A deep learning approach in detecting COVID-19, pneumonia, and normal cases from Chest X-Ray Images

Larissa Djoufack Basso
Department of Computer Science
Virginia Tech
Falls Church, Virginia
larissadb@vt.edu

Morva Saaty
Department of Computer Science
Virginia Tech
Blacksburg, Virginia
morvasaaty@vt.edu

ABSTRACT
The Coronavirus Disease 2019 (COVID-19) is a pandemic which has negatively impacted the health and well-being of the global population. There have been more than 155 millions cases of COVID-19 and 3 millions COVID-19 related deaths worldwide[8]. Significant efforts have been made by researchers, scientists, government entities,and health professionals to control this pandemic.Early detection via accurate testing and following the numerous COVID-19 prevention methods recommended by either the World Health Organization(WHO) or the Center of Disease Control(CDC) are crucial in mitigating the propagation of COVID-19. As such positive COVID-19 patients can receive proper treatment and isolate to reduce the spread of the disease. COVID-19 is a respiratory disease that attacks the respiratory track especially the lungs. With the daily increase of COVID-19 cases, the need for an instant and accurate detection of positive COVID-19 cases is crucial. Even though Computed tomography(CT) images provide more detailed information than Chest Xray images, Chest Xray images can be easily accessible in either the emergency room or inpatient wards to quickly triage COVID-19 positive patients.There are also a subset of patients that test positif for both COIVD-19 and pneumonia a phenomenon known as COVID-19pneumonia and can be severe [**?**] . Despite the exponential increase of respitorary disease especially COVID-19 a shortage of radiologists had been observed. In the United States,the shortage of radiologists could reach 42,000 by 2033 [**?**].With the current shortage of radiologists and the year-long health crisis the world has faced,we were motivated to build a multi-label classification model to detect COVID-19, pneumonia, and normal cases on chest x-ray images.We built three different models, and at the base of those models we used well known image classification pre-trained model such as VGG16, Resnet50, and densenet121 for tranfer learning and added five more layers as proposed in the Covnet architecture litterature[22]. We observed that the nCOVnet model worked best with Resnet50 as the transfer learning model compared to the one proposed in [22].As a result we were able to predict covid-19, pneumonia, and normal cases on chest xray images with high accuracy and other performance metrics.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

Author Keywords
COVID-19, Pneumonia, Chest X-Ray, Neural Network Models, Transfer Learning

INTRODUCTION
On December 2019, an outbreak of COVID19 was reported in Wuhan, China. On March 11th 2020, the WHO declared the novel coronavirus a global pandemic. The COVID-19 pandemic has become a leading cause of world-wide mortality rate, and negatively impacted patients' health. COVID-19 affects the patients' lungs by entering the human's respiratory tract and causing lung inflammation. Therefore, it is crucial to quickly and accurately detect COVID-19 cases [14]. It is important to note that the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the virus which causes COVID-19.
Of the various SARS-CoV-2 diagnostic testing approaches, some major ones are reverse transcript polymerase chain reaction(RT-PCR),imaging testing, and serology testing[14]. The standard and recommended test approach by the Center of Disease Control(CDC) is RT-PCR. RT-PCR is a nasopharyngeal swab used to detect certain segments of the SARS-COV-2 genome[21].Although RT-PCR is the preferred testing method, it can be time consuming, expensive, and supply dependent. Further, many factors can negatively impact the accuracy of RT-PCR. COVID-19 screening with radiography images is often conducted and analyzed by radiologists who look for visual indicators of COVID-19 positive cases. Some of the visual indicators used are: bilateral abnormalities in CXR images, ground-glass opacity, and interstitial abnormalities in CXR and CT images. The serologic testing consist of drawing blood samples and testing specific IgM and IgG to SARS-COVID-2. However, a positive result via serologic testing indicates that the tested subject was exposed to COVID-19 and does not necessarily mean that the subject is COVID-19 positive.
Image classification, recognition, and analysis are some of the key areas in which deep learning excels in and can be applied to many fields including the medical field. Further, X-ray images are used to scan various human organs in hospitals, and the interpretation of those images is often performed by radiologists. Many deep learning methods have been explored by scientists and researchers to automatically detect Covid-19 from Chest X-Ray images. However, the very small amount of available dataset has made the task of automatically detecting

COVID-19 cases a challenge. Also, some COVID19 patients can test positive for pneumonia. There have been many work and research conducted in pneumonia detection on CXR images. There is also an abundant amount of pneumonia and chest xray images freely available to the general public.
Motivated by the need to accurately and automatically detect positive COVID-19, pneumonia and normal cases in chest x-ray images, we propose CovNet models, including CovNet-VGG16, CovNet-ResNet50, and CovNet-Densenet121, which are multi-label models used to determine COVID-19, pneumonia and normal cases from chest xray images. The major contribution of this paper is as follows:

- We implemented the nCovnet model [14] by applying some changes in the model and compared it with our implementation of two other similar models but with different pre-trained based models (DenseNet-121, and ResNet50). We observed that using Resnet-50 as base performed better than the originally proposed nCOVnet model.
- We detected three different cases: COVID-19, pneumonia, and normal cases by using the CovNet models with an accuracy around 98%.
- We used F-1 score, ROC-curve, AUC score, precision, recall, and accuracy to evaluate our model performance.
- We showed the importance of data augmentation in small and imbalanced data.

The remainder of our paper will be structured as follows: First, we conduct some literature reviews on previous work. Second, we go over our dataset and the preprocessing techniques we have applied on the data. Next, we introduce our models and the experimental evaluation that we have done to measure the performance of the three models we implemented.

RELATED WORK

With the constant improvement in deep learning models, in the architecture design of deep learning models, literature review of previous or related work is crucial. Despite the novelty of COVID-19, many models have been proposed to detect COVID-19 cases from chest X-ray(CXR) images. We will review previous deep learning work in detecting images in general and COVID-19 cases from CXR images in particular.
Panwar et al. proposes nCOVnet which is a model whose architecture is composed of VGG16 as the base model to feature extraction in the transfer learning portion and the chest xray images are then passed into their proposed extra five layers architecture to detect covid-19 related indicators from CXR images. This model consists of 24 layers. The first layer takes as input an RGB image of dimension 224 x 224 x 3. The following 18 layers are a combination of 2d Convolutional Neural Network plus ReLU and Max Pooling. VGG-16 is a pre-trained image classification model that was originally trained on ImageNet and whose architecture includes the structure of the 19 layers aforementioned. In the 5 subsequent layers, the first layer is a 2d average pooling layer of size(4,4). The authors used fully connected layer with 64 units and a Rectified linear activation function (ReLU).They applied dropout with the threshold of 0.5. This model was then trained on sets of COVID19 CXR images [14].
Wang et al. proposed COVID-Net network architecture. ImageNet was the pretrained dataset of this model and the COVIDx dataset was later used for training with the Adam optimizer as the optimizer method. This model combines CNN, VGG19, and ResNet50. It extensively used a projection-expansion-projection design pattern in the COVID-Net architecture. Such utilization helped enhance the representational capacity and maintained the same computing resources efficiency. [22].

To detect COVID-19 on frontal chest radiographs, study [25] used deep CovidXR. Deep CovidXR is a weighted ensemble of Convolutional neural networks(CNN). The weighted ensemble of CNNs is composed of a sequence of the following pre-trained models: DenseNet-121, ResNet-50, InceptionV3, Inception-ResNetV2 , Xception , and EfficientNet-B2. In Figure 1, the input image to this model is first preprocessed to produce four separate images (cropped and uncropped images at resolutions of 224 x 3 x 224 pixels and 331 x 3 x 331 pixels). In the preprocessing stage, a set of images with and without a cropped square centered on the lung were downsampled to two distinct resolutions. Those resolutions are: 224x3x224 pixels and 331x3x331 pixels. These images were then passed in each of the six different validated CNN architecture. The data set used in this paper is a collection of different data collected from more than twenty well known institutions in the field to include the Northwestern Memorial Western Health system.

In [23], the authors proposed a human-machine collaborative strategy which uses deep neural network architecture to detect COVID-19 cases from chest X-RAY images. This model uses a lightweight projection-expansion-projection-extension (PEPX) design to reduce the need for high computational power and complexity. The paper also, introduced a large COVID-19 dataset which encompassed a diverse data of positive COVID-19, pneumonia, and neutral cases from chest X-Rays. Further, COVID-net learns features correlated to COVID-19 and provides a transparency where radiologists get to evaluate the performance of the model. COVID-net is an ensemble of lightweight residual (PEPX) design pattern. This model to get a lower dimension of input features, used first-stage projection by using 1 x 1 convolutions. Then, the model used an expansion of 1 x 1 convolutions to get higher dimensions, which are different from input features. The model also used an powerful 3D depth-wise CNN for learning spatial characteristics and reducing computational complexity while maintaining representational capacity. Second-stage of the model includes projection 1 x 1 convolutions for going back to lower dimension of features. Moreover, the model expand channel dimensionality to a higher dimension and generate the final features, use an 11 convolution extension.

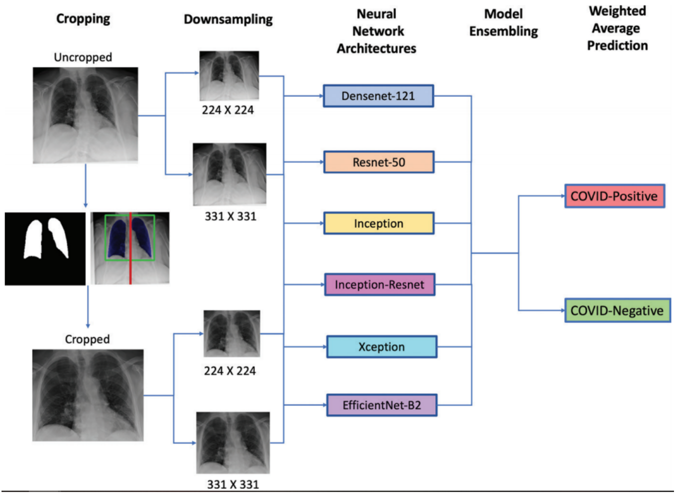

Figure 1. Overview of DeepCoviDXR model

METHOD AND RESULTS

Dataset

In the data gathering processing, our goal was to collect data available to the general public and that do not violate any of the the Health Insurance Portability and Accountability Act of 1996 (HIPAA) laws.As such, we collected data from numerous publicly available sources. The first dataset that we used is a collection of chest-xray images provided in kaggle[11]. This data set is a collection of both pneumonia and normal chest Xrays. Each data category is assigned to its labeled folder. In kaggle, this data page has been viewed by more than one million different users and has been downloaded more than one hundred thousand times. The pneumonia dataset is comprised of both viral and bacterial pneumonia. This original dataset is comprised of 5863 chest xray images. However, due to our limited computational resource we could not process all of this data and thus we only used 82 normal data cases and 390 datasets on pneumonia. The chest xray images were performed as part of patients routine medical care. The autors removed all low quality and unreadable chest-xray images before publishing this data. Three expert physicians verified the diagnosis of each images before clearing data for use in AI system research.The images in figure 2 is a screenshot which represent a subset of the data set for both pneumonia and normal cases.[11]

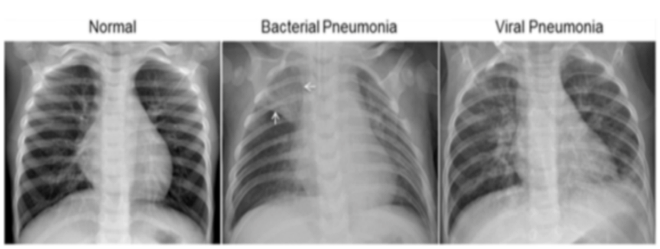

Figure 2. Image of Normal and Pneumonia in the dataset

The second data that we used is a collection of chest X-ray images of patients that tested positive for either COVID-19 and viral or bacterial pneumonia[5]. It is a dataset available to the general public. The data is a collection of data gathered from numerous online resources to include the Italian Society of Medical and In-terventional Radiology, Radiopedia.org, and numerous online publications[5].It is also comprised of three different chest xray images views: anteroposterior, posteroanterior, supine anteroposterior. Anteroposterior(AP) images are images taken from front to back. The posteroanterior(PA) image is a filming strategy where the xray beam are filmed from the back of the patient to the patient's front. The supine AP is often the prefered filming approach for patients unable to stand up.There are 25 different labels in this dataset. In each image for the different diseases in the labels, 1 is used to indicate positive case of that disease while 0 indicate negative cases. The labels repartition is highlighted in Figure 3 [5]. Our dataset is more imbalanced towards pneumonia classes. The first data that we collected only contained both viral and bacterial cases. However, in the second dataset, some chest xray images were positive for both pneumonia and covid-19, thus the imbalance.

Type	Genus or Species	Image Count
Viral	COVID-19 (SARSr-CoV-2)	468
	SARS (SARSr-CoV-1)	16
	MERS-CoV	10
	Varicella	5
	Influenza	4
	Herpes	3
Bacterial	*Streptococcus* spp.	13
	Klebsiella spp.	9
	Escherichia coli	4
	Nocardia spp.	4
	Mycoplasma spp.	5
	Legionella spp.	7
	Unknown	2
	Chlamydophila spp.	1
	Staphylococcus spp.	1
Fungal	*Pneumocystis* spp.	24
	Aspergillosis spp.	2
Lipoid	Not applicable	8
Aspiration	Not applicable	1
Unknown	Unknown	59

Figure 3. Label repartition in the COVID-19 dataset

Preprocessing

- Step 1.Reduce the size of the data. Due to computational limitations such as low RAM to process the data set and the various models we used, we had to reduce the size of the data.We only collected data with PA views, we reduced the size of both pneumonia and normal dataset but we included all the COVID-19 data from our data in PA view.

- Step2. Convert the images to arrays and then convert them to RGB images. The data was in grey scale images.

- Step3. Resize the image to dimensions 224x224. Due to the various imagery size in our dataset, we decided to convert all of our data to 224x224.

- Step4. Randomly flip the images horizontally with a probability of 50%

- Step 5 randomly rotated data at by 20 degrees

- Step5. Normalize the entire data

After preprocessing our data,figure 4 shows our data repartition.

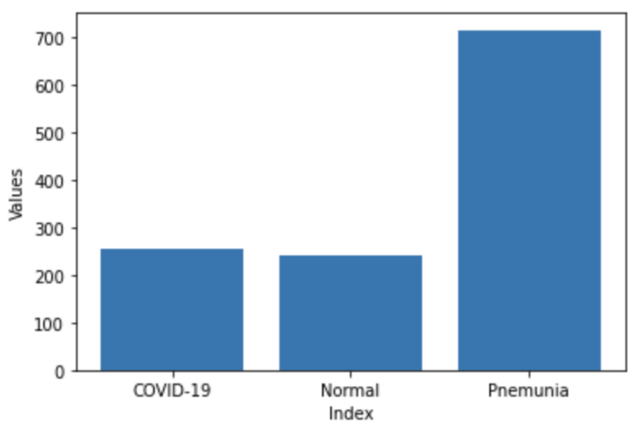

Figure 4. Dataset repartition

A sample of images of our data composition [15] can be seen on figure 5

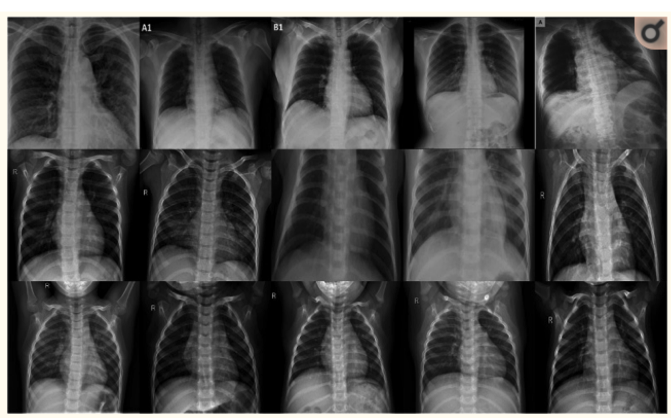

Figure 5. Chest-Xray Images: COVID-19 (row1), viral pneumonia(row2), normal(row3)

Data augmentation

After a thorough analysis of our data we observed that we had a small and imbalanced dataset.As such we augmented our data through a series of different techniques. Those are:

- Randomly flip the images horizontally with a probability of 50%

- We randomly rotated data by 20 degrees

Figure 6 is a snapshot of our data after the aforementioned transformations and preprocessing steps. The augmentation and normalization of our data greatly improved our results. Throughout this data we will show the results of our model before and after data augmentation.

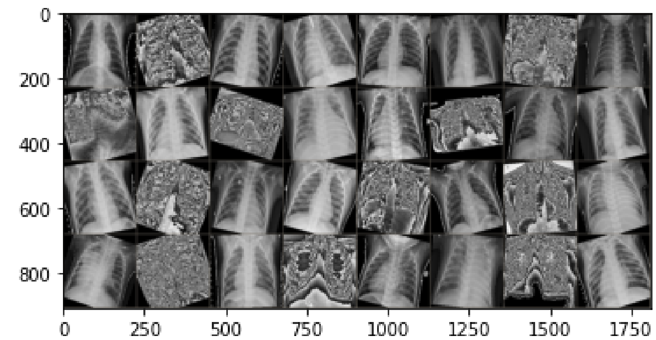

Figure 6. Data augmentation visualization

Model

In our model design process, in view of the nature of our collected data, we designed models compatible with our data. Since each image is associated with a multi-label list where 1 represents a positive case for a specific disease while a 0 represents a negative case, we realized that we were facing a multi-label problem. The difference between multi-class problems and multi-label problems is that in multi-class problems, the classes are mutually exclusive. Whereas, in multi-label problems, however each label indicates a different classification task consequently, the tasks are related in some ways. Therefore, it is better to deal with classes together rather than separately [7]. Patients can have many respiratory diseases at the same time. It is not uncommon to test positive for both the new coronavirus(COVID-19) and pneumonia. In our dataset, we have numerous chest-xray images with positive cases of both pneumonia and COVID-19. COVID-19 with pneumonia can be fatal [24]. Accurate and fast detection of these two diseases can be a life changing process which will enable patients to get treatment early on at the onset of the disease.

In this paper, we used the model proposed in [14] with some changes we will elaborate later. The proposed model is comprised transfer learning as our base model and Convolutional Neural Networks (CNN). Transfer learning is used to improve the performance of the model and speed up training. Transfer learning is a machine learning method that was created for one task, but it can be reused in other tasks such as in the beginning of a model and feature extractions [2].

We have used VGG16, DenseNet-121, and ResNet-50 for

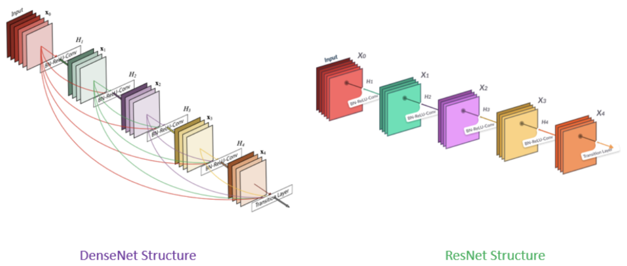

Figure 7. The structure of DenseNet and ResNet models [19].

transfer learning. The input layer of all the previously mentioned models get data of fixed size of 224 x 224 x 3 pixels, making it an RGB image.

- **VGG16** is a neural network that performed very well in the ImageNet. This model is a combination of Convolution + ReLU and Max Pooling layers and ends with a fully connected layer.
- **ResNet-50**, Residual Network, was the winner of ImageNet challenge in 2015[9]. ReNet first introduced the concept of skip connection [9], which avoids the problem of vanishing gradients. ResNet-50 includes 48 convolution layers along with one Max pool and one Average pool layer, and a fully connected layer at the end.
- In a **DenseNet-121** architecture, each layer is connected to every other layer, and the concatenation of feature maps from previous layers is the input of a layer. Each dense layer includes a combination of Batch Normalization, and Convolution + ReLU. Besides, this architecture also contains transition layers, which are included in Batch Normalization, Convolution + ReLU.Average Pooling, takes care of downsampling [17]. There is a fully connected layer at the end of DenseNet model for the classification purpose.

The model architecture of the DenseNet and ResNet models is indicated in Figure 7 [19].

ImageNet is a large visual database, which includes more than 14 million annotated high-resolution images with more than 20000 categories[12]. It is often considered to be the benchmark dataset in image classification model implementation. The images from ImageNet were originally collected from numerous online sources and manually labeled by humans with Amazon's Mechanical Turk crowd-sourcing tool. VGG16 was able to achieve a 92.7% accuracy, ResNet-50 achieved the accuracy of 92.9 %, and DenseNet-121 could achieve the accuracy of 91% when trained on ImageNet.

Proposed Model

In view of the pretrained image classification models we previously discussed, we designed a model, called CovNet. We used VGG16, ResNet-50, and DenseNet-121 in the first part of CovNet model to extract features. Based on the various sizes of the data we collected, we decided to downsample all of our images to dimensions 224x224 and none of our original image size was below 224x224. In our CovNet model

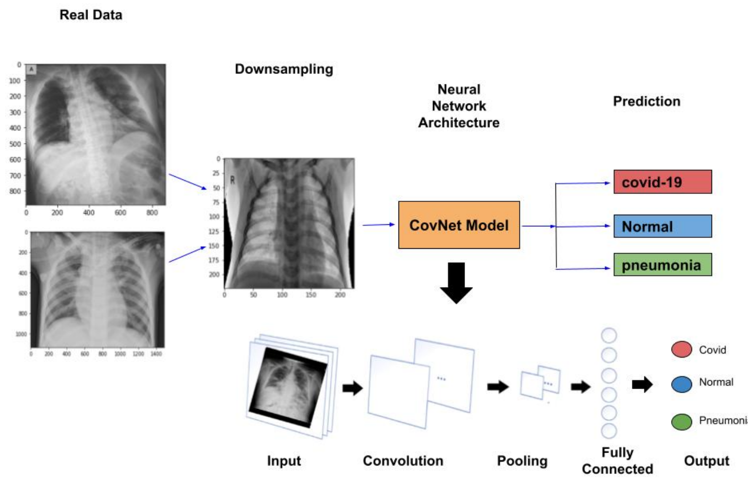

Figure 8. Overall CovNet architecture for classification and detection of COVID-19, Pneumonia, and Normal cases.

implementation, we passed a 32 batch images of dimensions 224 x 224 x 3 pixels. We first built a transfer learning model using 5 custom different layers for the classification purpose, and trained the model on Normal, Pneumonia, and COVID-19 datasets. Figure 8 represents the overall CovNet architecture for classification and detection of COVID-19, Pneumonia, and Normal cases.

We decided to choose the five layers approach(layers added at the end of the transfer learning base model) previously mentioned based on a model proposed in the paper [14]. In that paper they realized a training accuracy around 93-97%. These five layers contain an Average Pooling 2D layer with pool size of (4,4). Average pooling selects an average value of all the pixels in the batch, while Max pooling selects the maximum value of all the pixels in the batch. Then, we used a flatten layer to convert the two dimensional matrix into a vector. This vector will be fed into a fully connected neural network classifier within 64 units and a ReLU activation function. Next, we used dropout layer with a threshold of 0.5. Dropout simply ignores certain random units, taking in consideration of the threshold value, during training phase, to prevent neural networks from over-fitting. In fact, these units were not considered during a specific forward or backward pass[4]. Finally, our output layer is comprised of 25 units, which is equal to the size of our images' labels, and a Sigmoid activation function. We picked Sigmoid as an activation function, because it works well for multi-label problems and returns the value in the range of [0,1]. For instance, the summary of CovNet using VGG16 as a base model for transfer learning and five custom layers as head model is shown in Figure 9.
To better understanding what is happening through the mentioned CNN models layers by layers, we represented the features map of the first layer of CovNet-ResNet50, CovNet-VGG16, and CovNet-DenseNet121 respectively shown in Figures 16, 17, and 18 in the Appendix section.

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1         [-1, 64, 224, 224]           1,792
              ReLU-2         [-1, 64, 224, 224]               0
            Conv2d-3         [-1, 64, 224, 224]          36,928
              ReLU-4         [-1, 64, 224, 224]               0
         MaxPool2d-5         [-1, 64, 112, 112]               0
            Conv2d-6        [-1, 128, 112, 112]          73,856
              ReLU-7        [-1, 128, 112, 112]               0
            Conv2d-8        [-1, 128, 112, 112]         147,584
              ReLU-9        [-1, 128, 112, 112]               0
        MaxPool2d-10          [-1, 128, 56, 56]               0
           Conv2d-11          [-1, 256, 56, 56]         295,168
             ReLU-12          [-1, 256, 56, 56]               0
           Conv2d-13          [-1, 256, 56, 56]         590,080
             ReLU-14          [-1, 256, 56, 56]               0
           Conv2d-15          [-1, 256, 56, 56]         590,080
             ReLU-16          [-1, 256, 56, 56]               0
        MaxPool2d-17          [-1, 256, 28, 28]               0
           Conv2d-18          [-1, 512, 28, 28]       1,180,160
             ReLU-19          [-1, 512, 28, 28]               0
           Conv2d-20          [-1, 512, 28, 28]       2,359,808
             ReLU-21          [-1, 512, 28, 28]               0
           Conv2d-22          [-1, 512, 28, 28]       2,359,808
             ReLU-23          [-1, 512, 28, 28]               0
        MaxPool2d-24          [-1, 512, 14, 14]               0
           Conv2d-25          [-1, 512, 14, 14]       2,359,808
             ReLU-26          [-1, 512, 14, 14]               0
           Conv2d-27          [-1, 512, 14, 14]       2,359,808
             ReLU-28          [-1, 512, 14, 14]               0
           Conv2d-29          [-1, 512, 14, 14]       2,359,808
             ReLU-30          [-1, 512, 14, 14]               0
        MaxPool2d-31            [-1, 512, 7, 7]               0
           Linear-32                   [-1, 64]       1,605,696
             ReLU-33                   [-1, 64]               0
          Dropout-34                   [-1, 64]               0
           Linear-35                   [-1, 25]           1,625
================================================================
Total params: 16,322,009
Trainable params: 1,607,321
Non-trainable params: 14,714,688
----------------------------------------------------------------
Input size (MB): 0.57
Forward/backward pass size (MB): 218.40
Params size (MB): 62.26
Estimated Total Size (MB): 281.23
----------------------------------------------------------------
```

Figure 9. CovNet Architecture with VGG16 as base model and extra five custom layers as main model.

The model used Adam optimizer [11] for optimization. Adam incorporates the best features of the AdaGrad and RMSProp algorithms to build an optimization algorithm for noisy problems with sparse gradients [3] . Furthermore, because of the binary nature of our labels (value of our images' labels are 0 or 1), we used BCELoss() function to measure the Binary Cross Entropy between the predicted labels and annotated labels [16].

EXPERIMENTAL EVALUATION

We explored various pretrained state-of-the-art models in deep learning for image detection and classification. To evaluate the performance of the proposed models, we have used metrics such as Accuracy, Precision, Recall, and F1-score, Area Under Curve (AUC) of Receiver Operating Characteristics (ROC), and plotted the training and validation loss for the CovNet models.

Confusion Matrix

Confusion matrix is a measurement to investigate the effectiveness of models for deep learning classification. Confusion matrix is a square matrix where each cell represents the primary parameters known as True Positive (TP), False Positive (FP), False Negative (FN), and True Negative(TN). For instance, TP refers to the cases which are truly positive and also classified as positive, and FN refers to the cases which are truly positive, but classified as negative.
This matrix is useful in calculating Recall, Precision, Accuracy, F1-score, and most importantly to plot the ROC Curve and get the AUC of that curve.

- **Accuracy** determine the number of cases that are correctly identified, but it is not a good measure for scenarios with data imbalance among classes.

- **Precision** refers to the portion of predicted labels that are truly positive. Precision should ideally be near one or 100% for a good classifier.

- **Recall** calculates the number of the actual positives that have been predicted as positive by the model. An ideal recall score also should be near one for a good classifier. In fact, Recall is a metric that indicates how well our model can classify relevant data. It's also known as True Positive Rate or Sensitivity [18].

- **F1-score** is a harmonic mean of Precision and Recall, and it provides a more accurate picture of cases that were wrongly classified than the Accuracy Metric. F1-score might be a better measure to use if there is an uneven class distribution[10].

Based on [10], When True Positives and True Negatives are more significant, Accuracy is used, while F1-score is used when False Negatives and False Positives are critical. While testing our model, we had the following metrics: sensitivity of 86.64% and precision of 88.43% for the ResNet-50 model in real data. The results of Accuracy, Precision, Recall, and F1-score of CovNet models based on real data are indicated in Table 1.
But, we also applied data augmentation to our raw data and

Models	Accuracy	Precision	Recall	F1-score
VGG16	98.08 %	86.98 %	84.88 %	85.92 %
ResNet-50	98.30 %	88.43 %	86.64 %	87.53 %
DenseNet-121	91.62 %	78.72 %	43.34 %	55.91 %

Table 1. Classification Reports for detection of COVID-19, Pneumonia, and Normal Cases in Real Data

Models	Accuracy	Precision	Recall	F1-score
VGG16	98.27 %	87.60 %	86.89 %	87.24 %
ResNet-50	98.38 %	88.64 %	87.55 %	88.09 %
DenseNet-121	98.12 %	85.95 %	86.13 %	86.04 %

Table 2. Classification Reports for detection of COVID-19, Pneumonia, and Normal Cases after Using Data Augmentation

trained the model based on augmented data. We achieved an extremely better performance and increasing value in CovNet-Dense121 models in respect to passing real data to this model. We also got accuracy around 98%, Recall around 87%, and F1-score around 87% for all of the CovNet models. The classification reports of augmented data after passing to CovNet models can be seen in Table 2.

These value are indicative of the great performance of our model. This implies that we can detect COVID-19, Pneumonia, and Normal cases based on X-Ray images with 13.36% error regarding using real data and with 12.45% error regarding using augmented data. By calculating the accuracy of ResNet-50 model, we have achieved an overall accuracy of 98.38% , and F1-score of 88.09 % considering the uneven distribution of classes, and using augmented data.

Receiver Operating Characteristics (ROC)

ROC is a 2-dimensional graph showing the performance of a classification model at different threshold settings. A ROC curve plots the True Positive Rate (TPR), and False Negative Rate (FNR). In fact, ROC is a probability curve, and it can be defined as the trade-off between sensitivity and specificity [20]. We have plotted the ROC curve with False Positive Rate on the x-axis and True Positive Rate on the y-axis.

Area Under a Curve (AUC) calculation uses Trapezoidal rule, which is a integration numerical method for calculating the integral or region under the surface of a ROC curve [26]. AUC indicates how well the model can differentiate between classes [13]. In other words, the AUC represents how well the model predicts 0s as 0s and 1s as 1s. The AUC value ranges from 0.5 to 1, with 0.5 indicating a poor classifier and 1 indicating an outstanding classifier. For instance, when AUC is about 0.5, the model is unable to differentiate between positive and negative classes [13].The AUC of two or many models trained on the same data can be used as a tool to determine the best performing model overall or the best performing model based on certain values of the false positive rate. To conclude that a classifier is bad, good, or excellent, we need to consider the context and importance of the classification in that area. In medical diagnosis, we need great AUCs, ideally moran 0.90 [1].

The plotted ROC for our CovNet models, in two conditions of passing real data or augmentated data to the models,is shown in Figure 10. According to Figure 10, each row is dedicated to each CovNet models and left column is related to Real Data

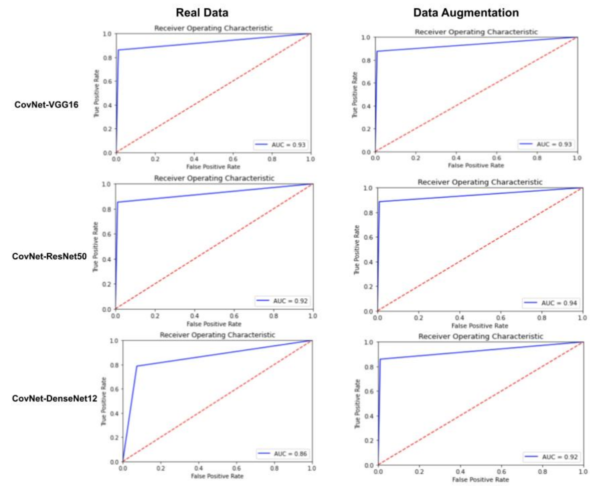

Figure 10. ROC curve Comparison between CovNet models for real data and after using Data Augmentation .

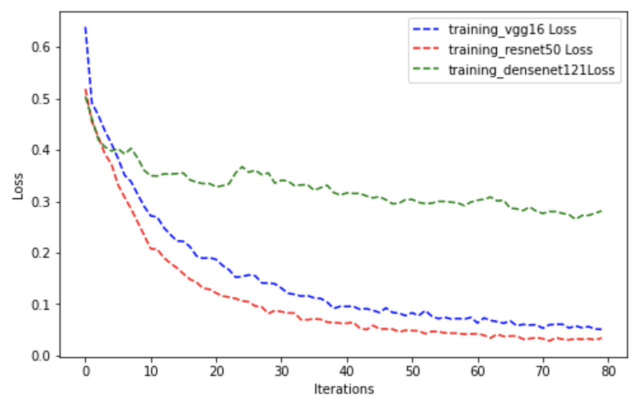

Figure 11. Training curve of loss for CovNet models.

and right column is related to Augmented data. Based on our findings, AUC of ROC is above the threshold for all CovNet models after and before using data augmentation. AUC is 0.92 for CovNet-VGG16, 0.93 for CovNet-ResNet50, and it is 0.86 for CovNet-DenseNet121 in the case of passing real data to the models. Whereas, the AUC values increased to 0.93 for CovNet-VGG16, 0.94 for CovNet-ResNet50, and 0.92 for CovNet-DenseNet121 when using augmented data. Therefore, CovNet-ResNet50 and CovNet-VGG16 are classified as "excellent" classifier in the field of medical diagnosis.

Training Loss

We have trained the customized models by using the base models of VGG16, ResNet-50, and DenseNet-121 with 80 epochs and set the learning rate at 0.0001. The training loss and validation loss of the mentioned models for raw data are shown in Figure 11. Our findings revealed that CovNet-ResNet50 performed better than CovNet-VGG16 in the training process, and the training loss is decreased over time. Whereas the loss value of CovNet-DenseNet121 is greater than two other models over time, during training.

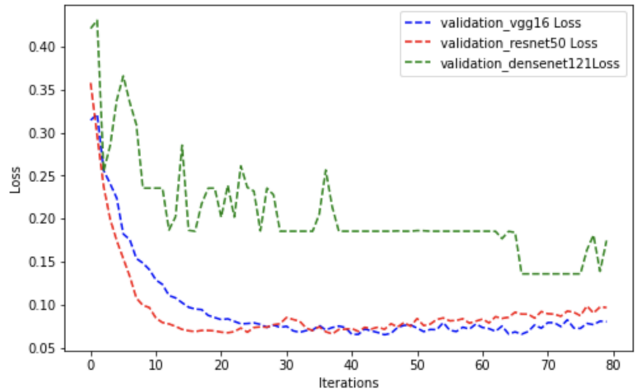

Figure 12. Validating curve of loss for CovNet models.

In the validation process of our model, we calculated the validation loss for the mentioned models in the case of using raw data. The results are shown in Figure 12. Based on the results, CovNet-VGG16 and CovNet-ResNet50 were better models in detecting and classifying the COVID-19, Pneumonia, and Normal cases.
Furthermore, we also compared training and validating loss of all CovNet models before and after using Data Augmentation. Figure 13 represents the training and validating loss graphs next to each other. According to our findings, data augmentation could mitigate the vanishing gradient problems and the loss values have been decreased over time. The loss values of CovNet-ResNet50 are lower than two other models. It means that CovNet-ResNet50 trained better than other models.

Feature Maps

The feature map is "the output of one filter applied to the previous layer" [6]. We showed the feature maps of our two top CovNet models (CovNet-ResNet50 and CovNet-VGG16) to get the better understanding of what is happening through the models, layer by layer. Figures 14 and 15 indicated the images of X-Ray images after applying the filter in two mentioned models.

CONCLUSION

In this paper, we implemented CovNet, a deep neural network designed to detect and classify COVID-19, Pneumonia, and Normal cases based on chest X-Ray which was proposed by [14].In addition, we explore that model implementation with not only VGG16 but also Restnet50, and densenet121.We leveraged transfer learning using VGG16, ResNet-50, and DenseNet-121 as our base models to CovNet to extract features. Then, we added five different custom layers with our chest xray image as the input dataset.We compared the three models we implemented with different metrics.

We evaluated the performance of our three implementation with metrics such as Accuracy, Precision, Recall, and F1-score. Also, we measured the effectiveness of CovNet model

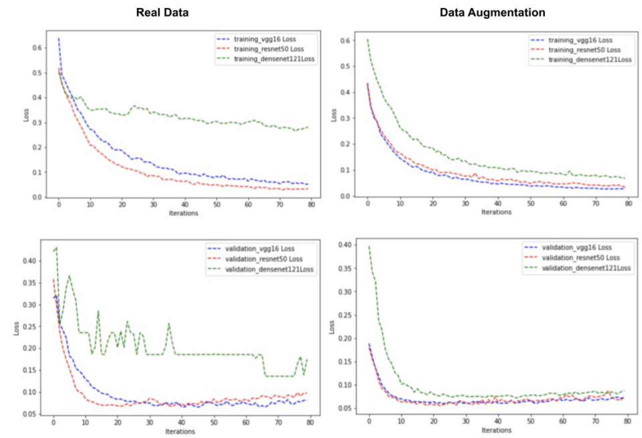

Figure 13. Training and Validating curve of loss for CovNet models for real data and after using data augmentation.

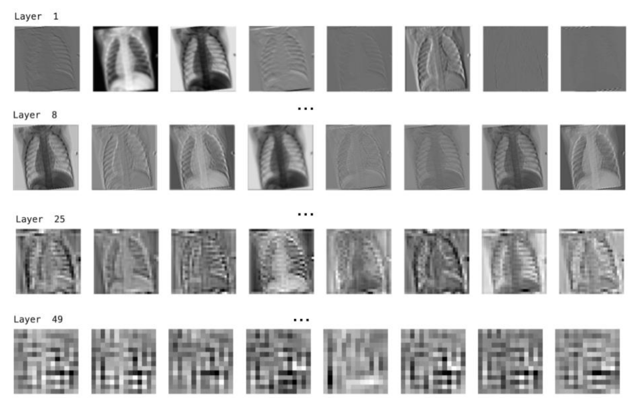

Figure 14. Images of feature maps layer by layer in ResNet50.

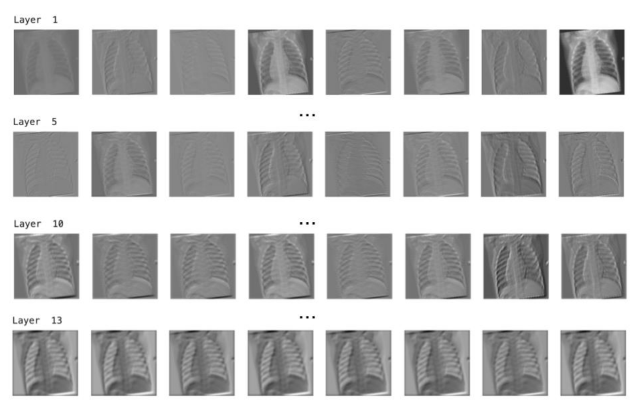

Figure 15. Images of feature maps layer by layer in VGG16.

with ROC, AUC, and Training loss. We observed that CovNet-VGG16 and CovNet-ResNet50 can detect and classify diseases with more accuracy around 98.3% and F1-score around 86-87.55%. According to the metrics' results, CovNet-ResNet50 was the best model.
We hope that CovNet-ResNet50 will help future researchers, doctors, and scientists detect COVID-19,pneumonia, and normal cases to from chest Xray images.

TEAM WORK CONTRIBUTION

We worked together and equally contributed in this project work. We met every week to discuss our project progress and ideas and any obstacles we faced in our project implementation.

REFERENCES

[1] Rosman AS and MA Korsten. 2007. Application of summary receiver operating characteristics (sROC) analysis to diagnostic clinical testing. *7 Reflections on the future of gastroenterology–unmet needs* 52 (2007), 76.

[2] Jason Brownlee. 2017. A Gentle Introduction to Transfer Learning for Deep Learning. (20 12 2017). `https://machinelearningmastery.com/transfer-learning-for-deep-learning/`.

[3] Jason Brownlee. 2021. Gentle Introduction to the Adam Optimization Algorithm for Deep Learning. (13 01 2021). `https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/`.

[4] Amar Budhiraja. 2016. Dropout in (Deep) Machine learning. (15 12 2016). `shorturl.at/bkFP0`.

[5] Joseph Paul Cohen, Paul Morrison, Lan Dao, Karsten Roth, Tim Q Duong, and Marzyeh Ghassemi. 2020. Covid-19 image data collection: Prospective predictions are the future. *arXiv preprint arXiv:2006.11988* (2020).

[6] ComputerScienceWiki. 2018. Feature maps (Activation maps). (2018). `shorturl.at/irV78`.

[7] Stackoverflow Contributors. 2015. Difference between Multi-label and Multi-Class. (2015).

[8] Worldometer Contributors. 2019. COVID-19 CORONAVIRUS PANDEMIC. (2019). `https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?`

[9] Priya Dwivedi. 2019. Understanding and Coding a ResNet in Keras. (04 01 2019). `https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33`.

[10] Purva Huilgol. 2019. Accuracy vs. F1-Score. (24 08 2019). `https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2#:~:text=Accuracy%20is%20used%20when%20the,and%20False%20Positives%20are%20crucial&text=In%20most%20real%2Dlife%20classification,to%20evaluate%20our%20model%20on.`

[11] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25 (2012), 1097–1105.

[13] Sarang Narkhede. 2018. Understanding AUC - ROC Curve. (26 06 2018). `https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5#:~:text=ROC%20is%20a%20probability%20curve,0s%20and%201s%20as%201s.`

[14] Harsh Panwar, PK Gupta, Mohammad Khubeb Siddiqui, Ruben Morales-Menendez, and Vaishnavi Singh. 2020. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. *Chaos, Solitons & Fractals* 138 (2020), 109944.

[15] Tuan D Pham. 2021. Classification of COVID-19 chest X-rays with deep learning: new models or fine tuning? *Health Information Science and Systems* 9, 1 (2021), 1–11.

[16] Pytorch. 2019. BCELOSS. (2019). `shorturl.at/jmuLQ`.

[17] Pablo Ruiz. 2018. Understanding and visualizing DenseNets. (2018). `https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a`.

[18] Koo Ping Shung. 2018. Accuracy, Precision, Recall or F1? (15 03 2018). `https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9`.

[19] Gaurav Singhal. 2020. Introduction to DenseNet with TensorFlow. (06 05 2020). `https://www.pluralsight.com/guides/introduction-to-densenet-with-tensorflow`.

[20] David L Streiner and John Cairney. 2007. What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry* 52, 2 (2007), 121–128.

[21] Daniel B Vinh, Xiao Zhao, Kimberley L Kiong, Theresa Guo, Yelda Jozaghi, Chris Yao, James M Kelley, and Ehab Y Hanna. 2020. Overview of COVID-19 testing and implications for otolaryngologists. *Head & neck* 42, 7 (2020), 1629–1633.

[22] L. Wang, Z. Q. Lin, and A Wong. 2020a. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports* (2020), 10(1), 1–12.

[23] Linda Wang, Zhong Qiu Lin, and Alexander Wong. 2020b. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports* 10, 1 (2020), 1–12.

[24] WebMD. 2020. Coronavirus and Pneumonia. (2020).

[25] Ramsey M Wehbe, Jiayue Sheng, Shinjan Dutta, Siyuan Chai, Amil Dravid, Semih Barutcu, Yunan Wu, Donald R Cantrell, Nicholas Xiao, Bradley D Allen, and others. 2021. DeepCOVID-XR: an artificial intelligence algorithm to detect COVID-19 on chest radiographs trained and tested on a large US clinical data set. *Radiology* 299, 1 (2021), E167–E176.

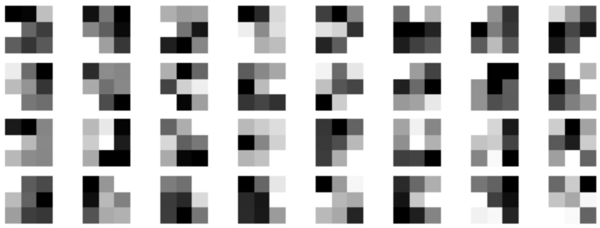

Figure 16. Feature map of resnet50

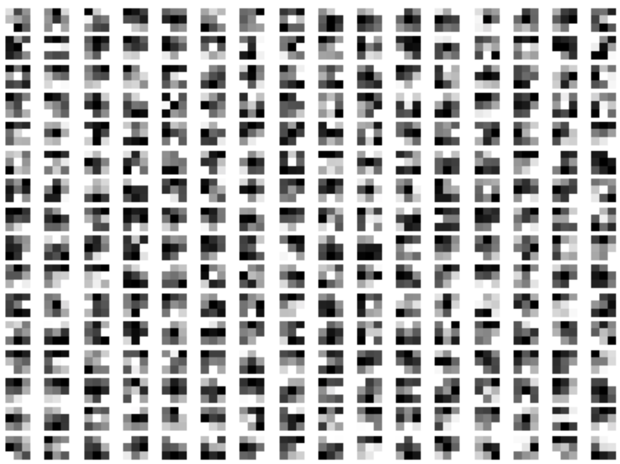

Figure 17. Feature map of vgg16

[26] Shi-Tao Yeh and others. 2002. Using trapezoidal rule for the area under a curve calculation. *Proceedings of the 27th Annual SAS® User Group International (SUGI'02)* (2002).

APPENDIX

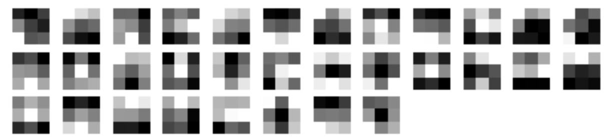

Figure 18. Feature map of Densenet121