2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR) | 978-1-6654-2521-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAITPR51569.2022.9844184

Small-Scale CNN-N model for Covid-19 Anomaly Detection and Localization From Chest X-Rays

Jagadeesh Marusani
Data Science and Machine Learning,
PES University,
Bengaluru, India
jagadeeshreddymarusani@gmail.com

Sudha B G
Data Science and Machine Learning,
Great Learning,
Bengaluru, India
sudha.bg@greatlearning.in

Narayana Darapaneni
Artificial Intelligence and Machine Learning,
Great Learning,
Bengaluru, India
narayana@greatlearning.in

Abstract*— Covid-19 has been posing a serious challenge to scientists and health organizations around the world in terms of detection and its treatment. Common methods are CT-Scans and X-rays to analyze the images of lungs for COVID-19. These days diagnosing covid-19 by manually looking at the reports has become difficult and challenging in the pandemic. Pneumonia and pulmonary infections along with covid-19 cause inflammation and fluids in the lungs. Covid-19 X-rays are very similar to viral and bacterial Pneumonia X-rays. So it becomes very difficult to differentiate between covid-19 and Pneumonia. In this paper we propose a computer vision model to detect the presence of covid-19 infection along with the localization of the infection in the lungs. 6337 images consisting of Negative for pneumonia, Typical Appearance, Intermediate Appearance and Atypical Appearance is considered. Although there are pre-trained CNN models which perform well on the data, this paper aims at reducing the size of the model and validate its performance on other datasets. Different image sizes are also considered. A small scale CNN model is built from scratch to detect and localize covid-19 abnormalities on chest radiographs using object detection algorithms like Yolov5 with different weights. There is a significant reduction in model size and parameters compared to many state of the art pre-trained models thereby ensuring efficient detection of covid-19 anomalies and show the region of infection to ensure timely treatment before it causes severe infection.

Keywords— ***Covid-19, Chest X-rays, diagnosing, Pneumonia, Negative for pneumonia, Typical Appearance, Intermediate Appearance, Atypical Appearance, pre-trained models, object detection algorithms, CNN, Yolov5***

I INTRODUCTION

Covid-19 declared by World Health Organization (WHO) as a pandemic mainly affect human lungs that causes viral and bacterial pneumonia. It can be life-threatening if not acted upon at the right time. Pneumonia causes inflammation of either one or two lungs which cause decrease in air ratio of lungs and filled with fluids. Chest X-rays are known for diagnosis of the human lung infection. This paper aims to automatically detect the four possible diagnosis such as negative pneumonia, typical appearances, intermediate appearances, and atypical appearances [1] from chest x-rays. At times, the diagnosis of chest x-rays is challenging for expert radiologists. The appearance of disease or infected lungs may confuse the radiologist which leads to patient getting wrong treatment. In this study, a deep learning model is built from scratch and compare its performance with various pre-trained models like VGG16, VGG19, Inceptionv3, Res-net50 and Efficient-net [2][3]. We perform object detection using yolov5 algorithm to predict the opacity by predicting bounding boxes around it [4].

The performance of pre-trained models with hyperparameter tuning, fine-tuning and re-training to unfreeze the top and bottom layers is compared with the smaller sized model built from scratch. The main aim of this study is to showcase the efficiency of the smaller sized model.

II. RELATED WORK

Pneumonia is detected using CNNs - Alex-Net, Res-Net, Dense-Net and Squeeze-Net in [3]. The authors took reported accuracy of 98% for normal and pneumonia classes and 95% for viral and bacterial Pneumonia using DenseNet201. Narin, et.al in their work [5] uses large and well-balanced data with four classes viral, bacterial, covid-19 and normal. They demonstrate the use of ResNet50 and do a binary classification of the data and show significant results. The use of deep learning architectures for covid-19 detection is showcased on a new dataset that contains 48,260 CT scan images from 282 normal persons and 15,589 images from 95 patients with COVID-19 infections [6]. They build a feature pyramid network on ResNet50V2 and showcase their results. In [7] the authors come up with an ensemble of pre-trained CNN architectures and experiment on different approaches such as fine tuning, re-training from scratch, with and without augmentation, etc. Modification to Inception v3 and MobileNetV2 have also been explored with good results. A Chest X-ray Network (CheXnet) was built in [8] which shows very good accuracy in real time. Parallel processing was carried out after the images were enhanced with contrast-limited adaptive histogram equalization and Butterworth bandpass filters. The yolov5 [4] algorithm was used in [9] for anomaly detection and localization.

III. PROPOSED METHODOLOGY

A. Data Description

The dataset was shared by Society for Imaging Informatics in Medicine (SIIM) on kaggle.com for competition [10]. The dataset consists of 6334 chest X-rays train images with annotations. For test data the annotations are hidden. Images are in DICOM format and all are labeled by radiologist.

1) Data preparation

DICOM images were converted to JPEG images and re-sized to 256x256 and 512x512 pixels using PYDICOM library[11].

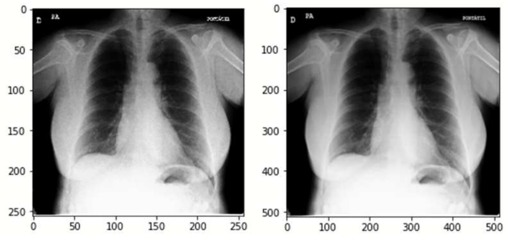

Fig 1: Re-sized images 256x256 and 512x512

Table 1: Dataset split into train and test

Labels	Total train Images	Train 80%	Test 20%
Negative for Pneumonia	1736	1389	347
Typical Appearances	3007	2407	601
Intermediate Appearances	1108	886	221
Atypical Appearances	483	387	96
Total	6334	5069	1265

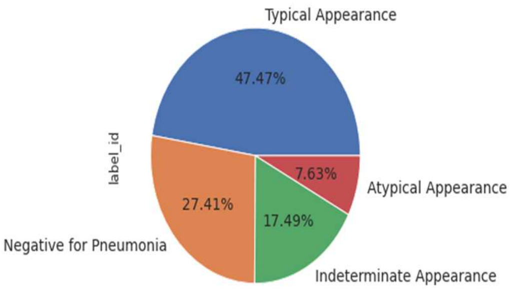

Fig 2: Pie chart for the distribution of Total train images

Characteristic used for anomaly detection: Opacity

Opacity: Pulmonary opacification is the decrease in the ratio of gas to soft tissue in lungs. For pneumonia affected persons the air in lungs are replaced with bacteria, virus and fluids which surface as the hazy gray areas in CT scans or X-rays of the lungs. The gray areas indicate increase in density inside the lungs. Opacities could be localized within the image as bounding boxes [12]. If an X-ray has three opacities, three boundary boxes could be drawn and the distribution is shown in Fig 3 and the images are shown in Fig. 4.

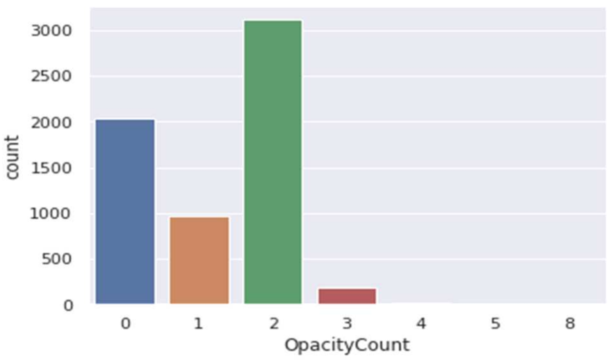

Fig 3: Opacity count of the images

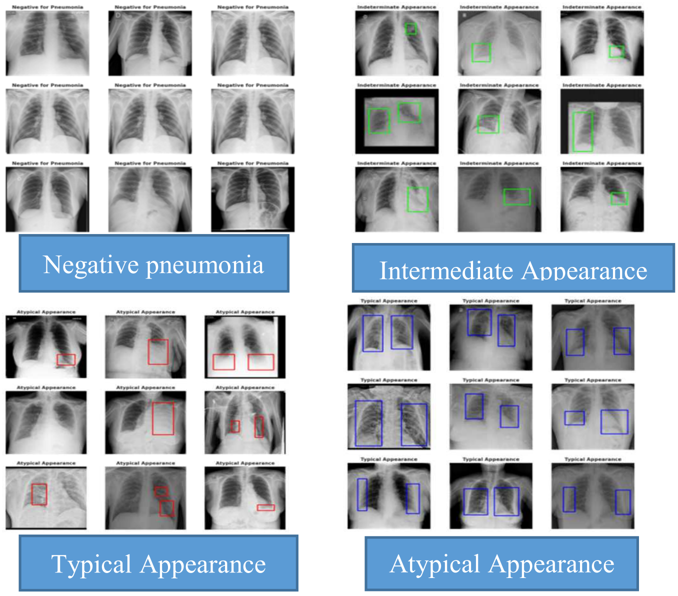

Fig 4: Opacity bounding boxes in CXR for different classes

B. Schematic diagram for classification

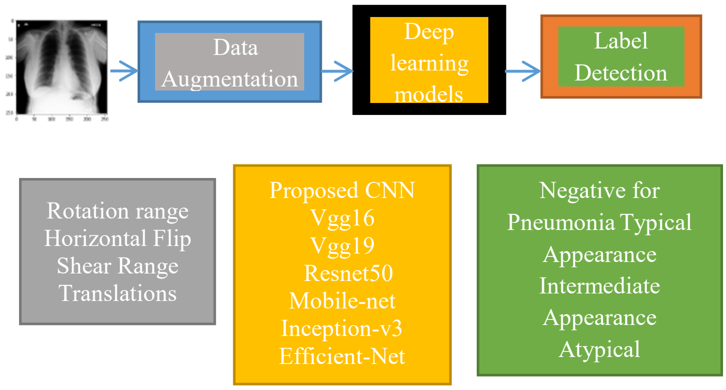

Fig 5: Schematic Diagram for classification

C. Data Augmentation

Data Augmentation is a technique used to generate the new training samples from the original data. For testing data we do not apply data augmentation. All pre-trained models are large to evaluate this data and may cause overfitting. To prevent this overfitting by adding some noise to this dataset for better performance like Rotation range, horizontal flip, shear range, width shift range, height shift range, zoom range [2][15].

1. Rotation Range

The following transformation is applied on the image to get rotation image. Transformation matrix = $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, Where θ range from 0 to 360 degrees[15].

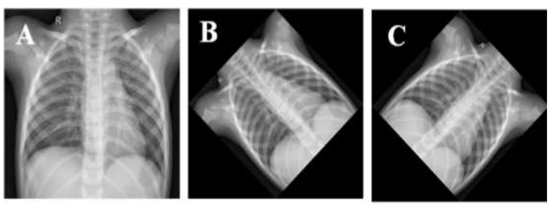

Fig 6: A. Actual image B. Counter clockwise rotated image. C. Clockwise rotated image

2. Horizontal Flip

Transformation matrix = ([[1, 0, 0], [0, -1, 0], [0, 0, 1]]) values given to Actual image to get horizontal image or reflected image.

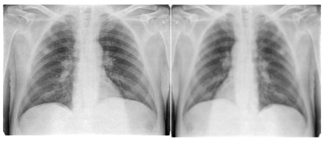

Fig 7: Actual image vs Reflected image

3. Shear

Transformation matrix = ([[1,0,0], [.5,1,0], [0,0,1]]) values given to Actual image to get shear image. Shearing in horizontal direction means that the upper part pixels has shifted to the right and the lower part pixels to the left.

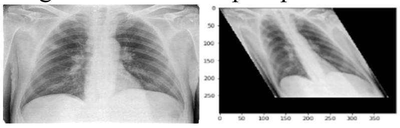

Fig 8: Actual image vs shear image

4. Translation

Translation of an image means pixel of the image will shift towards right, left, top and bottom without changing the dimension of the image.

Transformation matrix = ([[1,0,20],[0,1,20],[0,0,1]])

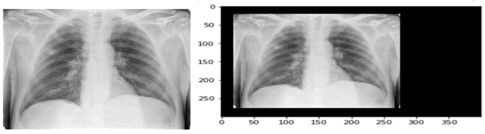

Fig 9: Actual image vs Width shift range

D. Proposed CNN Model

ur proposed CNN model named CNN-N has of 21 layers starting with less convolution layers followed by increasing and then reducing the layers. Two different filter sizes such as 3x3 and 4x4 are used. Batch normalization and activation function as rectified linear unit (RELU) are applied to the convolved images and one max pooling with size of 2x2 is used to reduce the dimensionality of feature maps. After flattening the data three more fully connected layers were added along with final layer activation function as SOFTMAX.

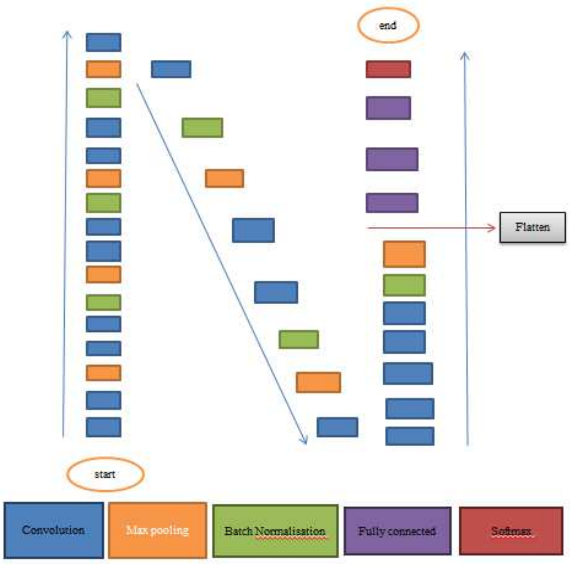

Fig 10. Proposed CNN model architecture

IV. Results and Discussion

Table 2: 256 x 256 label detection using deep learning models

Model	Training accuracy (%)	Testing accuracy (%)	Fine tuning training (%)	Fine tuning testing (%)	Parameters	Size (MB)
PROPOSED CNN	83.74	83.56	-	-	11,572,670	42
VGG16	80.12	80.68	82.67	83.63	15,387,268	60
VGG19	80.73	80.29	80.53	80.24	20,696,964	86
RESNET50	74.53	77.05	84.63	84.49	26,128,772	102
INCEPTIONV3	71.65	71.62	79.12	77.74	31,248,612	197
MOBILENET	71.63	71.62	84.26	82.86	74,32,068	40
EFFICIENTNETB0	77.05	80.43	82.24	83.41	14,535,975	139
EFFICIENTNETB1	78.56	81.62	85.01	83.81	17,061,643	65
EFFICIENTNETB2	78.61	82.29	85.84	83.73	30,838,525	118
EFFICIENTNETB3	79.43	82.14	86.46	84.16	23,367,091	123
EFFICIENTNETB4	82.18	83.13	87.81	84.37	18,043,299	179
EFFICIENTNETB5	83.52	82.64	88.02	84.42	32,708,475	161
EFFICIENTNETB6	82.27	82.25	86.43	83.47	41,821,267	171
EFFICIENTNETB7	84.50	82.90	88.91	84.50	67,129,563	287

Table 3: 512x512 pixel images label detection using deep learning models

Model	Training accuracy (%)	Testing accuracy (%)	Fine tuning training (%)	Fine tuning testing (%)	Parameters	size (MB)
Proposed CNN	83.95	83.65	-	-	11,965,886	45
VGG16	82.84	82.24	83.45	83.89	15,714,948	67
VGG19	82.97	82.48	83.15	83.25	21,024,644	90
RESNET50	83.38	83.06	86.54	84.50	26,357,444	125
INCEPTIONV3	67.71	65.71	79.07	78.37	47,501,796	387
MOBILENET	71.58	75.36	84.57	79.85	20,014,980	204
EFFICIENTNET-B0	84.28	83.48	86.78	84.55	25,029,99	95
EFFICIENTNETB1	82.99	82.87	88.34	84.17	27,555,659	78
EFFICIENTNETB2	83.36	83.04	87.17	84.39	11,005,245	55
EFFICIENTNETB3	83.34	82.89	86.44	83.94	14,167,667	68
EFFICIENTNETB4	83.59	83.04	87.23	84.02	21,352,867	99
EFFICIENTNETB5	82.73	83.39	89.24	84.01	31,312,539	145
EFFICIENTNETB6	81.97	81.92	88.15	83.95	44,204,787	197
EFFICIENTNETB7	82.33	83.15	88.04	83.44	67,457,243	291

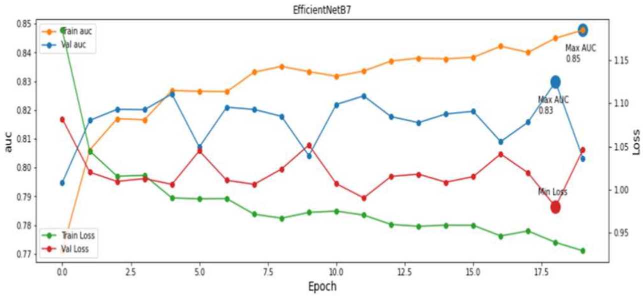

Fig 11: EfficientNetB7 Results (256x256)

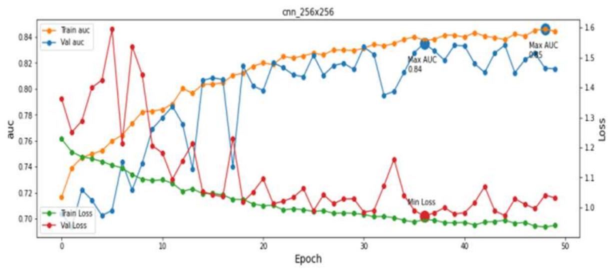

Fig 12: Proposed CNN Results (256x256)

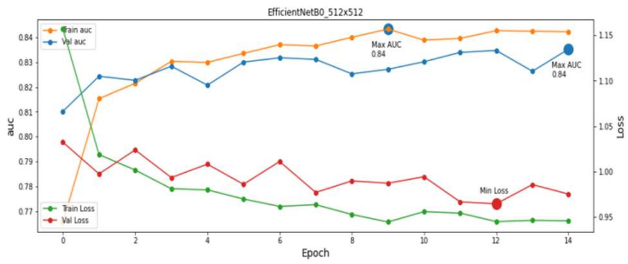

Fig 13: EfficientNetB0 Results (512x512)

The proposed CNN model has 11.5 and 11.9 million parameters with 21 convolutional layers whose size are 42 MB and 45 MB which is 6 times less compared to the best model Efficient-netB7. Thus the model runs in less time and performs well for the data. This is compared with the fine-tuned models such as VGG16, VGG19, ResNet50, InceptionV3, Mobile-Net and Efficient-Net B0 to EfficientNetB7. As per the results shown in Tables 2 and 3, both 256 and 512 images perform well but 512 images run slowly because of its size and information they possess and process. Performing fine tuning for all pre-trained models takes a lot of time to run the model. For the proposed CNN model, the hyperparameters used are Adam optimizer, ReLU activation function with a learning rate of 0.001 and with batch size as 64 although 0.0001 did not show much variation and hence not shown here.

A. Further Analysis and experiments on this data

1. Misclassified classes and Modeling with Augmentation:

On analysis of the misclassified classes, it was found that the class Atypical Appearance class led to the drop in overall accuracy. This is due to the fact that the number of observations are less. Hence to increase accuracy, data augmentation is done again to increase the minority class to see if the accuracy goes up. Table I shows that the data is imbalanced. By using image data generator new images are further created using augmentation techniques like rotation range, horizontal flip, shear range, translation like width shift range, height shift range. Here, only balanced atypical appearance class images are augmented so that the total images after augmenting is 1000.

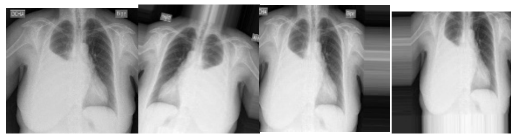

Fig 14: Generating new images using image data generator

Table 4: Balanced data

Labels	Total Train Images	Train (80%)	Test (20%)
Negative for Pneumonia	1736	1389	347
Typical Appearances	3007	2407	601
Intermediate Appearances	1108	886	221
Atypical Appearances	1000	800	200
Total	6851	5482	1369

Efficient-netB7 model was then applied to balanced data. After seeing the results for balanced data it doesn't perform well on testing data compared to imbalanced data and got training accuracy as 86% and testing accuracy as 78% which was not an improvement on the imbalanced data. So, further modeling is done using the data without augmentation.

2. Modeling without Augmentation:

Same data we perform the model without augmentation technique. Here, using Efficient-netB7 model got train accuracy as 96% and test accuracy as 76% shown in below figure 12. Clearly it is a overfitting model means data perform well on train compare to test.

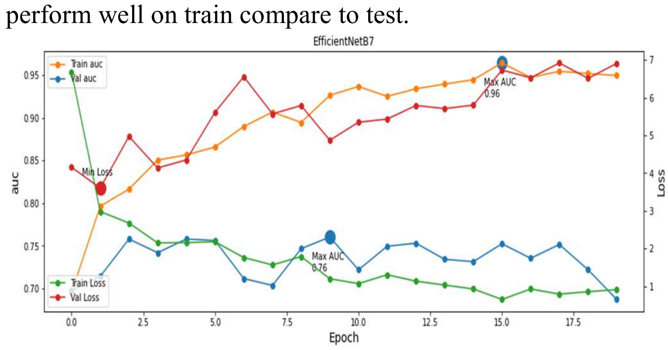

Fig 15: EfficientNetB7 Results for model without Augmentation (256 x 256)

3. Image Equalization using CLAHE:

Contrast limited Adaptive histogram equalization (CLAHE) is a pre-processing technique used to enhance contrast to improve quality [14].

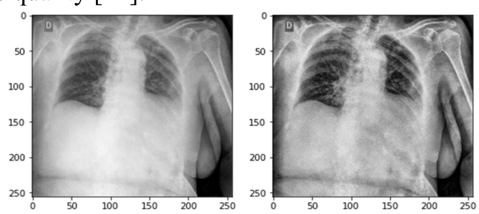

Fig 16: Normal image vs CLAHE image

The CLAHE images (256x256) are used in only the top three models from the Table 2.

Table 5: Classification results for CLAHE (256x256) and Normal images (256x256)

Model	CLAHE Train Accuracy (%)	CLAHE Test Accuracy (%)	Normal Train Accuracy (%)	Normal Test Accuracy (%)
Proposed CNN	85.64	83.60	83.74	83.56
Efficient-netB6	84.04	82.01	82.27	82.25
Efficient-netB7	84.01	82.22	84.50	82.90

From Table 5 it can be seen that the proposed CNN model slightly performs well on CLAHE images compared to normal images.

4. Validation of proposed CNN model on different datasets:

Two different datasets are taken to validate the performance of our proposed CNN model.

Dataset 1 Data of tomato leaf diseases classification from [15] [16] is used.

Table 6: Tomato leaf disease images count

Labels	Train	Test
Tomato bacterial spot	500	70
Tomato healthy	500	70
Tomato Target Spot	500	70
Tomato Yellow Leaf Virus	500	70

Table 7: Classification Results

Model	Training Accuracy (%)	Testing Accuracy (%)
Proposed CNN	97.61	96.87

Dataset 2 This dataset [17,18] consist of four different classes such as covid-19, Normal, Pneumonia and Tuberculosis.

Table 8: Train Test Split

Labels	Total Images	Train (80%)	Test (20%)
Covid-19	566	460	106
Normal	615	492	123
Pneumonia	890	712	178
Tuberculosis	551	441	110

Table 9: Classification Results

Model	Training Accuracy (%)	Testing Accuracy (%)
Proposed CNN	97.33	96.30

Dataset 3. The dataset was taken from kaggle [19] which has 4 different classes (covid-19, normal, viral and bacterial pneumonia).

Table 10: Train Test Split

Labels	Train	Test
Normal	880	176
Viral Pneumonia	412	82
Bacterial Pneumonia	650	130
Covid-19	60	15

Table 11: Classification Results

Model	Training Accuracy (%)	Testing Accuracy (%)
Proposed CNN	96.65	96.61

From the results of Table 7, 9, and 11, it is proven that the proposed CNN model performs well on other datasets as well.

5. Localization of pulmonary opacification:

Yolo (you only look once) is a real time object detection used for its accuracy and speed. Yolo applies a neural network to an image and then divide it into regions to predict bounding boxes and probabilities for each region. First, image is divided into a grid of say, 15*15 cells (s=15). Each of these cells is responsible for predicting minimum of five bounding boxes (b=5). A bounding box describes the rectangle that encloses an object and for each bounding box outputs a confidence score that tells us how good is the shape of the box.

Intersection over union (IOU)

IOU is a measure of overlapping between the ground truth boxes and predicted boxes. It ranges from 0 to 1 where 0 means no overlap and 1 means perfectly overlap having threshold value as 0.5.

$$\boxed{\text{IOU}} = \frac{\text{Area (Ground truth} \cap \text{Predicted boxes)}}{\text{Area (Ground truth} \cup \text{Predicted boxes)}}$$

In this paper Yolov5 algorithm for opacity detection is used for bounding boxes – as shown in (Fig 4). Using different yolov5 weights such as Yolov5s, Yolov5m, Yolov5l, Yolov5x [4]. A total of 4853 train images and 1211 test images. First, need to split the images in to five batches along with batch size as 24, epochs as 40. Here, using metrics as True positive and False Negative shown on below Table 12.

Table 12: Yolov5 results with different weights

Yolov5 weights	True positive (%)	False Negative (%)
Yolov5s	0.49	0.51
Yolov5m	0.53	0.47
Yolov5l	0.57	0.43
Yolov5x	0.52	0.48

True positive - If IOU ≥ 0.5, opacity bounding boxes were correctly predicted.

False Negative – when ground truth is present, model failed to detect the opacity.

From the above Table 12, Yolov5l weights gives better true positive compare to other yolov5 weights although there is scope for further improvement. Fig. 17 shows the results of Yolov5l results.

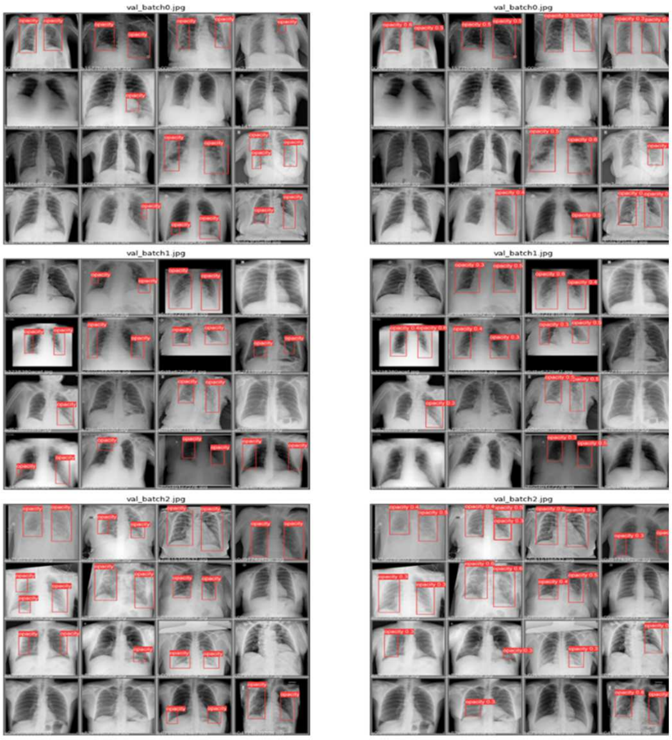

Fig 17: Yolov5l localization results. The left column ones are the ground truth and the right column image bonding boxes are the predicted ones.

V. CONCLUSION

This work presents a deep learning approach for automatic detection of covid-19 and its classes and opacity localizations. The CNN model proposed showed competent results on the chest X-ray dataset and on validation with different datasets, it shows good performance. This model is smaller in size with 6 times lesser parameters to train. It is comparable and sometimes even showed better results compared to the state-of-the-art EfficientNetB7 model. Comparison on different image resolutions 256x256 and 512x512 results shows good results in terms of accuracy. Data augmentation did not prove to be useful for this architecture and for other models as well. For opacity bounding box detection, Yolov5 algorithms with different weights are used. Yolov5l weights performs well compared to Yolov5s, Yolov5m and Yolov5x based on true positive, which makes the task of quantification of the opacification easier for the radiologist.

VI. FUTURE ENHANCEMENT

The CNN-N model would be further improved and validated with few other benchmark datasets to check if it would surpass the best pre-trained model. There is scope for improvement of the Yolov5l opacity localization results. Different set of hyperparameters could be further explored.

REFERENCES

[1] Simpson, Scott, et al. "Radiological Society of North America expert consensus document on reporting chest CT findings related to COVID-19: endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA." Radiology: Cardiothoracic Imaging 2.2 (2020): e200152.

[2] Alazab, Moutaz, et al. "COVID-19 prediction and detection using deep learning." *International Journal of Computer Information Systems and Industrial Management Applications* 12 (2020): 168-181.

[3] Rahman, Tawsifur, et al. "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray." *Applied Sciences* 10.9 (2020): 3233.

[4] Qu, Ruyi, Yi Yang, and Yuwei Wang. "COVID-19 Detection Using CT Image Based On YOLOv5 Network." *arXiv preprint arXiv:2201.09972* (2022).

[5] Narin, A., Kaya, C., & Pamuk, Z. (2021). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*, *24*(3), 1207-1220.

[6] Rahimzadeh, M., Attar, A., & Sakhaei, S. M. (2021). A fully automated deep learning-based network for detecting covid-19 from a new and large lung ct scan dataset. *Biomedical Signal Processing and Control*, *68*, 102588.

[7] Ahmad, Fareed, Muhammad Usman Ghani Khan, and Kashif Javed. "Deep learning model for distinguishing novel coronavirus from other chest related infections in X-ray images." *Computers in biology and medicine* 134 (2021): 104401.

[8] Al-Waisy, Alaa S., et al. "COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images." *Soft computing* (2020): 1-16.

[9] Qu, Ruyi, Yi Yang, and Yuwei Wang. "COVID-19 Detection Using CT Image Based On YOLOv5 Network." *arXiv preprint arXiv:2201.09972* (2022).

[10] SIIM-FISABIO-RSNA COVID-19 Detection | Kaggle 17-May-2021 to 9-August-2021 [online].

[11] https://pydicom.github.io/ (Last accessed: Dec 10, 2021].

[12] Muhammad, L. J., et al. "Deep Learning Models for Classification and Localization of COVID-19 Abnormalities on Chest Radiographs." 2021 4th International Conference on Computing & Information Sciences (ICCIS). IEEE, 2021.

[13] Hussain, Zeshan, et al. "Differential data augmentation techniques for medical imaging classification tasks." *AMIA annual symposium proceedings*. Vol. 2017. American Medical Informatics Association, 2017.

[14] K. Koonsanit, S. Thongvigitmanee, N. Pongnapang and P. Thajchayapong, "Image enhancement on digital x-ray images using N-CLAHE," 2017 10th Biomedical Engineering International Conference (BMEiCON), 2017, pp. 1-4, doi: 10.1109/BMEiCON.2017.8229130.

[15] https://www.kaggle.com/kaustubhb999/tomatoleaf

[16] Zhang, Keke, et al. "Can deep learning identify tomato leaf disease?." Advances in multimedia 2018 (2018).

[17] Mamalakis, Michail, et al. "DenResCov-19: A deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays." Computerized Medical Imaging and Graphics 94 (2021): 102008.

[18] https://www.kaggle.com/jtiptj/chest-xray-pneumoniacovid19tuberculosis?select=train

[19] https://www.kaggle.com/darshan1504/covid19-detection-xray-dataset?select=NonAugmentedTrain