COVID-DenseNet: A Deep Learning Architecture to Detect COVID-19 from Chest Radiology Images

Laboni Sarker¹, Md. Mohaiminul Islam², Tanveer Hannan³, and Zakaria Ahmed⁴

¹laboni@uap-bd.edu, ²mmiemon@uap-bd.edu, ³tanveer.hannan@campus.lmu.de, ⁴zakaria.ahmed@enosisd.com

^{1 2 3 4}B.Sc. from Department of CSE, Bangladesh University of Engineering and Technology

^{1 2}Lecturer, University of Asia Pacific, Bangladesh, ³M.Sc.Student, Ludwig Maximilian University of Munich, Germany

⁴Software Engineer, Enosis Solutions, Bangladesh

*All authors contributed equally and share the first-authorship of this paper.

Keywords: deep learning; CNN; DenseNet; COVID-19; transfer learning

Abstract—Coronavirus disease (COVID-19) is a pandemic infectious disease that has a severe risk of spreading rapidly. The quick identification and isolation of the affected persons is the very first step to fight against this virus. In this regard, chest radiology images have been proven to be an effective screening approach of COVID-19 affected patients. A number of AI based solutions have been developed to make the screening of radiological images faster and more accurate in detecting COVID-19. In this study, we are proposing a deep learning based approach using Densenet-121 to effectively detect COVID-19 patients. We incorporated transfer learning technique to leverage the information regarding radiology image learned by another model (CheXNet) which was trained on a huge Radiology dataset of 112,120 images. We trained and tested our model on COVIDx dataset containing 13,800 chest radiography images across 13,725 patients. To check the robustness of our model, we performed both two-class and three-class classifications and achieved 96.49% and 93.71% accuracy respectively. To further validate the consistency of our performance, we performed patient-wise k-fold cross-validation and achieved an average accuracy of 92.91% for three class task. Moreover, we performed an interpretability analysis using Grad-CAM to highlight the most important image regions in making a prediction. Besides ensuring trustworthiness, this explainability can also provide new insights about the critical factors regarding COVID-19. Finally, we developed a website that takes chest radiology images as input and generates probabilities of the presence of COVID-19 or pneumonia and a heatmap highlighting the probable infected regions. Code and models' weights are available. ¹

I. INTRODUCTION

Coronavirus disease (COVID-19) is a pandemic infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which is now a global issue as there are no specific vaccines or treatments for this. As it is able to infect people easily and can spread from person-to-person in an efficient and sustained way, the quick identification and isolation of the affected person is the very first step to

fight against this virus. Polymerase chain reaction (PCR) is the main method and gold standard for detecting COVID-19 cases which can detect SARS-CoV-2 RNA from respiratory specimens (collected through a variety of means such as nasopharyngeal or oropharyngeal swabs). [12] Though this method is the most effective one, it is very time consuming and intensive lab work is required after the collection of the samples to get result.

Another approach is examination of chest radiography imaging (e.g., Radiology or computed tomography (CT) imaging) which can be conducted faster but expert analysis is needed to interpret the subtle differences. To remove this bottleneck, many AI based systems have been proposed to detect COVID-19 from radiography images. Moreover, AI solutions are much faster than traditional system where radiologist need to examine the images by hand.

In our work, we have used Dense Convolutional Network (DenseNet) [6] of 121 layers as our model. DenseNet makes the training of deep learning models tractable by removing vanishing gradient problem, enhancing feature reuse, and increasing parameter efficiency. It has achieved the state-of-the-art performance in several computer vision tasks. Moreover, DenseNet has been used successfully in disease prediction from Radiology images. In paper [10], DenseNet-121 was used to detect 14 kinds of diseases from chest Radiology images (CheXNet) and better performance was achieved than practicing academic radiologists. Paper [4] also used DenseNet-121 for disease prediction from Radiology images of ChestRadiology-14 dataset and further improved the performance achieved by paper [10]. Motivated by the excellent performance of DenseNet on Radiology images (e.g. paper [10] and [4]), we used DenseNet-121 as our deep learning model. Moreover, we initialized our model's weights by the weights of CheXNet [10]. Our intuition of using this transfer learning technique was utilization of the information regarding Radiology images present in CheXNet pretrained model, since CheXNet was trained on ChestRadiology-14 [13] dataset containing 112,120 frontal view Radiology images from 30,805 unique patients.

We trained our model on CovidX dataset containing 13,800 chest radiography images across 13,725 patients. We tested our model for two class classification (COVID-19 vs non-COVID-

¹Code for reproducing results is available at <https://github.com/mmiemon/COVID-DenseNet> and models' weights can be found at <https://bit.ly/2YZwyk3>

19) and three class classification (COVID-19 vs Pneumonia vs Normal). We achieved 96.49% accuracy for two class and 91.85% accuracy for three class classification. These results show that our model is capable of differentiating COVID-19 Radiology images not only from the from the images of a healthy person but also from the images of other pneumonia patients. To check the robustness and consistency of our model, we performed 10-fold cross validation where no two fold contains COVID-19 images from same patients (patient-wise cross-validation) and achieved an average accuracy of 91.48%.

We used Gradient-weighted Class Activation Mapping (Grad-CAM) [11] to visualize how our model works. Using Grad-CAM, we created a heatmap for each input image highlighting the most important regions for which our model makes a certain prediction. This ensures interpretability as well as trustworthiness of our model. This works as a safeguard that our model is not making prediction based on inappropriate portions of the input Radiology image. Moreover, this will help doctors and clinicians to visualize the most significant features and give insights about the critical factors of COVID-19 patients.

It is important to develop a tool for allowing users to use our model and generate predictions easily. We developed a [web application](#) [3] which adapts our model to provide real-time predictions. We used TensorFlow.js for converting our model to work in the browser. The web application also generates heat maps of the Radiology images. A RESTful API is implemented using Flask micro web framework which is used to generate heat maps.

II. RELATED WORKS

A number of works has been done in this short period of time in detecting COVID-19 from radiography images. Different model architectures are used for accurate detection of the disease.

In paper [9], they have created a new model for COVID detection and named it as CovidNet using human-machine collaborative strategy. They have used two open source datasets, COVID Radiology data and Kaggle chest Radiology (pneumonia) dataset. In the final detection, they have used 4 class classification: Normal, Bacterial, Non-COVID19 Viral and COVID-19 Viral.

Table I: Sensitivity(Recall).

Normal	Bacterial	Non-Covid19 Viral	COVID19 Viral
73.9	93.1	81.9	100.0

Table II: Positive predictive value(Precision).

Normal	Bacterial	Non-Covid19 Viral	COVID19 Viral
95.1	87.1	67.0	80.0

From the table I and II, it is clear that the COVID-NET is very good in detecting Covid19 infection as sensitivity(Recall)

Figure 1: DenseNet-121 with 4 dense blocks and 3 transition layers

is 100%. A small portion of Radiology images can be misclassified as COVID-19. But for other classes, both the sensitivity(Recall) and PPV(Precision) rate is not that much good. So, there is a lot more to contribute to properly detect the COVID-19 infection from other respiratory infections as they are all very similar. The CovidNet model has acquired a test accuracy of 83%.

Paper [8] has distinguished COVID-19 from Community Acquired Pneumonia(CAP) from chest CT imaging. They have collected data from 6 hospitals. This dataset is not publicly accessible. A 3D deep learning framework referred as COVNet which consists of a RESTNet50 as the backbone is used in their work. They have segmented the lung region from the chest Radiology images using UNet. To train their model, they have used 1165 images of covid19, 1560 images from CAP and 1193 images of non-pneumonia CT scans. The reason to train the model with CAP and non-pneumonia CT images is to check the robustness that how efficiently the model can distinguish COVID-19 from other similar lung diseases.

Table III: Sensitivity(Recall) & Specificity.

	COVID-19	CAP	Non-pneumonia
Sensitivity	90	87	94
Specificity	96	92	96

Table III gives us an overview of the performance of their model which seems very promising, but not for public use.

III. METHODOLOGY

A. Model Architecture

For our model, we used a 121-layer densely connected Convolutional Network (DenseNet) [6]. Unlike traditional Convolutional Networks, in DenseNet every layer is directly connected with all other layers and each layer has direct access to loss functions and original input signals. Feature-maps of all preceding layers are concatenated and used as input for any particular layer and its own feature-maps are used as inputs into all subsequent layers. This special design improves information flow through the network and alleviates vanishing gradient problem. Moreover, DenseNet enhances feature reuse and parameter efficiency and provides each layer collective knowledge of the network. Another important reason for choosing DenseNet as our architecture is that dense connection has regularization effect and it reduces over-fitting on training with smaller data sets [6], which is our case.

DenseNet-121 has four dense blocks and a transition layers between each two dense blocks (Figure 1). Each dense block consists of several convolution layers and each transition layer consists of a batch normalization, a convolution and a average pooling layer. Finally, we have a fully connected layer with with soft-max activation function with three neurons for three-class classification and two neurons for two-class classification.

B. Data Generation

The Radiology images of COVID-19 infected patients are extremely rare. We used COVIDx Dataset assembled by [9]. They combined open source databases with chest Radiology or CT images from [5], [1], [2]. The total number of Corona infected Chest images are only 238. This is extremely small compared to the number of Radiology images available for Pneumonia infected and healthy persons which are 6045 and 8851 respectively. So the data is highly skewed because of the scarcity of images of COVID-19 patients. To deal with unbalanced data-set we augmented only the COVID-19 images in the training set. The following table IV shows the distribution of images before and after augmentation.

Table IV: Class Distribution.

Augmentation/ Class	Normal	Pneumonia	COVID-19
No	8851	6045	238
Yes	8851	6045	11416

The train and test split ratio is fixed at 0.1. We also stratified the train, validation and test split so that the proportion is maintained in each set. We augmented the training data in six different methods e.g. width shift, height shift, horizontal flip, rotation, brightness change, and zoom in or zoom out. We created 9 different images randomly for each category. So each COVID-19 Radiology images in the training set has a total of 54 augmentations. To validate the result, the data-set was prepared for 10-fold cross validation keeping the proportions of class label same for each fold. We maintained augmentation leakage by creating an index system so that the augmentation of images in one fold does not fall in another one. We also maintained index for patient ids' so that no two folds have images of the same patient. Each patient has variable number of images. So dividing the patients randomly among 10-folds would create imbalance in terms of number of images in each fold. So we had to maximize both the number of patients and images for each fold at the same time. We thus reduced correlation between train, and test images.

The COVID-19 data-set is currently growing and we created new data injection method to add new images to our data-set. This method also performs all the balancing acts to reduce the correlation of images between each fold.

C. Model Implementation

DenseNet-121 consists of 121 densely connected convolutional layers with a fully conncted(FC) layer of 1000 units as its final output layer. We removed the final layer and replaced

it with a FC layer with two neurons for two class classification and three neurons for three class classification. We initialized our models weights by the weights of CheXNet [10], which was trained on ChestRadiology-14 [13] dataset of 112,120 chest Radiology images. Since CheXnet was already trained to extract features from chest Radiology images, we used this transfer learning method to leverage the pretrained model.

The network was trained end-to-end with Adamax optimizer with standard parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) [7], and learning rate = .00001. Categorical cross-entropy was selected as loss function. The learning rate was reduced by the factor of 0.1 when validation loss plateaus. Early stopping with patience set to 5 was used to stop over-fitting. The train, validation and test split was set to 0.8, 0.1, and 0.1.

IV. EXPERIMENTAL RESULTS

Table V: Experiment-1 result.

Class/ Metric	Accuracy	Precision	Recall	f-score
Overall	0.94	0.94	0.94	0.94
COVID-19	0.87	0.87	0.87	0.87
Pneumonia	0.93	0.95	0.93	0.94
Normal	0.95	0.94	0.96	0.95

Table VI: Confusion matrix for experiment-1.

Predicted/ Actual	COVID-19	Pneumonia	Normal
COVID-19	27	3	1
Pneumonia	2	93	3
Normal	2	4	96

Table VII: Experiment-2 result.

Class/ Metric	Accuracy	Precision	Recall	f-score
Overall	0.93	0.92	0.92	0.92
COVID-19	0.86	0.77	0.85	0.81
Pneumonia	0.90	0.93	0.91	0.92
Normal	0.95	0.94	0.94	0.94

Table VIII: Experiment-3 result.

Class/ Metric	Accuracy	Precision	Recall	f-score
Overall	0.96	0.96	0.96	0.96
COVID-19	0.93	0.90	0.94	0.92
Non COVID-19	0.96	0.97	0.97	0.97

Table IX: Confusion matrix for experiment-3.

Predicted/ Actual	COVID-19	Non COVID-19
Corona	29	3
Non COVID-19	2	97

As the dataset for COVID-19 cases is not that much available, to be assured about the performance of our model we have performed both two class classification (COVID-19 and non-COVID-19) and three class classification (COVID-19, Pneumonia, Normal). Moreover, we performed patientwise

Figure 2: Accuracy vs epoch and loss vs epoch for train and validation set (Experiment 1).

10-fold cross validation to guarantee the robustness of our model. Finally, in qualitative analysis we analyzed the of decision making behaviour of our model to ensure interpretability and trustworthiness.

A. Quantitative Analysis

To show this particular analysis, we will analyze the test accuracy, precision, recall, and f-score of each experimental setup.

- **Experiment 1:** In this experiment we performed three class classification (COVID-19, Pneumonia, Normal). We split our dataset in train, validation, and test set in 80%-10%-10% ratio. There were no common image among three sets and augmentation was performed separately in each set. Results are shown in Table V and VI.
- **Experiment 2:** To check the robustness and performance consistency of our model, we have done patient wise 10-fold cross validation as the data-set contains multiple Radiology images of different days for same patient. Each fold has images of different patients and augmentation was performed separately in each of them. Results are shown in Table VII.
- **Experiment 3:** The same setup (train, validation, test split in 80%-10%-10% ratio) as the first experiment with only 2 class label Eg. COVID-19 and Non COVID-19 was used in this experiment. Results are shown in Table VIII and Table IX.

Figure 3: Actual Chest-Xray images along with heatmaps of a corona affected, a pneumonia affected, and a normal person (top to bottom respectively).

Overall 2 class classifier performed better as expected and the 10 fold result conformed with this as well.

B. Qualitative Analysis

To investigate how our model makes prediction we used Gradient-weighted Class Activation Mapping (Grad-CAM) [11], which produces a coarse localization map highlighting the important regions in the input image for making the prediction. In this approach, we computed the gradient score for the target class with respect to the feature maps of the final convolutional layer. These gradients are average-pooled to obtain the neuron importance weights and a weighted combination of the activation maps are followed by Rectified Linear Unit (ReLU). This results in a coarse heatmap of the input image. Figure 3 shows the actual Chest-Xray images

along with heatmaps of a corona affected, a pneumonia affected, and a normal person. We can see that our model is mainly emphasizing on the lung areas in detecting COVID-19 or Pneumonia.

This qualitative analysis is important for a number of factors:

- **Interpretability:** One of the major drawbacks of many deep learning models is lack of interpretability. With Grad-CAM, we tried to make our model interpretable and explainable. The generated heatmaps show us insights about how our model make predictions.
- **Trustworthiness:** From the heatmaps we can see the important regions of the images that leads to classification decision. Consequently, we can verify that our model is not making decision based on inappropriate regions of the Radiology image.
- **Possible critical factors:** Our approach can provide new insights and visual indicators about critical factors of COVID-19 disease.

V. CONCLUSION

In this work, we showed a novel transfer learning based approach to detect COVID-19. To assure that our model can differentiate COVID-19 radiology images from both healthy persons and pneumonia patients, we performed both two class and three class classification. To guarantee the robustness and consistency of our model we implemented patient-wise 10-fold cross validation. Moreover, we performed an explainability analysis to interpret and visualize how our model works. The open source data for COVID-19 radiology images is limited, if more data is available in future, our model can be tested against those data. How our model performs in detecting COVID-19 from other types of lung diseases can be a future research direction.

REFERENCES

- [1] figure1. <https://github.com/agchung/Figure1-COVID-chestxray-dataset>. Accessed: 2020-04-29.
- [2] rnsa. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>. Accessed: 2020-04-29.
- [3] Website. <https://plez.herokuapp.com>.
- [4] COHEN, J. P., BERTIN, P., AND FRAPPIER, V. Chester: A web delivered locally computed chest x-ray disease prediction system, 2019.
- [5] COHEN, J. P., MORRISON, P., AND DAO, L. Covid-19 image data collection. *arXiv 2003.11597* (2020).
- [6] HUANG G, LIU Z, V. D. M. L. W. K. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017).
- [7] KINGMA, D. P., AND BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [8] LIN LI, LIXIN QIN, Z. X. Y. Y. X. W. B. K. J. B. Y. L. Z. F. Q. S. K. C. D. L. G. W. Q. X. X. F. S. Z. J. X. J. X. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct.
- [9] LINDA WANG, A. W. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images. *Journal of Network & Computer Applications* (2020).
- [10] RAJPURKAR, P., IRVIN, J., ZHU, K., YANG, B., MEHTA, H., DUAN, T., DING, D., BAGUL, A., LANGLITZ, C., SHPANSKAYA, K., ET AL. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225* (2017).
- [11] SELVARAJU, R. R., COGSWELL, M., DAS, A., VEDANTAM, R., PARIKH, D., AND BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 618–626.
- [12] WANG. Detection of sars-cov-2 in different types of clinical specimens. *JAMA* (2020).
- [13] WANG, X., PENG, Y., LU, L., LU, Z., BAGHERI, M., AND SUMMERS, R. Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *IEEE CVPR* (2017).