Ensembling the Decisions from Deep Learning Models for Prediction of COVID-19 from Chest X-Ray Images

A project report

submitted in partial fulfillment of the requirement for the Degree of

Master of Computer Application

of

Department of Computer Science and Engineering of

Jadavpur University

by

ALISHA

Registration No.: 149865 of 2019-2020

Examination Roll No.: MCA226002

Under the guidance of

Prof. MAHANTAPAS KUNDU

Department of Computer Science and Engineering

Jadavpur University, Kolkata-700032

India

2022

Faculty of Engineering and Technology

Jadavpur University

Certificate of Recommendation

This is to certify that the project report entitled "Ensembling the Decisions from Deep Learning Models for Prediction of COVID-19 from Chest X-Ray Images" has been carried out by Alisha (University Registration No.: 149865 of 2019-2020, Examination Roll No.: MCA226002) under my guidance and supervision and be accepted in partial fulfillment of the requirement for the Degree of Master of Computer Application of the Department of Computer Science and Engineering, Jadavpur University. The research results presented in the report have not been included in any other paper submitted for the award of any degree in any other University or Institute.

Prof. MAHANTAPAS KUNDU (Project Supervisor)
Department of Computer Science and Engineering
Jadavpur University, Kolkata-32

Countersigned

Dr. Anupam Sinha
Professor and Head, Department of Computer Science and Engineering,
Jadavpur University, Kolkata-32

Prof. Chandan Mazumdar
Dean, Faculty of Engineering and Technology,
Jadavpur University, Kolkata-32

Faculty of Engineering and Technology

Jadavpur University

Certificate of Approval*

This is to certify that the project report entitled “Ensembling the Decisions from Deep Learning Models for Prediction of COVID-19 from Chest X-Ray Images” is a bonafide record of work carried out by Alisha in partial fulfillment of the requirements for the award of the degree of Master of Computer Application of the Department of Computer Science and Engineering, Jadavpur University during the period of January 2022 to June 2022. It is understood that by this approval the undersigned does not necessarily endorse, or approves any statement made, opinion expressed or conclusion drawn therein but approve the project report only for the purpose for which it has been submitted.

Signature of Examiner 1

Date:

Signature of the Examiner 2

Date:

*Only in the case the project report is approved.

Declaration of originality and compliance of academic ethics

I hereby declare that this project report entitled "Ensembling the Decisions from Deep Learning Models for Prediction of COVID-19 from Chest X-Ray Images" contains a literature survey and original research work by the undersigned candidate, as part of her degree of Master of Computer Application.

All information has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully referenced all materials and results that are not original to this work.

Name: Alisha

Registration No.: 149865 of 2019-2020

Exam Roll No.: MCA226002

Project Title: Ensembling the Decisions from Deep Learning Models for Prediction of COVID-19 from Chest X-Ray Images

Signature with date

ACKNOWLEDGEMENTS

The writing of the project report as well as the related work has been a long journey with input from many individuals, right from the first day till the development of the final project. I would like to express my deepest gratitude to my supervisor, Prof. Mahantapas Kundu, Professor, Department of Computer Science and Engineering, Jadavpur University for giving me the opportunity to do research and providing invaluable guidance throughout this work. His dynamism, vision, sincerity, and motivation have deeply inspired me. He has taught me the methodology to carry out the work and to present the works as clearly as possible. It was a great privilege and honor to work and study under his guidance.

I would also thank Dr. Ram Sarkar, Professor, Department of Computer Science and Engineering, Jadavpur University for his patience, guidance, suggestions and moral support in times of need.

I am also particularly thankful to Dr. Anupam Sinha, Professor and Head of the Department of Computer Science and Engineering, Jadavpur University for allowing us to carry out research in the department.

I would like to thank Aniruddha Maity for helping me in every possible way to complete this project.

I would also like to thank all the faculty members of the Department of Computer Science and Engineering of Jadavpur University for their continuous support.

This project would not have been completed without the inspiration and support of my family and friends and a number of wonderful individuals including my batchmates of Master of Computer Application at Jadavpur University — my thanks and appreciation to all of them for being part of this journey and making this project possible.

____

Alisha

Examination Roll No.: - MCA226002

Registration No.: - 149865 of 2019-2020

Class Roll No.: - 001910503002

CONTENTS

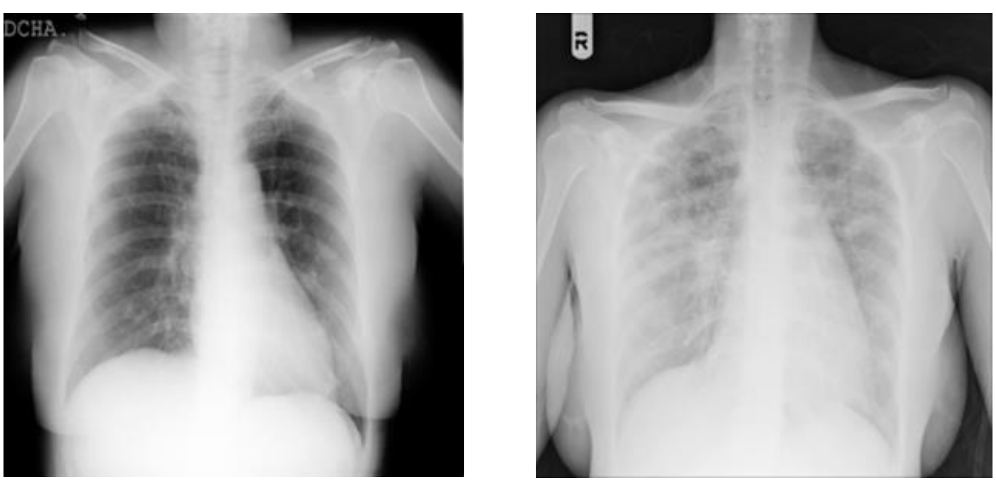
List of figures

List of tables

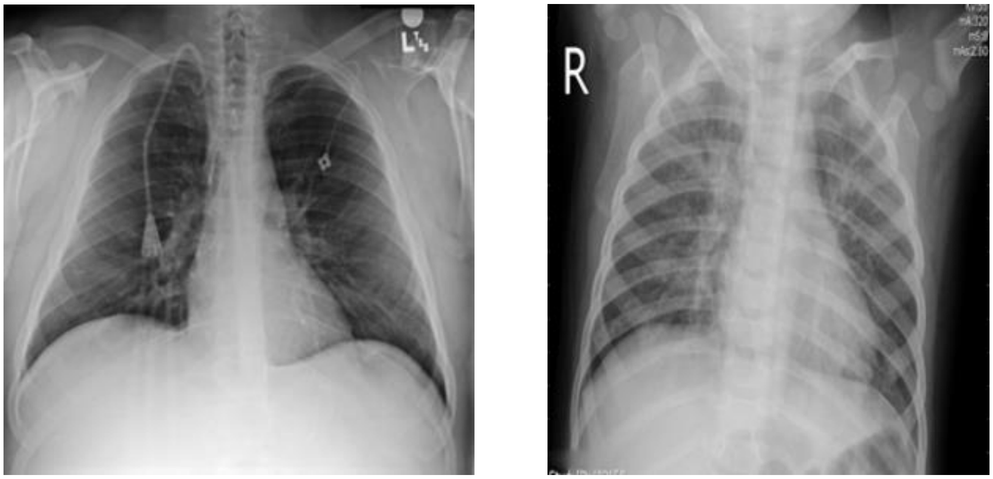
Chapter 1

Introduction

In early 2020, after a December 2019 outbreak of virus in China, the World Health Organization identified SARS-CoV-2 as a new type of coronavirus. The outbreak quickly spread around the world and was declared a pandemic by World Health Organization on 11 March 2020 [1]. Devastating economic consequences have been caused by the COVID-19 pandemic and have threatened human lives since it first emerged. As of May, 2022, COVID-19 has infected more than 517.6 million people across the globe, killing more than 6.26 million people and in India, it has affected more than 43.11 million people, killing more than 524 thousand people. COVID-19 is a disease caused by SARS-CoV-2 that can trigger a respiratory tract infection. It can affect one's upper respiratory tract (sinuses, nose, and throat) or lower respiratory tract (windpipe and lungs). It is one of seven types of coronaviruses, including the ones that cause severe diseases like Middle East respiratory syndrome (MERS) and sudden acute respiratory syndrome (SARS).



(a) (b)



(c) (d)

Figure 1.1. CXR images of people from COVID-19 radiography datasets, which categorizes CXR images into classes of (a) COVID-19 (b) Lung Opacity (c) Normal, and (d) Viral Pneumonia

1.1.Methods for COVID-19 detection

Reverse Transcription Polymerase Chain Reaction (RT-PCR) test on respiratory specimens is the highest quality level screening strategy for testing the COVID-19 patients. However, it is an expensive, manual, and time-consuming process.

Positive Radiographic images (Computed Tomography (CT)/ Chest Radiograph (CXR)) can be used as an alternative diagnosis tool. Bronchopneumonia causing fever, cough, dyspnea, and respiratory failure with acute respiratory distress syndrome (ARDS) is the clinical characteristic of severe COVID-19 infection. Readily available radiological imaging is an important diagnostic tool for COVID-19. CT scan has a higher sensitivity to pulmonary diseases but it has several limitations in the detection of COVID-19 at a large scale such as non-portability, long time scanning, and the risk of exposing the hospital staff whereas CXR imaging is portable, faster, more readily available, and can be performed within an isolated room while offering an acceptable accuracy in COVID-19 case detection [2].

1.2. The Research Trends

Due to the widespread infection of the coronavirus, several intelligence techniques, such as deep learning (DL) and machine learning (ML), have been created by various researchers to help the healthcare sector provide quick and accurate COVID-19 diagnosis. In this section, some of the study trends that have been used to detect or predict COVID-19 are discussed.

1.2.1. Machine learning Approaches

CXR images were used by Elaziz et al. [3] to create a visual diagnostic tool for distinguishing between COVID-19 and normal cases. Initially, features were extracted based on fractional multichannel exponent moments. But this process is time-consuming and costly. To accelerate the computational process a multi-core computational framework was utilized. This study proposed a new feature selection method and in order to find significant subset features from retrieved data, a modified manta-ray foraging optimization algorithm based on differential evolution (MRFODE) was applied. MRFODE generates a candidate subset of features iteratively, which is then evaluated using a KNN method. Raju et al. [4] proposed a method for Chest image classification and detection of COVID-19 using RELM classifier. Feature extraction is used for data cleaning and then RELM classifier is used for classification. Finally, ResNet50 is used for training and testing.

1.2.2. Deep Learning Approaches

1.2.2.1. Transfer learning and fine-tuning approaches

The main DL algorithms used for COVID-19 are CNN. In most of the papers, transfer learning is the preferred strategy. Transfer learning is done using pre-trained models that have been trained on the ImageNet database. Deep learning approaches such as mobileNet V2, Inception, Xception, InceptionResNet V2, and VGG19 that use transfer learning method is performed by Apostolopoulos et al. [5]. Other models like ResNet50, Inception V3, and InceptionResnet V2 are also utilized by Narin et al. [6]. Four prominent convolutional neural networks, including ResNet18, ResNet50, SqueezeNet, and DenseNet-121, were trained to identify COVID-19 disease in the examined chest X-Ray pictures using transfer learning [7]. Chowdhury et al. [8] did transfer learning on MobileNetv2, SqueezeNet, ResNet18, ResNet101 and DenseNet201. Stochastic Gradient Descent (SGD) optimization and Fivefold cross-validation was performed with and without data augmentation for two and three-class classification. Gifani et al. [9] introduced an automatic methodology for the detection of COVID-19. They used 15 pre-trained CNNs architecture and fine-tuned them and then the majority voting ensemble method of the best combination of deep transfer learning outputs was applied for improvement in the recognition performance. Pre-trained CNNs used are- EfficientNets(B0-B5), NasNetLarge, NasNetMobile, InceptionV3, ResNet-50, SeResnet 50, Xception, DenseNet121, ResNext50 and Inception_resnet_v2.

GAN with Deep transfer learning for COVID-19 detection was introduced by Loey et al. [10]. Three deep learning models were used- Alexnet, Googlenet, and Restnet18. The algorithm was tested on a dataset with 307 images. Bargshady et al. [11] proposed the CycleGAN-Inception model and used an extensive COVID-19 X-Ray and CT Chest Images Dataset. The proposed approach has 2 steps- the data augmentation technique in which Generative Adversarial Network (GAN) is coupled with trained, semi-supervised CycleGAN (SSA-CycleGAN), and then fine-tuned Inception V3 model is developed to train the algorithm.

1.2.2.2. Novel Architectures

Many researchers have focused on exploring Deep Convolutional Neural Networks (DCNNs), given the significant successes and state-of-the-art achieved in a variety of computer vision tasks. COVID-Net [12] is an open source and publicly available deep convolutional neural network design for detecting COVID-19 cases from CXR images. Tang et al. [2] proposed EDL-COVID, an ensemble deep learning model generated by combining multiple snapshot models of COVID-Net by employing an average ensemble method that is aware of different sensitivities of deep learning models on different classes and has achieved more accuracy than COVID-Net. COVID-CAPS [13] is a capsule network-based framework for detecting COVID-19 infection in CXR and CT scan images. This framework also uses transfer learning. Capsule network performs well when data are scarce. COVIDLite [14] uses Depth wise Separable Convolutional Neural Network (DSCNN) to classify CXR images for COVID-19 detection.

1.3. Motivation

Reverse Transcription Polymerase Chain Reaction (RT-PCR) is used for detecting COVID-19. However, it is an expensive, manual, and time-consuming process and is not easily available in several areas. Also, COVID-19 affects the lungs, so chest X-Ray and CT scan images can be used to predict COVID-19 and also X-Ray machines are easily available.

Machine learning in general and deep learning methods in particular have been shown to be effective for analyzing large, multi-dimensional medical images. CT or X-Ray findings of COVID-19 patients have similarities with other atypical and viral pneumonia diseases. As a result, machine and deep learning technologies may make it easier to distinguish COVID-19 from other pneumonia infections.

So, motivated by the above-mentioned facts implementation of some deep learning and ensemble approaches were tried in this work to detect COVID-19 from chest X-Ray images (discussed in chapters 3 and 4).

1.4. Scope of the Work

In this work two methods have been tried to predict COVID-19 using chest X-Ray images. In the first method, three deep learning models namely- ResNet50, Densenet169 and InceptionV3 have been trained using the concept of transfer learning. To obtain the better results the outputs of these models are combined using the majority voting ensemble method.

In the second method, DenseNet169 have been selected as the base learner as it gives best result among the above mentioned three models in the first method. Then the concept of snapshot ensembling is used, where DenseNet169 is trained for 125 epochs and 5 snapshot models have been generated. Later these snapshot models have been combined using the weighted average ensemble method.

1.5. Organization of the Project Report

The rest of the report is organized as follows: Chapter 2 presents Related Work. Chapter 3 consists of Deep Models and Classifier Combination Approaches used in the work. In Chapter 4 Proposed Work is discussed. Chapter 5 consists of Experimental Results and Analysis. Chapter 6 concludes the work and discusses some possible future extension of the proposed methods.

Chapter 2

Related Work

Recently, deep learning techniques such as transfer learning and convolutional neural networks (CNNs) are widely used for the classification of COVID-19 from chest X-Rays. In a short span of time researchers have done a lot of work. Various methods such as transfer learning, ensembling, etc. have been proposed in the literature to improve the performance of the DL models.

Khan et al. [15] proposed CoroNet, a deep CNN model based on Xception architecture. The authors used 1251 chest X-Ray images having 4 classes- 310 normal images, 330 Pneumonia Bacterial images, 327 Pneumonia Viral images, and 284 COVID-19 images and have achieved an overall accuracy of 89.6%.

Hussain et al. [16] proposed a novel 22-layer CNN model called CoroDet for automatic detection of COVID-19 by using raw chest X-Ray and CT scan images. They have achieved A classification accuracy of 99.1% for 2 class classification, 94.2% for 3 class classification, and 91.2% for 4 class classification.

Ayalew et al. [17] proposed a convolutional neural network (CNN) and histogram of oriented gradients (HOG) method for feature extraction. Then the features extracted using HOG and CNN are combined to classify using the SVM classifier. At first, they utilized a YOLO detector for the identification of chest X-Ray images and then applied a global thresholding method to segment out the key region from X-Ray images, which is used for feature extraction. They have used AlexNet as CNN and 6000 augmented images (3000 COVID-19 X-Ray images and 3000 Normal X-Ray images) and have achieved an accuracy of 99.67%.

Bargshady et al. [11] proposed the CycleGAN-Inception model. Their approach has two steps- the data augmentation technique, the model generates extra X-Ray images by applying CycleGAN. And then all data transfer to the image pre-processing section for normalization and resizing input data and then, all data transfer to the proposed and modified Inception V3 pre-trained model. They used

9544 X-Ray images of two classes (COVID-19 and Normal) and achieved an accuracy of 94.2%.

Masud et al. [18] proposed a lightweight convolutional neural network (CNN) model to classify COVID and Non-COVID patients. The designed CNN consists of 4 convolutional-max pooling layers and a fully connected layer of 64 units. In the first and second pair of convolutional layers, 32 and 64 kernels (of each size 3x3) are used respectively and the fully connected layer consists of 64 units. The max pooling layer is included after each convolutional layer to reduce the spatial dimension. Also, a dropout of 20% is included after every convolutional pair and a dropout of 50% is included after fully connected layer. They have used 5493 non-COVID and 3914 COVID images and have achieved 92.7% accuracy.

Kumar et al. [19] proposed SARS-Net, a CADx system by combining Graph Convolutional Networks and Convolutional Neural Networks for detecting abnormalities in CXR images for the presence of COVID-19 infection in a patient. They have used 15,254 CXR images (Normal: 8851 images, Pneumonia: 6045 images, COVID-19: 358 images) and have achieved an accuracy of 97.60%.

Bayram et al. [20] proposed a 3-stream fusion CNN-based approach to detect COVID-19 and 3 inputs X-Ray image, HOG image, and LBP image have been utilized. These three inputs are fed to the feature extraction module (FEM) which consists of five feature extraction layers. Each FEM layer consists of a convolutional layer (3x3 sized kernel) and an average pooling layer.

Ozturk et al. [21] proposed a new model for automatic COVID-19 detection using chest X-Ray images. They have used the DarkNet model as a classifier for you only look once (YOLO) real-time object detection. 17 convolutional layers were implemented and different filtering on each layer was introduced. The model provides accurate diagnostics for binary classification (COVID vs. No-Findings) and multi-class classification (COVID vs. No-Findings vs. Pneumonia) and has got an accuracy of 98.08% for binary classes and 87.02% for multi-class cases.

Oyelade et al. [22] proposed a framework named CovFrameNet., which consists of a pipelined image pre-processing method and a deep learning model for feature extraction, classification, and performance measurement. CNN consists of a convolution layer for extraction of features of the image and a fully connected layer to determine which class the input image belongs to the classification of the input image. For classification, they have applied the SoftMax function to the model's feature detection phase. They have achieved an accuracy of 0.1, recall/precision of 0.85, F-measure of 0.9, and specificity of 1.0.

Chapter 3

Deep Models and Classifier Combination Approaches

In computer vision, deep learning-based artificial intelligence studies provide state-of-art solutions. In different disciplines, CNN is now a preferred deep learning method in image recognition applications. Small details which are difficult to notice visually, can easily be distinguished using CNN. With minimal preprocessing, CNNs can recognize visual patterns directly from pixel images. For robust image classification, several CNN models are available. In this work, two methods have been used for COVID-19 prediction from chest X-Ray images, in the first method three models are trained- ResNet50, InceptionV3, and DenseNet169. Then majority voting ensemble method is used to combine the confidence scores of the three pre-trained CNN models to achieve better performance in detecting COVID-19 from CXR images. And in the second method, multiple snapshot models of DenseNet169 are combined by employing an average ensemble method that is aware of different sensitivities of deep learning models on different classes.

3.1. ResNet50

ResNet, which was proposed in 2015 by researchers at Microsoft Research introduced a new architecture called Residual Network. In 2015, it won both ILSVRC and COCO challenges in five main tasks, becoming popular and the new state-of-art. Resnet architecture was first proposed for dealing with the vanishing gradients problem that occurs when training deep networks.

It consists of skip connections in-between consecutive layers (highlighted in figure 3.1). ResNet showed appreciable performance even with a large number of layers due to the addition of skip connections.

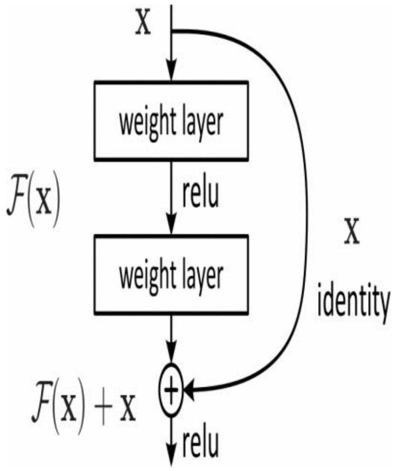


Figure 3.1. A pictorial representation of the skip connections in the ResNet Architecture[23]

An alternative route is provided by identity path for gradients to flow through. ResNet50 has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer and is a variant of the ResNet model.

3.2. InceptionV3

InceptionV3 is a CNN-based network for classification. The network was developed at Google and established new, state-of-art image classification and detection within the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14) [24]. The main feature of the inception model is the inception modules. It is 48 layers deep and uses inception modules, which comprise a concatenated layer with 1×1, 3×3, and 5×5 convolutions. By doing this, the number of parameters can be decreased and the training speed can be increased. Figure 3.2 shows the architecture of the InceptionV3 model.

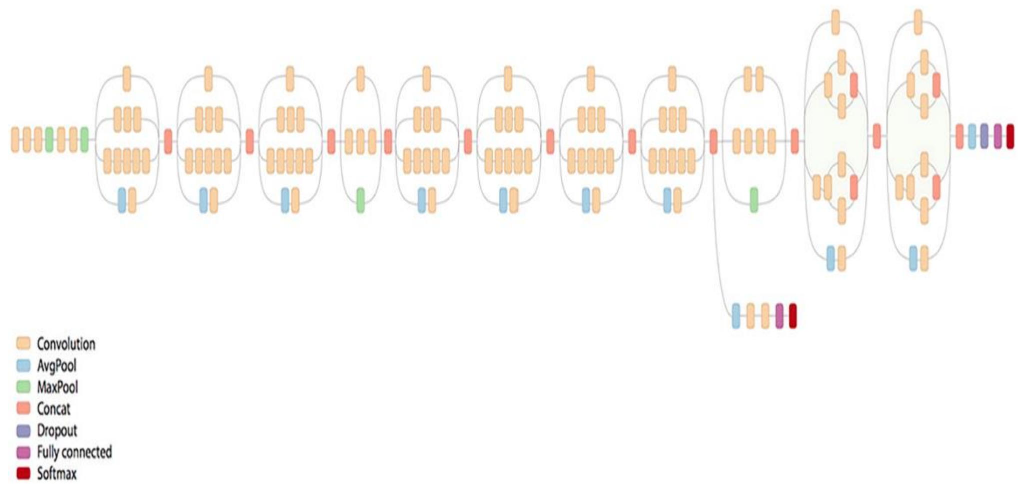


Figure 3.2. Architecture of Inception V3 Model [25].

3.3. DenseNet169

The DenseNet architecture is similar to Resnet architecture. Like ResNet adds skip connections between layers, DenseNet adds dense connections in-between layers. The output of a layer is directly connected to all subsequent layers of the network (highlighted in figure 3.3). These direct connections improve the parameter efficiency while reducing redundancy at the same time. Like ResNet, it also allows an improved flow of gradients through the network.

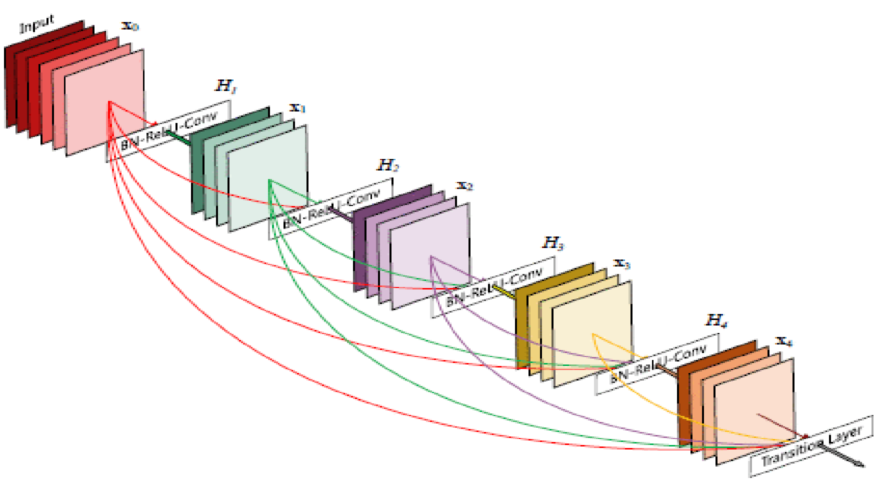


Figure 3.3. A pictorial representation of the dense connections in the DenseNet architecture [26]

3.4. Classifier Combination Methods

A combination or an ensemble of classifiers is a set of classifiers whose individual decisions are combined to classify new examples. Classifier combination is an effective and popular method to improve the predictive performance of classification models and is often much more accurate than the individual classifiers that make them up. The output of a classifier can be represented as a vector of dimension same as the number of classes the classifier is trained to predict. The task of classifiers combination can be seen as a problem of finding a combination function that accepts an N-dimensional score vector from each of M number of classifiers (fig. 3.4) and then produces a single N-dimensional classification score vector and the function should also minimize the number of misclassifications.

In order to combine classifiers output, 2 approaches can be considered to build an ensemble function *f*. The first approach is to run some machine learning

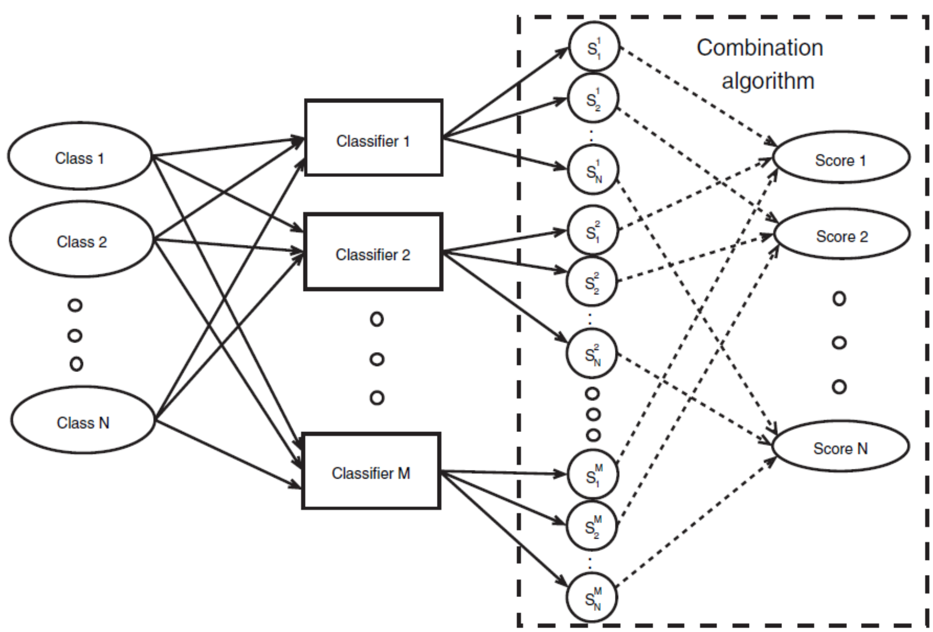


Figure 3.4. A pictorial description of the concept of the classifier combination in general[27]

algorithms on the outputs taken from the classifiers' and generate the final output vector. In other words, the ensemble function works as a secondary classifier that takes the outputs of the primary classifiers as input. Dar-Shyang Lee proposed neural networks [28] to generate the combination output using the output vectors generated from individual classifiers. So, for classification, *f* can be a neural network, support vector machine (SVM) or any other machine learning algorithm. The second approach is to define *f* as functions such as average, weighted average or sum. Therefore, instead of learning from the outputs of the classifiers, it considers the output score or even the output prediction (argmax of the output vector) to generate the combined output vector.

The sum, product, max, min rules, etc., use the output a posteriori probability of the classifiers. The summation of the posterior probabilities is used by the sum rule. Unlike, the sum rule, the product rule quantifies the likelihood of a hypothesis by combining the aposteriori probabilities generated by individual classifiers by means of the product rule. The max rule is an approximation of the sum rule and maximum of the a posteriori probability is taken and the min rule is an approximation of the product rule [29].

3.4.1. Majority Voting

The ensemble method used in the first method of this work is majority voting. Majority voting is a typical example of hard-level combiners and has found widespread use in the literature. Fig 3.5. shows a typical scheme for majority voting [29]. This is a voting method that only considers predicted classes of the classifiers and the most frequent class label from the whole output set is selected as the final output. A major drawback of this method is that it may result in a tie. Though Ho et al. [30] discussed tie-breaking methods, the number of classifiers is generally taken as odd while using this method.

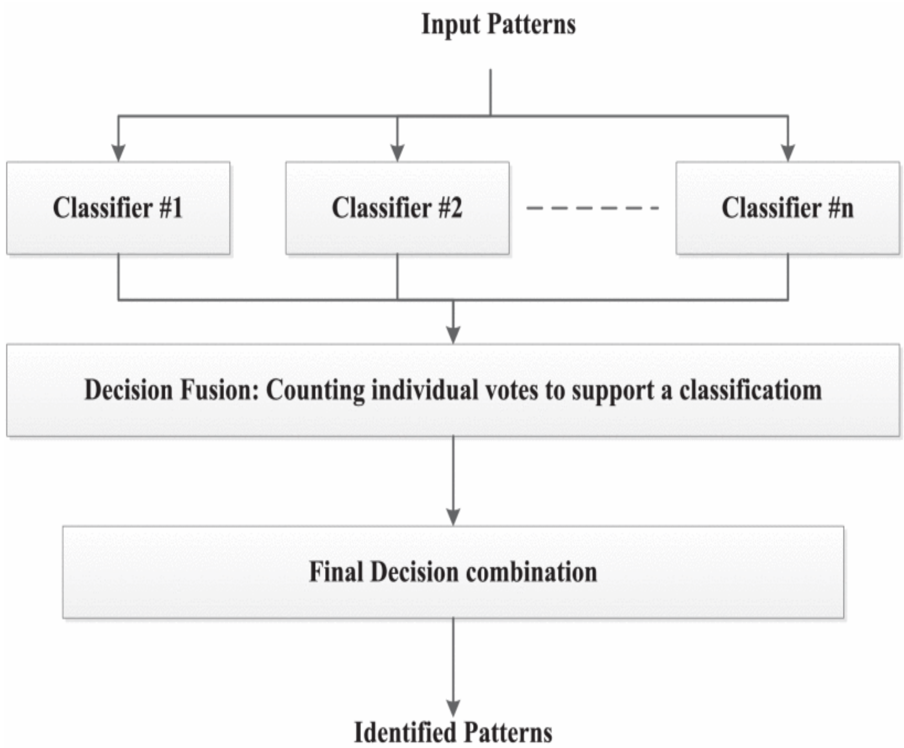


Figure 3.5. An illustration of classifier combination approach using majority voting.[29]

In this work there are four classes and three classifiers, so to break the tie the class with high probability is selected as the final output.

3.4.2. Weighted Average Ensemble (WAE)

The weighted average ensemble method is used in the second approach. A weighted ensemble is an extension of model averaging ensemble in which a member's contribution to the final prediction is weighted by the model's performance. In WAE, the average of the class probabilities for each class from all models respectively is taken as an input. Let N be the number of classifiers. Let $p_{m,k}(d_i)$ be the class probability of the k^{th} class output by the n^{th} classifier with respect to the input sample d_i. Let $a_{i,j}$ be the test accuracy of the i^{th} model for the j^{th} class. Let $w_{i,j}$ denote the normalized weight of the i^{th} model for the j^{th} class. Then we have

$$w_{i,j} = \frac{a_{i,j}}{\sum_{n=1}^{N} a_{n,j}} \quad \ldots\ldots (3.1)$$

For each input sample di, we first get the output of every class probability $p_{n,k}(d_i)$ from the nth model for $\forall n \in N$. Then we can estimate the class probability $p_k(d_i)$ by summing up the weighted class probabilities of all models for $\forall k \in K$.

The model weights are small positive values and the sum of all weights is equal to one, allowing the weights to indicate the percentage of trust or expected performance from each model.

Chapter 4

The Proposed Work

This chapter describes the methods used in this study to predict COVID-19 from chest X-Ray images.

4.1. Ensembling ResNet50, InceptionV3 and DenseNet169

In this method, the dataset was trained on pre-trained deep learning models- ResNet50, DenseNet169, and InceptionV3 (discussed in Chapter 3). The partitioning process in the dataset is considered for thorough training. The dataset was divided into three categories- training dataset, validation dataset, and test dataset consisting of 16,960, 2080, and 2144 CXR images respectively. Every image is resized to 256x256 pixels. All three models were trained and tested separately. The outputs of the classifiers are then combined using the majority voting ensemble method (discussed in Chapter 3). Figure 4.1 shows the workflow of the proposed ensemble method.

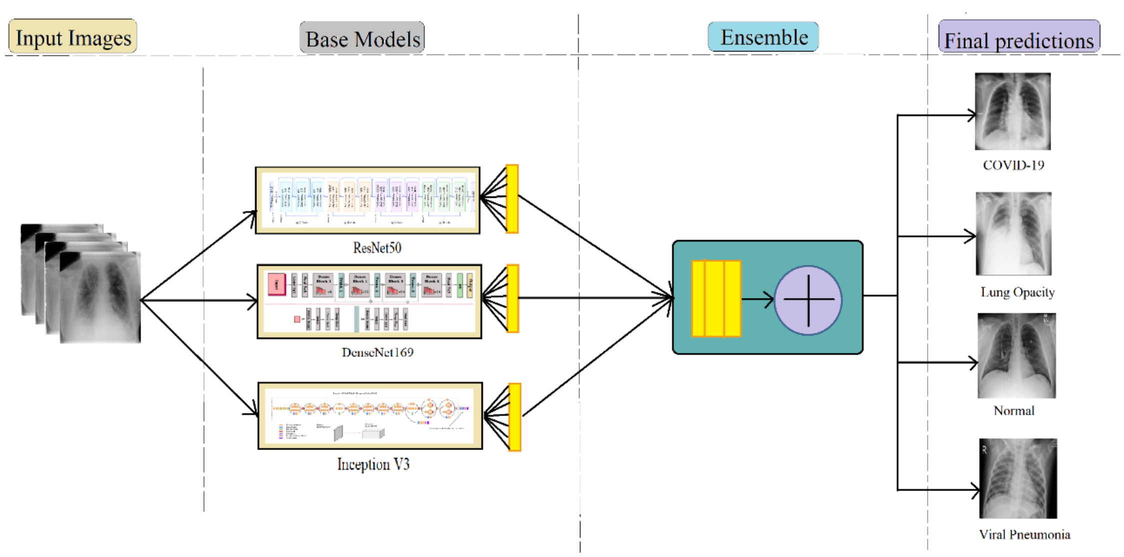


Figure 4.1. Overall Workflow of the proposed ensemble method.

ResNet-50 used in this experiment has over 23 million trainable parameters. Inception V3 has over 21 million trainable parameters and DenseNet169 has over

12 million trainable parameters. The input shape taken in all models is (256,256,1). For each model, the optimizer used is Adam. Each model is trained for 100 epochs with EarlyStopping callback which monitors “val_accuracy”.

The experimental result (Table 5.2.) indicates that the combination of CNN models using majority voting produces better results than the individual models themselves.

4.2. Ensembling Snapshots of DenseNet169

In this work, the method used is similar to the method proposed by Tang et al. [2]. They have proposed EDL-COVID, an ensemble deep learning model generated by combining multiple snapshot models of COVID-Net by employing an average ensemble method that is aware of different sensitivities of deep learning models on different classes. The same approach has been tried on pre-trained DenseNet169 instead of COVID-Net.

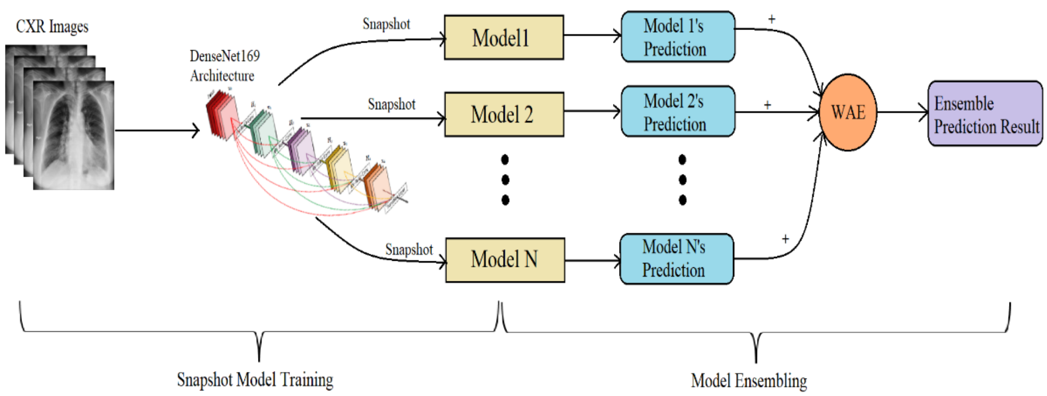


Figure 4.2. Overall flow of Ensembling Snapshots of DenseNet169

DenseNet169 is selected for this work as it gives better results than the three models used in the previous method. The partitioning process and image size are the same as in the previous method. At first, DenseNet169 is trained. The number of epochs taken is 125 and the number of snapshots taken is 5. Snapshot models are saved from a single deep learning network training so the generated snapshots tend to be similar which can produce similar predictions. To overcome this, a common approach used is to take an aggressive learning rate schedule during the training run which makes changes to model weights. To change the learning rate aggressively we have taken the Cosine annealing learning rate schedule proposed by Loshchilov et al. [31]. Different model weights are generated over training epochs in a systematic way by allowing the learning rate to start high and decrease to a minimum value near zero and then again increase to the maximum value based on the following formula [2]-

$$\alpha(t) = \frac{\alpha_0}{2}\left(cos\left(\frac{\pi mod\left(t-1, \lceil T/M \rceil\right)}{\lceil T/M \rceil}\right) + 1\right) \qquad \text{... (4.1)}$$

where α_0 is the maximum learning rate, α(t) is the learning rate at epoch t, M is the number of cycles, and T is the total number of epochs. Then each time a new model snapshot is produced after training a network for M cycles. Here, multiple model snapshots are trained by initializing α0 = 0.04, T= 125, and M = 5 for DenseNet169 network with COVID-19 Radiography Database. 5 snapshot models are obtained from DenseNet169. Now weighted averaging is used to ensemble the outputs of the snapshot models. The pseudocode of the applied Weighted Average Ensemble (WAE) method is given in Algorithm 1.

Algorithm 1: Pseudocode for WAE approach

Input:

N: the number of classifiers/deep learning models
S: Size of the CXR image dataset
K: number of classes
d_i: the i^{th} input sample
$a_{i,j}$: test accuracy of i^{th} model for j^{th} class
$w_{i,j}$: the normalized weight of the i^{th} model for WAE over the j^{th} class prediction ($w_{i,j}$ is calculated using equation i)

Output:

$c(d_i)$: predicted class index for i^{th} input sample
for i=1 to S **do**
  for n=1 to N **do**
    Get output of the nth model w.r.t. to the input sample d_i
  end for
  for k=1 to K **do**
    $P_k(d_i) = \sum_{n=1}^{N} p_{n,k}(d_i).w_{i,k}$

end for

$c(d_i) = \text{argmax}_{1<=k<=K}\{p_k(d_i)\}$

(Get class with maximum class probability for input sample)

end for

Chapter 5

Experimental Results and Analysis

5.1. Dataset used

In this study, the publicly available Kaggle dataset- COVID-19 Radiography Database [32] has been used. A team of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh along with their collaborators from Pakistan and Malaysia in collaboration with medical doctors have created the "COVID-19 Radiography Database" database of chest X-Ray images. The chest X-Ray images in the dataset has posterior-to-anterior (AP)/anterior-to-posterior (PA) view [19]. The dataset comprises a total of 21,164 CXR images consisting of 3,615 COVID-19 CXR images, 10,192 Normal CXR images, 6,012 Lung Opacity CXR images and 1,345 Viral Pneumonia CXR images. Dataset is randomly split into 80% for training (16,960 images), 10% for testing (2,144 images) and 10% for validation (2080 images). Table 5.1 shows the number of images of each class used for training, validation, and testing.

Table 5.1. Dataset Partition

Split	Normal	Viral Pneumonia	Lung Opacity	COVID-19
Train	8,135	1,058	4,862	2,905
Validation	1,059	134	548	339
Test	1,012	154	605	373

5.2. Performance Metrics

In this section, the performance metrics highlighted are used in the current work. Before defining metrics, true positives, true negatives, false positives, and false negatives are defined.

True Positives (TP) are the cases when the actual class of the data point was 1 (True) and the predicted is also 1 (True).

True Negatives (TN) are the cases when the actual class of the data point was 0 (False) and the predicted is also 0 (False).

False Positives (FP) are the cases when the actual class of the data point was 0 (False) and the predicted is 1 (True).

False Negatives (FN) are the cases when the actual class of the data point was 1 (True) and the predicted is 0 (False).

Accuracy is the most common performance metric used for classification algorithms. It may be defined as the number of correct predictions made as a ratio of all predictions made. It can be represented mathematically by equation (5.1).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{5.1}$$

Precision is the ratio of the number of correctly predicted items belonging to a class to the total number of items predicted belonging to the same class. It can be represented mathematically by equation (5.2).

$$Precision = \frac{TP}{TP+FP} \tag{5.2}$$

Recall is the ratio of the number of correctly predicted items belonging to a class to the total number of items belonging to the same class. It can be represented mathematically by equation (5.3).

$$Recall = \frac{TP}{TP+FN} \tag{5.3}$$

F1-Score is the harmonic mean of precision and recall. It can be represented mathematically by equation (5.4).

$$F1\ Score = 2\ \frac{Precision \times Recall}{Precision + Recall} \tag{5.4}$$

5.3. Results and Discussion

In this study, two methods have been used. For both the methods each image of the dataset is resized to 256x256. Also, the train, test and validation sets are the same for both experiments.

In the first proposed method, CXR images are trained on three pre-trained deep models – ResNet50, DenseNet169, and InceptionV3 separately and have achieved testing accuracy of 89.79%, 91.65%, and 91.00% respectively. This is a multi-class classification with the classes- ‘COVID-19’, ‘Lung_Opacity’, ‘Normal’, and ‘Viral Pneumonia’. The model accuracy, precision, and recall calculated are shown in table 5.2. The majority voting ensemble method is applied to the output obtained from the three classifiers. Since there are 4 classes and three classifiers, majority voting can result in a tie. To break this tie, the class with the highest confidence is chosen. Overall accuracy achieved using the ensemble method is 93%. Metrics calculated based on the output of the ensemble are shown in table 5.2. Figure 5.1 shows the confusion matrix of the models used in first proposed method. Figure 5.2 shows some of the Chest X-Ray images from the test dataset with their actual and predicted classes.

Table 5.2. Performance Evaluation of the proposed method 1

Models	Accuracy	Precision	Recall	F1-Score
ResNet50	89.79	90.74	88.79	89.69
InceptionV3	91.00	91.58	90.62	91.09
DeenseNet169	91.65	92.49	91.31	91.88
Ensemble	**93.00**	**93.03**	**94.13**	**93.57**

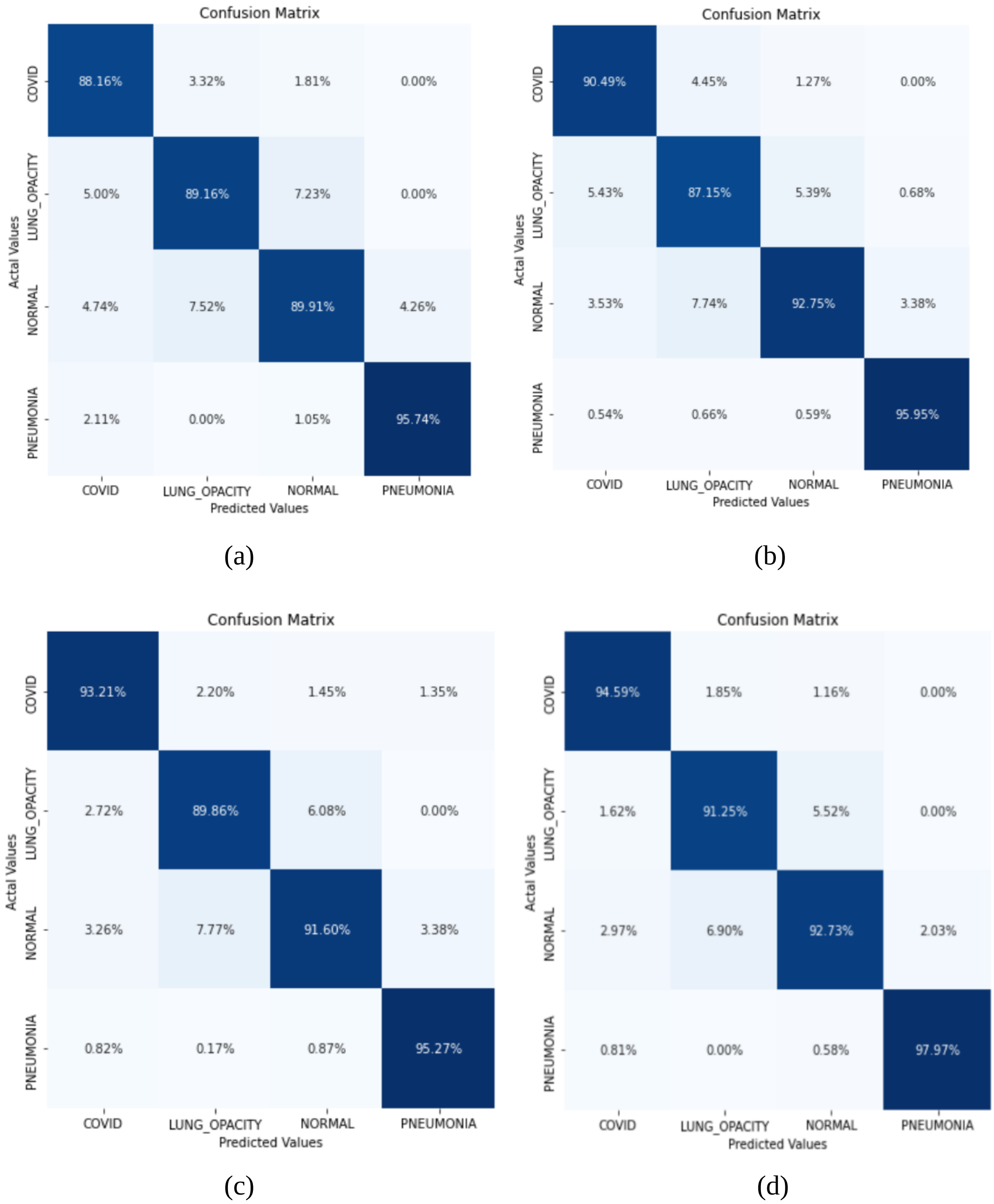


Figure 5.1. Confusion matrix of models used in proposed method 1 (a) ResNet50, (b) InceptionV3, (c) DenseNet169 and (d) Ensemble model

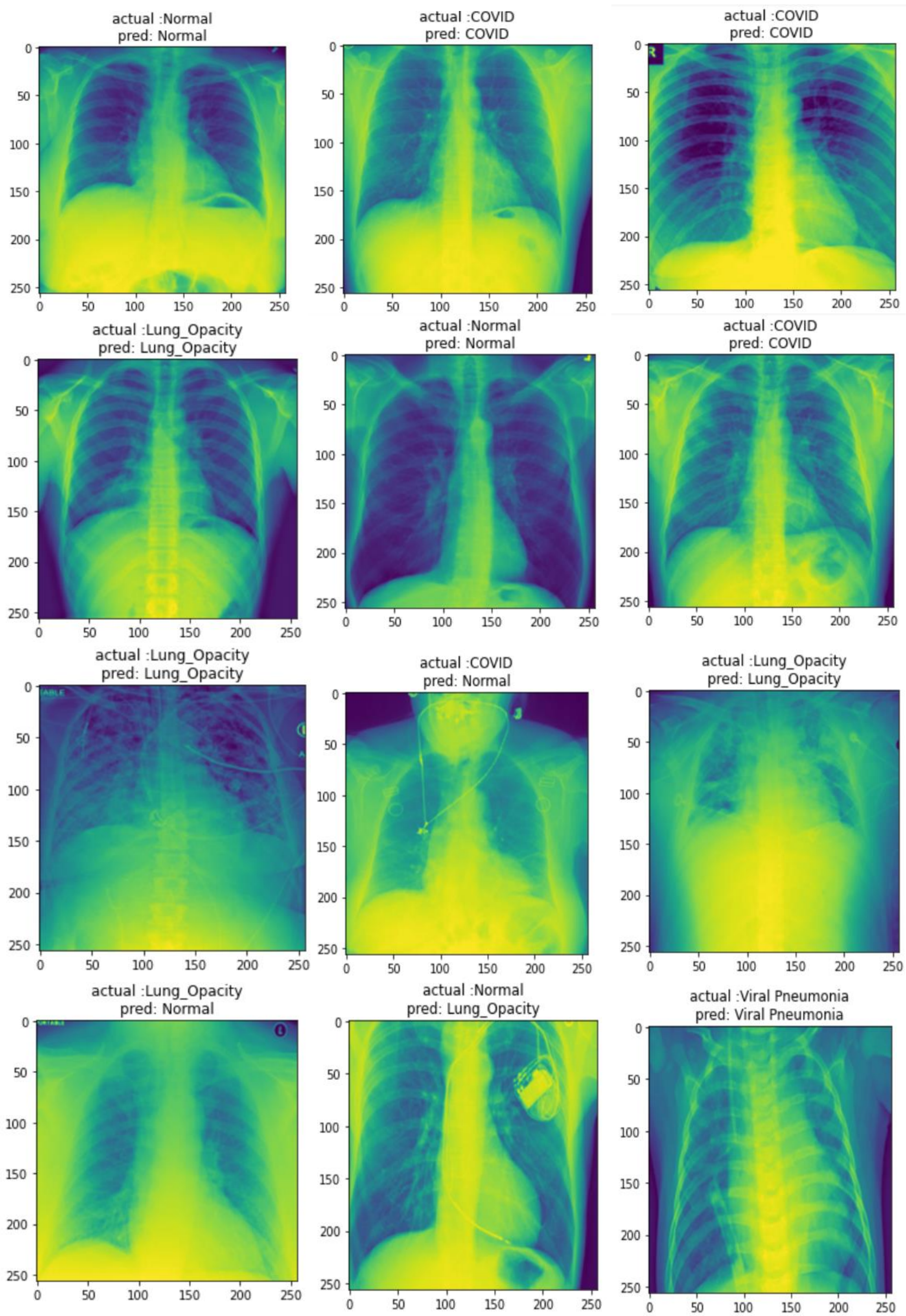


Figure 5.2. Chest X-Ray images from test dataset with their actual classes and the predicted classes of the proposed method 1.

In the second experiment, the dataset is trained on the pre-trained model DenseNet169. The model is trained for 125 epochs with an initial learning rate of 0.04 and 5 cycles. Learning rate is aggressively changed using the cosine annealing learning rate schedule proposed by Loshchilov et al. [31]. Therefore, 5 snapshots are produced from DenseNet169. Outputs of the snapshot models are then ensembled using the WAE method and achieved an accuracy of 92.86%. Tables 5.3 and 5.4 show the Precision and Recall of different models on each class. Table 5.5 shows the accuracy of each model. Figure 5.3 shows the confusion matrix of the ensemble method on test set.

Table 5.3. Precision of different models of the proposed method 2

Model	Normal	Lung opacity	Viral Pneumonia	COVID-19
Snapshot1	92.04	92.09	97.29	94.26
Snapshot2	92.23	92.11	97.29	94.50
Snapshot3	92.23	92.95	97.27	94.24
Snapshot4	92.23	91.81	97.27	94.47
Snapshot5	92.23	91.81	97.27	94.47
Ensemble	92.23	91.96	97.27	94.26

Table 5.4. Recall of different models of the proposed method 2

Model	Normal	Lung opacity	Viral Pneumonia	COVID-19
Snapshot1	94.96	86.61	93.51	96.76
Snapshot2	95.05	86.94	93.51	96.78
Snapshot3	95.05	86.94	92.86	96.51
Snapshot4	95.05	87.11	92.86	96.24
Snapshot5	92.23	87.11	92.86	96.24
Ensemble	94.96	86.94	92.86	96.78

Table 5.5. Accuracy of all models of proposed method 2

Models	Accuracy
Snapshot1	92.82
Snapshot2	92.96
Snapshot3	92.86
Snapshot4	92.86
Snapshot5	92.86
Ensemble	92.86

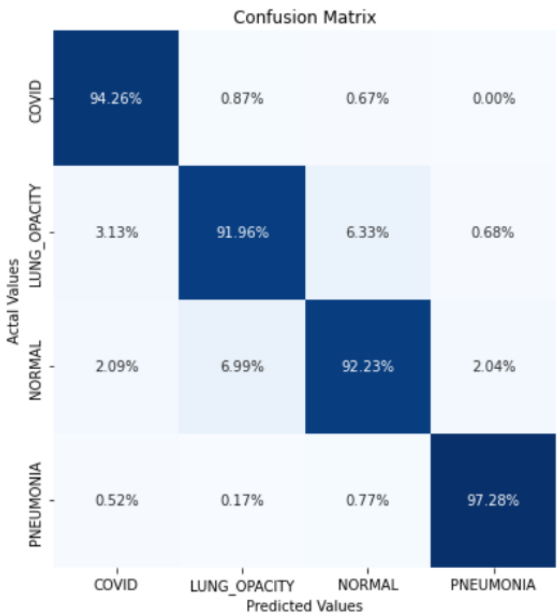


Figure 5.3. Confusion matrix of ensemble of snapshots

Both the proposed methods in this work used ensemble methods but the first proposed method got better results than the second proposed method. This is because the first method does an ensemble of three different models which brings the diversity in terms of the feature maps generated by the models while the second method does an ensemble of 5 snapshots generated from the single deep learning model which might consist of similar feature maps.

Chapter 6

Conclusion

In this work, two methods have been proposed to predict COVID-19 from chest X-Ray images. In the first method, three pre-trained models have been used to train on the "COVID-19 radiography Dataset" and combined them using the majority voting ensemble method. The experimental results indicate that the combination of CNN models using majority voting produces better results than the individual models themselves. In the second method, DenseNet169 has been trained on the available dataset and saved 5 snapshots of the network while training and has used a cosine annealing learning rate schedule to change the learning rate aggressively at each epoch. In section 5.3 it has been shown that the first method provides better results than the second method.

There are various limitations of the present work such as the CNN classifiers used may fail to detect COVID-19 in Chest X-Ray images of the patients in the early stage as minor or no artifacts may be present in the chest X-Ray images which CNN can detect as features. Also, the number of images in the dataset used is less due to which the models learn very limited features.

In future, different ensemble methods like sum rule, product rule, etc. can be implemented on the trained models. Models can be trained on different datasets. Also, to reduce the insufficiency of the data GAN can be used to learn about the data and generate synthetic images that augment the dataset.

Bibliography

[1] Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed. 2020 Mar 19;91(1):157-160. doi: 10.23750/abm.v91i1.9397. PMID: 32191675; PMCID: PMC7569573.

[2] S. Tang et al., "EDL-COVID: Ensemble Deep Learning for COVID-19 Case Detection From Chest X-Ray Images," in IEEE Transactions on Industrial Informatics, vol. 17, no. 9, pp. 6539-6549, Sept. 2021, doi: 10.1109/TII.2021.3057683.

[3] Elaziz, M. A., Hosny, K. M., Salah, A., Darwish, M. M., Lu, S., & Sahlol, A. T. (2020). New machine learning method for image-based diagnosis of COVID-19. PLoS One, 15(6), e0235187. https://doi.org/10.1371/journal.pone.0235187

[4] C. S. K. Raju, K. Pranitha, S. Sundari M, P. Samyuktha and J. Madhumathi, "Prediction of COVID 19- Chest Image Classification and Detection using RELM Classifier in Machine Learning," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), 2022, pp. 1184-1188, doi: 10.1109/ICACCS54159.2022.9785131.

[5] I.D. Apostolopoulos, T.A. Mpesiana, COVID -19: automatic detection from X-Ray images utilizing transfer learning with convolutional neural networks, Phys. Eng. Sci. Med. (2020) 1

[6] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (COVID - 19) using X-Ray images and deep convolutional neural networks, Pattern Anal. Appl. (2021) 1–14.

[7] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G.J. Soufi, Deep-covid: predicting covid-19 from chest X-Ray images using deep transfer learning, Med. Image Anal. 65 (2020), 101794.

[8] M. E. H. Chowdhury et al., "Can AI Help in Screening Viral and COVID-19 Pneumonia?," in IEEE Access, vol. 8, pp. 132665-132676, 2020, doi: 10.1109/ACCESS.2020.3010287.

[9] gifani, P., Shalbaf, A. & Vafaeezadeh, M. Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans. Int J CARS 16, 115–123 (2021). https://doi.org/10.1007/s11548-020-02286-w.

[10] Loey, M.; Smarandache, F.; M. Khalifa, N.E. Within the Lack of Chest COVID-19 X-Ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning. *Symmetry* 2020, *12*, 651. https://doi.org/10.3390/sym12040651

[11] Ghazal Bargshady, Xujuan Zhou, Prabal Datta Barua, Raj Gururajan, Yuefeng Li, U. Rajendra Acharya, Application of CycleGAN and transfer learning techniques for automated detection of COVID-19 using X-Ray images, Pattern Recognition Letters, Volume 153, 2022, Pages 67-74, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2021.11.020

[12] L. Wang, A. Wong, COVID-Net: a tailored deep convolutional neural network design for detection of COVID -19 cases from chest X-Ray images, arXiv preprint arXiv:2003 (2020), 09871.

[13] Afshar P, Heidarian S, Naderkhani F, Oikonomou A, Plataniotis KN, Mohammadi A. COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-Ray images. Pattern Recognit Lett. 2020 Oct;138:638-643. doi: 10.1016/j.patrec.2020.09.010. Epub 2020 Sep 16. PMID: 32958971; PMCID: PMC7493761.

[14] M. Siddhartha, A. Santra, COVIDLite: a depth-wise separable deep neural network with white balance and clahe for detection of COVID-19, arXiv preprint arXiv:2006 (2020) 13873.

[15] A. I. Khan, J. L. Shah, and M. M. Bhat, “CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-Ray images,” Computer methods and programs in biomedicine, vol. 196, no. 105581, p. 105581, 2020.

[16] E. Hussain et al., “CoroDet: A deep learning based classification for COVID-19 detection using chest X-Ray images,” Chaos Solitons Fractals, vol. 142, no. 110495, p. 110495, 2021.

[17] Melese, Aleka & Salau, Ayodeji & Abeje, Bekalu & Enyew, Belay. (2022). Detection and Classification of COVID-19 Disease from X-Ray Images Using Convolutional Neural Networks and Histogram of Oriented Gradients. Biomedical signal processing and control. 74. 103530. 10.1016/j.bspc.2022.103530.

[18] Masud, M. A light-weight convolutional Neural Network Architecture for classification of COVID-19 chest X-Ray images. *Multimedia Systems* (2022). https://doi.org/10.1007/s00530-021-00857-8

[19] Aayush Kumar, Ayush R Tripathi, Suresh Chandra Satapathy, Yu-Dong Zhang, SARS-Net: COVID-19 detection from chest X-Rays by combining graph convolutional network and convolutional neural network, Pattern Recognition, Volume 122, 2022, 108255, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2021.108255.

[20] Bayram, F., Eleyan, A. COVID-19 detection on chest radiographs using feature fusion based deep learning. *SIViP* (2022). https://doi.org/10.1007/s11760-021-02098-8

[21] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, U. Rajendra Acharya, Automated detection of COVID-19 cases using deep neural networks with X-Ray images, Computers in Biology and Medicine, Volume 121, 2020, 103792, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2020.103792.

[22] O. N. Oyelade, A. E. -S. Ezugwu and H. Chiroma, "CovFrameNet: An Enhanced Deep Learning Framework for COVID-19 Detection," in *IEEE Access*, vol. 9, pp. 77905-77919, 2021, doi: 10.1109/ACCESS.2021.3083516.

[23] K.He, X. Zhang, S. Ren, J. Sun, ‘’ Deep Residual Learning for Image Recognition’’, arXiv:1512.03385v1[cs.CV], 10 Dec 2015.

[24] Olga Russakovsky et al. “Imagenet large scale visual recognition challenge”. In: International Journal of Computer Vision 115.3 (2015), pp. 211–252.

[25] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016). Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818-2826

[26] Huang, G., Liu, Z., Van Der Maaten, L.,Weinberger, K.Q., 2017. Densely Connected Convolutional Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269. doi:10.1109/ CVPR.2017.243.

[27] Tulyakov, S., Jaeger, S., Govindaraju, V., Doermann, D. (2008). Review of Classifier Combination Methods. In: Marinai, S., Fujisawa, H. (eds) Machine Learning in Document Analysis and Recognition. Studies in Computational Intelligence, vol 90. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-76280-5_14.

[28]] Lee, D.S., Srihari, S.N., 1995. A theory of classifier combination: the neural network approach, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, IEEE. pp. 42–45.

[29] M. Mohandes, M. Deriche and S. O. Aliyu, "Classifiers Combination Techniques: A Comprehensive Review," in IEEE Access, vol. 6, pp. 19626-19639, 2018, doi: 10.1109/ACCESS.2018.2813079.

[30] Ho, T.K., Hull, J.J., Srihari, S.N., 1994. Decision combination in multiple classifier systems. IEEE transactions on pattern analysis and machine intelligence 16, 66–75.

[31] I. Loshchilov and F. Hutter, “SGDR: Stochastic gradient descent with warm restarts,” 2016, arXiv:1608.03983

[32] Rahman T. COVID-19 radiography dataset. Kaggle; 2020. Accessed 2 June 2020. Available from: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database/version/4