Premier University, Chittagong

Department of Computer Science & Engineering

Final Year Thesis Report

On

Diagnosis of Pneumonia from Chest X-ray using Deep Learning

By

Joy Das Shanta

Tasfia Azim

2020

Premier University, Chittagong

Department of Computer Science & Engineering

Diagnosis of Pneumonia from Chest X-ray using Deep Learning

Supervised By

Faisal Ahmed

Lecturer

Department of Computer Science & Engineering

Premier University, Chittagong

Thesis Submitted By

Joy Das Shanta

ID: 1402610200644

Tasfia Azim

ID: 1402610200668

Premier University, Chittagong

Department of Computer Science & Engineering

DECLARATION

It is here by declared that the content of this thesis is original and any part of it has not been submitted elsewhere.

……

Joy Das Shanta

ID: 1402610200644

Date:

……

Tasfia Azim

ID: 1402610200668

Date:

CERTIFICATE OF APPROVAL

This thesis titled "**Diagnosis of Pneumonia from Chest X-ray using Deep Learning**" submitted by Joy Das Shanta (ID: 1402610200644) and Tasfia Azim (ID: 1402610200668) has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering(CSE) as B.Sc. Engineering to be awarded by Premier University, Chittagong.

Prof. Dr. Taufique Sayeed

Chairman

Dept. Computer Science & Engineering

Premier University, Chittagong

Faisal Ahmed

Lecturer

Dept. Computer Science & Engineering

Premier University, Chittagong

Acknowledgement

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of people whose ceaseless corporation made it possible and also whose constant guidance and encouragement crown all efforts with success. First of all, we would like to thank the most merciful and gracious almighty who teaches and shows us the way of peace and prosperity. This major thesis would not have been possible without the valuable assistance of our thesis supervisor **Faisal Ahmed**, Lecturer of Department of Computer Science & Engineering of Premier University, Chittagong. We are also grateful to **Prof. Dr. Taufique Sayeed**, Chairman of Department of Computer Science and Engineering of Premier University, Chittagong for his persistent motivation and support. Finally, we are also grateful to our parents for their hidden help in terms of prayer and faith in us, which actually empowered us to fulfil this task. We also indebted to all of our teachers of the department of Computer Science & Engineering who helped us in various ways in our four years of long journey of the university.

Dedicated

To

Our Respected Parents & Teachers

AND

Our Beloved Friend

Late Ajay Chowdhury

26th batch, Department of CSE

Premier University

We Mourn

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Abstract

Pneumonia is answerable for around 28% of the deaths of youngsters under five years old in Bangladesh. One child dies in every 39 seconds due to pneumonia according to recent analysis. Pneumonia happens because of microscopic organisms, infections or growths, and leaves youngsters battling for breath as their lungs load up with discharge and liquid. Delay in looking for suitable consideration and access to different sources for treatment are the fundamental hazard factors for pneumonia demise in small kids in Bangladesh. To fight this problem, this country needed an easily accessible quick solution. The thesis "Diagnosis of Pneumonia from Chest X-ray using Deep Learning" has been developed in order to provide a simpler way to detect pneumonia in a short period of time. The model is developed with chest X-ray images taken from the frontal views to identify pneumonia. The CNN model is trained on a data set of 5,200 recently collected X-ray images. The dataset was divided into two classes, normal and pneumonia X-ray images. After trying and testing several models, one model is selected finally, which was showing the best possible outcome. This model can successfully detect pneumonia at an accuracy rate of approximately 98%. We trained our model on different size X-ray images to evaluate its performance and got similar results. Our plan is to collect a larger dataset consisting chest X-ray of patients and normal people from different parts of our country and test our model on it. As, now a days, people are more comfortable with mobile devices, we are planning to build a mobile application which will be easily accessible to them. We want to compare our model with existing models to evaluate and validate its effectiveness.

Chapter 1

1.1 Introduction

Pneumonia is an acute disease in Bangladesh, responsible for almost 28% child death under the age of 5 years, which is around 50000 children in number, 2500 a day. 80000 children get admitted to hospital with this viral respiratory disease each year, when total patient counts are much higher [1]. According to WHO, just household pollution is responsible for 4 million premature death each year, around the world, causing pneumonia [2]. The death rate of pneumonia is so high due to wrong or delayed treatment. In common scenario doctors often make mistakes to detect pneumonia on time and the response time for this disease is very narrow, less than two days [3]. So, a fast detecting process is very necessary to ensure proper treatment for a pneumonia patient. Pneumonia is detected by Chest X-Ray, which depends on the availability of expert radiologists or doctor, which leads to an acute chance of mistakes, ignorance and lack of medical resources and personals. A computerized clinical guideline is essential to get rid of those situations. We are developing a model to dig up pneumonia from a chest x-ray efficiently and precisely.

Deep Neural Network models are usually designed and checked by human experts experimenting them in real time, using trial and error method. A Deep Neural Network Architecture is used to reduce time to optimize classification of recourses which is very crucial for this model. This particular model is designed to classify pneumonia chest x-ray images classification tasks using convolution neural network algorithm, processing several related information on a given image and detecting relevant data from it. This model is aimed to reduce the processing cost and time comparing other conventional pneumonia classification processes. Deep learning algorithm, using CNN has become common in medical image classification now a days. CNN exhibit impressive results comparing human examined one. CNN deep learning architecture has two-part, first extraction and input image encoding using convolutional layers CNN and the second one shows a prediction model for the classification task using fully connected neural network classifier. Many hyper parameters including convolutional layers numbers, filters numbers and their respective sizes are optimized by CNN models.

1.2 Motivation

Pneumonia is the single leading cause of mortality in children under five and is a major cause of child mortality in every region of the world, with most deaths occurring in sub-Saharan Africa and South Asia. Pneumonia kills more children under five than AIDS, Malaria, and measles combined, yet increased attention in recent years have been on the latter diseases. [4]

In Bangladesh, pneumonia is responsible for around 28% of the deaths of children under five years of age. Around 50,000 children die of pneumonia every year. An estimated 80,000 children under five years of age are admitted to hospital with virus-associated acute respiratory illness each year; the total number of infections is likely to be much higher. [5]

Pneumonia kills 4 children every 3hrs in Bangladesh. Bangladesh ranked 14th in the list of total 15 countries who were listed for the high number of deaths due to pneumonia, followed by its South Asian neighbours India, ranked second with 127,000 and Pakistan, ranked third with 58,000 child deaths. Report further stated that pneumonia was the third major cause of child deaths in 2017 in Bangladesh. This infectious disease was responsible for the deaths of four children under five in 1,000 live births in 2018. [6]

Quick diagnosis is key to effective treatment of pneumonia, but this can be difficult. Different medical conditions like excess fluid, internal bleeding, lung cancer, and many others can also show similar opacities in a CXR. [7]

It takes time for a careful review of the CXR, and some clinicians in a short amount of time need to process a large amount of CXRs which is one of the major problems of treatment. The goal of our model is to predict whether a patient has pneumonia so that a clinic can take action as soon as possible and can prevent child from death.

1.3 Objective

By deep leaning we can easily learn a machine and can have a machine or software or anything related to deep learning or artificial intelligence to overcome our daily life problem.

Our latest studies show that latest improvements in deep learning models with the availability of huge datasets with availability of algorithms taking huge impacts on medical fields which deal with medical images tasks such as skin cancer classification [8], haemorrhage identification [9], arrhythmia detection [10], and diabetic retinopathy detection [11].

So, we build a model which deals with medical images. We work on chest X-ray images and build a model to classify those images by deep learning to identify whether pneumonia is present or not which takes a few moments to predict by our model. But for doctors and clinic it takes time to identify because this is very complex to identify.

Our theme was to reduce timing problem of identifying this disease because this disease takes a very short time to impact on patient a huge cause most of the patients of this disease is child below 5 years old. If we cannot take action in the very primary stage, we have to pay worse for it. And by overcome this problem, we can reduce the child death from pneumonia of our country.

Chapter 2

2.1 Literature Review

Researchers tried to find different ways to detect pneumonia using different kinds of computational methods.

A pneumonia detection system from Chest X-ray images using Supervised Learning was developed by Benjamin Antin, Joshua Kravitz and Emil Martayan [12]. They used NIH dataset, collected from Kaggle which contains 1,12,120 Chest X-ray images of 30,805 unique patients. Their dataset was labelled by different classes like Pneumonia, Fibrosis etc. and Healthy (If the patient has no disease of these classes). They use binary classification to detect from X-ray images whether a patient has one of these diseases or not. After trained the entire dataset using 32*32 dimension as input in their logistic regression, they achieved 0.60 score on AUC curve. Using 128*128 dimension their AUC curve received 0.58 score. Limitation was that model was unable to do a hyper parameter sweep on images larger than 32x32 and 224x224 where they were able to run logistic regression.

In paper [13], we can see, they have built a prototype with Chex Net containing 121-layer of CNN. In this system if an input chest X-ray image is given, the output is going to be the probability of pneumonia in addition to a heat map confining the areas of the image which highly indicate pneumonia. They have used the model Chex Net on ChestX-ray14 dataset containing 112,120 frontal-view chest X-ray images. These images are separately marked with up to 14 non-identical thoracic diseases, including pneumonia. They managed the optimization of this huge deep network using batch normalization and dense connections. There are two limitations in their model.
First one is, only frontal radiographs were submitted to the radiologists and model throughout diagnosis, but to get a proper result up to 15% of precise diagnosis require the lateral view. So, we know that this model can come up with a moderate estimate of performance. Secondly, radiologists and the model itself could not access patient history, as a result the performance in interpreting chest radiographs has decreased. Their Chex Net achieves an F1 score of 0.435 (95% CI 0.387, 0.481).

In [14], they cast pathology detection as a multi-label classification problem. All images $X=\{x\rightarrow 1,\ldots,no\}$, $x\rightarrow i\in X$ are associated with a ground truth label $y\rightarrow I$, while we seek a classification function $f\rightarrow:X\rightarrow Y$ that minimizes a specific loss function l using N training sample-label pairs ($x\rightarrow I$, $y\rightarrow i$), $i = 1 \ldots N$. In addition to the original ResNet-50 architecture, they involve two variants: Firstly, they reduce input channels to one (the ResNet-50 is drafted to process RGB images), which should speed up the training of an X-ray specific CNN. Secondly, the input size by a factor of two is increased (i.e. 448×448) which keeps the model architectures similar. Here the max-pooling layer used has the same parameters which the

"pooling1" layer (i.e. 3×3 kernel, stride 2, and padding) has. For an evaluation of the generalization performance, they execute 5 times re-sampling scheme20. Inside per split, the data is separated into 70% training, 20% testing, and 10% validation.
Without non-image features, the OTS FT 1channel large was 66.4 ± 2.7, 74.4 ± 1.6, 74.3 ± 1.5, 75.3 ± 2.2 and with non-image features it was 68.3 ± 2.3, 73.3 ± 1.3, 74.8 ± 1.5, 76.7 ± 1.5. Their limitation was they present a semantic approach for this problem on Chest-X ray 14 while we think satisfactory result can be obtain in ImageNet. Their model result was satisfactory enough to non-image data while it was not satisfactory in Image data.

In [15], they used an attention-guided mask inference algorithm or AG-CNN to locate salient image regions that stand indicative of pneumonia. They used JSRT dataset which contains 154 modules of lung. The characteristics of local and global network branches in this model are connected in a chain to evaluate the probability of the disease. The AUC reported 0.776 in pneumonia detection. Their accuracy fir ResNet-50 was 0.904 and in DenseNet-121 it was 0.912. Between 14 classes performance was amazing except pneumonia which was .067 which was the lowest performance rate on their proposed model.

No	Paper Name	Data Set Size	Epoch	Accuracy
1.	Detecting Pneumonia in Chest X-Rays with Supervised Learning	1,12,120 Images	70	0.60 or 60%
2.	CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning	1,12,120 Images	121 Layers with 50 epochs (14 Diseases)	0.633 or 63% in Pneumonia
3	Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification	494 Images split in 2 datasets	20	0.56 or 56%
4	Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification	112,120 Images	50	ResNet-50 was 0.904 or 90% and in DenseNet-121 it was 0.912 or 91%

5	Health of Things Algorithms for Malignancy Level Classification of Lung Nodules	4323 nodules	10	96.7 % for F-score measurements and precision in the first task and 74.5 % accuracy and 53.2 % F-score in the second task
6	Automated chest X-ray screening: Can lung region symmetry help detect pulmonary abnormalities?	(1) Montgomery County Hospital (MC)collection - 138 posteroanterior CXRs (2) Shenzhen Hospital, China (CH) collection - 682 CXRs (3) Indian (IN) collection - 156 posteroanterior CXRs	10	Accuracy and AUC are 91% (83% and 86%) and 0.96 (0.90 and 0.94), respectively, from the CH dataset

Table 1: Related Works

In paper [16], we see a Structural Co-occurrence Matrix (SCM) propose to categorize malignant nodules or benign nodules and also their amount of malignancy. The structural Co-occurrence matrix technique was applied to extract characteristics of the nodule images and classify them. For extracting features of the nodule images and categorize them, the structural co-occurrence matrix technique was used. Images of the Hounsfield unit was applied with four filters and SCM in gray scale which creates eight different configurations. The categorization stage utilized classifiers like support vector machine, multilayer perceptron, k-Nearest Neighbors algorithm and were used in two tasks (i) to categorize the nodule images as malignant or benign, (ii) to categorize the nodules pulmonary lesions at the amount of malignancy (1 to 5). O (Rodrigues et al., 2018) had an outcome of 96.7 % for F-score

measurements and precision in the first task and 74.5 % accuracy and 53.2 % F-score in the second task. Their limitation was in managing over fitting in proposed model. And their F1-score was 54% average while there were some models with only two CNN layers which gives 70% average score. For using high level and different dimension of images to their network the over fitting problem was high enough.

In paper [17], an idea is suggested that shows the change of the right and left lung region as to symmetry and programmed the chest X-ray system for the confirmation of tuberculosis. The suggested system is the study of radiological exams which result in bilateral comparisons in the lung field. Multilayer network perceptron, random forest and Bayesian network are the three classifiers that are used here. The accuracy of abnormality detection was 91% and in ROC curve the area under was 0.96. They compare the changes between the left and right lung while we think it is not a good approach to identify tuberculosis like this. Both should be checked to identify the disease properly.

In Paper [18], they proposed an model using Convolution Neural Network that describe comparative classification of Normal and Pneumonia stage. They compared it to the [6] which has 92.8% accuracy and their model has 95.30% accuracy. They used publicly available dataset (Kermany, 2018) from Kaggle which contains 5863 images with Pneumonia and Normal Classes. They use 300*300 image dimension as input of their network and with each Convolutional layer they used ReLU. After each 2 convolutional layer they use Pool Layer (Max Pool). And in the last layer they use Function SoftMax activation. Although their proposed model has 95 percent of accuracy it has some limitations. The pattern they followed in CNN layer it will face error from over fitting while a large number of dataset are applied in their network and also it is not very time and cost effective.

2.2 Features of Pneumonia Disease:

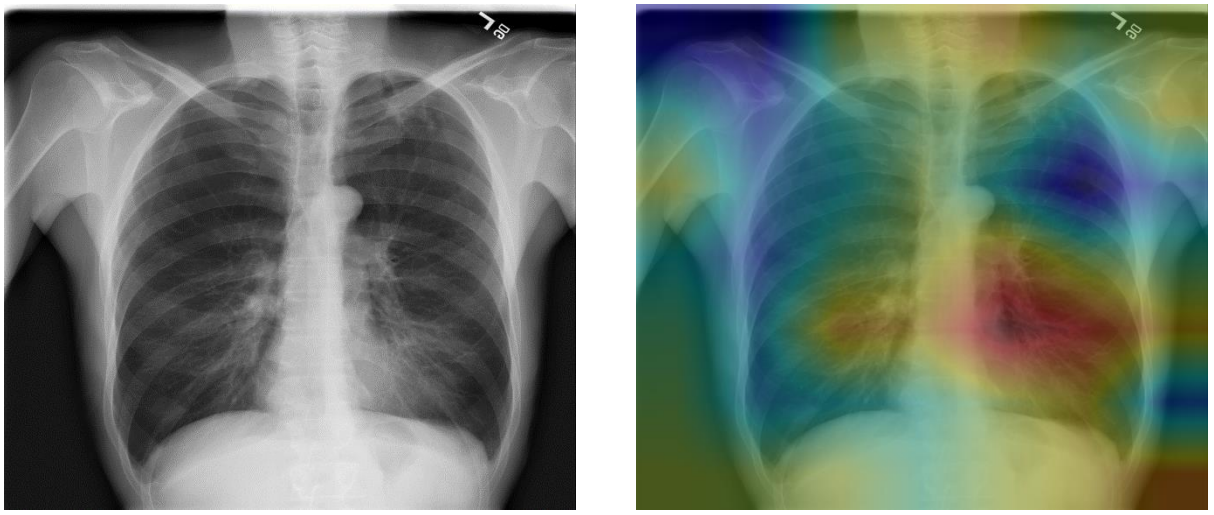

Figure 1: Features of Pneumonia Disease

Pneumonia is one of the major irresistible maladies responsible for noteworthy dismalness and mortality all through the world. Imaging assumes a vital job in the discovery and the board of patients with pneumonia.

This audit article talks about the distinctive imaging strategies utilized in the analysis and the executives of suspected pneumonic diseases. The imaging assessment ought to consistently start with customary radiography. At the point when the aftereffects of routine radiography are uncertain, figured tomography is required. A blend of example acknowledgement with information on the clinical setting is the best way to deal with the aspiratory irresistible procedures.

A particular example of contribution can propose an imaginable conclusion in numerous occurrences. In AIDS patients, diffuse ground-glass and interstitial penetrate are most regularly present in Pneumocystis carinii pneumonia though, in the non-immunosuppressed patients, a segmental lobar invade is reminiscent of bacterial pneumonia. Round pneumonia is regularly experienced in kids than grown-ups and is frequently brought about by Streptococcus pneumonia. Various mixes of parenchymal and pleural anomalies might be intriguing for extra findings.

At the point when an irresistible pneumonic procedure is suspected, information on the shifted radiographic appearances will limit the differential determination, assisting with coordinating extra demonstrative measures, and filling in as a perfect instrument for follow-up assessments.

Chapter 3
General Overview of methods used

3.1 Artificial Neural Network

In the modern sense of the word, artificial neural networks (ANNs) are networks of interconnected nodes inspired by the neurons in the brain. The nodes are typically grouped together in layers, and a series of layers make an ANN. While not a new invention, the use of ANNs has grown immensely in recent years and has seen applications in finance, marketing, production, monitoring, and more.

3.2 Convolution Neural Network

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

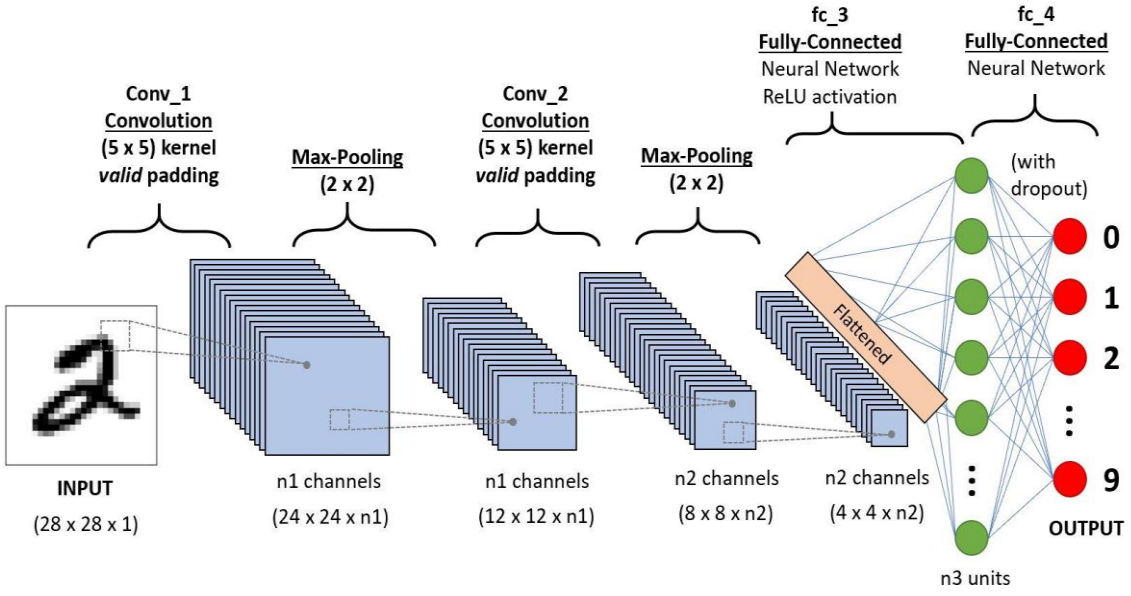

Figure 2: Basic CNN model architecture

rn of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlaps to cover the entire visual area.

There are three main layers in a traditional CNN network.

3.2.1 Convolution Layer

Convolution is the first layer to extract features from an input image. Convolution preserves the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel.

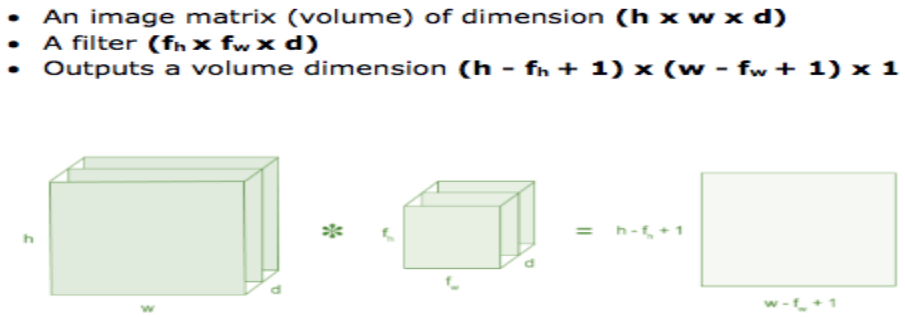

Figure 3: Convolution Layer

3.2.1.1 Strides

Stride is the number of pixels shifts over the input matrix. When the stride is 1 then we move the filters to 1 pixel at a time. When the stride is 2 then we move the filters to 2 pixels at a time and so on.

3.2.1.2 Padding

Sometimes filter does not fit perfectly fit the input image. We have two options:

1. Pad the picture with zeros (zero-padding) so that it fits
2. Drop the part of the image where the filter did not fit. This is called valid padding which keeps only valid part of the image.

3.2.1.3 Non Linearity (ReLU)

ReLU stands for Rectified Linear Unit for a non-linear operation. The output is $f(x) = max(0,x)$.

Why ReLU is important: ReLU's purpose is to introduce non-linearity in our ConvNet. Since, the real-world data would want our ConvNet to learn would be non-negative linear values.

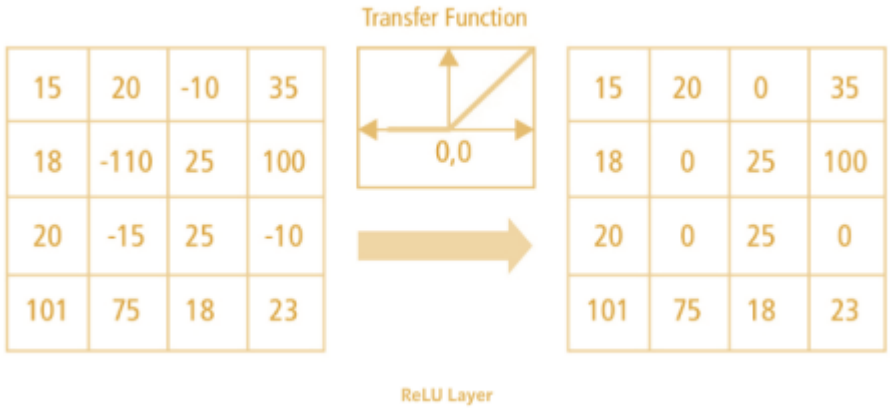

Figure 4: ReLU layer

There are other non-linear functions such as tanh or sigmoid that can also be used instead of ReLU. Most of the data scientists use ReLU since performance wise ReLU is better than the other two.

3.2.2 Pooling Layer

Pooling layers section would reduce the number of parameters when the images are too large. Spatial pooling also called subsampling or down sampling which reduces the dimensionality of each map but retains important information. Spatial pooling can be of different types:

- Max Pooling
- Average Pooling
- Sum Pooling.

3.2.2.1 MAX POOLING

Max pooling is a **sample-based discretization process**. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned.

This is done to in part to help over-fitting by providing an abstracted form of the representation. As well, it reduces the computational cost by reducing the number of parameters to learn and provides basic translation invariance to the internal representation.

Max pooling is done by applying a *max filter* to (usually) non-overlapping sub regions of the initial representation.

Let's say we have a 4x4 matrix representing our initial input. Let's say, as well, that we have a 2x2 filter that we'll run over our input. We'll have a **stride** of 2 (meaning the (dx, dy) for stepping over our input will be (2, 2)) and won't overlap regions.

For each of the regions represented by the filter, we will take the ***max*** of that region and create a new, output matrix where each element is the max of a region in the original input.

In order to make this super easy, with a nice pictorial representation - I give you this:

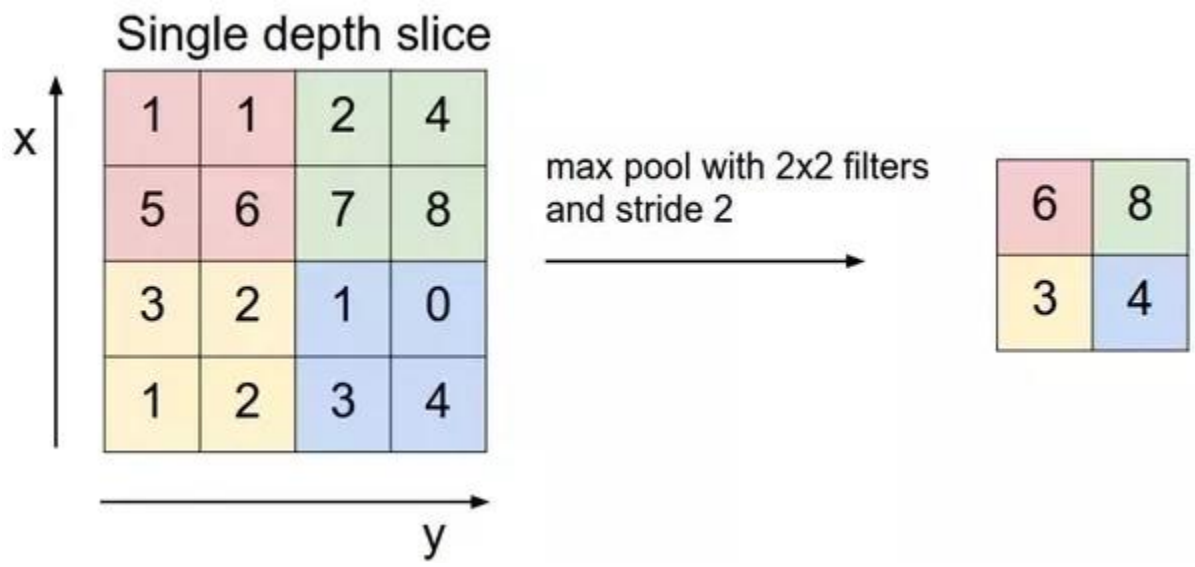

Figure 5: Max Pooling Layer

For a real example (note that the z dimension, the number of layers, remains unchanged in the pooling operation):

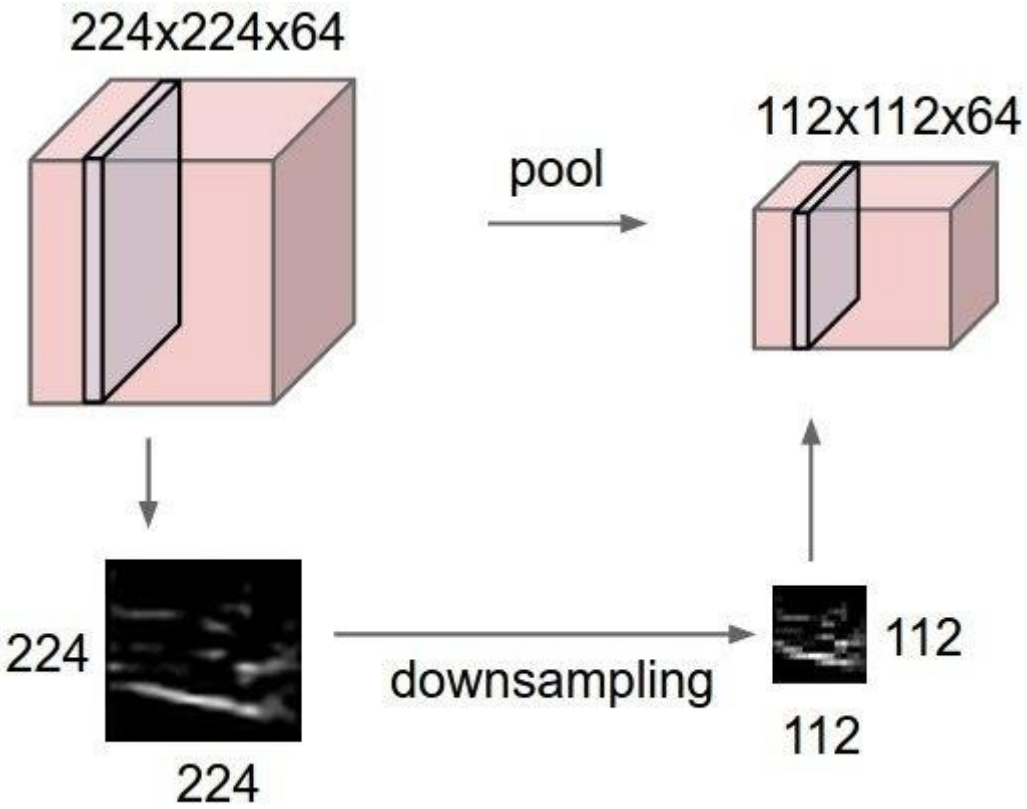

Figure 6: Down Sampling

However, the sub-regions don’t necessarily need to be non-overlapping.

3.2.2.2 AVG Pooling

Max Pooling is a down sampling strategy in Convolutional Neural Networks. Please see the following figure for a more comprehensive understanding.

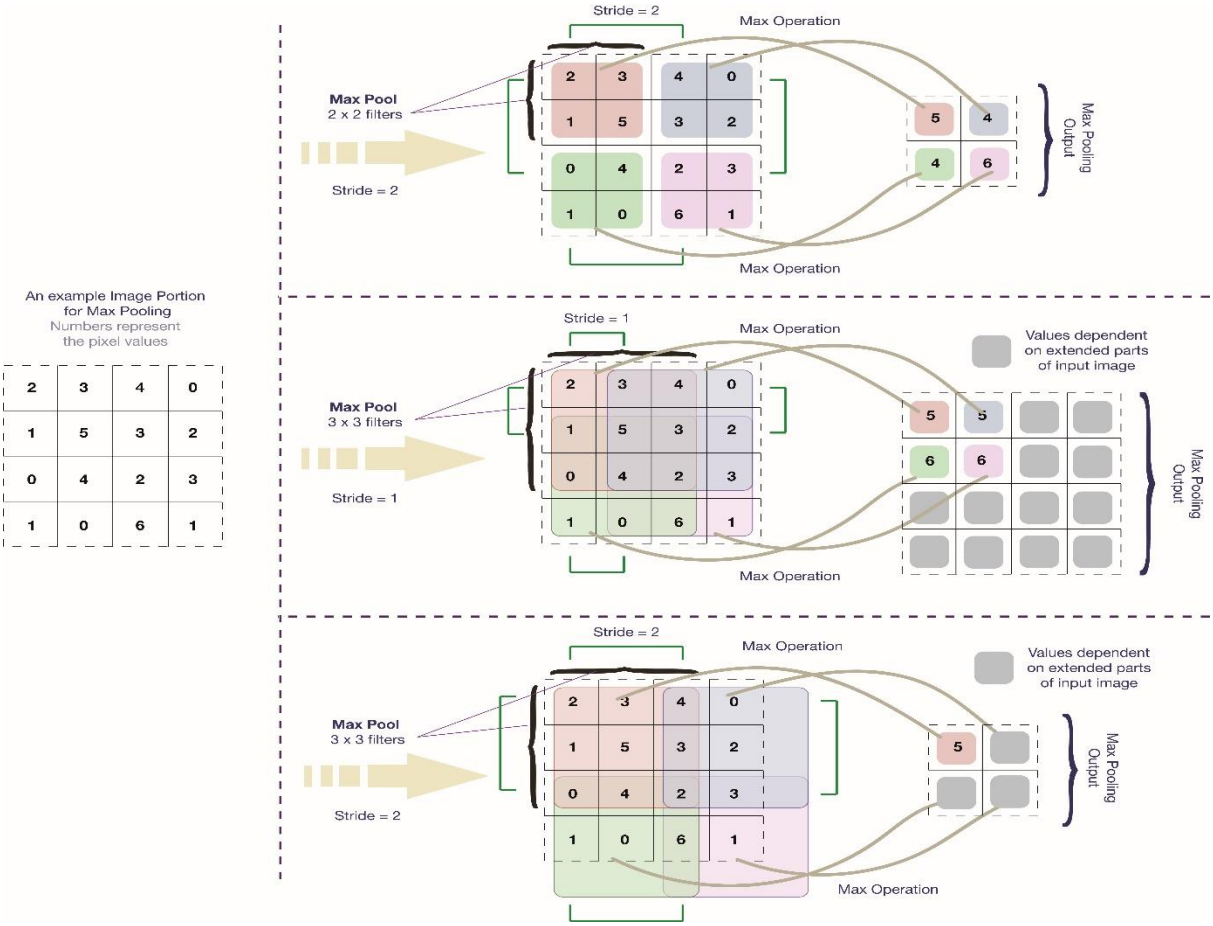

Figure 7: Average Pooling Layer

Here in the figure, we show the operation upon the pixel space. Alternatively, we can do a similar operation on some other mathematical space. Also, one can change the operation of taking 'Max' to something else, say taking an 'Average' (This is what is done in Average Pooling).

Generally, for pedagogical purposes, the depiction of max pooling is made for non-overlapping regions. This sometimes leads to a conjecture that max pooling is usually performed without overlaps. However, in reality, this notion is mostly not followed. In almost all of the famous CNN architectures, max pooling has been performed with overlapping regions. [Kernel Size, Stride] - Alex Net = [3x3, 2]; Google Net = [3x3, 2] , [3x3, 1]; VGG_CNN_S = [3x3,3], [2x2,2]; VGG_CNN_M and variants = [3x3, 2]; VGG_CNN_F = [3x3, 2]. We have thus shown in the figure all max pooling variants across the famous CNN architectures ([3x3,3] is similar in nature to [2x2,2]).

The pooling overlaps are in fact necessary in CNNs. As was pointed out by Hinton, that without overlaps, pooling operation may lose important information regarding the location of the object.

3.2.3 Fully Connected Layer

The layer we call as FC layer, we flattened our matrix into vector and feed it into a fully connected layer like a neural network.

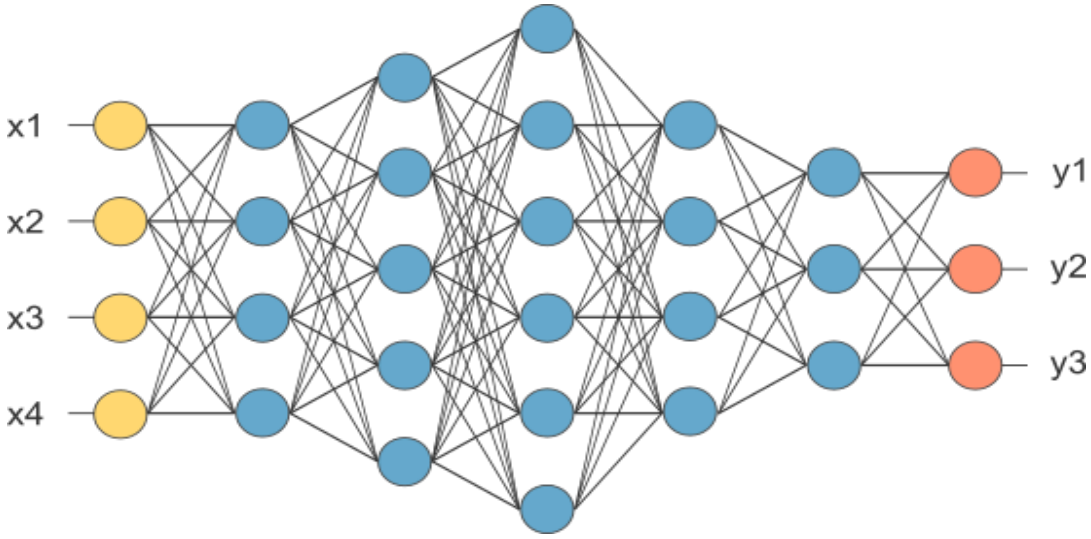

Figure 8: Fully Connected Layer

In the above diagram, the feature map matrix will be converted as vector (x1, x2, x3, …). With the fully connected layers, we combined these features together to create a model. Finally, we have an activation function such as SoftMax or sigmoid to classify the outputs as cat, dog, car, truck etc.

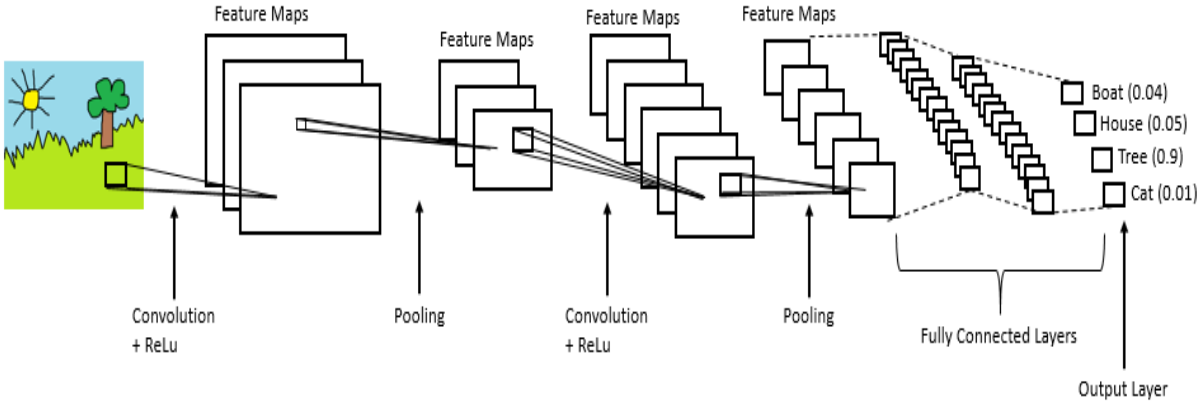

Figure 9: Overall architecture of a CNN model

3.2.4 Overfitting

One of the problems that occur during neural network training is called overfitting. The word overfitting refers to a model that models the training data too well. Instead of learning the general distribution of the data, the model learns the expected output for every data point. The error on the training set is driven to a very small value, but when new data presented in network the error is large. The network has memorized the training examples, but it has not learned to generalize to new situation.

As stated above overfitting is characterized by the in ability of the model to generalize. Too test this ability a simple method consists in splitting the dataset into two parts the training set and test set. When selecting models, we might want to split the dataset. With this split we can check the performance of the model on each set to gain insight on how the training process is going and spot overfitting when it happens.

3.2.5 Back Propagation

Backpropagation is the process of adjusting the weights of an ANN so that it can be trained to solve a problem and is the typical process that is repeated when training a network. This is done by passing the input through the network, comparing the output to ground truth, i.e. expected output, and then calculating an error rate via a specialised loss function. This error rate is then cascaded backwards through the network to modify the weights in the instance of CNN's filters are the weights) so that the next time the same or a similar image is passed through the network, the prediction should be closer to the ground truth. A series of parameters define limits on how drastically the network can learn, i.e. how drastically weights may be changed at a time to attempt generalisability. The theory is that with enough small changes and tweaks to weights, a network of nodes is able to learn any arbitrary mapping of input to output.

Weights are usually not updated after only passing a single image through the network, but rather after passing a batch of images through. The size of the batch is usually constant when training a network, however, it is not standardised across AI frameworks.

After one batch has been passed through, and the error rate backpropagated, it is called a single epoch. Note that the term epoch varies in definition and use, and epoch as a term in this thesis follows the given definition.

Chapter 4
Proposed Method

4.1 Model selection

We have developed two CNN models and evaluated their performance for their training.

4.1.1 Model-1

In our model we used 3x3 filter in the original image which gives 32 feature maps. Each feature map is reduced to 222x222. Now 2x2 max pooling layer is used for down sampling the feature matrix to 111x111. Thus the 1st convolution layer is completed. The 2nd convolution layer will generate 64 feature maps. Again, a 2x2 max pooling operation is performed to down sample data which produces 64 feature maps. Then we flattened and got 186624 neurons. Then each of them are connected to 128 neurons. At the last layer we have 2 neurons as we have 2 classes. Our proposed model is illustrated in Fig. 1.

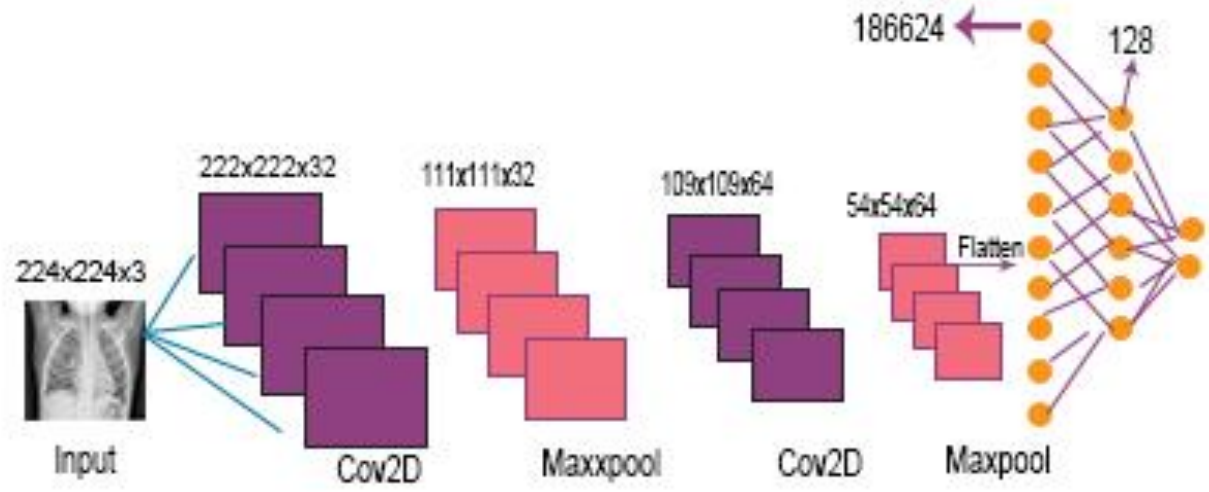

Figure 10: CNN Model-1 for classification

4.1.2 Model-2

Our overall proposed CNN model is shown in Model. Figure 13 which contains two major parts one is feature extractor and another is a classifier (sigmoid activation function). The proposed model consists of convolution, classification layer and max pooling. The feature extractors are composed of cov3x3, 32; cov3x3, 64; cov3x3, 128; cov3x3, 128. And it has max pooling layer of shape 2x2 with a RELU activator between them. All the results are arranged in a 2D plane called feature maps. The size of feature maps for the convolution operation are

148x148x32, 72x72x64, 34x34x128 and 15x15x128. And the size of the feature maps for pooling operation are 74x74x32, 36x36x64, 17x17x128 and 7x7x128. The size of our input image is 150x150x3. All the data are shown in Figure 15.

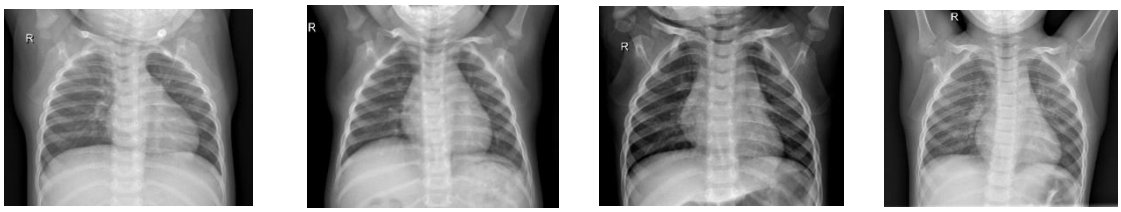

Figure 11:Sample Images without Pneumonia

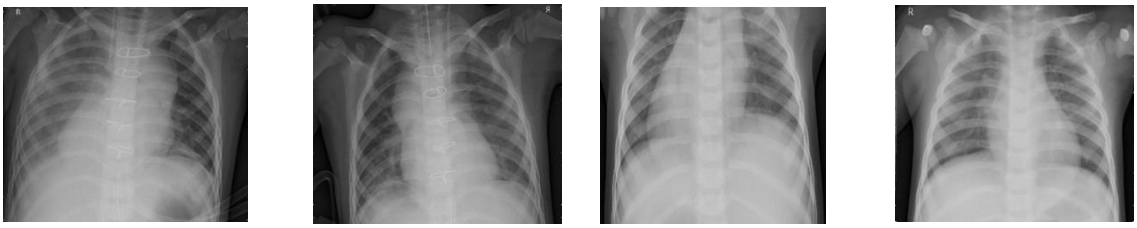

Figure 12: Sample Images with Pneumonia

From the proposed model it is seen that the classifier sigmoid activation function is placed the end of the CNN model. The output now is converted to 1D and the process is called flattening where the output of the convolution layer is flattened to generate a feature vector for the dense layer. Next, we used a dropout of size 0.5 and two dense layers of sizes 512 and 1 respectively with a RELU between them. We have included some x-ray images with and without pneumonia for better understanding.

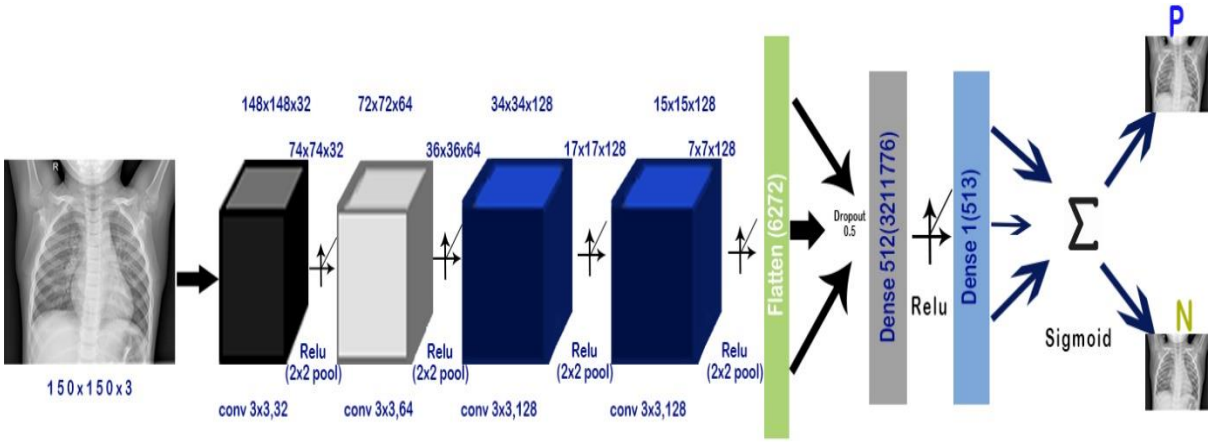

Figure 13: Proposed CNN Model 2 for classification

4.2 Model Summary

After applying deep convolution neural network algorithm, we found out the summary of two models that we developed.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 222, 222, 32)      896
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 111, 111, 32)      0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 109, 109, 64)      18496
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 54, 54, 64)        0
_________________________________________________________________
flatten_1 (Flatten)          (None, 186624)            0
_________________________________________________________________
dense_1 (Dense)              (None, 128)               23888000
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 129
=================================================================
Total params: 23,907,521
Trainable params: 23,907,521
Non-trainable params: 0
_________________________________________________________________
```

Figure 14: Model-1 Summary

Layer (type)	Output Shape	Param #
conv2d_13 (Conv2D)	(None, 148, 148, 32)	896
max_pooling2d_13 (MaxPooling	(None, 74, 74, 32)	0
conv2d_14 (Conv2D)	(None, 72, 72, 64)	18496
max_pooling2d_14 (MaxPooling	(None, 36, 36, 64)	0
conv2d_15 (Conv2D)	(None, 34, 34, 128)	73856
max_pooling2d_15 (MaxPooling	(None, 17, 17, 128)	0
conv2d_16 (Conv2D)	(None, 15, 15, 128)	147584
max_pooling2d_16 (MaxPooling	(None, 7, 7, 128)	0
flatten_4 (Flatten)	(None, 6272)	0
dropout_4 (Dropout)	(None, 6272)	0
dense_7 (Dense)	(None, 512)	3211776
dense_8 (Dense)	(None, 1)	513

Total params: 3,453,121
Trainable params: 3,453,121
Non-trainable params: 0

Figure 15: Model-2 Summary

In Model-1, the total parameter we received was 23,907,521 and all of them were trainable. There was no non-trainable value.

In Model-2, the total parameter that we obtained was 3,453,121 and all of them were trainable parameters. We got no non-trainable parameter in Model-2 as well.

So, from Model-1 to Model-2 the difference in parameter was 20,454,400.

Chapter 5
Experimental Analysis

5.1 Dataset

By CNN with TensorFlow framework we tried to identify the disease pneumonia from Chest X-ray images. We have collected these images from an open source dataset source named Kaggle [1]. Our dataset consists with 5,863 X-ray images in the format of JPEG. These images were performed as regular routine of clinical care. Images are divided in two classes. All values are mentioned in Table below.

Class	Training Images	Testing Images
Pneumonia	3330	560
Normal	770	560
Total	4100	1100

Table 2: Data Set

5.2 Performance Metrics

5.2.1 AUC - ROC curve

AUC - ROC curve is a performance measurement for classification problem at various thresholds settings. ROC is a probability curve and AUC represent degree or measure of separability. It tells how much model is capable of distinguishing between classes. Higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s. By analogy, Higher the AUC, better the model is at distinguishing between patients with disease and no disease.

The ROC curve is plotted with TPR against the FPR where TPR is on y-axis and FPR is on the x-axis.

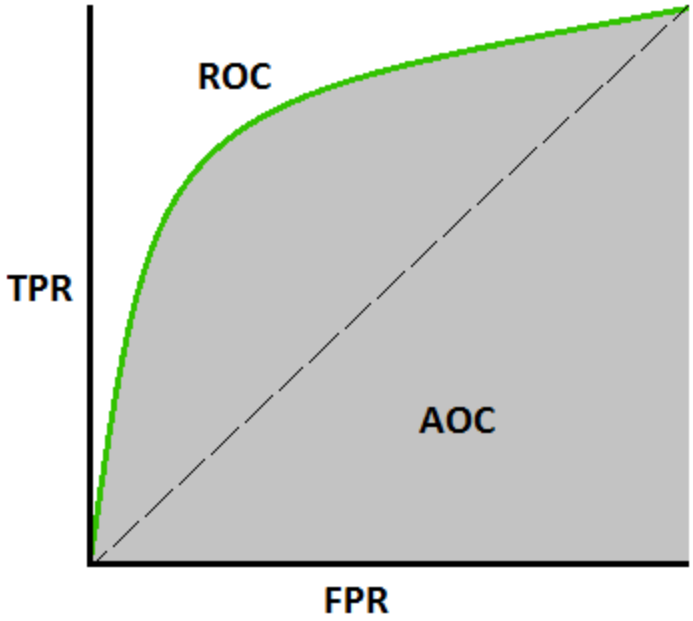

Figure 16: AUC - ROC Curve

1. True Positive (TP): True Positive is the number of truly classify as a positive.
2. False Positive (FP): False Positive is the number of truly classify as a negative.
3. True Negative (TN): True Negative is the number of truly classify as a negative.
4. False Negative (FN): False Negative is the number of falsely classified as negative.

5.2.2 Accuracy

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \qquad (1)$$

5.2.3 Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

5.2.4 Recall

Recall is the ratio of correctly predicted positive observations to the all observations in actual class.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

5.2.5 F1 score

F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy.

$$F1\ score = \frac{2*(Recall*Precision)}{(Recall+Precision)} \quad (4)$$

5.3 System Configuration

Platform and tools that we used for developing our model are –

5.3.1 Tools Required

1. NumPy
2. Matplotlib
3. Scikit-learn
4. Keras
5. TensorFlow

5.3.2 Operating System

1. Ubuntu

5.3.3 IDE

1. Spyder
2. Jupiter Note Book
3. Google Co-Lab

We performed our experiment on compiler Spyder and Jupiter Notebook integrates with some numbers of packages with an Intel® Core™ i7-8565U CPU 1.8 GHz (8M Cache, up to 4.6

GHz), powered with NVIDIA GeForce MX250. The GeForce MX250 has 4 cores, and 2GB GDDR5 memory. The frequency is 2.4 GHz to 4.6 GHz. In table below more information about our machine is presented. We used Google Co-Lab for the final compilation. All data are mentioned in Table below.

Configuration Item	**Value**
Machine Name	ASUS Vivo-Book S15 S531FL
CPU	Intel® Core™ i7-8565U Processor
Clock Speed	1.8 GHz up to 4.6 GHz
Cache Memory	8MB
Graphics processor units	NVIDIA GeForce MX250 GDDR5 2GB
Operating system	Windows 10 (64bit)
Memory	8GB DDR5
Hard disk	SATA 1TB 5400RPM 2.5’ HDD + PCIEG3x2 NVME 256GB M.2 SSD

Table 3: System Configuration

5.4 Selected Model

We worked on two models here. We used the same dataset and same system configuration for both models to evaluate their performance. Here, 2 models provide almost similar AUC but Model-2 has 20,454,400 parameters less than Model-1. So, we have chosen Model-2 and the rest of the experiments are performed using that model.

5.5 Model Validation

As we received the best output from Model-2, we trained this model on different image sizes as inputs to evaluate and validate its performance on loss, accuracy, precision, recall and f1-score.

We applied

- 100 by 100 image shape in input,
- 150 by 150 image shape in input,
- 200 by 200 image shape in input,
- 250 by 250 image shape in input,
- 300 by 300 image shape in input in our proposed model.

But there was no noticeable difference in loss, accuracy, precision, recall, f1-score. Also, in ROC curve.

Chapter 6

6.1 Result

6.1.1 Model-1

6.1.1.1 Accuracy graph of Model-1

The figure below is the graph of model accuracy per epoch during training and test of Model-1. We see that the training accuracy is stable during training while the test or validation accuracy sometimes dropped.

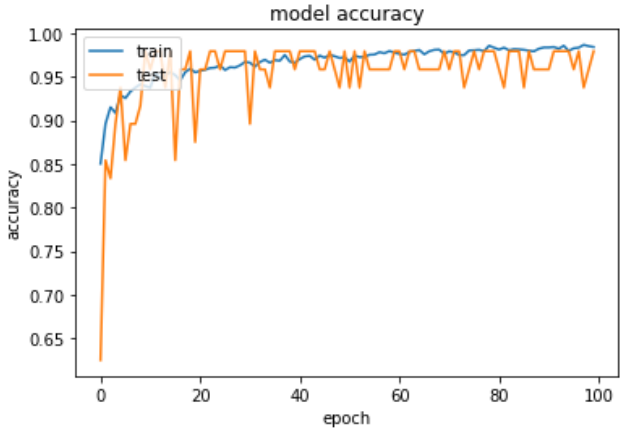

Figure 17: Model-1 Accuracy Graph

6.1.1.2 Loss graph of Model-1

The figure below is the graph of model loss per epoch during training and test of Model-1. We see that the test loss is higher than the training loss.

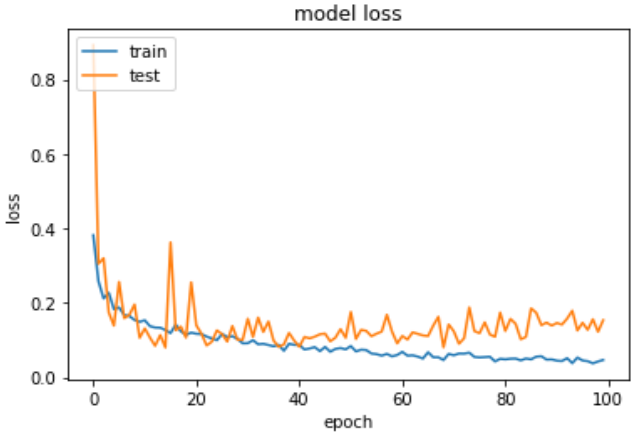

Figure 18: Model-1 Loss Graph

6.1.1.3 Confusion Matrix of Model-1

We found out the confusion matrix for our model-1. The true positive and True Negative was higher. But also, the False positive rate was higher than usual.

True Positive: 526	False Positive: 102
False Negative: 14	True Negative: 458

Table 4: Confusion Matrix of Model-1

6.1.1.4 Precision, Recall & F1-Score of Model-1

In model-1 precision, recall & F1-score for pneumonia were 0.97, 0.82, 0.89 whereas for normal 0.84, 0.97, 0.90. Training accuracy was 0.9313 and validation accuracy was 0.9125

	precision	recall	f1-score	support
pneumonia	0.97	0.82	0.89	560
normal	0.84	0.97	0.90	540
accuracy			0.89	1100
macro avg	0.90	0.90	0.89	1100
weighted avg	0.91	0.89	0.89	1100

Figure 19: Performance Measurement of Model-1

6.1.1.5 ROC curve of Model-1

The ROC area for model-1 is 0.983. The ROC area is as we excepted.

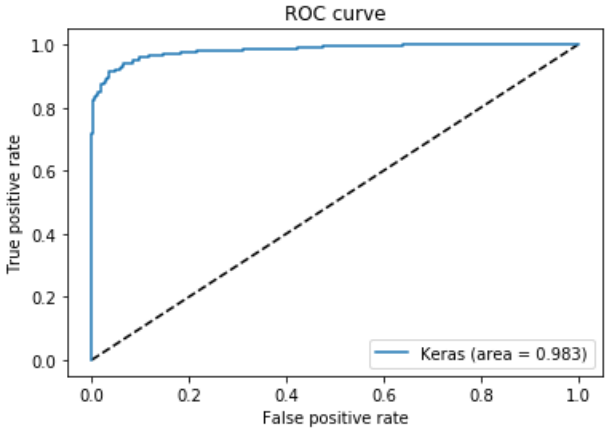

Figure 20: ROC area of Model-1

6.1.2 Model-2

6.1.2.1 Accuracy graph of Model-2

The figure below is the graph of model accuracy per epoch during training and test of Model-2. Here, training and test accuracy was low at the beginning and getting higher during training.

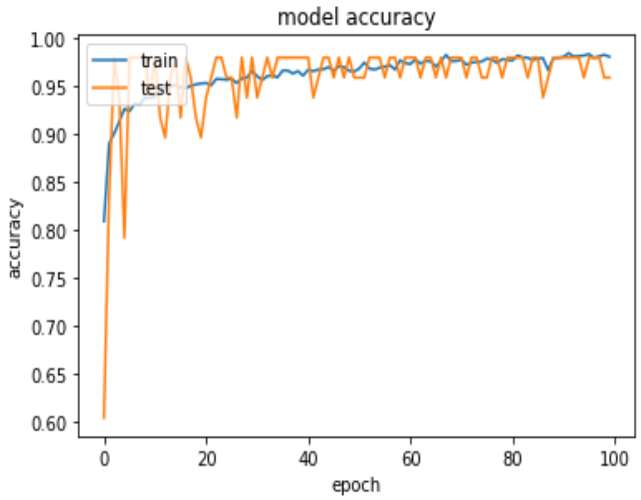

Figure 21: Model-2 Accuracy Graph

6.1.2.2 Loss graph of Model-2

The figure below is the graph of model loss per epoch during training and test of Model-2. We see that the test loss is higher than the training loss. But between epoch 10 to 20 the train loss was lower than the test.

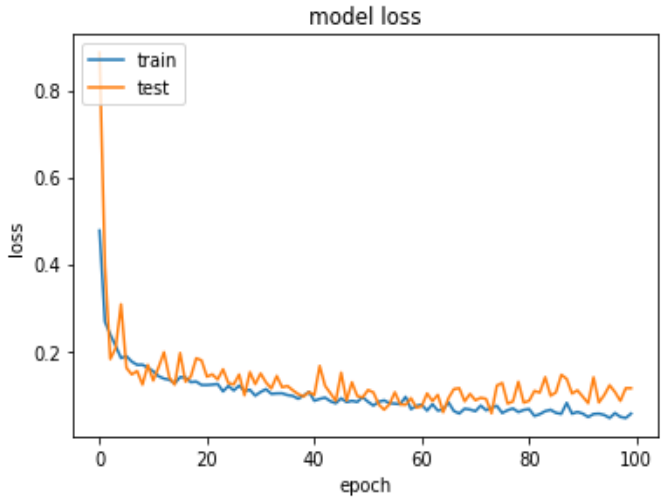

Figure 22: Model-2 Loss Graph

6.1.2.3 Confusion Matrix of Model 2

We found out the confusion matrix for our model-1. The true positive and True Negative was higher. But also, the False positive rate was higher than usual but the False Negative value was excepted.

True Positive: 534	False Positive: 113
False Negative: 6	True Negative: 447

Table 5: Confusion Matrix of Model-2

6.1.2.4 Precision, Recall & F1-Score of Model-2

In model-2 precision, recall & F1-score for pneumonia were 0.99, 0.80, 0.88 whereas for normal 0.83, 0.99, 0.90. Training accuracy was 0.9597 and validation accuracy was 0.9712.

	precision	recall	f1-score	support
pneumonia	0.99	0.80	0.88	560
normal	0.83	0.99	0.90	540
accuracy			0.89	1100
macro avg	0.91	0.89	0.89	1100
weighted avg	0.91	0.89	0.89	1100

Figure 23: Performance Measurement of Model-2

6.1.2.5 ROC Curve of Model-2

The ROC area for model-2 is 0.988. The ROC area is higher than Model-1.

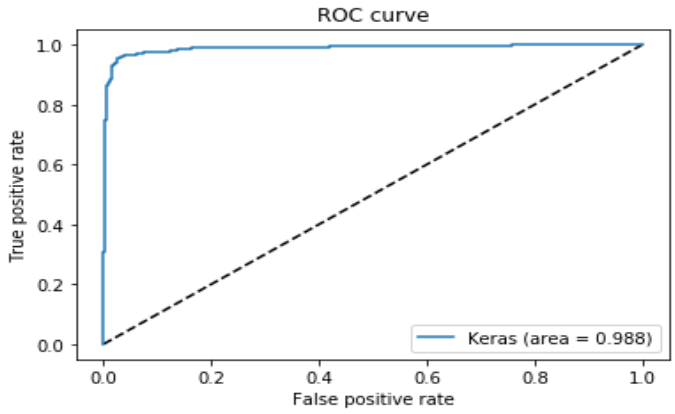

Figure 24: ROC area of Model-2

6.1.3 Combined ROC of Model-1 & Model-2

We combined and compare ROC curve of Model-1 & Model-2. We observed that the Model-2 has 0.988 ROC area which is higher than the Model-1 ROC area which is 0.983.

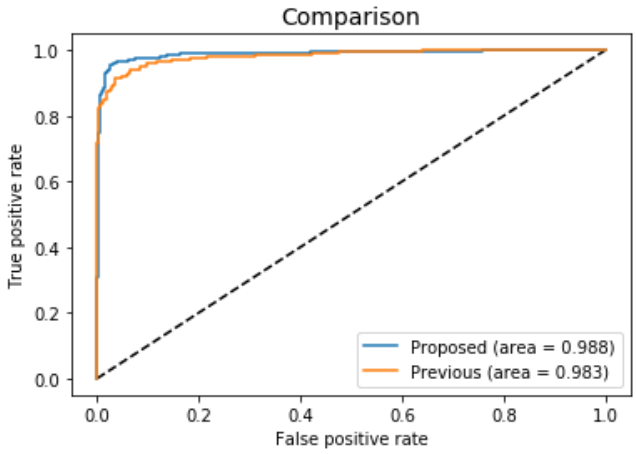

Figure 25: Combined ROC Curve of model-1 & Model-2

6.1.4 100 * 100 Image size:

6.1.4.1 Accuracy Graph:

The figure below is the graph of model accuracy per epoch during training and test. Here the test accuracy was not stable while the training accuracy was stable.

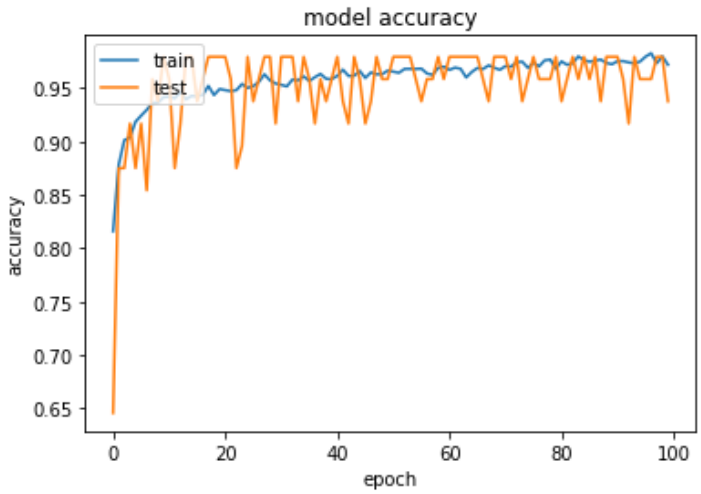

Figure 26: Accuracy Graph of 100 * 100

6.1.4.2 Loss Graph:

The figure below is the graph of model loss per epoch during training and test. Here the train loss is higher than the test.

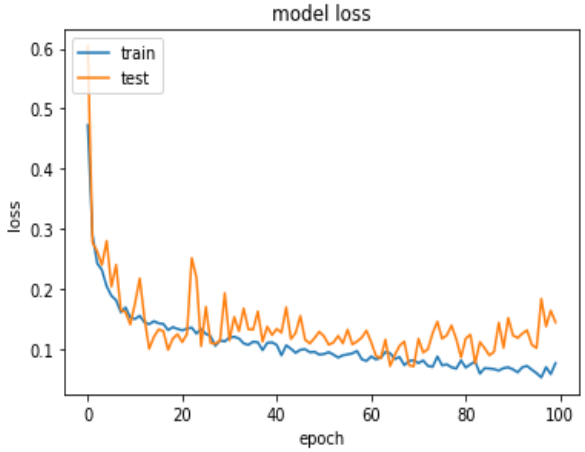

Figure 27: Loss Graph of 100 * 100

6.1.4.3 Confusion Matrix:

The true positive and True Negative was higher. But also, the False positive rate & False Negative rate was excepted.

True Positive: 530	False Positive: 14
False Negative: 10	True Negative: 546

Table 6: Confusion Matrix of 100 * 100

6.1.4.4 Precision, Recall & F1 score:

Here, precision, recall & F1-score for pneumonia were 0.98, 0.97, 0.98 whereas for normal 0.97, 0.98, 0.98. Values were satisfied enough.

	precision	recall	f1-score	support
pneumonia	0.98	0.97	0.98	560
normal	0.97	0.98	0.98	540
accuracy			0.98	1100
macro avg	0.98	0.98	0.98	1100
weighted avg	0.98	0.98	0.98	1100

Figure 28: Performance Measurement of 100 * 100

6.1.4.5 ROC curve:

The ROC area for this was 0.995. ROC area was satisfied.

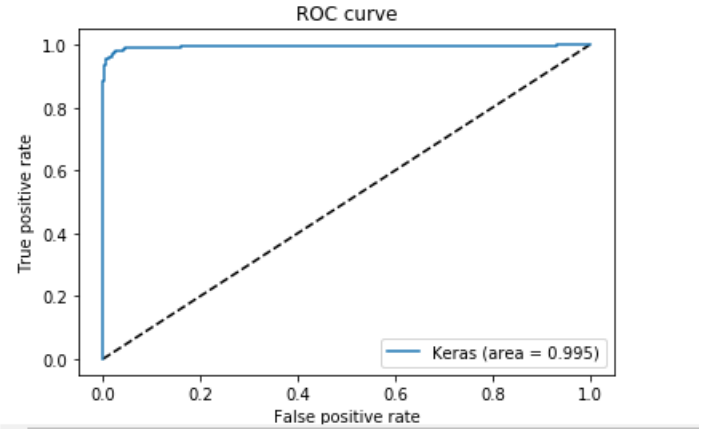

Figure 29: ROC of 100 * 100

Overall, in this shape as input Training accuracy was 0.9701 and validation accuracy was 0.9745

6.1.5 150 * 150 Image size:

6.1.5.1 Accuracy Graph:

The figure below is the graph of model accuracy per epoch during training and test. Here the training accuracy was stable while the test accuracy was not stable.

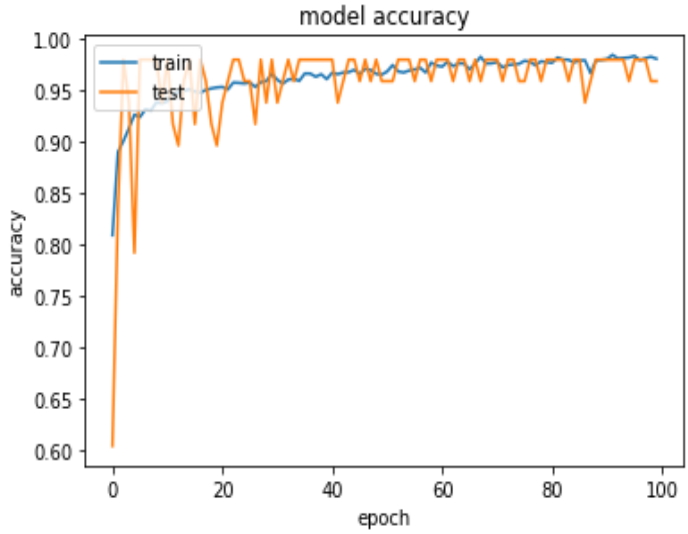

Figure 30: Accuracy Graph of 150 * 150

6.1.5.2 Loss Graph:

The figure below is the graph of model loss per epoch during training and test. Here the test loss is higher than the training.

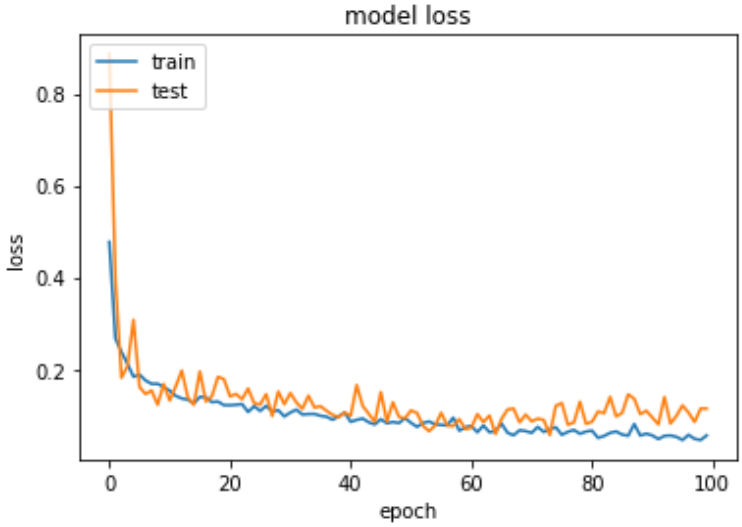

Figure 31: Loss Graph of 150 * 150

6.1.5.3 Confusion Matrix:

The true positive and True Negative was higher than any other model and input shape. But also, the False positive rate was 0 & False Negative rate was 18. This is the best confusion matrix we ever found.

True Positive: 522	False Positive: 0
False Negative: 18	True Negative: 560

Table 7: Confusion Matrix of 150 * 150

6.1.5.4 Precision, Recall & F1 score:

Here, precision, recall & F1-score for pneumonia were 0.97, 1.0, 0.98 whereas for normal 1.0, 0.97, 0.98. Values were satisfied enough and also this is the best result we ever found during our experimental time.

	precision	recall	f1-score	support
pneumonia	0.97	1.00	0.98	560
normal	1.00	0.97	0.98	540
accuracy			0.98	1100
macro avg	0.98	0.98	0.98	1100
weighted avg	0.98	0.98	0.98	1100

Figure 32: Performance Measurement of 150 * 150

6.1.5.5 ROC Curve:

The ROC area for this was 0.998. ROC area was higher than any other ROC area than different input shape.

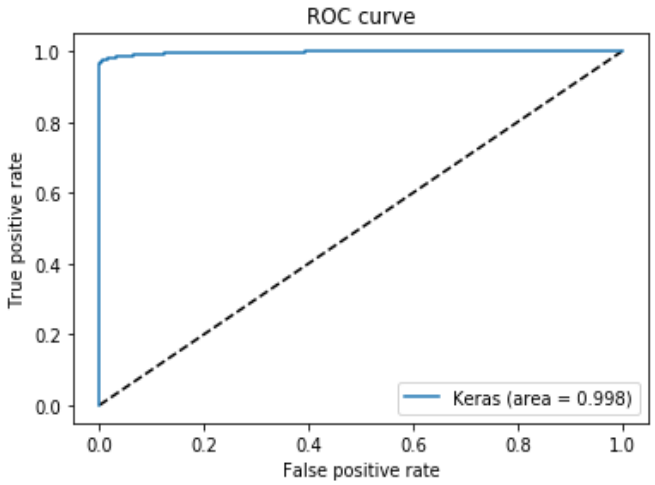

Figure 33: ROC of 150 * 150

Here, training accuracy was 0.9802 and validation accuracy was 0.9887. These values are highest among all.

6.1.6 200 * 200 Image size:

6.1.6.1 Accuracy Graph:

The figure below is the graph of model accuracy per epoch during training and test. Here the test accuracy was not stable at all while the train accuracy was stable than that and it was continuously increased.

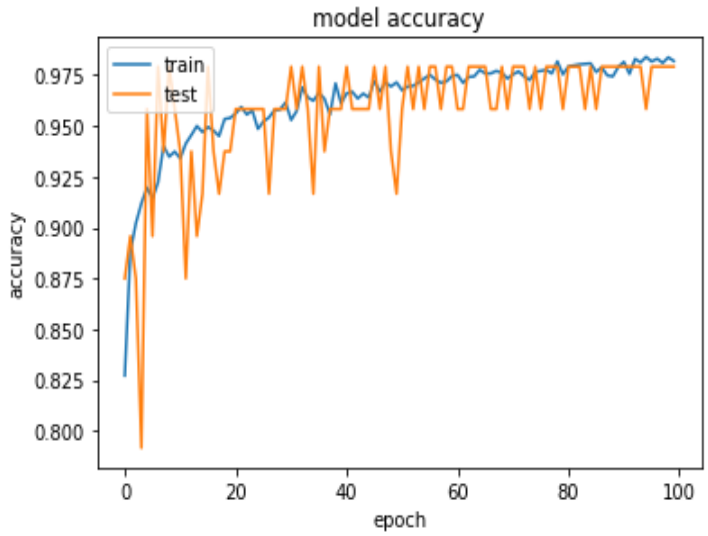

Figure 34: Accuracy Graph of 200 * 200

6.1.6.2 Loss Graph:

The figure below is the graph of model loss per epoch during training and test. Here the test loss is higher than the training and also it was higher than any other input shape.

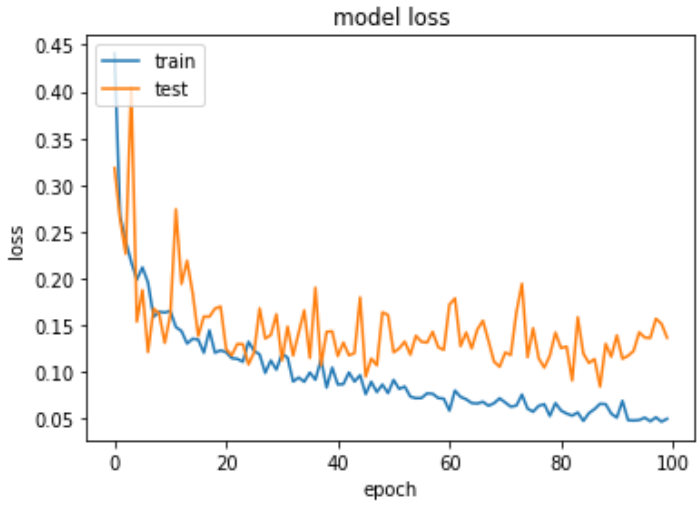

Figure 35: Loss Graph of 200 * 200

6.1.6.3 Confusion Matrix:

The true positive and True Negative was higher. But also, the False positive rate & False Negative rate was excepted.

True Positive: 529	False Positive: 9
False Negative: 11	True Negative: 551

Table 8: Confusion Matrix of 200 * 200

6.1.6.4 Precision, Recall & F1 score:

Here, precision, recall & F1-score for pneumonia were 0.98, 0.98, 0.98 whereas for normal 0.98, 0.98, 0.98. Values were satisfied enough. But here all values were 0.98.

	precision	recall	f1-score	support
pneumonia	0.98	0.98	0.98	560
normal	0.98	0.98	0.98	540
accuracy			0.98	1100
macro avg	0.98	0.98	0.98	1100
weighted avg	0.98	0.98	0.98	1100

Figure 36: Performance Measurement of 200 * 200

6.1.6.5 ROC curve:

The ROC area for this was 0.997. ROC area was satisfied.

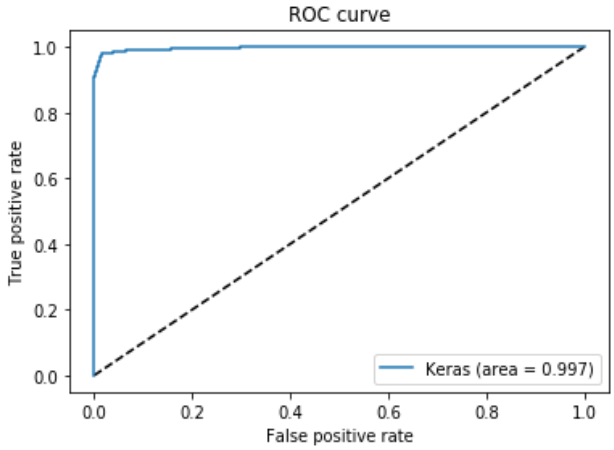

Figure 37: ROC of 200 * 200

Here, training accuracy was 0.9687 and validation accuracy was 0.9703.

6.1.7 250 * 250 Image size:

6.1.7.1 Accuracy Graph:

The figure below is the graph of model accuracy per epoch during training and test. Here the test accuracy was not stable at all while the train accuracy was stable than that and it was continuously increased.

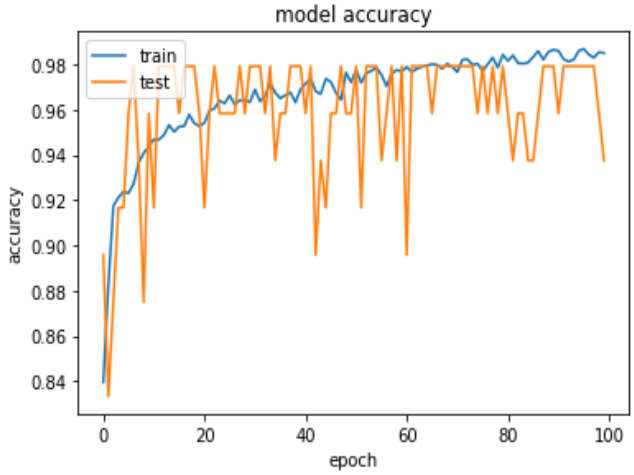

Figure 38: Accuracy Graph of 250 * 250

6.1.7.2 Loss Graph:

The figure below is the graph of model loss per epoch during training and test. Here the test loss is higher than the training.

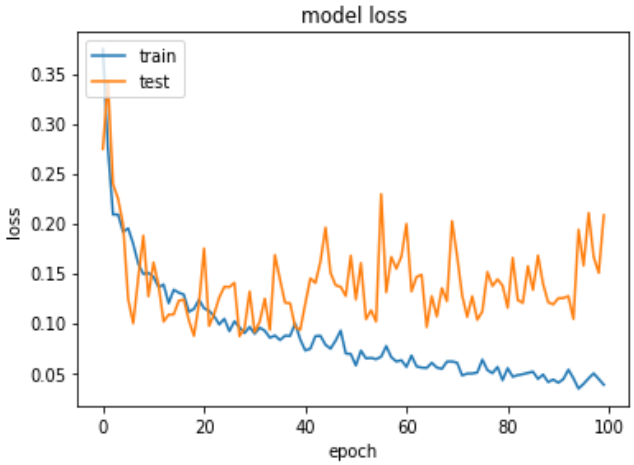

Figure 39: Loss Graph of 250 * 250

6.1.7.3 Confusion Matrix:

The true positive and True Negative was higher. But also, the False positive rate & False Negative rate was excepted.

True Positive: 527	False Positive: 7
False Negative: 13	True Negative: 553

Table 9: Confusion Matrix of 250 * 250

6.1.7.4 Precision, Recall & F1 score:

Here, precision, recall & F1-score for pneumonia were 0.98, 0.99, 0.98 whereas for normal 0.99, 0.98, 0.98. Values were satisfied enough.

	precision	recall	f1-score	support
pneumonia	0.98	0.99	0.98	560
normal	0.99	0.98	0.98	540
accuracy			0.98	1100
macro avg	0.98	0.98	0.98	1100
weighted avg	0.98	0.98	0.98	1100

Figure 40: Performance Measurement of 250 * 250

6.1.7.5 ROC curve:

The ROC area for this was 0.997. ROC area was satisfied.

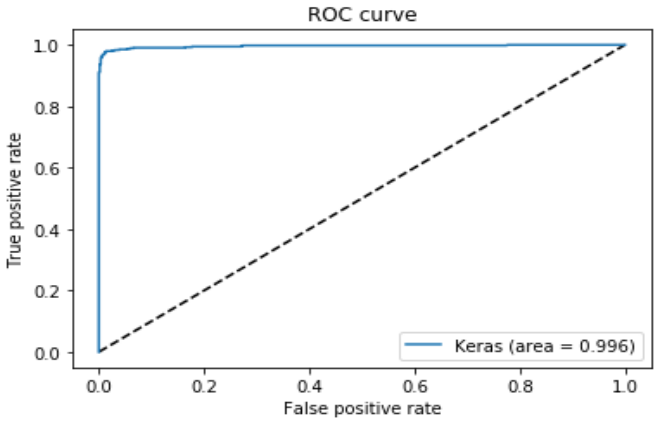

Figure 41: ROC of 250 * 250

Here, training accuracy was 0.9543 and validation accuracy was 0.9634. We noticed that this area value is decreased after 150 by 150 continuously.

6.1.8 300 * 300 Image size:

6.1.8.1 Accuracy Graph:

The figure below is the graph of model accuracy per epoch during training and test. Here the test accuracy was not stable at all while the train accuracy was stable than that and it was continuously increased.

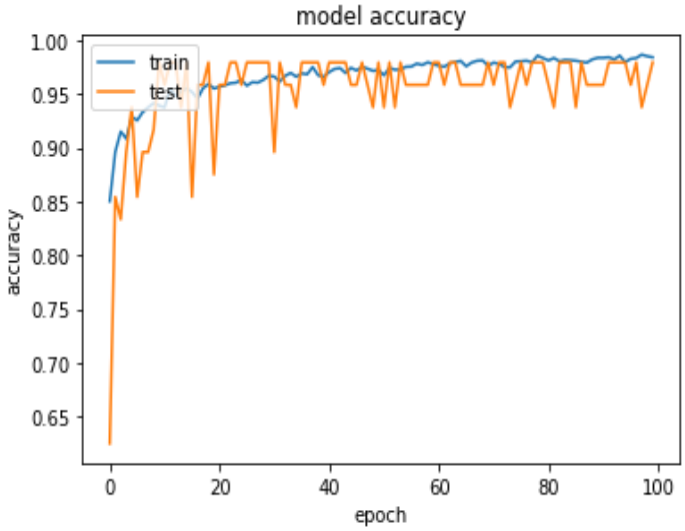

Figure 42: Accuracy Graph of 300 * 300

6.1.8.2 Loss Graph:

The figure below is the graph of model loss per epoch during training and test. Here the test loss is higher than the training.

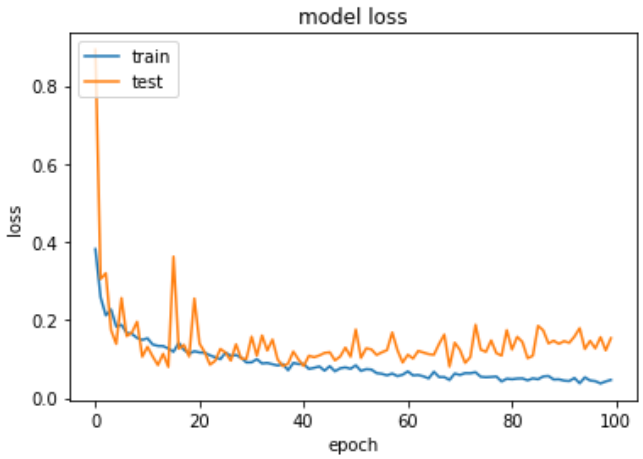

Figure 43: Loss Graph of 300 * 300

6.1.8.3 Confusion Matrix:

The true positive and True Negative was higher. But also, the False positive rate & False Negative rate was excepted.

True Positive: 518	False Positive: 3
False Negative: 22	True Negative: 557

Table 10: Confusion Matrix of 300 * 300

6.1.8.4 Precision, Recall & F1 score:

Here, precision, recall & F1-score for pneumonia were 0.96, 0.99, 0.98 whereas for normal 0.99, 0.96, 0.98. Values were satisfied enough.

	precision	recall	f1-score	support
pneumonia	0.96	0.99	0.98	560
normal	0.99	0.96	0.98	540
accuracy			0.98	1100
macro avg	0.98	0.98	0.98	1100
weighted avg	0.98	0.98	0.98	1100

Figure 44: Performance Measurement of 300 * 300

6.1.8.5 ROC curve:

The ROC area for this was 0.995. ROC area was satisfied.

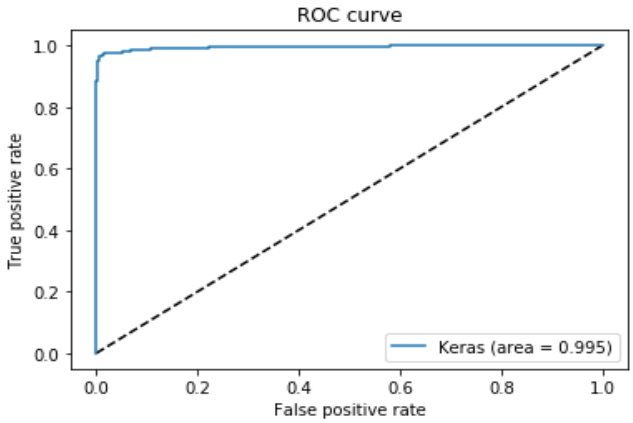

Figure 45: ROC of 300 * 300

Here, training accuracy was 0.9647 and validation accuracy was 0.9721.

6.1.9 Comparison:

6.1.9.1 Accuracy Comparison of Training:

The figure below the is combined training accuracy graph of above five input shapes. Here, we can see that, in 100 it was lowest and in 150 it was more stable among all.

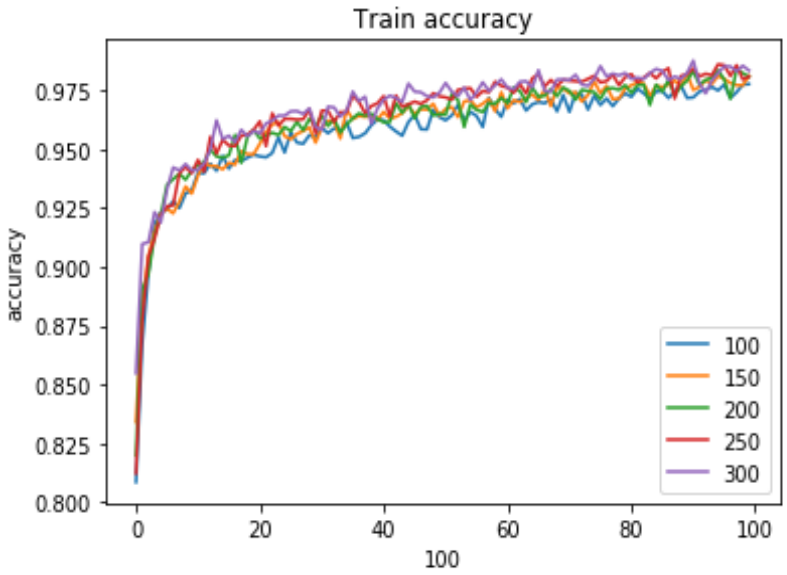

Figure 46: Combined Train Accuracy Graph

6.1.9.2 Loss Comparison of Training:

The figure below the is combined loss graph of training above five input shapes. Here, we can see that, in 100 it was highest and in 150 it was more stable among all.

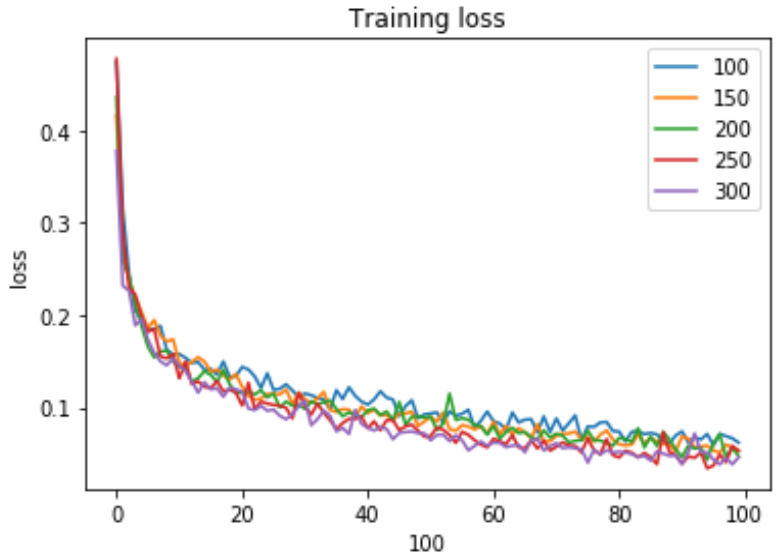

Figure 47: Combined Train Loss Graph

6.1.9.3 Accuracy Comparison of Test:

The figure below the is combined test accuracy graph of above five input shapes. Here, we can see that, in 100 it was lowest and in 150 and 250 it was higher among all.

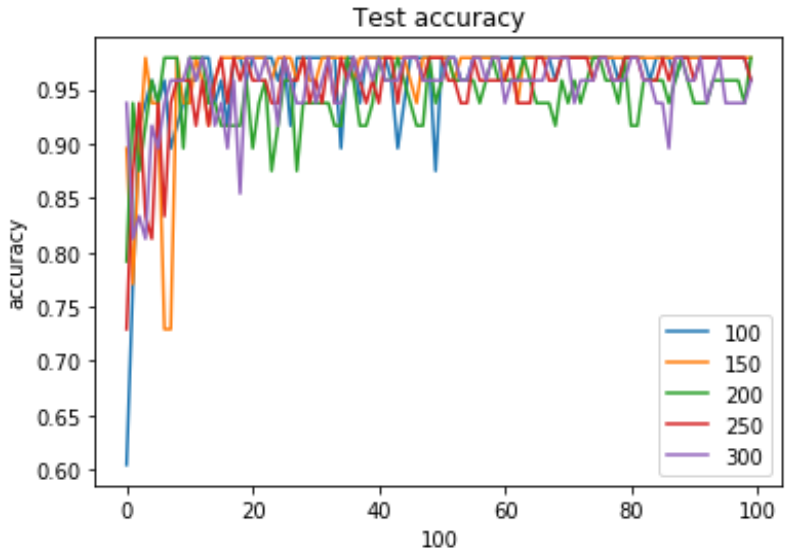

Figure 48: Combined Test Accuracy Graph

6.1.9.4 Loss Comparison of Test:

The figure below the is combined test loss graph of above five input shapes. Here, we can see that, in 300 it was highest and in 150 it was lowest among all.

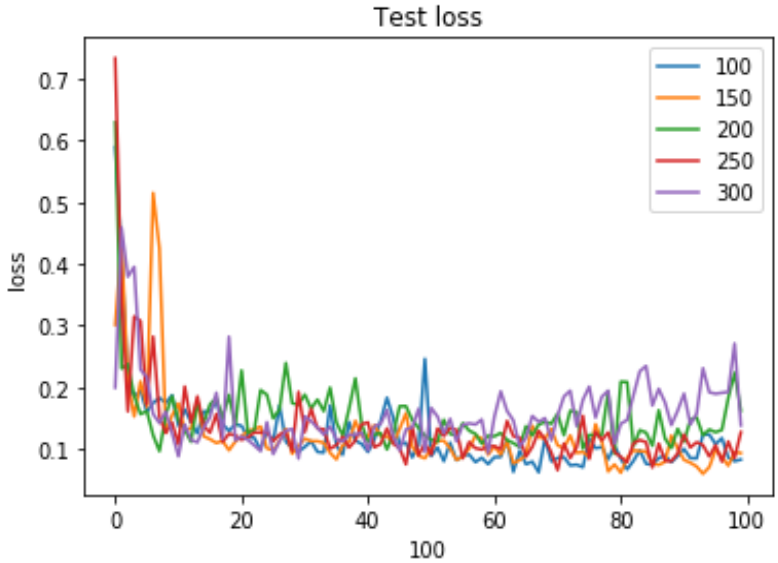

Figure 49: Combined Test Loss Graph

6.1.9.5 Combined ROC:

The figure below is the combined ROC curve of above five models or input shapes. Here, we can see that, for 100 * 100 and for 300 * 300 it was same which is 0.995 and these two are lowest of all. For 150 * 150 it was 0.998 the best ROC area we ever got during our experiment.

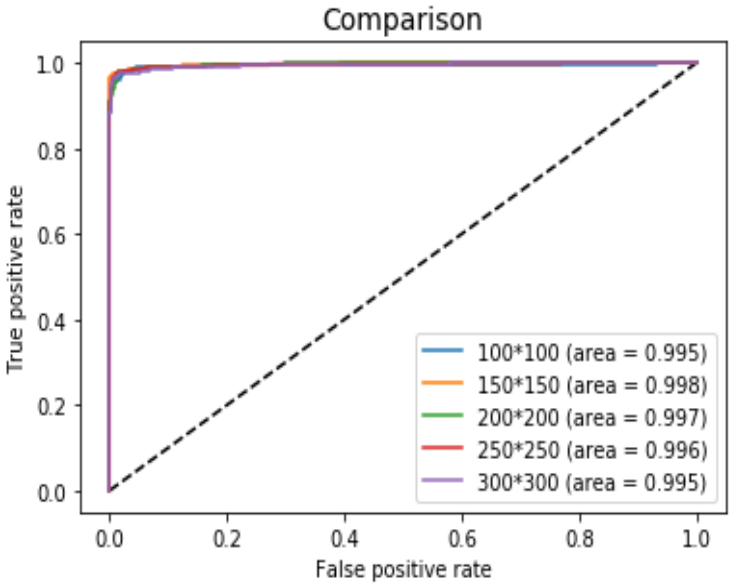

Figure 50: Combined ROC

6.1.10 Accuracy Table:

The table below is the overall training accuracy and validation accuracy table. For 150 * 150 we got highest values which is our best result we got. The average training & validation accuracy for all five input shapes are 0.9647 & 0.9721.

Data Size	Training Accuracy	Validation Accuracy
100	0.9701	0.9754
150	0.9802	0.9887
200	0.9687	0.9703
250	0.9543	0.9634
300	0.9502	0.9630
Average	0.9647	0.9721

Table 11: Training & Test Accuracy table

6.2 Discussion:

We developed this algorithm to detect pneumonia from chest X-ray images taken from frontal views at high validation accuracy as shown in table 7.

The algorithm begins with transforming X-ray images into smaller sizes. Then for detecting and classifying pneumonia we used convolution neural network framework, which extracted features from the images.

As in result section we have trained our proposed model with different image sizes as we mentioned in result section. After couple of experiments we found that, our model is so stable that the difference among them in any performance area is not so mentionable, although we found that for the 150 by 150 image size we got the best result in all performance area which is higher than any other approaches we mentioned in related work chapter and also from different image shapes and also from model 1. We repeated our experiment several times to check the accuracy and every time we received the same result.

We believe, if we have huge access to data and can train the model with radiological data from patients and normal people from different places, we could bring significant changes to this model.

Chapter 7

7.1 Conclusions:

We have exhibited a model to segregate positive or negative pneumonia data from a number of X-ray images. Our model is distinguishable from other models that depend mainly on transfer learning approach. This model also can be used in detecting and classifying X-ray images consisting of lung cancer and pneumonia. We hope this model will be very useful in rural areas where expert doctors are not available and people are deprived of medical help.

7.2 Future Work:

Our plan is to collect a larger dataset consisting chest X-ray of patients and normal people from different parts of our country and test our model on it. As now-a-days people are more comfortable with mobile devices, we are planning to build a mobile application which will be easily accessible to them. We want to compare our model with existing models to evaluate and validate its effectiveness.

References:

[1] https://www.icddrb.org/news-and-events/press-corner/media-resources/pneumonia-and-other-respiratory-diseases [Last access time: 19 November, 2019]

[2] World Health Organization, Household Air Pollution and Health [Fact Sheet], WHO, Geneva, Switzerland, 2018, http://www.who.int/newa-room/fact-sheet/detail/household-air-pollution-and-health [Last access time: 25 November, 2019]

[3] Pneumonia mortality and healthcare utilization in young children in rural Bangladesh: a prospective verbal autopsy study, https://tropmedhealth.biomedcentral.com/articles/10.1186/s41182-018-0099-4 [Last access time: 1 December, 2019]

[4] Pneumonia: The forgotten killer of children. Geneva, UNICEF/WHO, 2006. [Last access time: 5 December, 2019]

[5] https://www.icddrb.org/news-and-events/press-corner/media-resources/pneumonia-and-other-respiratory-diseases [Last access time: 6 December, 2019]

[6] https://www.dhakatribune.com/bangladesh/2019/11/13/report-pneumonia-kills-4-children-every-3-hours-in-bangladesh [Last access time: 8 December, 2019]

[7] Kaggle. RSNA Pneumonia Detection Challenge. www.kaggle.com/c/rsna-pneumonia-detectionchallenge/data [Last access time: 22 December, 2019]

[8] E. Andre, K. Brett, A Roberto et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017. [Last access time: 3 December, 2019]

[9] M. Grewal, M. M. Srivastava, P. Kumar, and S. Varadarajan, "Radiologist level accuracy using deep learning for hemorrhage detection in CT scans," 2017, http://arxiv.org/abs/1710. 04934. [Last access time: 21 December, 2019]

[10] R. Pranav, Y. H. Awni, H. Masoumeh, B. Codie, and Y. N. Andrew, "Cardiologist-level arrhythmia detection with convolutional neural networks," 2017, http://arxiv.org/abs/ 1707.01836. [Last access time: 6 December, 2019]

[11] G. Varun, P. Lily, C. Marc et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, 2017, vol. 316, no. 22, pp. 2402–2410. [Last access time: 29 December, 2019]

[12] Detecting Pneumonia in Chest X-Rays with Supervised Learning Benjamin Antin, Joshua Kravitz, and Emil Martayan. https://pdfs.semanticscholar.org/bbc7/49a5c9139dc642a78647c1dfed1df71bba07.pdf [Last access time: 1 January, 2019]

[13] CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Robyn L. Ball, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, Andrew Y. Ng [Last access time: 9 December, 2019]

[14] Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification Ivo M. Baltruschat, Hannes Nickisch, Michael Grass, Tobias Knopp & Axel Saalbach. [Last access time: 3 January, 2019]

[15] Guan, Q.; Huang, Y.; Zhong, Z.; Zheng, Z.; Zheng, L.; Yang, Y. Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification. arXiv. 2018. Available online: https://arxiv.org/abs/1801.09927v1 (accessed on 17 June 2018) [Last access time: 10 January, 2019]

[16] Rodrigues, M. B., Nobrega, R. V. M. D., Alves, S. S. A., ´ Filho, P. P. R., Duarte, J. B. F., Sangaiah, A. K., and Albuquerque, V. H. C. D. (2018). Health of things algorithms for malignancy level classification of lung nodules. IEEE Access, 6:18592–18601. [Last access time: 17 January, 2019]

[17] Santosh, K. C. and Antani, S. (2018). Automated chest x-ray screening: Can lung region symmetry help detect pulmonary abnormalities? IEEE Transactions on Medical Imaging, 37(5):1168–1177. [Last access time: 3 February, 2019]

[18] Classification of Images of Childhood Pneumonia using Convolutional Neural Networks- A. A. Saraiva1 a , N. M. Fonseca Ferreira2,3,4 b , Luciano Lopes de Sousa5 c , Nator Junior C. Costa5 d , Jose Vigno Moura Sousa ´ 5,6 e , D. B. S. Santos5 f , Antonio Valente2,7 g and Salviano Soares [Last access time: 8 February, 2019]

[19] https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia [Last access time: 27 November, 2019]