Detecting Of SARS-CoV-2 From Chest X-Ray Using Artificial Intelligence

Madhuri Thote[1], Prof. Sonali Sethi[2]

[1] *Student, Computer Engineering, NBN Sinhgad School of Engineering Pune, Maharashtra, India*
[2] *Professor, Computer Engineering, NBN Sinhgad School of Engineering Pune, Maharashtra, India*

ABSTRACT

Chest radiographs (X-rays) combined with Deep Convolutional Neural Network (CNN) methods have been demonstrated to detect and diagnose the onset of COVID-19, the disease caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). However, questions remain regarding the accuracy of those methods as they are often challenged by limited datasets, performance legitimacy on imbalanced data, and have their results typically reported without proper confidence intervals. Considering the opportunity to address these issues, in this study, we propose and test six modified deep learning models, including VGG16, InceptionResNetV2, ResNet50, MobileNetV2, ResNet101, and VGG19 to detect SARSCoV-2 infection from chest X-ray images. Results are evaluated in terms of accuracy, precision, recall, and f- score using a small and balanced dataset (Study One), and a larger and imbalanced dataset (Study Two). With 95% confidence interval, VGG16 and MobileNetV2 show that, on both datasets, the model could identify patients with COVID-19 symptoms with an accuracy of up to 100%. We also present a pilot test of VGG16 models on a multi-class dataset, showing promising results by achieving 91% accuracy in detecting COVID-19, normal, and Pneumonia patients. Furthermore, we demonstrated that poorly performing models in Study One (ResNet50 and ResNet101) had their accuracy rise from 70% to 93% once trained with the comparatively larger dataset of Study Two. Still, models like InceptionResNetV2 and VGG19's demonstrated an accuracy of 97% on both datasets, which posits the effectiveness of our proposed methods, ultimately presenting a reasonable and accessible alternative to identify patients with COVID-19.

Keyword : - *Artificial intelligence, COVID-19, coronavirus, SARS-CoV-2, deep learning, chest X-ray, imbalanced data, small data.*

1. INTRODUCTION

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), previously known as the Novel Coronavirus, was first reported in Wuhan, China and rapidly spread around the world, pushing the World Health Organization (WHO) to declare the outbreak of the virus as a global pandemic and health emergency on March 11, 2020. According to official data, 19 million people have been infected worldwide, with the number of deaths surpassing 700, 000, and 12 million recovery cases reported by August 6, 2020 [1]. In the United States, the first case was reported on January 20, 2020, which evolved into a current number of confirmed cases, deaths, and recovered patients reaching more than 5 million, 162, 000, and 2.5 million, respectively (August 6, 2020 data) [1]. COVID-19 can be transmitted in several ways. The virus can spread quickly among humans via community transmission, such as close contact between individuals, and the transfer of respiratory droplets produced via coughing, sneezing, and talking. Several symptoms have been reported so far, including fever, tiredness, and dry cough as the most common. Additionally, aches, pain, nasal congestion, runny nose, sore throat, and diarrhea have also been associated with the disease [2], [3]. Several methods can be followed to detect SARS-CoV-2 infection [4], including: • Real-time reverse transcription polymerase chain reaction (RT-PCR)-based methods • Isothermal nucleic acid amplification-based methods • Microarray-based methods. Health authorities in most countries have chosen to adopt the RT-PCR method, as it is regarded as the gold-standard in diagnosing viral and bacterial infections at the molecular level [5]. However, due to the rapidly increasing number of new cases and limited healthcare infrastructure, rapid detection or mass testing is required to lower the curve of infection. Recent studies claimed that chest Computed Tomography (CT) has the capability to detect the disease promptly. Therefore, in China, to deal with many new cases, CT scans were used for the initial screening of patients with COVID-19 symptoms [6]–[9]. Similarly, chest radiograph (X-ray) image-based diagnosis may be a more attractive and readily available method for detecting the onset of the disease due to its low cost and fast image acquisition procedure. In our study, we investigate recent literature on the topic and tackle the opportunity to present an effective deep learning-based screening method to detect patients with COVID-19 from chest X-ray images. Developing deep learning models using small image datasets often results in the incorrect identification of regions of interest in those images, an issue not often addressed in the existing literature. Therefore, in the present work, we have analyzed our models' performance layer by layer and chose to select only the best-performing ones, based on the correct identification of the infectious regions present on the X-ray images. Also, previous works often do not demonstrate how their proposed models perform with imbalanced datasets which is often challenging. Here, we diversify the analysis and consider small, imbalanced, and large datasets while presenting a comprehensive description of our results with statistical measures, including 95% confidence intervals, p-values, and t-values. A summary of our technical contributions is presented below:

• Modification and evaluation of six different deep CNN models (VGG16, InceptionResNetV2, ResNet50, MobilenetV2, ResNet101, VGG19) for detection of COVID-19 patients using X-ray image data on both balanced and imbalanced datasets; and

• Verify the possibility to locate affected regions on chest X-rays incorporated with heatmaps, including a cross-check with a medical doctor's opinion.

2. LITERATURE REVIEW

In the recent past, the adoption of Artificial Intelligence (AI) in the field of infectious disease diagnosis has gained a notable prominence, which led to the investigation of its potential in the fight against the novel coronavirus [10]–[12]. Current AI-related research efforts on COVID-19 detection using chest CT and X-ray images are discussed below to provide a brief insight on the topic and highlight our motivations to research it further.

A. CT SCAN BASED SCREENING

To date, several efforts in detecting COVID-19 from CT images have been reported. A recent study by Chua et al. (2020) suggested that the pathological pathway observed from the pneumonic injury leading to respiratory death can be detected early via chest CT, especially when the patient is scanned two or more days after the development of symptoms [13]. Related studies proposed that deep learning techniques could be beneficial for identifying COVID-19 disease from chest CT [12]. For instance, Shi et al. (2020) introduced a machine learning-based method for the COVID-19 screening from an online COVID-19 CT dataset. Similarly, Gozes et al. (2020) developed an automated system using artificial intelligence to monitor and detect patients from chest CT [16]. Chua et al. (2020) focused on the role of Chest CT in the detection and management of COVID-19 disease from a high incidence region (United Kingdom) [13]. Ai et al. (2020) also supported CT-based diagnosis as an efficient approach compared to RT-PCR testing for COVID-19 patients detection with a 97% sensitivity [17], [18]. Due to data scarcity, most preliminary studies considered minimal datasets. For example, Chen et al. (2020) used a UNet++ deep learning model and identified 51 COVID-19 patients with a 98.5% accuracy. However, the authors did not mention the number of healthy patients used in the study. Ardakani et al. (2020) used 194 CT images (108 COVID-19 and 86 other patients) and implemented ten deep learning methods to observe COVID-19 related infections and acquired 99.02% accuracy. Moreover, a study conducted by Wang et al. (2020) considered 453 CT images of confirmed COVID-19 cases, from which 217 images were used as the training set, and obtained 73.1% accuracy, using the inception-based model. The authors, however, did not explain the model network and did not show the mark region of interest of the infections. Similarly, Zheng et al. (2020) introduced a deep learning-based model with 90% accuracy to screen patients using 499 3D CT images. Despite promising results, a very high performance on small datasets often raises questions about the model's practical accuracy and reliability. Therefore, a better way to represent model accuracy is to present it with an associated confidence interval. However, none of the work herein referenced expressed their results with confidence intervals, which should be addressed in future studies. As larger datasets become available, deep-learning-based studies taking advantage of their potential have been proposed to detect and diagnose COVID-19. Xu et al. (2020) investigated a dataset of 618 medical images to detect COVID-19 patients and acquired 86.7% accuracy using ResNet23. Li et al. (2020) utilized an even larger dataset (a combination of 1296 COVID-19 and 3060 Non-COVID-19 patients CT images) and achieved 96% accuracy using ResNet50. With larger datasets, it is no surprise that deep learning-based models predict patients with COVID-19 symptoms with accuracies ranging from 85% to 96%. However, obtaining a chest CT scan is a notably time consuming, costly, and complex procedure. Despite allowing for comparatively better image quality, its associated challenges inspired many researchers to propose X-ray-based COVID-19 screening methods as a reliable alternative way.

B. CHEST X-RAY BASED SCREENING

Preliminary studies have used transfer learning techniques to evaluate COVID-19 and pneumonia cases in the early stages of the COVID-19 pandemic. However, data insufficiency also hinders the ability of such proposed models to provide reliable COVID-19 screening tools based on chest X-ray [12]. For instance, Hemdan et al. (2020) proposed a CNN-based model adapted from VGG19 and achieved 90% accuracy using 50 images [32]. Ahsan et al. (2020) developed a COVID-19 diagnosis model using Multilayer Perceptron and Convolutional Neural Network (MLP-CNN) for mixed-data (numerical/categorical and image data). The model predicts and differentiates between 112 COVID-19 and 30 non-COVID-19 patients, with a higher accuracy of 95.4%. Sethy & Behera (2020) also considered only 50 images and used ResNet50 for COVID-19 patients classification, and ultimately reached 95% accuracy. Also, Narin et al. (2020) used 100 images and achieved 86% accuracy using InceptionResNetV 2 [12]. As noted, these studies use relatively small datasets, which does not guarantee whether their proposed models would perform equally well on larger datasets. Also, the possibility of a model overfitting is another concern for larger CNN-based networks when trained with a small datasets.

3. RESEARCH METHODOLOGY

We propose three separate studies, wherein three distinct datasets were used, as detailed below:

1) Study One – smaller, balanced dataset: chest X-ray images of 25 patients with COVID-19 symptoms, and 25 images of patients with diagnosed pneumonia, obtained from the open-source repository shared by Dr. Joseph Cohen [43].

2) Study Two – larger, imbalanced dataset: chest X-ray images of 262 patients with COVID-19 symptoms, and 1583 images of patients with diagnosed pneumonia, obtained from the Kaggle COVID-19 chest X-ray dataset [44].

Figure 1 presents a set of representative chest X-ray images of both COVID-19 and pneumonia patients from the aforementioned datasets. Table 1 details the overall assignment of data for training and testing of each investigated CNN model. In both studies, six different deep learning approaches were investigated: VGG16 , InceptionResNetV2, ResNet50, MobileNetV2, ResNet101 and VGG19. A. USING PRE-TRAINED CONVET A pre-trained network is a network that was previously trained on a larger dataset which, in most cases, is enough to learn a unique hierarchy to extract features from. It works more effectively on small datasets. A prime example is the VGG16 architecture, developed by Simoyan and Zisserman (2014). Figure 2 shows a sample architecture of the pre-trained model procedure. All models implemented in this study are available as a pre-package within Keras. Figure 3 demonstrates a fine-tuning sequence on the VGG16 network. The modified architecture follows the steps below:

1) Firstly, the models were initiated with a pre-trained network without a fully connected (FC) layer.
2) Then, an entirely new connected layer added a pooling layer and ‘‘softmax’’ as an activation function, appended it on top of the VGG16 model.
3) Finally, the convolution weight was frozen during the training phase so that only the FC layer should train during the experiment.

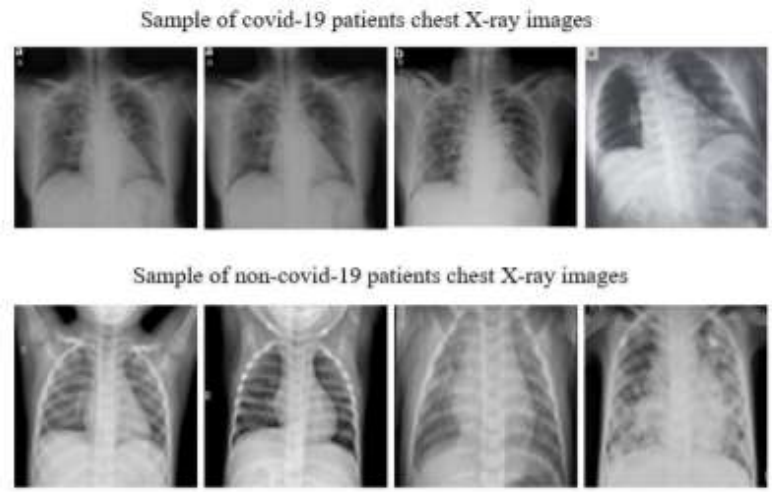

Figure 1: Representative samples of chest X-ray images from the open-source data repositories [43] used in our proposed studies.

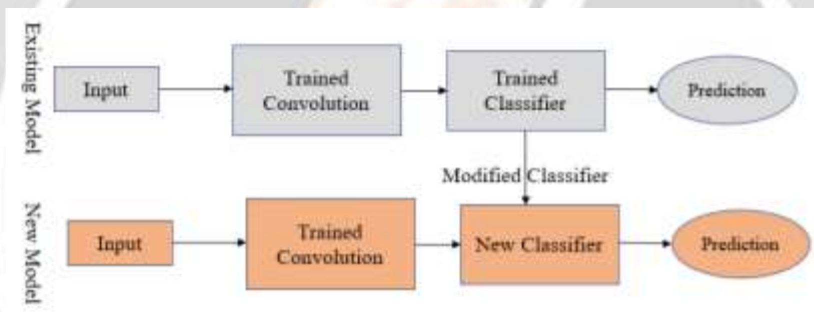

Figure 2: Modified architecture with new classifier [51]

We used the grid search method, which is commonly used for parameter tuning. Initially, we randomly selected the following: Batch size = [4, 5, 8, 10] Number of epochs = [10, 20, 30, 40] Learning rate = [.001, .01, 0.1] For Study One, using the grid search method, we achieved better results with the following: Batch size = 8 Number of epochs = 30 Learning rate = .001 Similarly, for Study Two, the best results were achieved with: Batch size = 50 Number of epochs = 50 Learning rate = .001 Finally, during Study Three, best performance was achieved with: Batch size = 50 Number of epochs = 100 Learning rate = .001 We used the adaptive learning rate optimization algorithm (Adam) as an optimization algorithm for all models due to its robust performance on binary image classification . As commonly adopted in data mining techniques, this study used 80% data for training, whereas the remaining 20% was used for testing. Each study was conducted twice, and the final result was represented as the average of those two experiment outcomes, as suggested by Zhang et al. (2020). Performance results were presented as model accuracy, precision, recall, and f-score. Accuracy = tp + tn tp + tn + fp + fn (1) Precision = tp tp + fp (2) Recall = tp tn + fp (3) F-score = 2 × Precision × Recall Precision+Recall (4) where, • True Positive (tp) = COVID-19 patient classified as patient • False Positive (fp) = Healthy people classified as patient • True Negative (tn) = Healthy people classified as healthy • False Negative (fn) = COVID-19 patient classified as healthy.

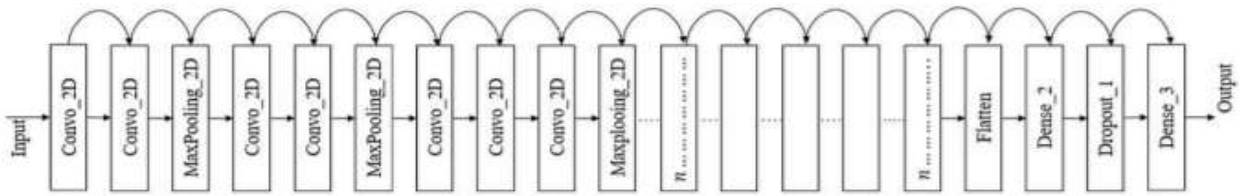

Figure 3: VGG16 architecture used during this experiment.

4. RESEARCH METHODOLOGY

A. STUDY ONE

The overall model performance for all CNN approaches was measured both on the training (40 images) and test (10 images) sets using equation 1, 2, 3, and 4. Table 2 presents the results of the training set. In this case, VGG16 and MobileNetV2 outperformed all other models in terms of accuracy, precision, recall, and f score. In contrast, the ResNet50 model showed the worst performance across all measures. Table 3 presents the performance results for all models on the test set. Models VGG16 and MobileNetV2 showed 100% performance across all measures. On the other hand, ResNet50, ResNet101, and VGG19 demonstrated significantly worse results.

1) CONFUSION MATRIX

Confusion matrices were used to better visualize the overall performance of prediction. The test set contains 10 samples (5 COVID-19 and 5 other patients). In accordance with the performance results previously presented, Figure 4 shows that the VGG16, InceptionResNetV2, and MobileNetV2 models correctly classified all patients. In contrast, models ResNet50, and ResNet101 incorrectly classified 3 non-COVID-19 patients as COVID-19 patients, and models VGG19 classified 2 non-COVID patients as COVID-19 patients while also classifying 1 COVID-19 patient as non-COVID-19.

2) MODEL ACCURACY

Figure 5 shows the overall training and validation accuracy during each epoch for all models. Models VGG16 and MobileNetV2 demonstrated higher accuracy at epochs 25 to 30, while VGG19, ResNet50, and ResNet101 displayed lower accuracy which sporadically fluctuated between epochs 10.

3) MODEL LOSS

Both training loss and validation loss were reduced following each epoch for VGG16, InceptionResNetV2, and MobileNetV2. In contrast, for VGG19, both measures are scattered over time, which is an indicative of poor performance. B. STUDY TWO For Study Two, on the training set, most model accuracies were measured above 90%. Table 4 shows that 100% accuracy, precision, recall, and f score were achieved using MobileNetV2. Among all other models, ResNet50 showed the worst performance across all measures.

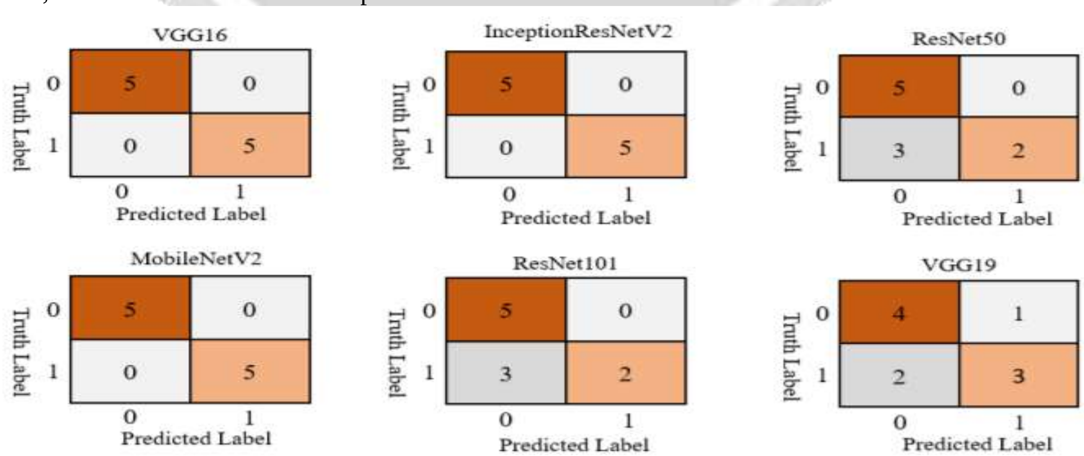

Figure 4: Study one confusion matrices for six different deep learning models applied on the test set.

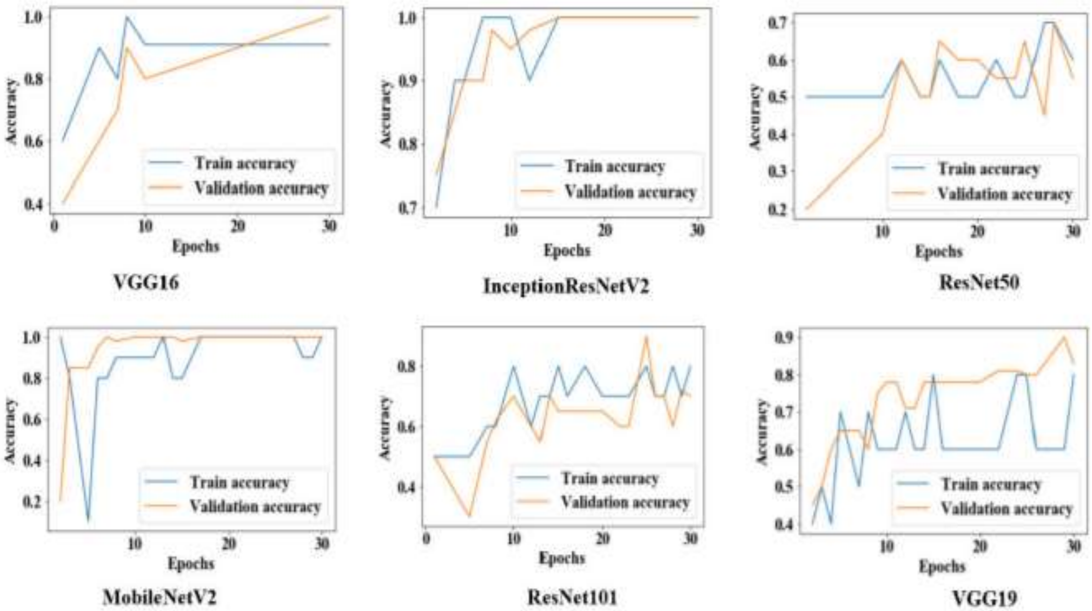

Figure 5: Training and validation accuracy throughout the execution of each model in study one

5. CONCLUSIONS

Our study proposed and assessed the performance of six different deep learning approaches (VGG16, InceptionResNetV2, ResNet50, MobileNetV2, ResNet101, and VGG19) to detect SARS-CoV-2 infection from chest X-ray images. Our findings suggest that modified VGG16 and MobileNetV2 models can distinguish patients with COVID-19 symptoms on both balanced and imbalanced dataset with an accuracy of nearly 99%. Our model outputs were crosschecked by healthcare professionals to ensure that the results could be validated. We hope to highlight the potential of artificial-intelligence-based approaches in the fight against the current pandemic using diagnosis methods that work reliably with data that can be easily obtained, such as chest radiographs. Some of the limitations associated with our work can be addressed by conducting experiments with extensively imbalanced big data, comparing the performance of our methods with those using CT scan data and/or other deep learning approaches, and developing models with explainable artificial intelligence on a mixed dataset.

6. REFERENCES

[1] Worldometer. COVID-19 World Meter. Accessed: Aug. 6, 2020. [Online]. Available: https://www.worldometers.info/coronavirus.

[2] H. A. Rothan and S. N. Byrareddy, ‘‘The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak,’’ J. Autoimmunity, vol. 109, May 2020, Art. no. 102433, doi: 10.1016/j.jaut.2020.102433.

[3] L. Pan, M. Mu, P. Yang, Y. Sun, R. Wang, J. Yan, P. Li, B. Hu, J. Wang, C. Hu, Y. Jin, X. Niu, R. Ping, Y. Du, T. Li, G. Xu, Q. Hu, and L. Tu, ‘‘Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: A descriptive, cross-sectional, multicenter study,’’ Amer. J. Gastroenterol., vol. 115, no. 5, pp. 766–773, May 2020, doi: 10.14309/ajg.0000000000000620.

[4] M. Shen, Y. Zhou, J. Ye, A. A. AL-Maskri, Y. Kang, S. Zeng, and S. Cai, ‘‘Recent advances and perspectives of nucleic acid detection for coronavirus,’’ J. Pharmaceutical Anal., vol. 10, no. 2, pp. 97–101, Apr. 2020, doi: 10.1016/j.jpha.2020.02.010.

[5] A. Tahamtan and A. Ardebili, ‘‘Real-time RT-PCR in COVID-19 detection: Issues affecting the results,’’ Expert Rev. Mol. Diag., vol. 20, no. 5, pp. 453–454, Apr. 2020, doi: 10.1080/14737159.2020.1757437.

[6] Z. Hu, C. Song, C. Xu, G. Jin, Y. Chen, X. Xu, H. Ma, W. Chen, Y. Lin, Y. Zheng, J. Wang, Z. Hu, Y. Yi, and H. Shen, ‘‘Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in nanjing, China,’’ Sci. China Life Sci., vol. 63, no. 5, pp. 706–711, Mar. 2020, doi: 10.1007/s11427-020-1661-4.

[7] H. Bai, B. Hsieh, Z. Xiong, K. Halsey, J. Choi, and T. Tran, ‘‘Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT,’’ Radiology, vol. 296, no. 2, pp. 46–54, Mar. 2020, doi: 10.1148/radiol.2020200823.

[8] Y. Li and L. Xia, ‘‘Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management,’’ Amer. J. Roentgenol., vol. 214, no. 6, pp. 1280–1286, Jun. 2020, doi: 10.2214/ajr.20.22954.

[9] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z. Fayad, N. Zhang, K. Diao, and S. Li, ‘‘Chest CT findings in coronavirus disease-19 (COVID-19): Relationship to duration of infection,’’ Radiology, vol. 295, no. 3, Feb. 2020, Art. no. 200463, doi: 10.1148/radiol.2020200463.

[10] K. C. Santosh, ‘‘AI-driven tools for coronavirus outbreak: Need of active learning and cross-population train/test models on multitudinal/multimodal data,’’ J. Med. Syst., vol. 44, no. 5, pp. 1–5, Mar. 2020, doi: 10.1007/s10916-020-01562-1.

[11] B. Ghoshal and A. Tucker, ‘‘Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection,’’ 2020, arXiv:2003.10769. [Online]. Available: http://arxiv.org/abs/2003.10769

[12] A. Narin, C. Kaya, and Z. Pamuk, ‘‘Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks,’’ 2020, arXiv:2003.10849. [Online]. Available: http://arxiv. org/abs/2003.10849

[13] F. Chua, D. Armstrong-James, and S. R. Desai, ‘‘The role of CT in case ascertainment and management of COVID-19 pneumonia in the UK: Insights from high-incidence regions,’’ Lancet Respiratory Med., vol. 8, no. 5, pp. 438–440, May 2020, doi: 10.1016/s2213-2600(20)30132-6.

[14] MD MANJURUL AHSAN 1 , MD TANVIR AHAD 2 , (Graduate Student Member, IEEE), FARZANA AKTER SOMA 3 , SHUVA PAUL 4 , (Member, IEEE), ANANNA CHOWDHURY 5 , SHAHANA AKTER LUNA 6 , MUNSHI MD. SHAFWAT YAZDAN 7 , AKHLAQUR RAHMAN 8 , (Member, IEEE), ZAHED SIDDIQUE 2 , AND PEDRO HUEBNER 1. Detecting SARS-CoV-2 From Chest X-Ray Using Artificial Intelligence