Deep Learning-based Diagnosis of Pneumonia using X-Ray Scans

Ansh Saxena[1], Shiva Singh Tomar[2], Gaurav Jain[3]

[1]Student, Amity Business School, Amity University, Noida, Uttar Pradesh
[2]Student, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh
[3]Student, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh

---***---

Abstract - *Since the beginning of 2020, the coronavirus infection which infected the world in 2019 (also known as COVID-19) has widely spread all over the globe. As it is popularly said that early detection of any disease may lead to more cures or longer survival. We established the formulations which the several models trained which is rapid and gives precise detection of viral pneumonia using chest X-ray can be very important on a large-scale epidemic testing and prevention.*

In this research-paper, several models have been trained based on three distinct chest x-ray datasets. The first dataset which has transpired from the publicly repository ie., Kaggle being publicly available, consisting of 5863 images, the second Dataset which consists of around 8000 images of pneumonia and the third dataset had around 22000 images. We devise the problem of distinguishing the Chest X-Ray pictures that are consisting of Pneumonia and non-pneumonia as one-class classification detection problem. It was intended to experiment with different optimizers, loss function, adding a dense layer with dropout or having a global average layer. Then, it was vital to go for the best combination of these characteristics and train transfer learning models such as inception, Xception, VGG-19, Efficient Net, InceptionResnetv2. The essential place of our task is to use "deep" learning methods and detect the diseases in scope from the designed pictures to get consistent, and high precision level of results. Nonetheless, generative, and discriminative learning are extremely novel. The motivation behind the exploration of a machine learning solution that utilizes a flexible, high-capacity CNN architecture while being efficient and fast result. Here, the description of different model choices has been given that has been found to be essential for obtaining competitive performing results. During the pandemic period artificial Intelligence (AI) has from the very beginning, been busily operating behind the screen helping the limits of human information on this huge endeavor. As we all know, machine learning is the main driving force behind AI. Then a sincere effort was invested in developing a supporting application which could help the models be utilized in a way which simulates the process of diagnosing diseases using image classification.

Key Words: Deep Leaning, Pneumonia, X-Ray Scans, Pandemic, Artificial Intelligence

1.INTRODUCTION

Pneumonia is an infection related to lungs with a range of possible causes, it can be serious and normally it is caused due to bacterial or fungal infection. Around 3.7 million people died in the USA because of this disease, all causing between the duration of January 2020 to February 2021. The possible common symptoms include Cough, Shortness of breath, Fever - sweating and shaking chills, Fatigue, Chest pain, Nausea, vomiting or diarrhoea, Confusion (especially in older adults). As it is known that clinical imaging is considered as a fundamental strategy to help doctors assess the infection and to advance anticipation and control measures. Clinically, chest X-ray remains the most typically utilized imaging methodology in the symptomatic workup of patients with thoracic irregularities, due to its quick imaging speed, low radiation, and minimal effort.

1.1 Motivation and the Issue with the Current Systems

The population Density of India is around 464 People per Km sq. according to the – Indian Consensus. The vulnerability of the Medical Situation of India was exposed during the Covid crisis, and it was remarked that the ratio of Doctors in India to patients was nearly 1:1456. It results in very critical and crucial time getting wasted ranging from collection of reports, appointment, diagnosis, expert advice to deciding the future implications. Hence, automation is required to speed up the process of detection and figuring the procedure to be followed.

2. Methodology

2.1 Dataset Description

In brief, the study for detection of pneumonia was done in three parts i.e., One with the data set 1 comprising of around 5800 images a subset of the Data set used in during the study of ChexNet, second dataset comprised of 8000 images and third one had 22000 images.

Firstly, talking about the first part of the study when we trained the models on the data set encompassing of 5800 images out of which 4685 images fitting to two classes were in the training set and 1171 (approximately 20%) images were in the testing set.

Table 1 Pneumonia Dataset – 1

Dataset - 1	**Images**	**Training Set**	**Testing Set**	**Source 1** [1]
Pneumonia	4273	3418	854	4273
Normal	1583	1266	316	1583
Total	**5856**	**4684**	**1170**	**5856**

Secondly, talking about the second part of the study when we trained the models on the data set encompassing of around 8500 out of which 7518 images have its place to two classes were in the training set and 1024 (approximately 12%) images were in the testing set. Here, while merging the datasets from different sources, it was observed that many files were already present in the dataset being considered previously and hence they were ignored to avoid redundancy in the image dataset.

Dataset - 2	**Images**	**Training Set**	**Testing Set**	**Source 1** [1]	**Source 2** [2]
Pneumonia	5618	4943	674	4273	1695
Normal	2924	2573	350	1583	1524
Total	**8542**	**7516**	**1024**	**5856**	**3219**

Table 2 Pneumonia Dataset – 2

Thirdly, talking about the third part of the study when we trained the models on the data set encompassing of around 22000 images out of which 20055 images be appropriate to two classes were in the training set and 2734 (approximately 12%) images were in the testing set. Here, while merging the datasets from different sources, it was observed that many files were already present in the dataset being considered previously and hence they were ignored to avoid redundancy in the image dataset.

Table 3 Pneumonia Dataset – 3

Dataset - 3	**Images**	**Training Set**	**Testing Set**	**Source 1** [1]	**Source 2** [3]	**Source 3** [2]	**Source 4** [4]
Pneumonia	14964	13168	1795	4273	4657	1695	5690
Normal	7825	6886	939	1583	3270	1524	3685
Total	**22789**	**20054**	**2734**	**5856**	**7927**	**3219**	**9375**

2.2 Image Augmentations and Pre-Processing

The pictures in the data set were utilized in the input shape of 224 x 224 pixels while implementing models of transfer learning and the shape of images used for training CNN (sequential model) was 64 x 64 pixels to decrease the number of convolution layers and the concealed layers. The several augmentations used for the study to prepare the dataset and prevent the model being over fit were Rotation in range of 20, Shear Range of 0.2, zoom range of 0.2 and horizontal flip.

2.3 Model Training and Architecture

After the literature review, we knew that the CNN model provided an accuracy of nearly about 85% and the RESNET had the second-best score of around 80% and then it was the CheXNet [5] Model of 121 layers to score around 78% while testing.

So, we wanted to experiment and get even better possible scores and for that we started experimenting with the optimizers, loss function, making the layers trainable or not, adding together a dense layer with drop out or getting a global average pooling layer instead. When we tried these on RESTNET 50 model the results did not enhance considerably but the change was significant. Then it was the need to go for the greatest potential combinations of these characteristics out of the possible ones, and train as many transfer learning models as possible. The combination that was chosen was to go for a dense layer with a dropout of 0.5 which is a simple way to reduce the over-fitting in the models being trained with binary cross entropy for the binary classification and categorical cross entropy for the categorical classification as optimizer functions and the loss function being Adam.

The models trained during the study were CNN (using sequential model), and various Transfer Learning models comprising of RESNET-50, INCEPTION V3, VGG19, VGG16, XCEPTION, EfficientNetB0, RESNET101V2, DeenseNet-201 and InceptionResnetV2. The models have been trained for 30 epochs each with early stopping being configured to each model.

Convolutional Neural Network (CNN – Sequential) - For the training of convolutional neural network, Sequential model was utilized with the help of TensorFlow and Keras libraries in python. The input shape for the model was taken as 64x64 pixels and each matrix had 3 dimensions associated with it. There were two Convolutional 2D layers each with a MaxPooling2d Layer supported by a Flatten layer to flatten every pooled feature map into a Vector.

Then there was an addition of 2 hidden layers used for the model each with ReLU (Rectified Linear Unit) activation function. Furthermore, a dense layer was added with Sigmoid as the activation function before compiling the model with Adam as the optimizer being used, and the loss function being binary Cross Entropy.

Transfer Learning Models

During the process of training for Transfer Learning Models, the input shape of the pictures was taken as 224 x 224 pixels and each matrix had 3 dimensions associated with it. The top layer of every model was not included, and the layers were restricted from being trained to reduce the parameters being trained onto, thus helping us reduce the training time of the model. Then, a flatten later was added with a dropout layer of 0.5 and then a dense layer to finally end the classification task. The loss function entertained was binary cross entropy and the optimizer function being Adam.

Training Process

Moreover, a reduction in learning rate of the layers was used with a factor of 0.3 and a level of patience of 2 epochs which signifies that the learning rate pf the model would tend to decrease by a factor of 0.3 after it encounters two epochs with same validation accuracy and when the accuracy increases, the model's weights are being saved in a separate HDF5 file. Furthermore, early stopping was utilized monitoring the validation accuracy with a patience of 5 epochs which signifies that in the situation of the model not being able to give better scores (or a flat line) in the validation accuracy, there is very less chance that it will yield better results in the coming epochs. So, to diminish the consumption of the resources and to save time, the model training stops after 5 epochs on the trot.

3. RESULTS AN INFERENCES

In the first Part of the study, several different transfer learning models have been trained for the problem domain out of which Resnet-50, Inception, Xception, VGG16 and VGG19 are some of the prominent names. The table demonstrated overhead are the findings and outcomes of the training and testing process for several different models trained. it was observed that VGG-19, Inception, Xception models were very close with their Testing Accuracy and the VGG-19 model outperformed its counterparts based upon the records of Testing Loss. Also, VGG-19 took the maximum time to get trained which is nearly 12 hours and 15 minutes. it was also observed that the CNN-Sequential Model was trained in merely 32 minutes which yielded Testing Accuracy of nearly 89%. Moreover, on the dataset VGG-16, Resnet101V2 and InceptionResnetV2 showed a flat line (Validation Accuracy did not improve more than the threshold point of 72%, 46% and 51% respectively) which signifies that it's not a decent notion to make these models work on the dataset. We can see the characteristics mentioned above from the table shown below as well.

Table 4 Cumulative Results for all the models which are trained for Pneumonia - Dataset 1

Cumulative Results for Pneumonia							
Model / Metrics	**Training Accuracy**	**Training Loss**	**Testing Accuracy**	**Testing Loss**	**Epochs**	**Training Time (Seconds)**	**Training Time (Hours)**
CNN - Sequential	0.96	0.11	0.90	0.31	30	1966.85	0.32
InceptionV3	0.92	1.83	0.94	1.24	30	8282.65	2.18
VGG-16	0.72	0.72	0.73	0.64	11	12456.00	3.27
RESNET-50	0.82	0.73	0.88	0.41	30	18735.96	5.12
Xception	0.94	1.06	0.93	1.00	30	17744.96	4.55
VGG-19	0.93	0.28	0.96	0.12	30	44563.99	12.22
EfficientNetB0	0.57	0.79	0.73	0.60	30	10646.36	2.57
ResNet101V2	0.46	2.40	0.46	1.38	4	2573.00	0.42
DenseNet-201	0.56	0.99	0.62	0.69	30	26565.64	7.22
InceptionResnetV2	0.50	1.36	0.51	0.91	4	2672.00	0.44

In the Second Part of the study, several different transfer learning models have been trained for the problem domain out of which Resnet-50, Inception, Xception, VGG16 and VGG19 are some of the prominent names.

The table exhibited above are the outcomes of the training and testing process for several different models trained. it was observed that VGG-19, InceptionV3, Xception and CNN (Sequential) models were very close with their Testing Accuracies and the VGG-19 model and CNN(Sequential) models outperformed their counterparts based on the scores of Testing Loss.

Also, VGG-19 took the maximum time to get trained which is nearly 8 hours or 535 minutes whereas the CNN-Sequential Model was trained in merely 34 minutes, yet it yielded a Testing Accuracy of nearly 92%. Moreover, on the dataset DenseNet-201, Resnet101V2 and InceptionResnetV2 showed a flat line (Validation Accuracy did not improve more than the threshold point of 72%, 46% and 51% respectively) which signifies that it is not a good idea to make these models work on the dataset. We can see the characteristics mentioned above from the graphs below as well.

Table 5 Cumulative Results for all the models that have been trained for Pneumonia - Dataset 2

Cumulative Results for Pneumonia (30 Epochs)								
Model / Metrics	**Training Accuracy**	**Training Loss**	**Testing Accuracy**	**Testing Loss**	**Epochs**	**Early Stopping**	**Training Time (Seconds)**	**Training Time (Hours)**
CNN - Sequential	0.94	0.15	0.92	0.21	18	Yes	2045.24	0.34

InceptionV3	0.94	0.24	0.93	0.32	10	Yes	5499.74	1.31
VGG-19	0.95	0.13	0.91	0.22	13	Yes	32104.02	8.55
RESNET-50	0.89	0.29	0.88	0.30	13	Yes	9933.26	2.45
Xception	0.95	0.16	0.90	0.39	9	Yes	8468.49	2.21
EfficientNetB0	0.54	0.76	0.66	0.65	6	Yes	3027.67	0.5
ResNet101V2	0.42	1.89	0.37	1.63	7	Yes	8584.25	2.23
DenseNet-201	0.56	0.80	0.61	0.66	8	Yes	9732.85	2.42
InceptionResnetV2	0.37	1.86	0.34	1.59	10	Yes	11527.97	3.12
InceptionV2 (2nd Attempt)	0.93	1.74	0.95	0.89	30	No	16260.01	4.31

Table 6 Cumulative Results for all the models that have been trained for Pneumonia - Dataset 3

			Cumulative Results for Pneumonia (30 Epochs)					
Model / Metrics	**Training Accuracy**	**Training Loss**	**Testing Accuracy**	**Testing Loss**	**Epochs**	**Early Stopping**	**Training Time (Seconds)**	**Training Time (Hours)**
CNN - Sequential	0.95	0.11	0.96	0.11	16	Yes	8734.81	2.25
InceptionV2	0.94	0.36	0.95	0.26	12	Yes	22203.38	6.1
VGG-19	0.95	0.12	0.94	0.14	9	Yes	78742.48	21.52

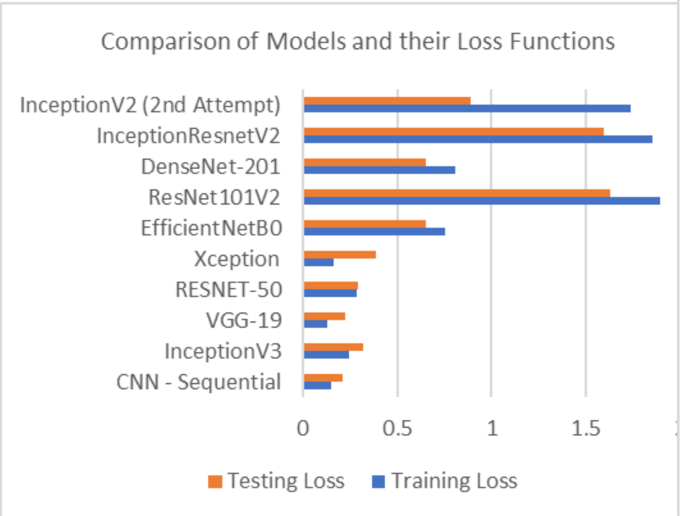

Figure 2 Comparison of Models and their Losses attained in second study.

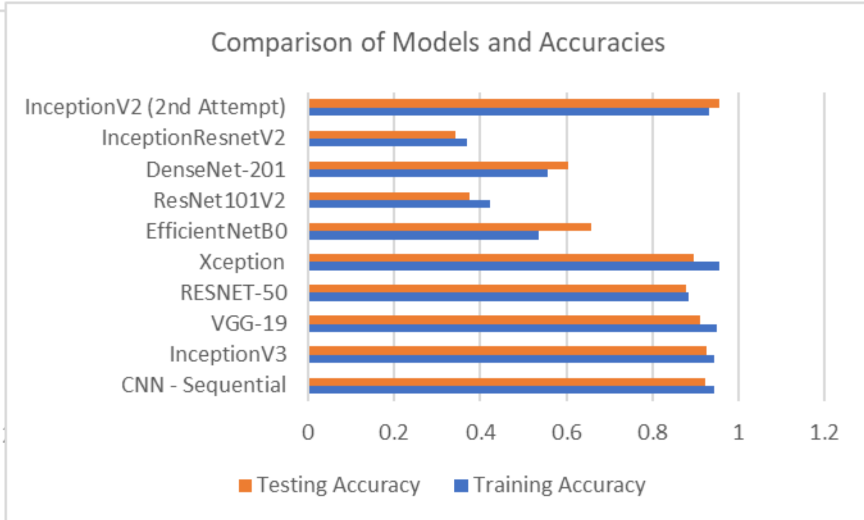

Figure 3 Comparison of Models and their Accuracies attained in second study.

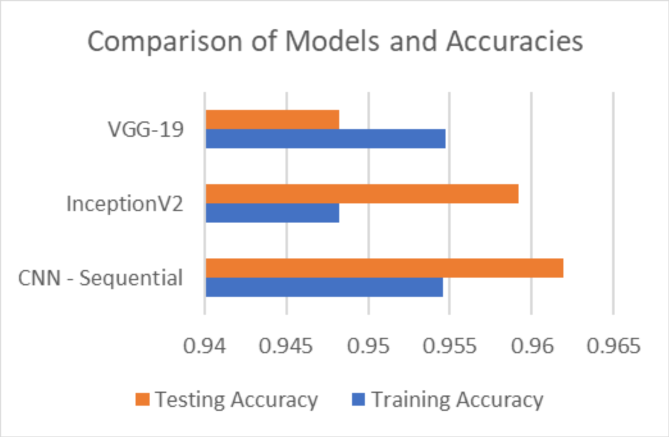

Figure 4 Comparison of Models and their Accuracies attained in third study.

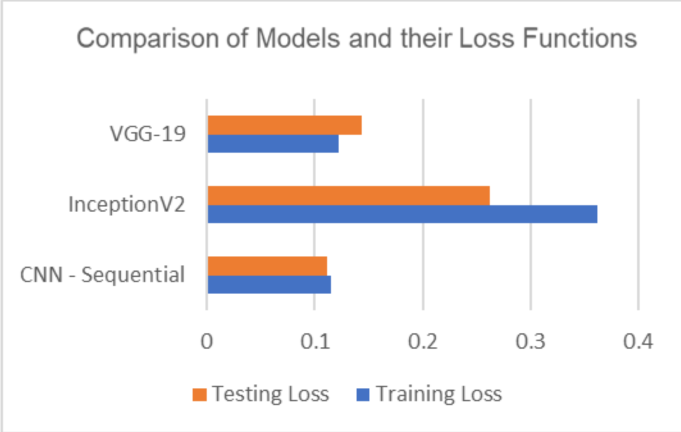

Figure 5 Comparison of Models and their Loss Functions attained in third study.

The Third Part of the study, several different transfer learning models which were obtained from the results previously attained have been trained for the conundrum domain out of which CNN(Sequential), Inception and VGG19 are some of the prominent names. The table exhibited above have been the outcomes of the training and testing process for several different models trained. It was observed that VGG-19, InceptionV3, and CNN (Sequential) models were very close with their Testing Accuracies and the CNN Sequential model outperformed their counterpart based on the scores of Testing Loss. Also, VGG-19 took the maximum time to get trained which is nearly 22 hours or 1312 minutes whereas the CNN-Sequential Model was trained in merely 145 minutes or 2 hours 25 minutes, yet it yielded a Testing Accuracy of nearly 96%.

Validation Results

Table 7 Validation results shown above for all the models trained for Pneumonia.

S.No.	Name	Dataset	Input Dims	Recall	Precision	F1-Score	Accuracy	True Positives	False Positives	True Negatives	False Negatives
1	**RESNET-50**	1	224	1.97	3.45	2.50	23.43	59	1653	1347	2941
2	**RESNET-50**	2	224	88.40	95.22	91.69	91.98	2652	133	2867	348
3	**Xception**	1	224	5.43	5.39	5.41	5.07	163	2859	141	2837
4	**Xception**	2	224	91.37	98.63	94.86	95.05	2741	38	2962	259
5	**InceptionV3**	1	224	100.00	50.00	66.67	50.00	3000	3000	0	0
6	**InceptionV3**	2	224	2.53	2.64	2.59	4.53	76	2804	196	2924
7	**InceptionV3**	3	224	93.90	97.98	95.90	95.98	2817	58	2942	183
8	**VGG19**	1	224	3.53	3.57	3.55	4.00	106	2866	134	2894
9	**VGG19**	2	224	90.87	98.70	94.62	94.83	2726	36	2964	274
10	**VGG19**	3	224	90.67	99.13	94.71	94.93	2720	24	2976	280
11	**CNN-Sequential**	1	64	95.07	95.99	95.53	95.55	2852	119	2881	148
12	**CNN-Sequential**	2	64	91.60	96.83	94.14	94.30	2748	90	2910	252
13	**CNN-Sequential**	3	64	92.67	97.75	95.14	95.27	2780	64	2936	220

4. CONCLUSIONS

It has been pragmatic that the automation is required to speed up detection and figuring the procedure to be followed consequently the research was directed with Pneumonia.

Firstly, the study for Pneumonia was conducted in 3 stages, the 1st part comprises of about 5800 images, the 2nd part comprises of about 8500 images and the 3rd part comprises of about 22000 images which were used to train CNN Sequential, Inception V3, VGG-16, VGG-19, RESNET-50, Efficient NET 80, Resnet101V2, DEnseNet201, Inception-Resnet V2. The best training accuracy was given by the VGG-19 model of 92% and training loss of 0.28 and testing accuracy and loss of about 96% and 0.12. In the 2nd part the best training accuracy was given by the CNN Sequential model of 92% and training loss of 0.14 and testing accuracy and loss of 92% and 0.21. In the 3rd part the best training accuracy was given by the CNN Sequential model of 95% and training loss of 0.11 and testing accuracy and loss of 96% and 0.11. The model of CNN was trained on 64x64 parameters while the VGG-19 model took 224x224 parameters, so it was better to go with the VGG-19 model because of better consideration.

As of now there wasn't any system which could assist the users/patients in detecting diseases harnessing the intelligence of computer systems. The domain of AI has been utilized over several field of applications such as computer vision, decision making, etc. but it has not been used in the realm of medical diagnostic of diseases efficiently. Hereby we planned to study and develop application which could facilitate the public/patients towards early diagnosis, hence better treatment opportunities. This would ultimately save lives using state of the art technology and give back to the society which has given us so many things.

Future Scope

It has been discovered that the automation is required to expedite the process of detection because typically, the Physicians spend around 14-15 minute to analyse the X-Ray scans and with the help of machine learning we can save the time and can predict patient is infected or not in a few seconds of time.

As it has been observed previously that the VGG-19 Model has provided the best possible results in the case of detection of Pneumonia, it must be tried with other combinations as well to get a lower training time somehow. It has taken a lot of time to get trained and therefore it must be reduced.

Moreover, for the application part, in future it can be built upon, and several types of diseases can be included in the scope such as Skin Cancer, Cancer Detection using X-Rays, specific eye related diseases like Glaucoma etc to aid the patients in recognising diseases as early as possible and save lives ultimately. The application in future can be re-established and released to public where people could use the platform to aid the patients by serving with correct and efficient results in the form of correct diagnosis.

REFERENCES

[1] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," University of California San Diego, doi: 10.17632/rscbjbr9sj.2.

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

[3] U. SAIT et al., "Curated Dataset for COVID-19 Posterior-Anterior Chest Radiography Images (X-Rays).," Indian Institute of Science, PES University, M S Ramaiah Institute of Technology, Concordia University.

[4] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, pp. 3462–3471, doi: 10.1109/CVPR.2017.369.

[5] P. Rajpurkar et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," Nov. 2017, [Online]. Available: http://arxiv.org/abs/1711.05225.