An automated approach for classification of Covid-19 patients from chest X ray images using transfer learning through an ensemble of pretrained models

Project & Thesis-I
CSE 4100

A thesis Report
Submitted in partial fulfillment of the requirements for the Degree of
Bachelor of Science in Computer Science and Engineering

Submitted by

A.M Asif Iqbal	**170104040**
Ishraqul Munzerin	**170104068**
Tanima Islam	**170104085**
Faiza Anan Noor	**170104093**

Supervised by
Mr. Emam Hossain

Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Dhaka, Bangladesh

June 27,2021

ABSTRACT

The Corona Virus pandemic is proliferating rapidly and substantially over the whole world since 2019 creating huge havoc, loss of life and damage to the economy. An automated means for predicting the virus is of utmost importance to help the medical personnel to detect patients, prepare reports and produce results fast and impeccably so that people can get early treatment and prevent future transmissions. Chest X-ray images can be used as they are cheap, in comparison to CT-scans, for distinguishing a Covid positive patients from a negative one with the help of different deep learning methods. We performed 1 main experiment and 2 additional experiments for comparison in this study, using 3 different publicly available datasets of Covid positive and Covid negative patients. Firstly, we trained pretrained models InceptionV3, VGG19, Resnet50, Densenet201, on our first dataset "Covid-19 Radiography Dataset" and got accuracy of 94.5%, 94.7%, 61.3%, 93% respectively and then we ensembled these 4 models and got an accuracy of 96%. Secondly, we trained the InceptionV3 and VGG19 models from our first experiment on our second dataset "Covid and normal xray". Thirdly, we analyzed the models trained in our second experiment on our third dataset "Chest-xray-for-covid19-detection". Then we analyzed the results of individual models and compared between their performances on different datasets.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The coronavirus pandemic is one of the greatest challenges that we have faced and it has placed a catastrophic effect on human lives ever since its outbreak. According to the latest estimate of the World Health Organization(WHO), as of July 9th, 2020, the number of cases has been almost 111Million have been infected with this virus and the number of deaths is close to 2.46Million [1].

Due to this pandemic, not only have we faced a tremendous loss of life but it has also led to an unprecedented socio-economic crisis and the aftermath of such are hard to undo. Covid-19 will continue to take a toll on more lives and will go on creating havoc on health care systems,financial markets and industries if appropriate and fast measures are not taken to flatten the steep curve of the pandemic by hindering or stopping its fast growth.

Early Detection of the CoronaVirus plays a notable role in treatment of prevalent cases and for prevention of all future complications. We really need a fast and effective mechanism to detect the virus because the situation is aggravating and the cases are going out of control for mere doctors alone without the aid of a machine that can ease out their task. An affected patient might inflict others because of the slow pace in detection and for lack of sufficient means to cope with newly appearing cases. In many cases there exists a lack of sufficient kits and medical personnel when it comes to handling so many cases. In some cases the medical kits also might be costly and the procedures might be time consuming. So this calls for an automatic disease detection framework which can play vital roles in eradicating the work burden of doctors. With the exponential rate at which the Virus is spreading, it has become highly crucial for an automated approach for detecting and diagnosing Corona in patients efficiently and effectively because otherwise it would cause an excruciating workload for the doctors to deal with the newly emerging cases alone.

1.1 Motivation

Our main motivation for our thesis is to curtail the devastating impact of the virus with the help of automated mechanisms.Our motivation stems from the desire to help doctors and medical staff in easing out and simplifying their tasks of virus detection with accurate and consistent results. Early identification and treatment is our main goal as it will help to slow down and stop the growth of the virus. We want to tackle the virus by providing something that can be used by the doctors and medical personnel for proper screening, tracking and predicting the current and future patients. AI is used for the development of drugs and vaccines, and the reduction of workload of healthcare workers. So our main motive is to detect Covid 19 in patients using their chest x-ray images to assist doctors in diagnosing the disease so that we can get fast and consistent results.

The idea and reason behind implementing an automated covid 19 detection mechanism is to solve problems regarding the following things:

- **Slow and Inaccurate Results**

  In many cases it is seen that the existing mechanisms to detect Covid are time consuming and the doctors can't keep up with the rate at which the patients are being affected because of how fast the virus is spreading.And in many cases there are instances of erroneous diagnosis due to human prone errors and so sometimes it doesn't yield fully accurate results. So we think that an automated based technique can solve these kinds of problems.

- **High cost and availability of medical kits**

  The kits and tools used for diagnosis are mostly not available in many hospitals and might not be sufficient in quantity and in many cases they are deemed to be quite costly [1]. So we opted for creating a system that can use a cheaper, more available alternative that so that's why we thought detection or identification of covid19 using chest X Rays would be the best choice.

- **Elimination of workload and detection time reduction**

  The medical techniques and procedures used to identify corona are very time consuming and since the virus is spreading vastly everyday, doctors are facing a great hassle in coping up with the new cases. Also the results of the manual procedures take some time to reach the patients so in this case one patient knowingly or unknowingly might affect others. So this inspired us to create something that can help in the reduction of the workload because they can produce results in a very short period of time and also ease out tasks for doctors.

- **Absence of trained staffs**

  In many cases hospitals can be burdened with a lack of trained or available staff and that is when an automated tool can be of great use. Giving training to radiologists or hospital staff in such a short amount of time is very unlikely and not practically possible. So this is why we want to implement some kind of automated system that can solve this absence of trained staff.

- **Risk Factor**

  Many people who are involved in Covid situations in the health sector are at huge risk.Especially those who are involved in collecting swab.There is some body contact with the suspected patients.Though they are covered with protective gear but many of them unconsciously put off their protective gears then they can be affected.In that way many Covid warrior infected with this horrible disease. If they will be affected then who will treat or help us in that situation.

- **Human Prone Errors**

  Normally humans are prone or susceptible to different kinds of errors and they take a lot of time in detecting whether or not a patient has been detected when compared with automated machines. For example, if we consider reading a chest x-ray image , it takes at least 6-7 minutes. It is quite impractical to lessen this amount of time for a doctor because this is a very systematic process.So to make sure that we can do fast and accurate predictions we have come up with the idea of our proposed system.

1.2 Objectives

The main objective of this paper is to create a mechanism or technique for fast detection of Corona Virus using automated approaches.The main target was to place a substantial value in the medical field by giving rise to a computer aided analysis which can help doctors detect the disease. We want to create something that can be used by the medical staff for proper screening, analyzing, prediction and tracking of current patients and possible future patients in a fast and efficient manner. Our objective is to build something that is capable of producing the following:

- **Fast and Accurate Results**

  There are many deep learning based techniques that have caused a substantial effect and contribution to accurate and swift diagnosis and identification of diseases in the medical field [2]. We want to minimize the amount of time needed in detection of the virus and at the same time we want to achieve great accuracy in doing so. We aim at creating a Deep learning based technique using CNN or convoluted neural networks because they can be used in extracting features very fast and for identifying between different classes. This also leads to fast results that are quite impeccable too.

- **Cost Reduction and availability of chest X Ray images**

  We are aiming to use chest x ray images because the kits used for diagnosis are mostly not available and in many cases costly. So using chest X ray images of the patients can be helpful in this regard. Mainly Chest x-ray and chest CT are the one of the more efficient imaging techniques for diagnosing medical issues concerning lungs. A chest x-ray has a much lower cost process in comparison to chest CT and hence we want to use chest X-ray for our research in detecting covid 19.

- **Elimination of workload and detection time reduction**

  We believe that our proposed technique will help the doctors to a great extent and will help them in dealing with the emerging new cases everyday as manual based techniques are quite time consuming and burdensome. So the assistance of an automated system can be of great help in eliminating the workload because they can produce results in a very short period of time.

- **Solving the absence of trained staff**

  In many cases hospitals can be burdened with a lack of trained or available staff and that is when an automated tool can be of great use.

  Deep Learning Convoluted Neural Networks(CNN) has played a tremendous role in the medical field for extracting features and detection of cases or identification of diseases. There are many deep learning based approaches which have been followed previously and paved the way for future work which we can implement in order to fulfill our motivation and that constitutes our objective.

Chapter 2

Literature Review

2.1 Related Works

In Paper [3], the main goal of Md. Zabirul Islam, Md. Milon Islam and Amanullah Asraf. was to detect Covid 19 in patients using their chest x-ray images to assist doctors in diagnosing the disease so that we can get fast and consistent results. A deep learning based system was made using LSTM, which has the capability of learning from experiences and could store them and CNN which is capable of learning complex features. Preprocessing of images was done to eliminate noise to a great extent. They measured the loss and accuracy after each epoch and the following performance metrics were used: F1. CNN was used as a feature extractor and LSTM was used as a classifier for those features which could then distinguish between covid and pneumonia accurately. The CNN-LTSM architecture produces faster ,more accurate results than traditional mechanisms. The AUC for the system was 99.9% whereas for plain CNN architectures it was 99.8%. The dataset was collected from various sources like Github, radiopaedia, TCIA,etc. The model for this paper had an approximate 99.4% accuracy, specificity of 99.2%, sensitivity of 99.3%, F1-Score of 98.9% .

The main objective of Khandaker Foysal Haque; Fatin Farhan Haque; Lisa Gandy and Ahmed Abdelgawad for paper [4] was to save time, money and energy of the doctors by using CNN based positive cases of covid 19. This adds up an extra layer of validation due to the fact that it mostly guarantees 100% accuracy. The images for training, testing were mainly collected from kaggle and github repositories at kohen et al. Then they were divided into training and testing images in a ratio of 80% and 20% and they were all in post anterior view. For Standardization ,they were converted to 224x224 pixels. 4 main components of this model were : input layers, convolutional layers, fully connected layers,output layers. The tuned and preprocessed data set was fed into the input layers of the model and then the model was used to differentiate between positive and negative cases with high precision and accuracy. Furthermore, overfitting of the models was avoided by using data augmentation. This basically includes random cropping and random horizontal flipping. The designed model was trained for 25 epochs with 10 steps per epoch with the training dataset and this model was referred to as Model 1. For comparative analysis, they designed and trained two more models with three and five convolutional layers. These two models were referred to as Model 2 and Model 3 respectively. Sufficient comparison studies were conducted and they show that model 1 showed better F1-score and overall performance than the other two. The model for this research resulted in an accuracy of 97.56% and a precision of 95.34% along with an ROC curve area of 0.976% and F1-score of 97.61%.

In paper [5], Asmaa Abbas, Mohammed M. Abdelsamea and Mohamed Medhat Gaber used DeTraC which is short for Decompose,Transfer, and Compose, for the classification of COVID-19 chest X-ray images. The main objective of this paper was to use CNN architecture to detect COVID19 in patients and to provide an effective mechanism to slow down and control the spread of the disease. In this research a combination of two datasets were used. 80 samples of normal CXRs were collected from the Japanese Society of Radiological Technology and another image set contains 105 and 11 samples of COVID-19 and SARS respectively. Then different preprocessing and augmentation techniques were applied like flipping up, flipping down and right/left. Translation and rotation was done using random five different angles. The dataset was divided into 2 parts, 70% for training the model and 30% for testing. An ImageNet pre-trained network called ResNet18 was used in this experiment. ResNet18 is composed of 18 layers. The input image size was 224×22. Six classes could be identified using the last fully-connected layer.The learning rate or Lr for all the CNN layers was fixed to 0.0001 except for the last fully connected layer. For this the Lr was 0.01 with the purpose of accelerating the learning procedure. Here the batch size was 64 and it had a minimum of 100 epochs.In Order to prevent overfitting, 0.0001 was set for the weight decay.The momentum value was 0.95 and after every 5 epochs,the schedule of drop learning rate was set to 0.95 .The model had a high accuracy of 95.12%, a sensitivity of 97.91%, a specificity of 91.87%, a precision of 93.36%.

A new classification technique based on CNN was proposed by Saddam Hussain Khan, Anabia Sohail and Asifullah Khan in paper [6] for automated Covid-19 detection from chest X-rays. The main objective is to build the fastest way for early detection of infected individuals. In this paper they developed a new CNN model which can detect COVID-19 patterns in a very systematic way. The model developed named “STM-RENet” is based on the novel split-transform-merge (STM). For training and evaluation of these models, a Holdout cross-validation scheme is used. 3 publicly available datasets of chest X-rays are assembled and named as CoV-Healthy-6k, CoV-NonCoV-10k, and CoV-NonCoV-15k, respectively. In these datasets, 80% is used for training and validation of the model whereas 20% for testing the model. The 20% test set was used for the final evaluation. They evaluated the two proposed “STM-RENet” techniques (both with and without Channel boosting) and compared the results. The training of models was done throughout the whole process. Moreover, cross-entropy loss was also reduced using SDG. MATLAB 2019b is used in building CNN models, and DL library is used for performing all kinds of simulations. Comparison between the proposed technique and the new technique, “STM-RENet” and “STM-RE-CBNet” respectively is done. The result shows how effective the discrimination ability of the new proposed technique is on the basis of accuracy, MCC, and F-score.

In paper [7], a new method has been developed by Terry Gao using deep convolutional neural network (CNN) to separate the pneumonia patients who are infected with covid-19 from the pneumonia patients who are not infected with covid-19 by analyzing their chest x-ray images. They have designed a VGG-19 based convolutional neural network. Here the convolution layers, the elementary units of this network maintain a hierarchical structure. The feature maps generated by the input space are considered as inputs for the following convolution layers. The set of feature maps generated by the convolutional layers are further fed to the pooling layer where ReLU, an activation function, is applied to filter out relevant features.To regularize the model they have used a dropout factor of value 0.5 to their dropout layer. Then these relevant features are transformed into the feature vector, which is a one-dimensional array. With the help of this feature vector the fully connected layer generates the output of the CNN. Their dataset consists of a total number of 1400 chest x-rays as training set, and 400 chest x-rays as test set. Training set images are distributed into four categories, such: normal images, pneumonia infected by bacteria images, pneumonia infected by other virus images, and the last one is, pneumonia infected by COVID-19 images. First three categories have 400 images each, and the last category, which includes covid-19 patients, has 200 chest x-ray images. For the test set, it has 100 images of each category. Furthermore, for the validation purpose they used 100 images, amongst those 50 images were COVID-19 infected x-rays. The precision, recall, and F1-score of their CNN model on pneumonia patients infected by covid-19 are 0.95, 0.98, and 0.97 respectively. On the other hand, the precision, recall, and F1-score on pneumonia patients infected by normal viruses are 0.98, 0.96, and 0.97 respectively. Lastly, on normal patients, the precision, recall, and F1-score of this model are 0.97, 0.94, and 0.95 respectively.

In paper [8], Prabira Kumar Sethy , Santi Kumari Behera , Pradyumna Kumar Ratha , Preesat Biswas for the detection of coronavirus infected patient support vector machine (SVM) used . Deep Learning based classifiers need a large dataset for training and validation. So, Deep feature and SVM are suggested as an alternative for COVID-19 detection. ‘Novel’ is used as a prefix of coronavirus as it is a new strain of virus, which is spreading really fast all over the world. At present, identification of coronavirus has become the most essential task. Detection of COVID-19 to stop the regular multiplying deaths and spread of this virus is the main motivation behind this paper. Deep feature extraction is done from different 13

CNN models from a particular layer. Then classification of the deep features is done by the SVM classifier and the performance is measured. The function 'fit class error-correcting output codes ('fitcecoc') is used for the training of SVM. The response of every layer of CNN is collected in a detailed way. In this paper a difference has been created by using deep features and SVM for the classification of COVID-19 instead of deep learning based classifiers. The accuracy, sensitivity, FPR and F1 score achieved by RsNet50 plus SVM are respectively 95.33%,95.33%,2.33% and 95.34%.

The perspective of Linda Wang, Zhong Qiu Lin Alexander Wong for study [9] is to introduce the first open source network design using deep convolutional neural networks for COVID 19 detection using chest X Ray images at the time to the general public . Radiography examination can be conducted faster and is a standard procedure for respiratory patients, so to develop the solution urgently a DNN architecture to detect COVID-19 lightweight projection-expansion-projection-extension (PEPX) design as it enhances representational capacity rather than computational complexity. In this study ,a human- machine collaborative design strategy combining human driven principle network machine -driven design exploration and a COVIDx dataset were proposed .The proposal was pretrained on ImageNET dataset trained on COVIDx dataset using Adam optimizer.The dataset tested all DNN architecture for which images were cropped to 8%. Later on,the test accuracy,sensitivity and positive predictive value were investigated.In the process CoVID-Net achieved 93.3% accuracy to detect covid case with 98.9% PPV It can also leverage visual indicators in its decision making process because of its transparency new insight discovery . The researchers investigated how COVID-Net makes predictions using an explainability method in an attempt to gain deeper insights into critical factors associated with COVID cases, which can aid clinicians in improved screening as well as improve trust and transparency when leveraging COVID-Net for accelerated computer-aided screening.

In paper [10], to save the people from the threat of Covid 19 Xavier P. Burgos-Artizzu distinguished between affected and normal people by using Polymer Chaining Reaction. It works with the nasal swab and if there is a little presence of that virus it results positively. Though it is the most accurate system, it is costly and lengthy as well.It is not available at all the places.We can also detect Covid by using radiation like MRI,CXR etc.But we should use CXR techniques. It is the most common radiative technique.Though it is comparatively less effective than PCR.If we manually detect the affected one by seeing their CXR it will be a very lengthy process. So we can make a system which automatically detects Covid by image processing and a deep learning process.

At first they took a data set which is equally distributed by normal data and affected data.Then they train the system with some data and then test the system with some data. If they select a feasible algorithm then they can get the highest accuracy. If we get the highest accuracy then we can use it . By using that we can distinguish normal and affected ones.They work with covid, bacterial pneumonia and healthy people.Their result is 98.6% accurate (detection rate of 91.8% at 1.1% false positive rate). So,We can work with their results.

The main idea of Amit Kumar Das, Sayantani Ghosh,Samiruddin Thunder, Rohit Dutta , Sachin Agarwal and Amlan Chakrabarti of paper [11] was to help medical personnel in detecting corona through an automated mechanism using the help of chest X ray images. The main objective of this paper was on site and fast detection of CoronaVirus using automated approaches. In this paper a GUI was also made which can take in image inputs directly and detect the virus on spot. The main target was to place a substantial value in the medical field by giving rise to a computer aided analysis which can help doctors detect the disease. Deep Learning based solution and CNN was used in this case where a number of positive and negative cases of X Ray images were taken and they were further divided into testing and training set for training our CNN model. Here we have ensembled 3 deep learning models which are: DenseNet201,Resnet50v2 and Inceptionv3. Each of them worked independently and produced their own predictions, then the models were combined and the weighted average was done to predict a class. Here they collected the images from various sources and then only retained the frontal portion since lungs can't be examined that well when viewed laterally. Then they did appropriate resizing and then the images were divided into testing and training sets and the models were trained using these. After that the models were made to predict outcomes by applying them on test images. Lastly, the weighted average of the predictions made by the 3 models was found and a class value was assigned. Here 3 state of the art deep learning models were used and ensembled in classifying between Positive and negative cases and 98 out of 100 patients were found to be measured accurately using CXR(chest X Ray or chest radiograph). So for this paper, the accuracy was 95.7% and sensitivity was 98%. Even without human intervention this would give nearly full accurate results.

In paper [12], a classifier called DarkNet, and a real time object detection system named YOLO, were used by Tulin Ozturk, Muhammed Talo ,Eylul Azra Yildirim, Ulas Baran Baloglu,Ozal Yildirim and U.Rajendra Acharya to build a model that can do both binary classification (COVID vs. No-Findings) and multi-class classification (COVID vs. No-Findings vs. Pneumonia) by analyzing chest X-ray images. They named their architecture as DarkCovidNet, which has 17 convolutional layers, and it's based on Darknet-19. Darknet-19 is a classifier model with 19 convolutional layers and five pooling layers. Darknet-19 uses a real-time object detection system named You-Only-Look-Once (YOLO) as its backbone network. The darknet (DN) layer does 2 major operations, and has one single convolution layer. One of those operations is the batch normalization operation which makes the model more stable, and another one is LeakyReLU operation that is a variation of the ReLU operation which prevents the neurons from dying. Moreover, in their model they have downsized the inputs by using a Maxpool method. The dataset researchers used in developing their model had

1125 chest x-ray images, among those 125 chest x-rays were from covid-19 patients, 125 chest x-rays were from pneumonia patients, and another 125 chest x-rays were from normal patients. DarkCovidNet achieved an accuracy of 98.08% on binary classification, and an accuracy of 87.02% on multi-class classification.

The objective of Ali Narin, Ceren Kaya and Ziynet Pamuk for study [13] is to implement an automatic detection system for Covid-19 as a quick alternative diagnosis option to prevent it from spreading rapidly among people by using chest X-ray radiographs. To ease the process And decrease the death rate of people make doctors ,nurses, caregivers out of risk Computer-Aided diagnosis(CAD) introduced diagnosis of pneumonia from chest Xray more quickly accurately using Deep learning medical Imaging .It helps to cope with datasets exceeding human potential workload of medical persons. For a healthy differentiation on COVID 19 detection pre-trained transfer models has been adopted using small image datasets .For Healthy chest X-ray image "ChestX-ray8" database and for bacterial vs viral pneumonia chest X-ray image "Chest X-Ray Images (Pneumonia)" has been used .Along with that InceptionV3, ResNet50, ResNet101, ResNet152 and Inception-ResNetV2. While 80% of the data is reserved for training, the remaining 20% is reserved for testing. For the cross validation method, they chose the K-fold method and results were obtained according to 5 different k values .It was realized end-to-end directly with raw data. After applying all the model they have decided that by yielding ResNet50 pre-trained model accuracy can be gained upto 96.1% for dataset-1 ,99.5% for Dataset-2 99.7% for dataset-3 even after working with very limited data.Their study also showed that the classification performance of different CNN models can be tested by increasing the number of COVID-19 Chest X-ray images in the dataset. As a result it will help to detect COVID 19 at an early age though they are hoping that in future the features will be extracted using image processing methods on X-ray CT images.

The objective of Polat, Çağín Karaman ,Onur Karaman, Ceren | Korkmaz, Güney Balcı, Mehmet Can Kelek and Sevim Ercan for study [14] is to implement a debiasing-data-loader approach with the help of nCoV-NET transfer learning model to avoid bias problems ,reduce possible spreading risk of virus obtain high accuracy value of detecting COVID-19 cases from Chest x-ray images. To train evaluate the system "Chest X-ray 14","COVID-19 image data collection" "Chest X-ray collection from Indiana University" were used. To solve the probable bias problem due to the unbalanced cases in two classes of the datasets, ResNet, DenseNet, and VGG architectures were re-trained in the fine-tuning stage of the process to distinguish COVID-19 classes using a transfer learning method. During the process previously trained datasets using ImageNet were frozen only fully connected layers were re- trained. In the backpropagation process the "stochastic gradient descent with restarts" technique is used to determine the minimum points on the loss function more accurately ."Cosine annealing "method was repeatedly used to update the learning rate value of iterations progress. Lastly, the optimized final nCoV-NET model was applied to the testing dataset to verify the performance of the proposed model. In the evaluation process ,all CNN architecture s found same 8 data of COVID-19 as false negative.Due which all of them gave a close result . But in the process the performance of fine-tuning : learning stage 1 2 and of NCoV-Net were calculated using VGG-16,DenseNet-161 Res-Net-50 architecture in Accuracy ,Recall precision metrics. Among all architectures ,Dense-Net -161 showed the highest accuracy of about 82% in all performance metrics.

Antonios Makris, Ioannis Kontopoulos and Konstantinos Tserpes's objective for study [15] is to evaluate the effectiveness of several state-of-the-art pre-trained convolutional neural networks regarding the automatic detection of COVID-19 disease from chest X-Ray images. The used data set in this system is a combination of Chest X-ray/CT images of confirmed COVID-19,ARDS,MERS,SARS pneumonia datasets, Dr. Joseph Cohen's Github repository ,112 Posterior-Anterior (PA) X-ray images of lungs .For feature extraction an arbitrary feature extractor was employed on the top of pre-trained model for new samples. In addition, fine-tuning was used to increase the performance weights, identify classify different classes of covid ,pneumonia normal cases using VGG16 network architecture. To evaluate the result precision,recall,F1-Score ,specificity metrics were developed. In all classes VGG16 VGG19 showed upto 98% 100% COVID detection rate respectively for all metrics whereas Inception V3 DenseNEt201 showed the lowest rate of sensitivity with 4% 12% . In this study it was demonstrated that despite having shortage in the number of available data ,VGG16 VGG19 methods showed encouraging results to achieve high sensitivity precisions which can be used directly with help of clinical diagnosis.

In paper [16], to differentiate affected people nowadays Sifat Ahmed, Tonmoy Hossain, Oishee Bintey Hoque, Sujan Sarker, Sejuti Rahman, Faisal Muhammad Shah has used the Polymer Chaining Reaction test.But PCR test is lengthy and costly as well. There are kinds of barriers to differentiate the affected people because this process is lengthy. To prevent the scattering of Covid we should break this length barrier. For breaking that we can use CX-ray for detecting Covid-19. First of all we take a computed CX-ray image and then We'll analyze the image by machine learning.This technique is less costly and too fast compare to PCR lab testing. But There are some difficulties in that system.There are some lung diseases like bronchitis,pneumonia,tuberculosis etc.We have to differentiate between them and Covid. We will try to solve this problem. If we can solve that problem then we will able to stop the scattering of Covid rapidly.

In paper [17] ,the main objective of Boran Sekeroglu and Ilker Ozsahin was early detection of Covid in cases where laboratory kits were in many cases costly,insufficient in quantity and also not often quite accurate in coping with the large number of cases arising everyday. The main purpose was served by conducting different kinds of experiments like by using machine learning models,pre-trained networks for transfer learning,CNN etc.In this paper CNN was used without pre-processing and mainly minimized layers played the part of detecting COVID-19 in imbalanced, chest X-ray images. In this paper different kinds of methodologies and experiments were followed, some of them being10 experiments using 5 machine learning models, 14 or more experiments using state of the art transfer learning networks,38 experiments using Convolutional Neural Networks. The experiments overall results were of mean sensitivity of 93.84%, mean specificity of 99.18%, mean accuracy of 98.50%,AOC scores of 96.51%. In this study, several experiments were performed for the correct detection of CoronaVirus in chest X-ray images using ConvNets. Main combinations used for classification were:COVID-19/Normal, COVID-19/Pneumonia,COVID-19/Pneumonia/Normal. Different image dimensions and network architectures, state-of-the-art pre-trained networks, and machine learning models were manifested and tested using the gathered images and statistical data.

Rahul Kumar,Ridhi Arora,Vipul Bansal,Vinodh J Sahayasheela, Himanshu Buckchash,Javed Imran, Narayanan Narayanan, Ganesh N Pandian and Balasubramanian Raman used ResNet152 in paper [18] for the classification of COVID-19 patients on chest X-ray images. Moreover, for the balance of data points of COVID-19 and Normal patients SMOTE algorithm is encompassed. The model also uses predictive machine learning classifiers for effective classification of COVID-19. In this paper they used two datasets. First one has chest X-ray images of covid patients, pneumonia patients and normal patients. Second one has chest x-rays of only pneumonia patients and normal patients. The first dataset, which includes the x-rays of covid patients, has a total number of 2748 images of which 1833 were in the training set, and other 915 images were in the test set. The second dataset contains 5840 x-rays, where the training set has 5216 images. The proposed model was applied on these two datasets. All the chest X-ray images were resized into 224x224 sized resolution. Then feature extraction is done on these resized X-ray images. At first, the training of ResNet152 is done for the classification of COVID-19 and pneumonia patients. Then, the last three layers of ResNet152 are replaced by ReLU, an FC layer, and an output layer. Now, to balance the imbalanced data points of the classes synthetic minority oversampling technique (SMOTE) is used. Here SMOTE plays a significant role because in the datasets the class of COVID-19 patients has fewer data points comparing other classes, and SMOTE brings symmetry between the classes. As a result, an equal number of samples are created for each class. Then the dataset is fitted into different machine learning predictive classifiers. The different machine learning predictive classifiers Logistic Regression, Nearest Neighbors, Decision Tree, Random Forest, AdaBoost Classifier, Naive Bayes, and XGB Classifier has accuracy of 96.6%, 94.7%, 93.1%, 97.3%, 92.1% and 88.9%, 97.7% respectively. Among these, Random forest and XGB classifiers have the highest accuracy.

Table 2.1: Summary of Related Works

Paper Reference	Dataset	Methodology	Result
[3]	NIH dataset & images from different websites	CNN- LSTM, CNN	Accuracy:99.4%, Specificity : 99.2%, Sensitivity :99.3% F1-score : 98.9%
[17]	Covid-19 X-ray image collection,images from Cohen &Kermany.	APPN & ReLU approach , ConvNet, Transfer learning	Accuracy : 98.50% Specificity : 99.18% Sensitivity :93.84%, AOC scores : 96.51%.
[11]	CheXpert, Chest-Xray-pneumonia, Covid-19 image data collection, Covid-Chest-X-ray dataset	DenseNet201, Resnet50v2 and Inceptionv3	Accuracy: 95.7% Sensitivity : 98%.
[4]	Chest X-ray Images (pneumonia), different chest X-ray images	CNN	Accuracy : 97.56% , Precision : 95.34% , ROC curve area : 97.6% , F1-score : 97.61%.
[5]	Lung Segmentation in CXR, Automatic Tuberculosis screening, covid-chestxray-dataset	DeTraC Architecture, CNN, ResNet18	Accuracy : 95.12%, Sensitivity : 97.91%, Specificity : 91.87%, Precision : 93.36%.
[6]	CoV-Healthy-6k, CoV-NonConV-10k, CoV-NonConV-15k	Deep CNN, Channel Boosted STM-RENet	Accuracy: 96.53%, F-score : 95%, MCC : 93%, AUC : 98%
[7]	The new England Journal of Medicine, Kaggle	CNN, Bonferroni correction, VGG-19	Precision 95%, Recall 98%, F1-score 95%

Table 2.1: Summary of Related Works

Paper Reference	Dataset	Methodology	Result
[18]	Chest X-ray Images (pneumonia), COVID-19 public dataset from Italy	SMOTE, ResNet 152	Accuracy : 97% Logistic Regression 96.6%, Nearest Neighbors 94.7%, Decision Tree 93.1%, Random Forest 97.3%, AdaBoost Classifier 92.1%, Naive Bayes 88.9%, XGB Classifier 97.7%.
[8]	Covid-chext-xray dataset, conVid19-X-rays	CNN, ResNet50, SVM	Accuracy :98.66% Sensitivity:95.33% F1 Score:95.34% FPR:2.33%
[12]	X-ray image dataset, Chest Xray8 database	DarkCovidNet	Accuracy:98.08%, Specificity:95.3%, Sensitivity:95.13%, Precision:98.03%, F1-Score:96.51%.
[13]	Chest X-ray 8 database, Chest X-Ray Images (Pneumonia)	InceptionV3, ResNet50, ResNet101, ResNet152, Inception-ResNetV2	Accuracy : Dataset-1 : 96.1% , Dataset-2 : 99.5% , Dataset-3 : 99.7% .
[9]	COVIDx, Chest CXR images, COVID-Net github respiratory	VGG-19, ResNet-50, CNN	Accuracy : 93.3%, PPV : 98.9% .
[14]	ChestX-ray14, COVID-19 imagedata collection, Chest X-ray collection from Indiana University.	ResNet, DenseNet, VGG architectures , nCoV-NET , Grad-CAM	Accuracy: 97.10%

Table 2.1: Summary of Related Works

Paper Reference	Dataset	Methodology	Result
[15]	Dr. Joseph Cohen's Github repository, chest-xray-pneumonia	VGG16, VGG19, MobileNet V2, Inception V3, Xception, InceptionResNet V2, DenseNet201, ResNet152 V2 and NASNetLarge.	VGG-16 ; VGG-19 Accuracy: 95.88% ; 95.03%, Specificity: 98% ; 100%, Sensitivity: 96% ; 92%, Precision: 96% ; 100%, F1-Score: 96% ; 96%;
[16]	X-ray images (chest), SIRM COVID-19 Database, nCOVID-19 Dataset, Chest X-Ray Images (Pneumonia).	ResNet-50	Accuracy : 96.9%, Specificity : 100%, Sensitivity : 100%, Precision : 100%, F1-score : 100%.
[10]	COVID chest x ray dataset COVIDEEP, ChestX-ray8database,	DNN, SqueezeNet, Inception-V3, VGG16, MobileNet, Xception, VGG19+MobileNet	Accuracy:90%, Precision:100%
[19]	(RSNA),Radiopaedia, and (SIRM)	HRnet	Accuracy : 99.26%, Sensitivity: 98.53%, Specificity : 98.82% .
[20]	U.S. National Library of Medicine Tuberculosis Datasets, TB_portals	CNN	Accuracy : 98.6% , AU : C90%

2.2 Used Tools

2.2.1 Convolutional Neural Network (CNN)

In the field of image processing Convolutional Neural Network (CNN) is vastly used. In modern day deep learning models which are used for image classification are mostly built upon convolutional neural networks. In CNN the input image has to go through a set of convolution layers followed by pooling layers and fully-connected layers, then an activation-function, such as softmax or sigmoid, is applied to the image to perform image classification. The convolution layers are the very basic of CNN. These layers are used for feature extraction, and to apply filters to the input image. Pooling layers are used to reduce the total number parameters from the image and create feature maps. There are different kinds of pooling, such as: max pooling, average pooling and sum pooling. The fully-connected layers are some of the last layers of a CNN. The output we get from pooling layers or convolutional layers, is used as the input to the fully-connected layers. Lastly, applying an activation function we get our final classification output.

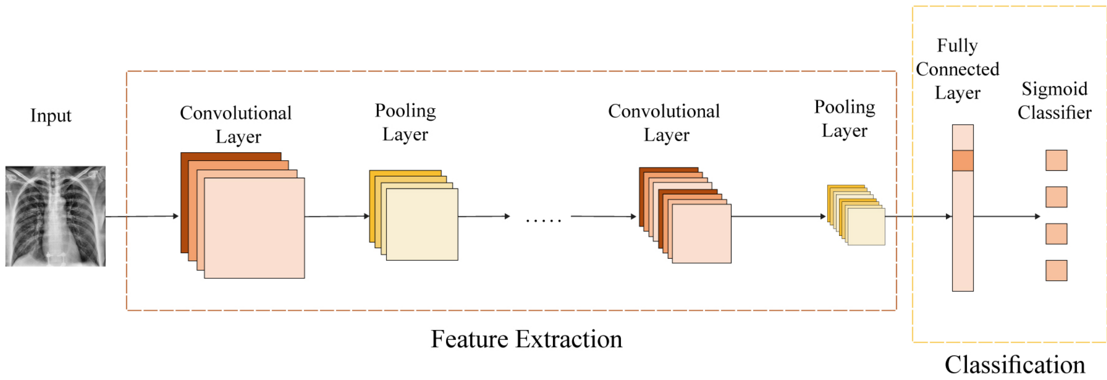

Figure 2.1: Architecture of CNN

2.2.2 Transfer learning

To build our model we used transfer learning. We used a model that was pre-trained on the 'ImageNet' dataset, and we modified the last layer of that model, so it can easily align with our purpose. Now-a-days transfer learning approach is becoming quite popular as it saves training time and researchers don't need to make models from scratch.

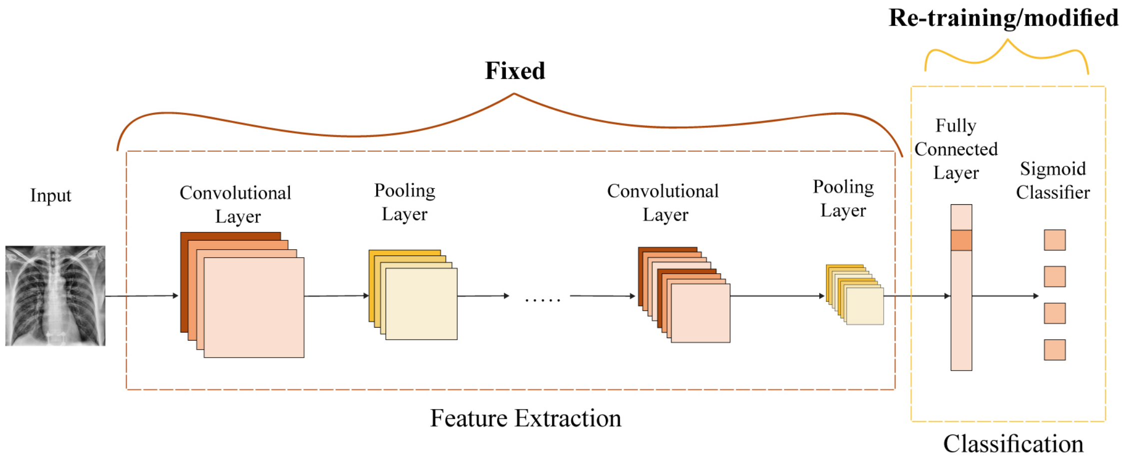

Figure 2.2: Architecture of Transfer learning with CNN

Chapter 3

Methodology

For our research we are going to use the steps : image acquisition,image processing,Feature extraction,training and testing of the datasets and models, detect and classify the models to detect the disease accurately and classify accordingly.

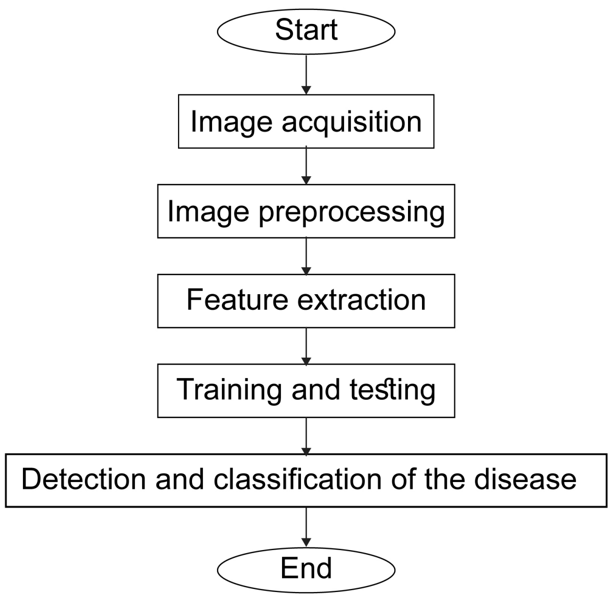

Figure 3.1: Flow chart of methodology

The system will mainly consist of the following steps:

- **Image acquisition**

  The images or dataset needed for training and testing on our system will be mainly chest x ray images collected from patients.

- **Image preprocessing**

  This method will process our dataset and images and convert it into a suitable form on which different algorithms can be trained and tested upon. This includes resizing, rescaling, zoom ranging ,shuffling etc This step is done to reduce the noise and to make the images suitable for training our model. We transformed all of our image data into 299x299x3 in order to feed it into our model.

- **Feature extraction**

  We used several pre-trained CNN models training and testing. CNN helps to extract the basic features first by learning the basic shapes in the starting layers. After that it learns more advanced features as it goes deeper into the network. This will help to cause a much better image classification.

- **Training and testing**

  Our images from our first dataset were divided into training and testing sets for the training of our model and for testing whether our model performs well or if it gives correct predictions.For this particular dataset, 80% of the images were used for training purposes, 10% were used for validation and the remaining 10% were used for testing. The second dataset was already found to be divided into a ratio of 0.8: 0.2 for training and testing images. And the third dataset was only used for testing for analyzing the performance from the models trained from the second dataset.

- **Detection and classification of the disease**

  This step is the ultimate step and it plays the actual part of detection and classification of the virus. It searches out and classifies or identifies between positive and negative cases of covid19 in patients. We did the detection and classification using different kinds of models like InceptionV3, Resnet50, VGG19, Densenet201 etc and also ensemble the results of the models. We also compared and contrasted between different datasets for models InceptionV3 and VGG19

3.1 Datasets used

We used 2 datasets for our training purpose and testing purposes which are off the shelf and are collected from kaggle which are [21] and [22]. We also had one additional dataset [23] which was used just for testing purposes only.

3.1.1 Covid-19 Radiography Dataset

The first dataset we used for training was the "COVID-19 Radiography Dataset" or [21]. This database consists of 3616 COVID-19 positive cases along with 10,192 Normal, 6012 Lung Opacity (Non-COVID lung infection) and 1345 Viral Pneumonia images. For our purpose we needed the Normal images and the Covid positive images. But since there exists an imbalance in the dataset as the amount of Normal images is almost thrice than that of Covid positive, so we took 3616 images from the Normal Image directory only. Among these we only used 3616 Normal images and the Covid positive images. We hope to fix this imbalancing problem in our future works.

The data distribution for this dataset is summarized below:

Table 3.1: COVID-19 Radiography Dataset

	Covid + images	Covid - images
Original Dataset	3616	10200
Minimized Dataset	3616	3616

3.1.2 Covid and normal xray v1

The second dataset we used was "covid and normal xray v1" or [22].This dataset consisted of two directories called Train and Test. The training set consisted of 1073 Covid -ve images and 960 Covid +ve images. The testing set consisted of 268 Covid -ve images and 240 Covid +ve images.

Table 3.2: Covid and normal xray v1 Dataset

	Total	Covid + images	Covid - images
Training Set	2033	960	1073
Testing Set	508	240	268

3.1.3 Chest-xray-for-covid19-detection

The third dataset is called “chest-xray-for-covid19-detection” or [23]. This originally has 3 folders called Prediction, Train and Val (Validation) containing 23, 288 and 60 photos respectively and they were divided equally for the classes “Covid” and “Normal” But we used the Validation set of this dataset to see how well our trained dataset worked.

Table 3.3: Chest-xray-for-covid19-detection Dataset

	Total	Covid + images	Covid - images
Testing Set	60	30	30

After collecting our datasets, we applied different kinds of preprocessing on the images using keras DataGenerator. We did resizing, shuffling, rescaling,horizontal flipping,shear and zoom ranging.

3.2 Model Formation

We performed mainly 3 kinds of experiments. The first experiment was conducted on Dataset 1 and it involved the making use of transfer learning using the models Inceptionv3, VGG19, Resnet50 and Densenet201. Then we ensemble the results of the models by using weighted average and then it surpassed the overall performance of the model. The second experiment was done on the Dataset 2 on which we trained our Inceptionv3 and VGG19 models prepared earlier and then the validation was done on this dataset. The third experiment included validating the results from the model trained on dataset 2 on the validation set of dataset 3.

3.2.1 Model formation for experiment 1

For the first experiment We took our first dataset called "Covid-19 Radiography Dataset" and then split the Covid +ve and Covid -ve image folders into Training, Testing, and Validation images using a ratio of 0.8:0.1:0.1. So our main image folder consisted of 7232 images in total and 3616 images were from Covid +ve patients and 3616 images were from Covid -ve patients. Then the 3616 for each of the cases were subdivided into 2158 training, 736 testing, 722 validation images. The summarization of the splitting of the first dataset:

Table 3.4: Dataset Information for experiment 1

	Training Set	Validation Set	Testing Set
Dataset 1	2158	722	736

Then we used transfer learning using pretrained models like inceptionv3, vgg19, densenet201 and resnet50. For training the models we used 40 epochs for Inceptionv3, 55 epochs for resnet50, 15 epochs for vgg19 and 40 epochs for densenet201 The optimiser used for each of the models is the Adam optimizer and then the models are saved as .pb files after every epoch. The time taken for model training are - 1234 seconds / epoch for Resnet50, 115seconds / epoch for Inceptionv3, 5667 seconds / epoch for VGG19 and 127 seconds / epoch for DenseNet201.

The total time needed for each model is summarized below:

Table 3.5: Time needed for training of experiment 1 models

Model Name	Number of epochs needed	Time needed per epoch (sec)	Total training time (hours)
InceptionV3	40	115	1.27
VGG19	15	5667	23.61
Resnet50	55	1234	18.85
Densenet201	40	127	1.41

We used different kinds of hyperparameter settings for training our individual models.

Table 3.6: Hyperparameter settings for models of experiment 1

	Model Name			
Parameters	InceptionV3	VGG19	Resnet50	Densenet201
Learning Rate	0.0001	0.0001	0.001	0.0001
Number of epochs	40	15	55	40
Training batch size	64	64	64	16
Validation batch size	64	64	64	16
Total number of parameters	25M	58M	24M	38M
Optimizer	Adam	Adam	Adam	Adam

The loss function used for all the cases was binary cross entropy. After training the models individually, we used weighted average based ensemble technique to bolster the results of each individual model. So we combined the results of InceptionV3, VGG19, Resnet50 and Densenet201 to improve the performance of each model. Then we compared and contrasted between each of their results.

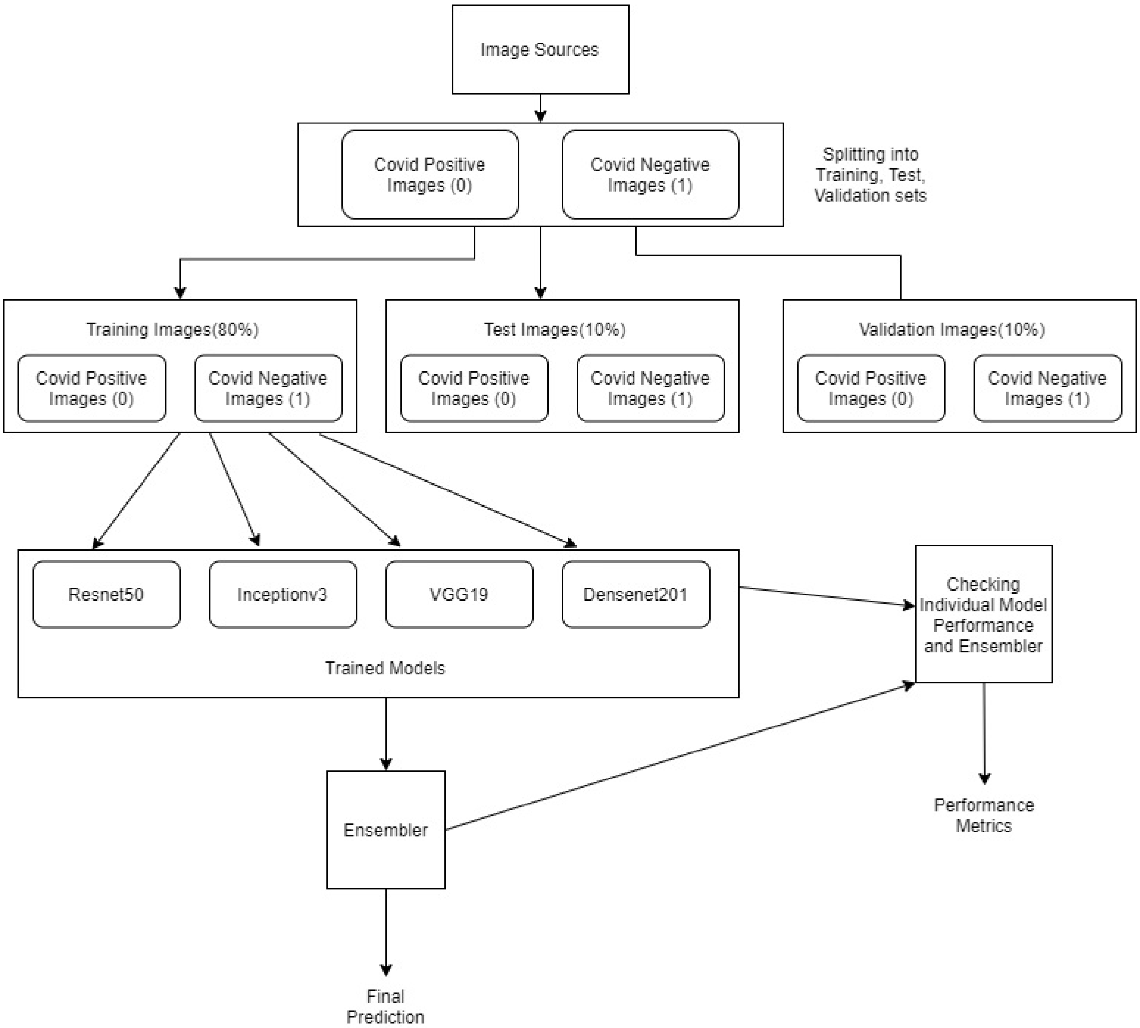

Figure 3.2: Flow Chart of Ensemble Learning

3.2.2 Model formation for experiment 2

The second experiment used the second dataset called "Covid and normal xray" which consisted of two directories called Train and Test. The training set consisted of 1073 Covid -ve images and 960 Covid +ve images. The testing set consisted of 268 Covid -ve images and 240 Covid +ve images.

Table 3.7: Dataset Information for experiment 2

	Total	Covid +ve	Covid -ve
Training set	2033	960	1073
Testing set	508	240	268

For training the models we used 10 epochs for Inceptionv3,10 epochs for vgg19. The optimiser used for each of the models is the Adam optimizer and then the models are saved as .pb files after every epoch. The time taken for model training are - 80 seconds / epoch for Inceptionv3, 490 seconds / epoch for VGG19.

Table 3.8: Time taken for training of experiment 2 models

Model Name	Number of epochs needed	Time needed per epoch (sec)	Total training time (hours)
Inceptionv3	10	80	0.22
VGG19	10	490	1.36

We trained and tested using two pretrained models using transfer learning. These two models are Inceptionv3 and VGG19. The hyperparameters are summarized below:

Table 3.9: Hyperparameter settings for models of experiment 2

	Model Name	
Parameters	InceptionV3	VGG19
Learning Rate	0.0001	0.0001
Number of epochs	10	10
Training batch size	32	32
Validation batch size	32	32
Total number of parameters	25M	58M
Optimizer	Adam	Adam

The loss function used for all the cases was binary cross entropy. After training the models we compared and contrasted between their results.

3.2.3 Model formation for experiment 3

We performed an additional validation on the model trained in experiment 2 with our 3rd dataset called "Chest-xray-for-covid19-detection" which has 3 folders called Prediction, Train and Val (Validation) containing 23, 288 and 60 photos respectively and they were divided equally for the classes “Covid” and “Normal” . But we used the Validation set of this dataset to see how well our trained dataset worked or to see how well our model can correctly identify a Covid +ve case as Covid +ve or a Covid -ve case as Covid -ve. The models used for experiment 3 were the trained InceptionV3 and VGG19 models from Experiment 2. So basically the trained models were evaluated on our third dataset.

Table 3.10: Dataset Information for experiment 3

	Total	Covid +ve	Covid -ve
Validation set	60	30	30

3.3 Evaluation Matrics

We are going to evaluate the system based on how well the system can correctly classify between Covid +ve and Covid -ve images. For evaluating our system, we mainly used the following performance metrics where TP stands for True Positive, FP for False Positive, FN for False Negative and TN for True Negative. For our case, the number of Covid-19 +ve cases that have been correctly classified by the model is called True Positive and if those Covid-19 +ve cases are incorrectly classified as Covid -ve, then they are termed as False Positive. Likewise, the Covid -ve patients classified accurately are called True Negative and if they are incorrectly classified as Covid +ve then they are called False Negative.

Confusion Matrix

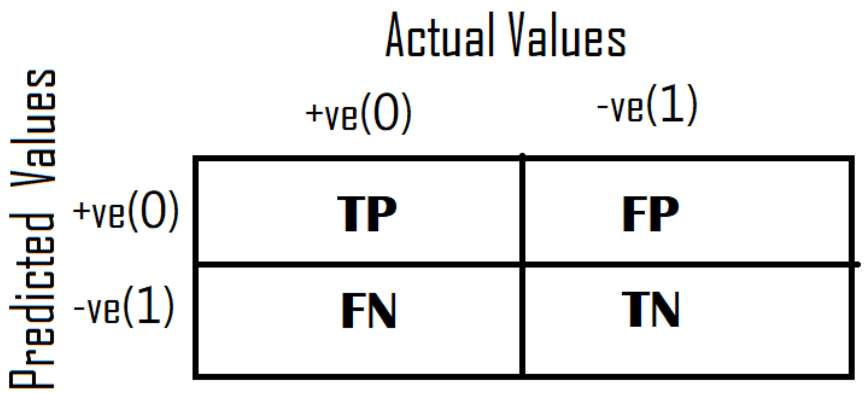

Figure 3.3: Confusion Matrics

Accuracy

Accuracy can be defined as something that is the ratio of the number of accurate or correct predictions made as a ratio of all the predictions.

$$\textbf{Accuracy=(TP+TN) / (TP+FP+FN+TN)} \tag{3.1}$$

Classification Report

Classification report is made up of the scores of Precisions, Recall, F1 and Support. These elements are described below:

Precision

Precision refers to the number of correct or precise documents returned by our model.

$$\textbf{Precision= TP / (TP + FP)} \tag{3.2}$$

Recall or Sensitivity

Recall refers to the number of positive results returned by our model.

$$\textbf{Recall=TP / (TP + FN)} \tag{3.3}$$

Specificity

Specificity refers to the number of negative results returned by our model.

$$\textbf{Specificity=TN /(TN+FP)} \tag{3.4}$$

Support

Support is something that refers to the number of actual samples or actual occurrences of a particular class in a particular dataset.

F1 Score

F1 Score refers to the harmonic mean or average of recall and precision. In other words, F1 score is the weighted average of the precision and recall.

$$\textbf{F1 = 2 * (recall * precision) / (recall + precision)} \tag{3.5}$$

AUC (Area Under ROC curve)

AUC (Area Under Curve)-ROC (Receiver Operating Characteristic) denotes the capability or the overall effectiveness of a particular model in distinguishing the classes or for figuring out which class an object falls under. If the AUC is high, it means that the model performs well. Likewise, If the AUC is low, it means that the model does not perform well in distinguishing or classifying between the classes. It is based on different threshold settings. Normally, the plotting of an AUC-ROC curve can be done by mapping TPR (True Positive Rate) vs FPR (False Positive Rate) at different threshold settings.

Chapter 4

Result Analysis

We analyzed the results obtained from our models for the 3 datasets from our 3 experiments using the evaluation metrics specified in section 3.4.

4.1 Results from Experiment 1

To analyze the results from experiment 1, we first trained and validated our models for a number of epochs using our training set and validation set respectively. After that we evaluated how well our model performs using the test set from our dataset 1. The results of the models InceptionV3, VGG19, Resnet50 and Densenet201 were analyzed.

4.1.1 InceptionV3

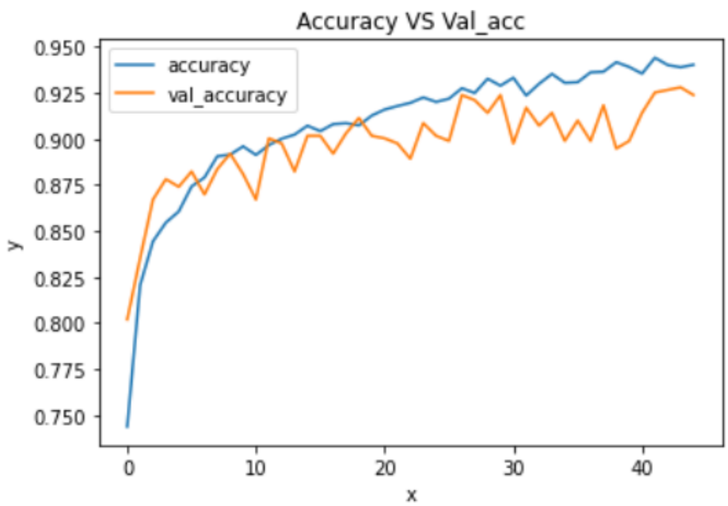

Figure 4.1: Training and validation accuracy of InceptionV3

Figure 4.2: Training and validation loss of InceptionV3

The Figure 4.1 shows the Training and validation accuracy of InceptionV3 and Figure 4.2 shows Training and validation loss of InceptionV3 for our first dataset over 40 epochs. The accuracy gradually increases as the model learns over the successive epochs and the loss gradually decreases overall. But the average Loss came out to be 0.132 and the accuracy came out as 0.945 on an average.

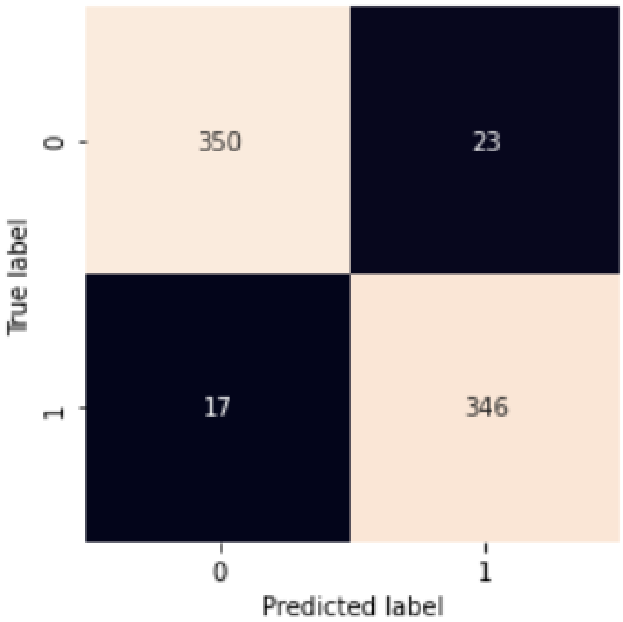

Figure 4.3: Confusion Matrix

The confusion matrix, depicted in Fig 4.3 shows that out of 736 patients, 350 Covid +ve patients have been accurately classified as having Covid, and 346 Covid -ve patients have been accurately classified as not having Covid. On the other hand, 23 Covid -ve patients have been inaccurately classified as having Covid and 17 Covid Covid +ve patients have been inaccurately classified as not having Covid.

	precision	recall	f1-score	support
0	0.95	0.94	0.95	373
1	0.94	0.95	0.95	363
accuracy			0.95	736
macro avg	0.95	0.95	0.95	736
weighted avg	0.95	0.95	0.95	736

Figure 4.4: Classification report of InceptionV3

In the above classification report we can see precision ,recall and f1-scoring is 95%,94%,95% for class-0 and 94%,95%,95% for class-1 respectively.The accuracy , macro average and weighted average is almost 95% for InceptionV3.

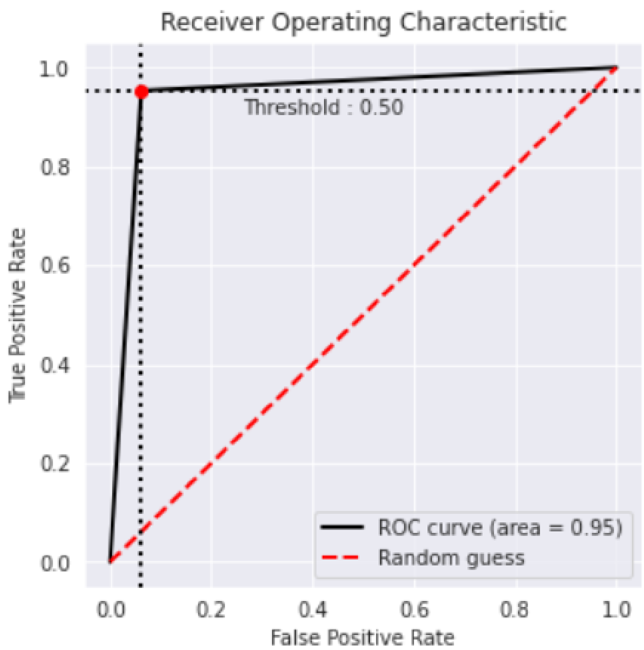

Figure 4.5: ROC curve of InceptionV3

Here the area under the ROC curve (AUC) was calculated to be approaximately 95% having 97% sensitivity and threshold value at 0.5 .

4.1.2 VGG19

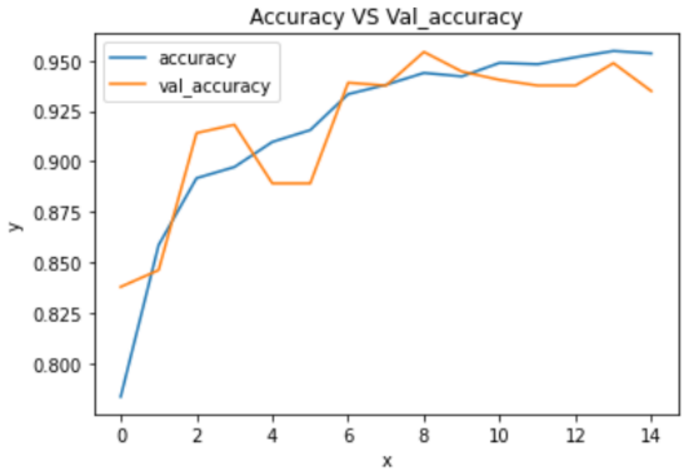

Figure 4.6: Training and Validation accuracy of VGG19

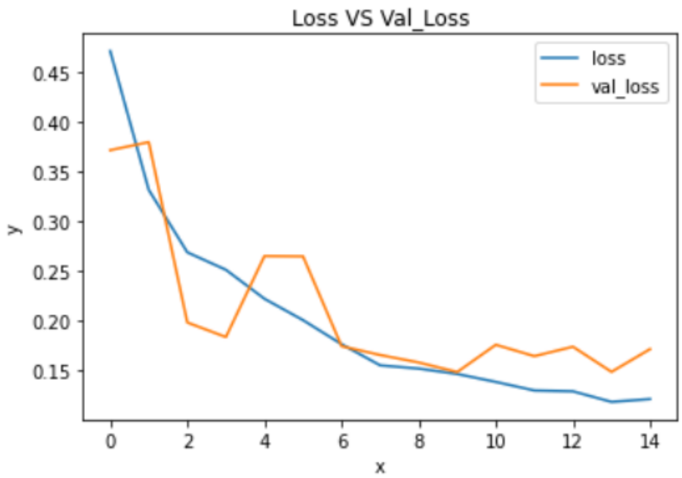

Figure 4.7: Training and validation loss of VGG19

The diagram no 4.6 shows the Training and validation accuracy of VGG19 and diagram no 4.7 shows Training and validation loss of VGG19 for our first dataset over 15 epochs. The accuracy gradually increases as the model learns over the successive epochs and the loss gradually decreases overall. But there was a sudden loss increase at epochs 4 to 6 but then it went down on the subsequent epochs and the accuracy increased. But the average loss is 0.142 and accuracy is 0.947 for the validation set of our first dataset.

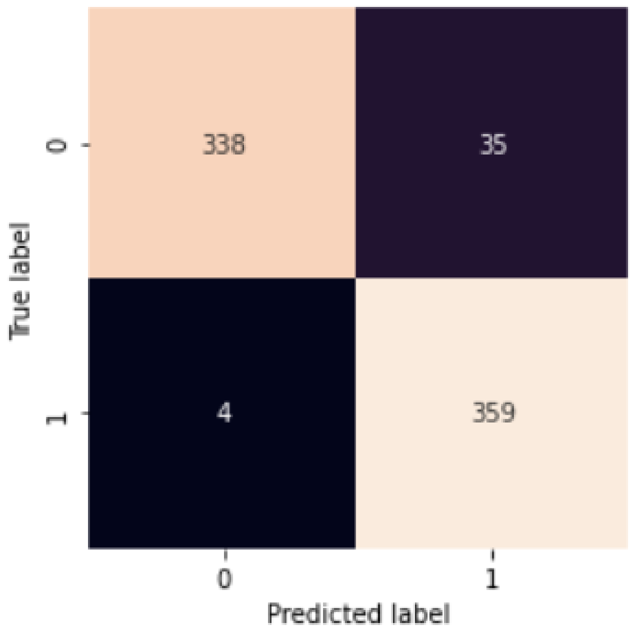

Figure 4.8: Confusion Matrix of VGG19

The confusion matrix, depicted in Figure 4.8 shows that out of 736 patients, 338 Covid +ve patients have been accurately classified as having Covid, and 359 Covid -ve patients have been accurately classified as not having Covid. On the other hand, 35 Covid -ve patients have been inaccurately classified as having Covid and 4 Covid Covid +ve patients have been inaccurately classified as not having Covid.

	precision	recall	f1-score	support
0	0.99	0.91	0.95	373
1	0.91	0.99	0.95	363
accuracy			0.95	736
macro avg	0.95	0.95	0.95	736
weighted avg	0.95	0.95	0.95	736

Figure 4.9: Classification report of VGG19

In the classification report we can see precision ,recall and f1-scoring is 99%,91%,95% for class-0 and 91%,99%,95% for class-1 respectively.The accuracy , macro average and weighted average is almost 95% for VGG19.

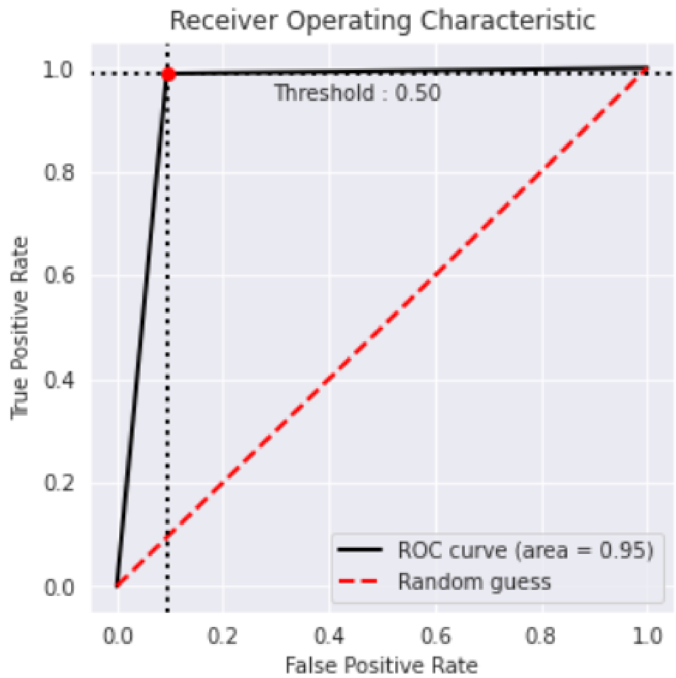

Figure 4.10: Roc Curve of VGG19

Here the area under the ROC curve (AUC) was calculated to be approaximately 95% having sensitivity of 99% and threshold value at 0.5 .

4.1.3 Resnet50

The diagram no 4.11 shows the Training and validation accuracy of Resnet50 and diagram no 4.12 shows Training and validation loss ofResnet50 for our first dataset over 55 epochs. The accuracy gradually increases as the model learns over the successive epochs and the loss gradually decreases overall. There were sudden increases and decreases of accuracies over the 55 folds or epochs but the average accuracy came out to be 0.677 but the average Loss is: 0.613.

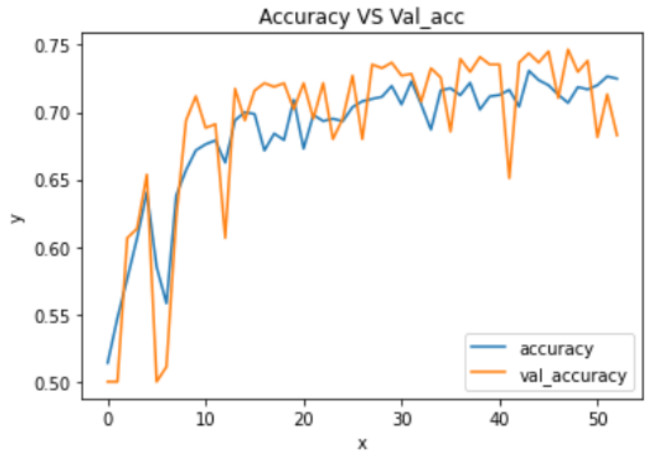

Figure 4.11: Training and validation accuracy of Resnet50

Figure 4.12: Training and validation loss of Resnet50

The diagram no 4.11 shows the Training and validation accuracy of Densenet201 and diagram no 4.12 shows Training and validation loss of Densenet201 for our first dataset over 40 epochs. The accuracy gradually increases as the model learns over the successive epochs and the loss gradually decreases overall. But the average loss is 0.13 and accuracy is 0.93 for the validation set of our first dataset.

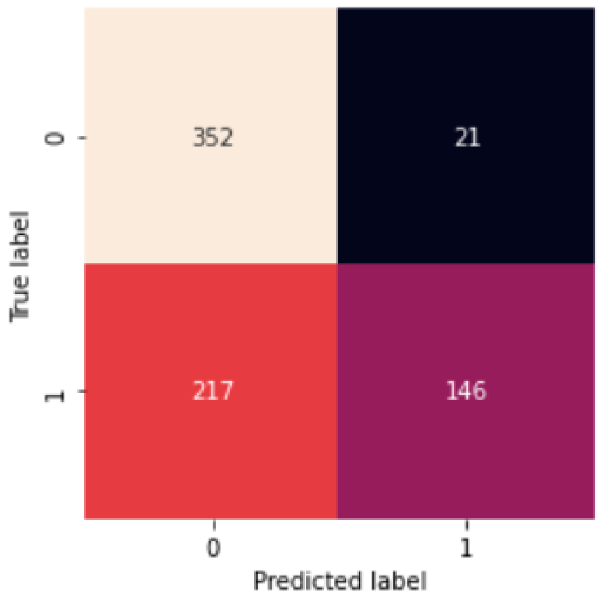

Figure 4.13: Confusion Matrix of Resnet50

The confusion matrix, depicted in Fig. 4.13 shows that out of 736 patients, 338 Covid +ve patients have been accurately classified as having Covid, and 359 Covid -ve patients have been accurately classified as not having Covid. On the other hand, 35 Covid -ve patients have been inaccurately classified as having Covid and 4 Covid +ve patients have been inaccurately classified as not having Covid.

	precision	recall	f1-score	support
0	0.62	0.94	0.75	373
1	0.87	0.40	0.55	363
accuracy			0.68	736
macro avg	0.75	0.67	0.65	736
weighted avg	0.74	0.68	0.65	736

Figure 4.14: Classification Report of Resnet50

In the classification report we can see precision ,recall and f1-scoring is 62%,94%,75% for class-0 and 87%,40%,55% for class-1 respectively.The accuracy is 68%, macro average value is 75%,67%,65% and weighted average is 74%,68%,65% respectively for Resnet50.

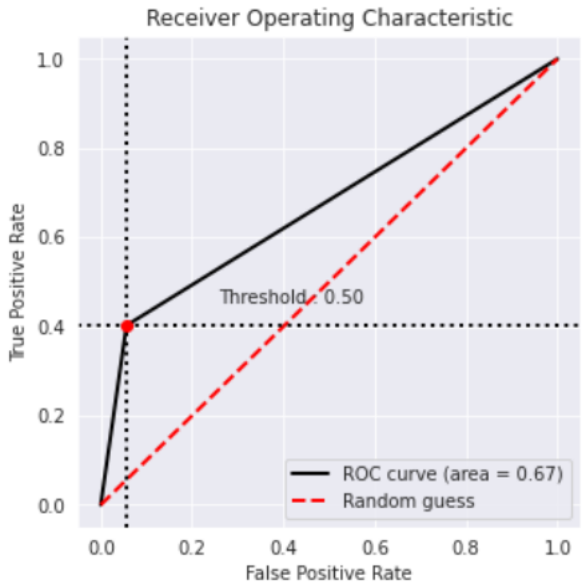

Figure 4.15: ROC Curve of Resnet50

Here the area under the ROC curve (AUC) was calculated to be approximately 67% having confidence interval 0.42-1 and threshold value 0.5 .

4.1.4 Densenet201

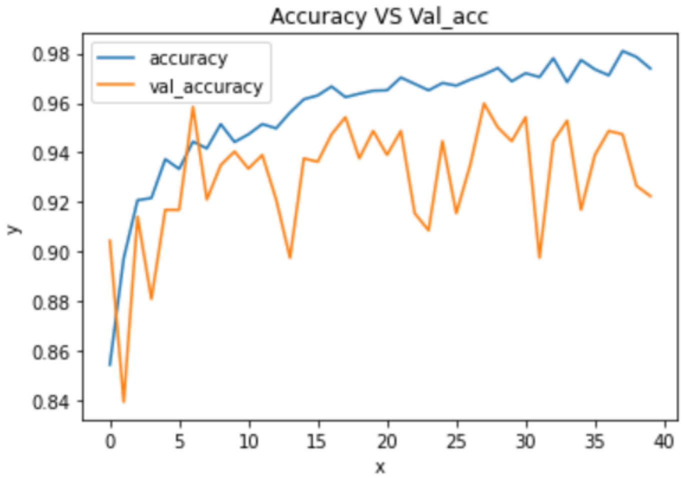

Figure 4.16: training and validation accuracy of Densenet201

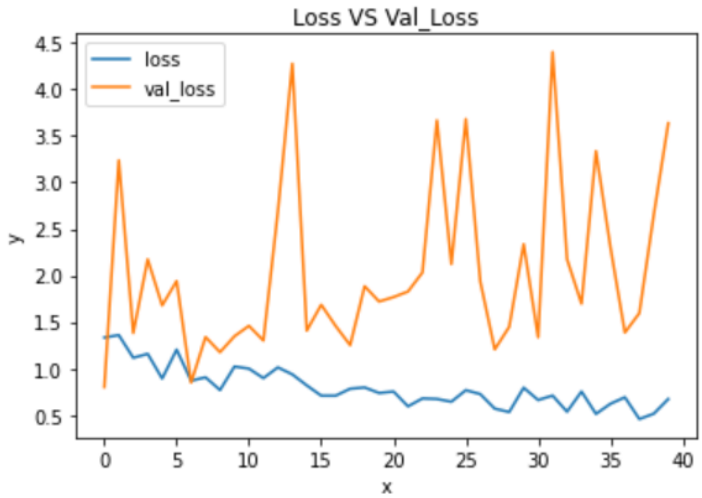

Figure 4.17: training and validation loss of Densenet201

```
Confusion Matrix
Confusion matrix,without normalization
[[348  25]
 [ 27 336]]
```

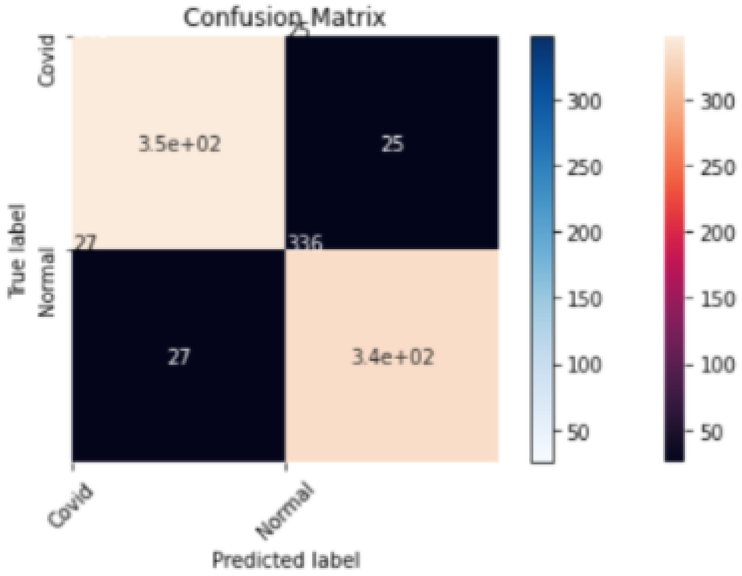

Figure 4.18: Confusion Matrix of Densenet201

The confusion matrix, depicted in Fig 4.18 shows that out of 736 patients, 348 Covid +ve patients have been accurately classified as having Covid, and 336 Covid -ve patients have been accurately classified as not having Covid. On the other hand, 25 Covid -ve patients have been inaccurately classified as having Covid and 27 Covid Covid +ve patients have been inaccurately classified as not having Covid.

	precision	recall	f1-score	support
class 0	0.93	0.93	0.93	373
class 1	0.93	0.93	0.93	363
accuracy			0.93	736
macro avg	0.93	0.93	0.93	736
weighted avg	0.93	0.93	0.93	736

Figure 4.19: Classification Report of Densenet201

In the classification report we can see precision ,recall and f1-scoring for class-0 , class-1 ,accuracy ,macro average and weighted average is almost 93% for Densenet201.

4.1.5 Ensemble of InceptionV3, VGG19, Resnet50, Densenet201

Here we combined the results of the models mentioned in section 4.1.1, 4.1.2, 4.1.3, 4.1.4 by average weight ensemble technique to bolster the efficiency of the standalone models.

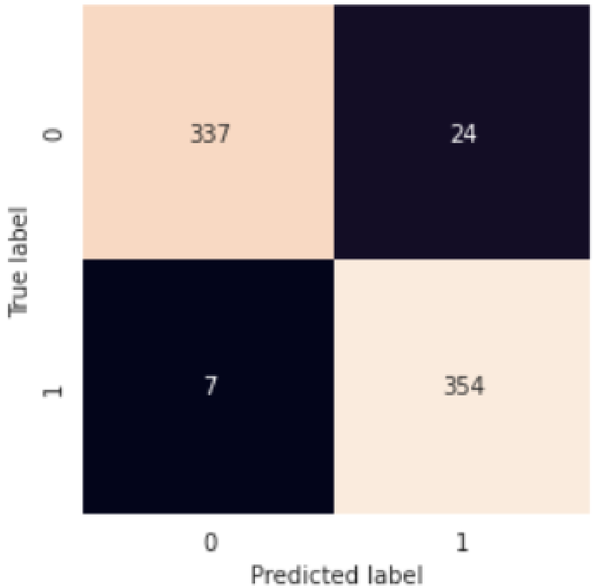

Figure 4.20: Confusion Matrix of Ensemble Learning

The confusion matrix, depicted in Fig. 4.21 shows that out of 736 patients, 358 Covid +ve patients have been accurately classified as having Covid, and 349 Covid -ve patients have been accurately classified as not having Covid. On the other hand, 15 Covid -ve patients have been inaccurately classified as having Covid and 14 Covid Covid +ve patients have been inaccurately classified as not having Covid.

	precision	recall	f1-score	support
0	0.98	0.93	0.96	361
1	0.94	0.98	0.96	361
accuracy			0.96	722
macro avg	0.96	0.96	0.96	722
weighted avg	0.96	0.96	0.96	722

Figure 4.21: Classification report of Ensemble learning method

In the classification report we can see precision ,recall and f1-scoring for class-0 , class-1 ,accuracy ,macro average and weighted average is almost 96% for Ensembling of Inceptionv3, VGG19, Resnet50, Densenet201.

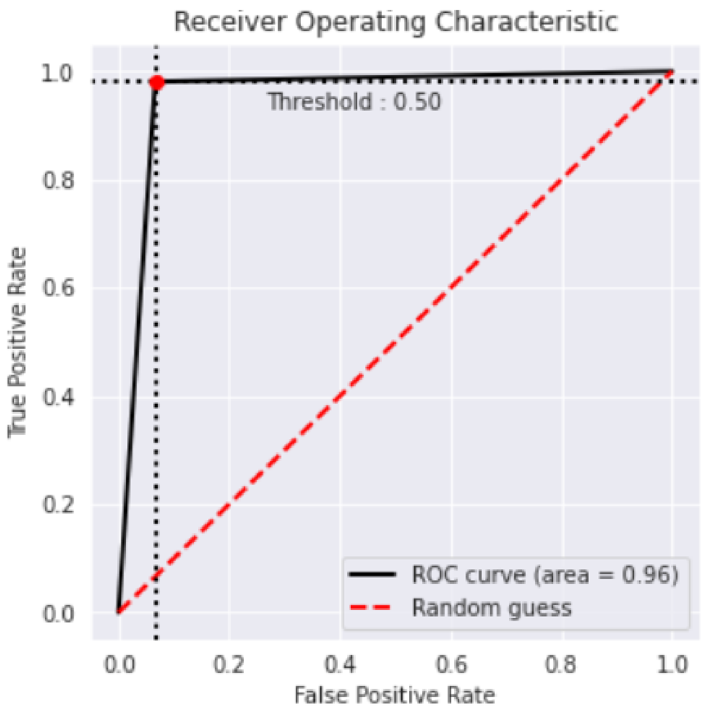

Figure 4.22: ROC curve of Ensemble learning method

Here the area under the ROC curve (AUC) was calculated to be approximately 96% having 99% sensitivity and threshold value 0.5 .

4.2 Results from Experiment 2

To analyze the results from experiment 2 we first trained and validated our models for a number of epochs using our training set and validation set respectively. After that we evaluated how well our model performs using the test set from our dataset 2. The results of the models Inceptionv3, VGG19 were analyzed.

4.2.1 InceptionV3

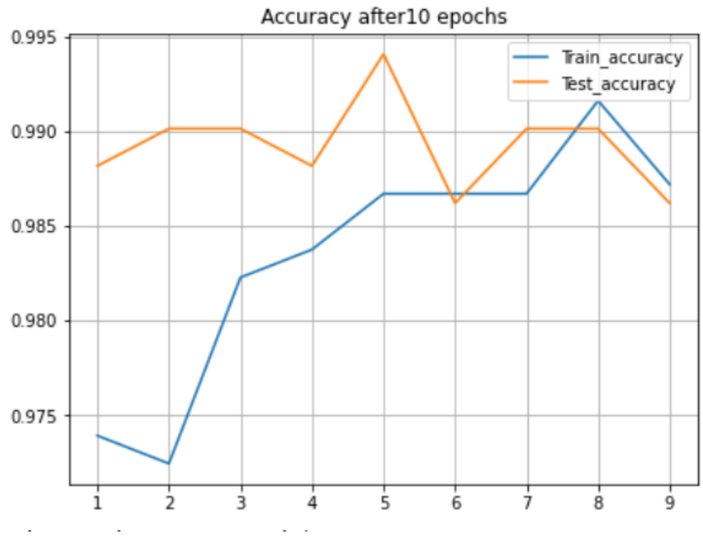

Figure 4.23: Training and Testing Accuracy Graph of InceptionV3 method after 10 epochs

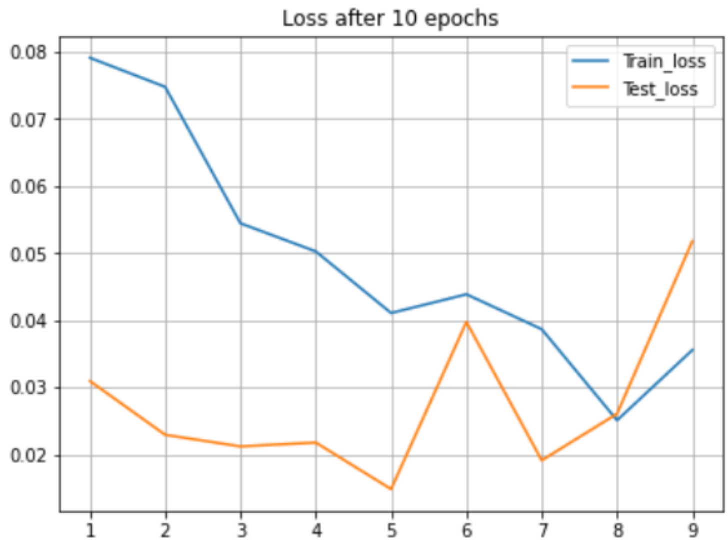

Figure 4.24: Training and testing Loss Graph of InceptionV3 method after 10 epochs

The diagram no 4.24 shows the training and validation accuracy of Inceptionv3 and diagram no 4.25 shows training and validation loss of Inceptionv3 for our second dataset over 10 epochs. The training accuracy gradually increases as the model learns over the successive epochs and the training loss gradually decreases overall for the validation set of our second dataset. But there were fluctuations in the validation accuracy and validation loss. But the average Loss is: 0.0216 and accuracy is: 0.995.

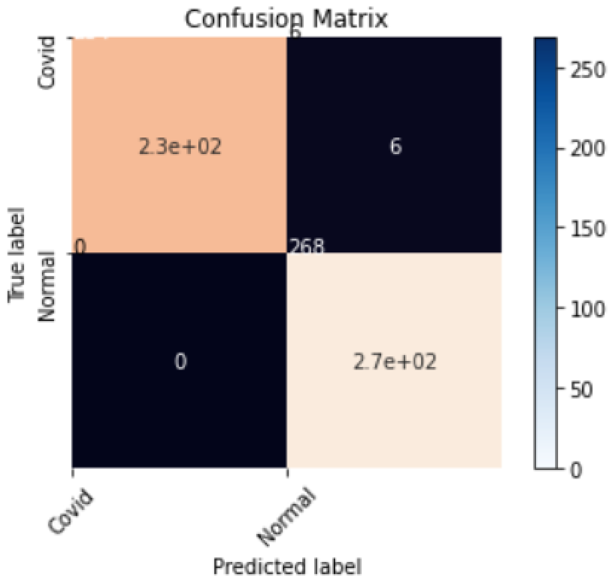

Figure 4.25: Confusion Matrix of InceptionV3 method after 10 epochs

The confusion matrix, depicted in Fig. 4.26 shows that out of 508 patients, 234 Covid +ve patients have been accurately classified as having Covid, and 268 Covid -ve patients have been accurately classified as not having Covid. On the other hand, 6 Covid -ve patients have been inaccurately classified as having Covid and 0 Covid Covid +ve patients have been inaccurately classified as not having Covid.

	precision	recall	f1-score	support
class 0	1.00	0.97	0.99	240
class 1	0.98	1.00	0.99	268
accuracy			0.99	508
macro avg	0.99	0.99	0.99	508
weighted avg	0.99	0.99	0.99	508

Figure 4.26: Classification Report of InceptionV3 method after 10 epochs

In the above classification report we can see precision ,recall and f1-scoring is 100%,97%,99% for class-0 and 98%,100%,99% for class-1 respectively. The accuracy , macro average and weighted average is almost 99% for InceptionV3.

4.2.2 Vgg19

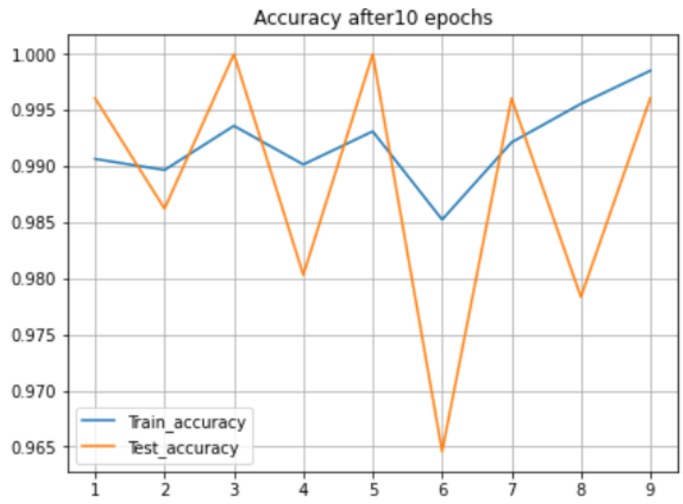

Figure 4.27: Training and Testing Accuracy Graph of Vgg19

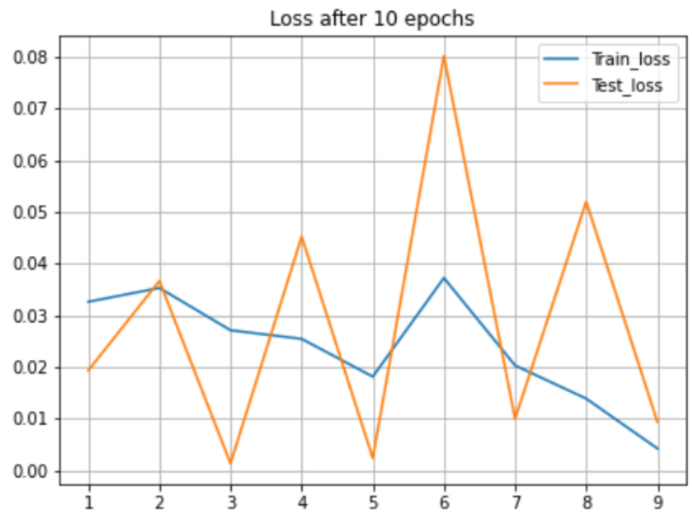

Figure 4.28: Training and testing Loss Graph of Vgg19

The diagram no 4.28 shows the training and validation accuracy of VGG19 and diagram no 4.29 shows Training and validation loss of VGG19 for our second dataset over 10 epochs. The training accuracy gradually increases as the model learns over the successive epochs and the training loss gradually decreases overall for the validation set of our second dataset.But there were fluctuations in the validation accuracy and validation loss. But the average Loss is: 0.0216 and accuracy is: 0.996.

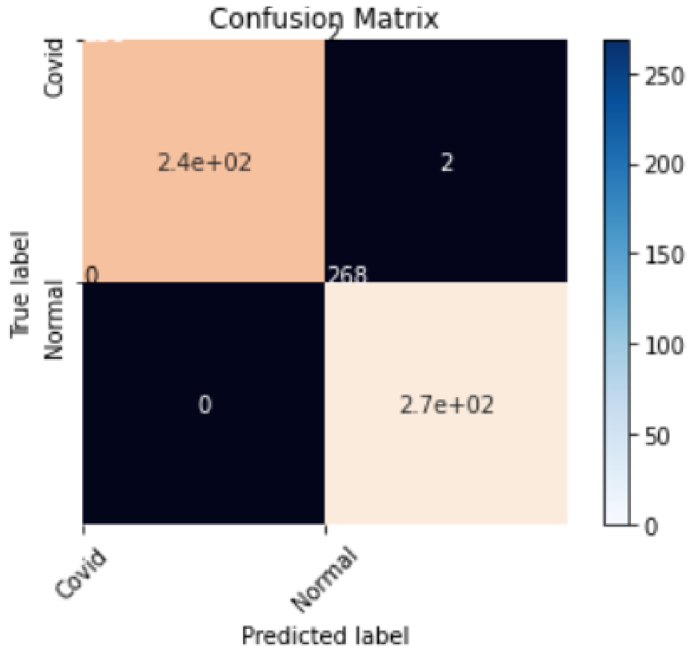

Figure 4.29: Confusion Matrix of Vgg19 method after 10 epochs

The confusion matrix, depicted in Fig.4.30 shows that out of 508 patients, 238 Covid +ve patients have been accurately classified as having Covid, and 268 Covid -ve patients have been accurately classified as not having Covid. On the other hand, 2 Covid -ve patients have been inaccurately classified as having Covid and 0 Covid Covid +ve patients have been inaccurately classified as not having Covid.

	precision	recall	f1-score	support
class 0	1.00	0.99	1.00	240
class 1	0.99	1.00	1.00	268
accuracy			1.00	508
macro avg	1.00	1.00	1.00	508
weighted avg	1.00	1.00	1.00	508

Figure 4.30: Classification Report of Vgg19 method after 10 epochs

In the above classification report we can see precision ,recall and f1-scoring is 100%,99%,100% for class-0 and 99%,100%,100% for class-1 respectively. The accuracy , macro average and weighted average is 100% for VGG19.

4.3 Results from Experiment 3

Here we evaluated how well the model trained from dataset 2 performs using the test set from our dataset 3. The results of the models Inceptionv3, VGG19 were analyzed.

4.3.1 InceptionV3

The confusion matrix, depicted in Fig. 12 shows that out of 60 patients, 29 Covid +ve patients have been accurately classified as having Covid, and 30 Covid -ve patients have been accurately classified as not having Covid. On the other hand, 1 Covid -ve patients have been inaccurately classified as having Covid and 0 Covid Covid +ve patients have been inaccurately classified as not having Covid.

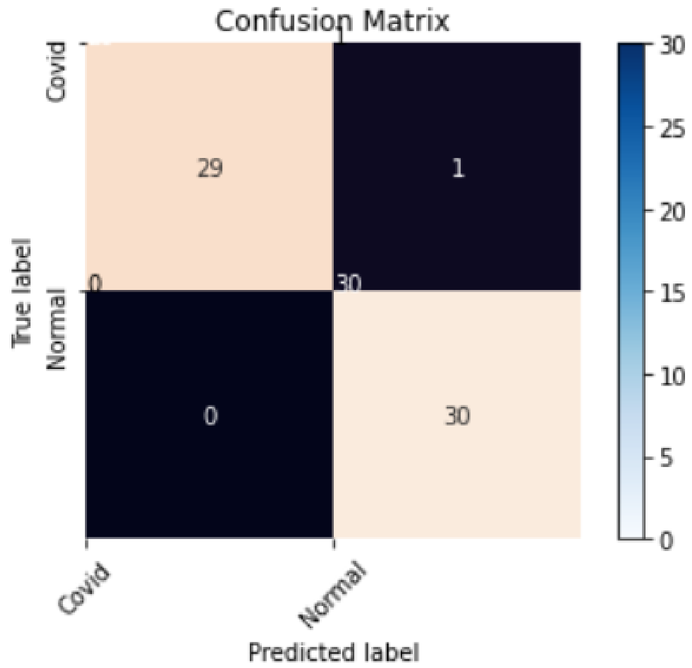

Figure 4.31: Confusion Matrix of InceptionV3

	precision	recall	f1-score	support
class 0	1.00	0.97	0.98	30
class 1	0.97	1.00	0.98	30
accuracy			0.98	60
macro avg	0.98	0.98	0.98	60
weighted avg	0.98	0.98	0.98	60

Figure 4.32: Classification Report of InceptionV3

In the above classification report we can see precision ,recall and f1-scoring is 100%,97%,98% for class-0 and 97%,100%,98% for class-1 respectively. The accuracy , macro average and weighted average is almost 98% for InceptionV3.

4.3.2 Vgg19

The confusion matrix, depicted in Fig.4.34 shows that out of 60 patients, 29 Covid +ve patients have been accurately classified as having Covid, and 30 Covid -ve patients have been accurately classified as not having Covid. On the other hand, 1 Covid -ve patients have been inaccurately classified as having Covid and 0 Covid Covid +ve patients have been inaccurately classified as not having Covid.

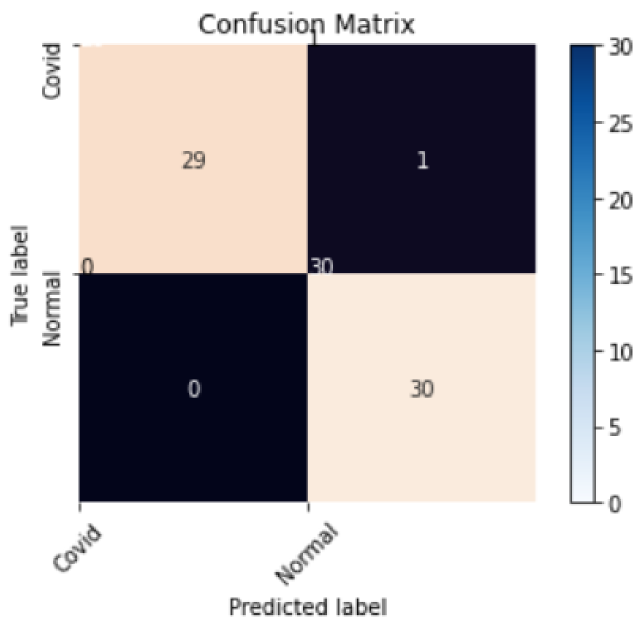

Figure 4.33: Confusion Matrix of Vgg19

```
              precision    recall  f1-score   support

     class 0       1.00      0.97      0.98        30
     class 1       0.97      1.00      0.98        30

    accuracy                           0.98        60
   macro avg       0.98      0.98      0.98        60
weighted avg       0.98      0.98      0.98        60
```

Figure 4.34: Classification Report of Vgg19

In the classification report we can see precision ,recall and f1-scoring for class-0 class-1 is 100%,97%,98% and 97%,100%,98% respectively. The accuracy for macro average and weighted average is almost 98% for VGG-19.

4.4 Comparative Results and Discussions

By analyzing the results found from section 4.3, we can compare and contrast between the different models by seeing how well they can perform on each dataset in each experiment. For the first experiment we found out that Resnet50 had the least efficiency and its accuracy rose to 62.7% only after 55 epochs. On the other hand Inceptionv3, Densenet201 and VGG19 exhibited similar performances showing an accuracy of 94.5%, 93.0% and 94.7% respectively. But the ensemble of these 4 models yielded an accuracy of 96%.

Table 4.1: Summary of results from experiment 1

Model	TP	TN	FP	FN	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
InceptionV3	350	346	23	17	94.5	93.8	95.4	93.8
VGG19	338	359	35	4	94.7	90.6	98.8	91.1
Resnet50	352	146	21	217	67	94.4	61.9	87.4
Densenet201	348	336	25	27	93	93.3	92.8	93.1
Ensemble	358	349	15	14	96	96	96.2	95.9

For the second experiment, both Inceptionv3 and VGG19 exhibited similar performances and yielded 99.6% and 99.5% accuracies respectively. But this high accuracy was caused by overfitting due to such a small dataset. To improve this we needed a much larger dataset or we needed to reduce the complexity of our model.

Table 4.2: Summary of results from experiment 2

Model	TP	TN	FP	FN	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
InceptionV3	234	268	6	0	99.5	97.5	100	97.8
VGG19	238	268	2	0	99.6	99.2	100	99.3

For the third experiment, both Inceptionv3 and VGG19 exhibited the same performance and yielded 99.6% accuracy. Similar to the previous case this high accuracy was caused by overfitting due to such a small dataset. To improve this we needed a much larger dataset or we needed to reduce the complexity of our model.

Table 4.3: Summary of results from experiment 3

Model	TP	TN	FP	FN	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
InceptionV3	29	30	1	0	99.6	96.7	100	96.8
VGG19	29	30	1	0	99.6	96.7	100	96.8

Chapter 5

Future Works

From the beginning we faced a problem with the dataset.It was imbalanced and for that we couldn't do our job properly. So,we worked with our dataset to balance it. Then we got good results from before.But we can do more things. So,We want to work with our dataset more precisely and I think after doing that we will obtain better result using developed dataset. Here is also a problem which is under utilization of our dataset. We can't use our data properly because we can't utilize all of the images of our dataset. We want to use all the images of our dataset. By preventing this problem we hope we will get a better result than before. The main scope of improvement is to fix the imbalancing in the first dataset mentioned by using some kind of balancing technique, either by overbalancing or underbalancing. We mainly hope to use overbalancing techniques like SMOTE or ensembling for this purpose. We also plan to explore the capabilities of newer and more efficient pretrained models like Efficient NetB4 and improve the performance of our model by incorporating segmentation in our images.

Chapter 6

Conclusion

Fast and accurate detection of covid19 is absolutely crucial to keep the virus under control so that it doesn't aggravate the situation any further. Taking all of the limitations and constraints in mind we have come up with the idea of our proposed system which is to detect covid19 using chest x ray images. We mainly want to implement the system by using the help of various techniques and approaches of deep learning. The only plausible way to combat the disease and to stop or halt the spread of this virus is early detection of covid19 and we believe that our proposed system will go a long way in achieving this goal and will help doctors and medical personnel and will add a significant value to the medical field.

References

[1] wikipedia, "Covid-19 Pandemic." https://en.wikipedia.org/wiki/COVID-19_pandemic, 2020. [Online; accessed 24 Feb 2021].

[2] R. S. Kaplan and M. E. Porter, "How to solve the cost crisis in health care," *Harv Bus Rev*, vol. 89, no. 9, pp. 46–52, 2011.

[3] M. Z. Islam, M. M. Islam, and A. Asraf, "A combined deep cnn-lstm network for the detection of novel coronavirus (covid-19) using x-ray images," *Informatics in medicine unlocked*, vol. 20, p. 100412, 2020.

[4] K. F. Haque, F. F. Haque, L. Gandy, and A. Abdelgawad, "Automatic detection of covid-19 from chest x-ray images with convolutional neural networks," in *2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, pp. 125–130, IEEE, 2020.

[5] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network," *Applied Intelligence*, vol. 51, no. 2, pp. 854–864, 2021.

[6] S. Hussain Khan, A. Sohail, and A. Khan, "Covid-19 detection in chest x-ray images using a new channel boosted cnn," *arXiv e-prints*, pp. arXiv–2012, 2020.

[7] T. Gao, "Chest x-ray image analysis and classification for covid-19 pneumonia detection using deep cnn," *medRxiv*, 2020.

[8] P. K. Sethy, S. K. Behera, P. K. Ratha, and P. Biswas, "Detection of coronavirus disease (covid-19) based on deep features and support vector machine," 2020.

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

[10] X. P. Burgos-Artizzu, "Automated covid-19 detection from frontal chest x-ray images using deep learning: an online feasibility study," *medRxiv*, 2020.

[11] A. K. Das, S. Ghosh, S. Thunder, R. Dutta, S. Agarwal, and A. Chakrabarti, "Automatic covid-19 detection from x-ray images using ensemble learning with convolutional neural network," *Pattern Analysis and Applications*, pp. 1–14, 2021.

[12] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in biology and medicine*, vol. 121, p. 103792, 2020.

[13] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arxiv 2020," *arXiv preprint arXiv:2003.10849*, 2003.

[14] Ç. Polat, O. Karaman, C. Karaman, G. Korkmaz, M. C. Balcí, and S. E. Kelek, "Covid-19 diagnosis from chest x-ray images using transfer learning: Enhanced performance by debiasing dataloader," *Journal of X-Ray Science and Technology*, no. Preprint, pp. 1–18.

[15] A. Makris, 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*, pp. 60–66, 2020.

[16] S. Ahmed, T. Hossain, O. B. Hoque, S. Sarker, S. Rahman, and F. M. Shah, "Automated covid-19 detection from chest x-ray images: A high-resolution network (hrnet) approach," *SN computer science*, vol. 2, no. 4, pp. 1–17, 2021.

[17] B. Sekeroglu and I. Ozsahin, "<? covid19?> detection of covid-19 from chest x-ray images using convolutional neural networks," *SLAS TECHNOLOGY: Translating Life Sciences Innovation*, vol. 25, no. 6, pp. 553–565, 2020.

[18] R. Kumar, A. Arora, B. Bansal, V. Sahayasheela, H. Buckchash, J. Imran, N. Narayanan, G. Pandian, and B. Raman, "Accurate prediction of covid-19 using chest x-ray images through deep feature learning model with smote and machine learning classifiers. medrxiv 2020.04. 13.20063461."

[19] F. A. Saiz and I. Barandiaran, "Covid-19 detection in chest x-ray images using a deep learning approach," *International Journal of Interactive Multimedia and Artificial Intelligence, InPress (InPress)*, vol. 1, 2020.

[20] S. Sakib, M. A. B. Siddique, M. M. R. Khan, N. Yasmin, A. Aziz, M. Chowdhury, and I. K. Tasawar, "Detection of covid-19 disease from chest x-ray images: A deep transfer learning framework," *medRxiv*, 2020.

[21] T. Rahman, "covid19-radiography-database." https://www.kaggle.com/tawsifurrahman/covid19-radiography-database, 2021. [Online; accessed May 2021].

[22] K. Dinleyici, "covid and normal xray v1." https://www.kaggle.com/kamildinleyici/cov-dataset, 2020. [Online; accessed May 2021].

[23] F. Fenta, "Chest Xray for covid-19 detection." https://www.kaggle.com/fusicfenta/chest-xray-for-covid19-detection, 2020. [Online; accessed May 2021].

Generated using Undegraduate Thesis LaTeX Template, Version 1.4. Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh.

This thesis was generated on Monday 20th December, 2021 at 5:58am.