble Plot

Around 690 images from all three classes were included for the test, the calculated accuracy for polling-based ensemble is 0.9058, individual F1 score for covid-19, normal and pneumonia is as follows 0.87, 0.81 and 0.94 respectively. F1 score as the formula goes

$$F1 = 2^* \frac{(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})}$$

helps us to strike a balance among Recall and Precision.[4] The Precision and Recall for covid-19 was 1 and 0.77, for normal it was 0.68 and 1.00 and for pneumonia it was 1.0 and 0.89, although the precision for covid-19 and pneumonia was very efficient but for normal it was not up to mark which raises concern.

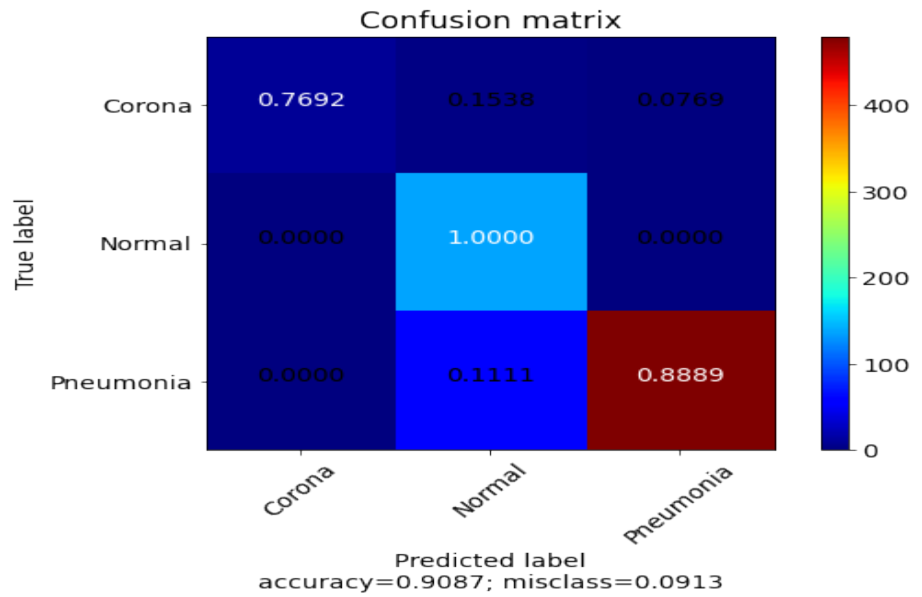


Figure 11: Weight based Ensemble Plot

For our weight-based polling ensemble, the accuracy only improved by a margin of 0.5% as seen in Figure 11 but considering the amount of data we had it was predictable, also since deep neural network tends to improve their performance overtime with new information, with this rate we can expect the weight-based ensemble to outperform the former when trained with more input data.

[4] https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

4 Results and Discussion

In this section we would be discussing the outcome of the project based on the evaluation metrics seen above. Foremost, we loaded the dataset and performed certain image augmentation techniques so as to have good amount of input data to train, the training time for each model used in the ensemble was around 2 hours on an average, considering the accuracies of our models trained individually on the dataset we got most of the models with accuracy of close to 90% and above, this is an indication that our models are able to understand and distinguish the difference between covid-19 and pneumonia in most cases. As seen in the evaluation metrics, our ensemble networks gave an accuracy of about 90%.

Ensemble networks being the novelty of this project over chest x-ray classification for covid-19, pneumonia and normal images we successfully implemented the mentioned approach and built a reliable model. However, since we are working with medical data and considering the fast-changing nature of medical research we (C. Hickie, 2020) might have some ethical concerns which are stated in the coming sections.

5 Conclusion and Ethical Concerns

Automated classification of covid-19 and pneumonia x-rays can help save a lot of time and in return can benefit the patient to recover quickly. To conclude, in this project we successfully created a reliable and one-of-a-kind ensemble network for covid-19 classification to the best of my knowledge as covid-19 being a very new disease there aren't any models with such approach available.

In future, the as more x-ray images related to covid-19 gets available, the models can be retrained to improve the performance further also as mentioned earlier ensemble networks have a reputation of winning image classification competitions hence same approach can be applied to create automated detections for brain and heart related diseases. Even other lung related diseases can be included, and models retrained to make this very ensemble network a generic model for detection of any lung related disease (Rahman, et al., 2020). Considering we were working on medical data, the factor of ethical concerns remains a point as even a small misclassification or error can lead to loss of human life hence results from the automated model must be reported to a doctor who in the end will take final decision as to how to treat the patient for speedy recovery.

6 Acknowledgement

I would sincerely like to thank and convey gratitude to my supervisor Professor Rashmi Gupta for her tremendous support and guidance throughout the project phase, also like to thank other faculties from National College of Ireland who helped me as and when required and authors and researchers who's papers I've referred and cited for this project as development in science and research is an incremental step. I've stood over shoulders of these giants and tried giving my drop in the ocean of scientific endeavours.

7 Bibliography

Wang, X. et al., n.d. *ChestX-rays8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases.* Bethesda, IEEE.

Krizhevsky, A., Sutskever, I. & Hinton, G., 2012. ImageNet Classification with Deep Convolution Neural Networks. *Neural Information Processing Systems,* Volume 25.

Baltruschat, I. M. et al., 2018. Comparison of Deep Learning Approaches for Multi-Label Chest X ray Classification. *Scientific Reports,* Volume 2, pp. 1-11.

Maguolo, G. & Nanni, L., 2020. *A Critic Evaluation of Methods for COVID-19 Automatic Detection from X-ray Images,* Padua: s.n.

Tawsifur Rahman, M. E. H. C. A. K. K. R. I. F. I. Z. B. M. M. A. K. S. K., 2020. Transfer Learning with Deep Convolutional Neural Network(CNN) for Pneumonia Detection Using Chest X-ray. *MDPI*, 6 May, p. 17.

Chollet, F., 2020. *Grad-CAM class activation visualization.* [Online]
Available at: https://keras.io/examples/vision/grad_cam/
[Accessed 10 December 2020].

Gao Huang, Z. L. L. v. d. M. K. Q. W., 2018. Densely Connected Convolutional Networks. 28 January, pp. 1-9.

Simonyan, K. & Zisserman, A., 2015. *VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION.* San Diego, ICLR.

Chollet, F., 2017. Xception: Deep Learning with Depthwise Separable Convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 1800-1807.

Selvaraju, R. R. et al., 2019. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *IEEE International Conference on Computer Vision (ICCV),* pp. 618-626.

K. He, X. Z. S. R. a. J. S., 2016. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 770-778.

Juan Manuel Carrillo-de-Gea, G. G.-M. J. L. F.-A. J. L. H.-H., 2016. A Computer-Aided Detection System for Digital Chest Radiographs. *Journal of healthcare engineering.*

Mohammad Tariqul Islam, M. A. A. ,. A. T. M. ,. K. A., 2017. Abnormality Detection and Localization in Chest X-Rays using Deep Convolutional Neural Networks.

Zoph, B. &. V. V. &. S. J. &. L. Q., 2018. Learning Transferable Architectures for Scalable Image Recognition. *IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 8697-8710.

C. Hickie, C. M. W. T., 2020. Imaging of Covid-19; an Irish Perspective. *Irish Medical Journal,* 113(Ir Med J), p. 47.

Grossi, E. &. B. M., 2008. Introduction to artificial neural networks. *European journal of gastroenterology & hepatology,* Volume 19, pp. 1046-54.

Weiss, K. K. T. &. W. D., 2016. A survey of transfer learning. *Journal of Big Data,* Volume 3, pp. 1-40.

Saraiva, A. et al., 2019. *Models of Learning to Classify X-ray Images for the Detection of Pneumonia using Neural Networks..* Prague, In Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies.

Ayan, E. & Unver, H., 2019. *Diagnosis of Pneumonia from Chest X-ray Images Using Deep Learning.* Istanbul, Electrical-Electronics & Biomedical Engineering and Computer Science.
Joanne Cleverley, J. P. M. M. J., 2020. The role of chest radiography in confirming covid-19 pneumonia. *BMJ,* Volume 370, pp. 1-9.

Pranav Rajpurkar, A. Y. H. M. H. C. B. A. Y. N., 2019. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. *Nature Medicine,* Volume 25, p. 65–69.

Ioannis D. Apostolopoulos, T. A. M., 2020. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine,* Volume 43, p. 635–640.

Rajpurkar. Pranav, I. J. Z. K. Y. B. M. H. D. T. D. D. B. A. L. C. S. K. L. M. N. A., 2017. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.

Saad, M. N., Muda, Z., Sahari, N. & Hamid, H. A., 2014. *Image Segmentation for Lung region in Chest Xray Images using Edge Detection and Morphology.* Penang, IEEE.