

Received February 14, 2021, accepted February 16, 2021, date of publication February 23, 2021, date of current version March 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3061621

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

MD MANJURUL AHSAN¹, MD TANVIR AHAD², (Graduate Student Member, IEEE), FARZANA AKTER SOMA³, SHUVA PAUL⁴, (Member, IEEE), ANANNA CHOWDHURY⁵, SHAHANA AKTER LUNA⁶, MUNSHI MD. SHAFWAT YAZDAN⁷, AKHLAQR RAHMAN⁸, (Member, IEEE), ZAHED SIDDIQUE², AND PEDRO HUEBNER¹

¹School of Industrial and Systems Engineering, The University of Oklahoma, Norman, OK 73019, USA

²School of Aerospace and Mechanical Engineering, The University of Oklahoma, Norman, OK 73019, USA

³Holy Family Red Crescent Medical College & Hospital, Dhaka 1000, Bangladesh

⁴School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 0250, USA

⁵Z. H. Sikder Women's Medical College & Hospital, Dhaka 1212, Bangladesh

⁶Dhaka Medical College & Hospital, Dhaka 1000, Bangladesh

⁷Civil and Environmental Engineering Department, Idaho State University, Pocatello, ID 83209, USA

⁸School of Industrial Automation and Electrical Engineering, Engineering Institute of Technology, Melbourne, VIC 3000, Australia

Corresponding author: Md Tanvir Ahad (ahadtanvir@gmail.com)

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 SARS-CoV-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.

INDEX TERMS Artificial intelligence, COVID-19, coronavirus, SARS-CoV-2, deep learning, chest X-ray, imbalanced data, small data.

I. 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

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott¹.

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.

II. 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], [14]. For instance, Shi *et al.* (2020) introduced a machine learning-based method for the COVID-19 screening from an online COVID-19 CT dataset [15]. 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 [19]–[21]. For example, Chen *et al.* (2020) used a UNet++ deep learning model and identified 51 COVID-19 patients with a 98.5% accuracy [19]. 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 [20]. 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 [22]. Similarly, Zheng *et al.* (2020) introduced a deep learning-based model with 90% accuracy to screen patients using 499 3D CT images [21]. 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 [23]. 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 [24]. 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 [25]. 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 [26], [27].

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 [28]–[31]. However, data insufficiency also hinders the ability of such proposed models to provide reliable COVID-19 screening tools based on chest X-ray [12], [32], [33]. For instance, Hemandan *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% [34]. Sethy & Behera (2020) also considered only 50 images and used ResNet50 for COVID-19 patients classification, and ultimately reached 95% accuracy [33]. 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.

In view of these issues, recent studies proposed model training with larger datasets and reported a better performance compared to smaller ones [35]–[38]. Chandra *et al.* (2020) developed an automatic COVID screening system to detect infected patients using 2088 (696 normal, 696 pneumonia, and 696 COVID-19) and 258 (86 images of each category) chest X-ray images, and achieved 98% accuracy [39]. Sekeroglu *et al.* (2020) developed a deep learning-based method to detect COVID-19 using publicly available X-ray images (1583 healthy, 4292 pneumonia, and 225 confirmed COVID-19), which involved the training of deep learning and machine learning classifiers [40]. Pandit *et al.* (2020) explored pre-trained VGG-16 using 1428 chest X-rays with a mix of confirmed COVID-19, common bacterial pneumonia, and healthy cases (no infection). Their results showed an accuracy of 96% and 92.5% in two and three output class cases [41]. Ghosal & Tucker (2020) used 5941 chest X-ray images and obtained 92.9% accuracy [11]. Brunese *et al.* (2020) proposed a modified VGG16 model and achieved 99% accuracy with a dataset of 6505 images. However, they have used fairly balanced data with a 1 : 1.17 ratio; 3003 COVID-19 and 3520 other patients.

TABLE 1. Assignment of Data Used for Training and Testing of Deep Learning Models.

Study	Label	Train	Test
One	COVID-19	20	5
	Normal	20	5
Two	COVID-19	210	52
	Normal	1266	317
Three	COVID-19	176	43
	Normal	1073	268
	Pneumonia	1076	269

It is not immediately clear how their model would perform on an imbalanced dataset [42]. On the other hand, Khan *et al.* (2020) developed a model based on Xception CNN techniques considering 284 COVID-19 patients and 967 other patients (data ratio 1 : 3.4). Partially as an effect of a more imbalanced dataset, their reported accuracy was comparatively low, reaching 89.6% [38]. On imbalanced datasets, there is a higher chance that the model may be biased on significant classes and might affect the overall performance of the model.

III. 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].
- 3) Study Three – multiclass dataset: chest X-ray images of 219 patients with COVID-19 symptoms, 1345 images of patients with diagnosed pneumonia and 1073 images of normal patients, also obtained from the Kaggle COVID-19 chest X-ray dataset [45].

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 [46], InceptionResNetV2 [47], ResNet50 [48], MobileNetV2 [49], ResNet101 [50] and VGG19 [46].

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) [51]. Figure 2 shows a sample architecture of the

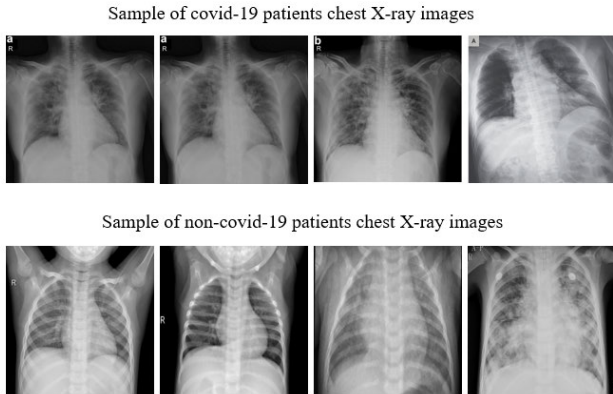


FIGURE 1. Representative samples of chest X-ray images from the open source data repositories [43] used in our proposed studies.

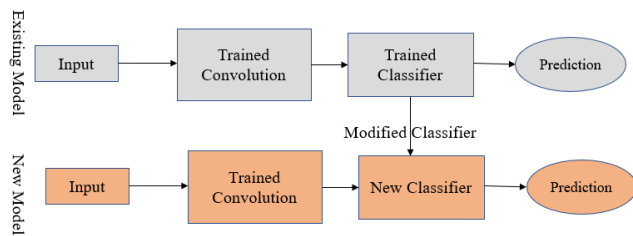


FIGURE 2. Modified architecture with new classifier [51].

pre-trained model procedure. All models implemented in this study are available as a pre-package within Keras [51].

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.

The same procedure was followed for all other deep learning techniques. In this experiment, the additional modification of the model for all CNN architectures was constructed as follows: *AveragePooling2D(Poolsize = (4, 4))* → *Flatten* → *Dense* → *Dropout(0.5)* → *Dense(Activation = “softmax”)*. As it is known, most pre-trained models contain multiple layers which are associated with different parameters (i.e., number of filters, kernel size, number of hidden layers, number of neurons) [52]. However, manually tuning those parameters is considerably time consuming [53], [54]. With that in mind, in our models, we have optimized three parameters: batch size,¹ epochs,² and learning rate³

¹Batch size characterizes the number of samples to work through before updating the internal model parameters [55]

²It defines how many times the learning algorithm will work through the entire dataset [55]

³It is a hyper-parameter that controls the amount to change the model in order to calculate the error each time the model weights are updated [56]

(inspired by [57], [58]). We used the grid search method [59], which is commonly used for parameter tuning. Initially, we randomly selected the following:

$$\text{Batch size} = [4, 5, 8, 10]$$

$$\text{Number of epochs} = [10, 20, 30, 40]$$

$$\text{Learning rate} = [.001, .01, 0.1]$$

For Study One, using the grid search method, we achieved better results with the following:

$$\text{Batch size} = 8$$

$$\text{Number of epochs} = 30$$

$$\text{Learning rate} = .001$$

Similarly, for Study Two, the best results were achieved with:

$$\text{Batch size} = 50$$

$$\text{Number of epochs} = 50$$

$$\text{Learning rate} = .001$$

Finally, during Study Three, best performance was achieved with:

$$\text{Batch size} = 50$$

$$\text{Number of epochs} = 100$$

$$\text{Learning rate} = .001$$

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 [60], [61]. As commonly adopted in data mining techniques, this study used 80% data for training, whereas the remaining 20% was used for testing [62]–[64]. 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) [65]. Performance results were presented as model accuracy, precision, recall, and f-score [66].

$$\text{Accuracy} = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (1)$$

$$\text{Precision} = \frac{t_p}{t_p + f_p} \quad (2)$$

$$\text{Recall} = \frac{t_p}{t_n + f_p} \quad (3)$$

$$\text{F-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

where,

- True Positive (t_p) = COVID-19 patient classified as patient
- False Positive (f_p) = Healthy people classified as patient
- True Negative (t_n) = Healthy people classified as healthy
- False Negative (f_n) = COVID-19 patient classified as healthy.

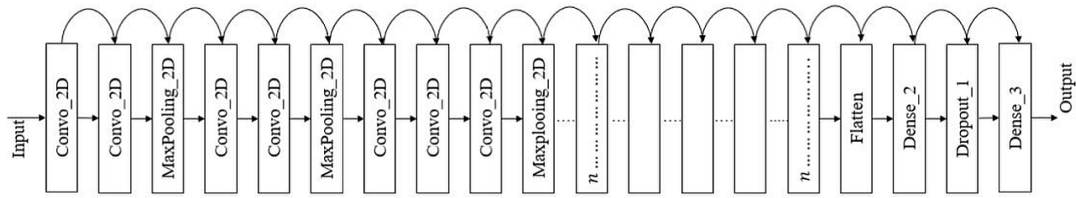


FIGURE 3. VGG16 architecture used during this experiment.

TABLE 2. Study One Model Performance on Train Set.

Model	Accuracy	Precision	Recall	F-score
VGG16	100%	100%	100%	100%
InceptionResNetV2	97%	98%	97%	97%
ResNet50	70%	81%	70%	67%
MobileNetV2	100%	100%	100%	100%
ResNet101	80%	83%	80%	80%
VGG19	93%	93%	93%	92%

TABLE 3. Study One Model Performance on Test Set.

Model	Accuracy	Precision	Recall	F-score
VGG16	100%	100%	100%	100%
InceptionResNetV2	97%	98%	97%	97%
ResNet50	70%	81%	70%	67%
MobileNetV2	100%	100%	100%	100%
ResNet101	70%	81%	70%	67%
VGG19	70%	71%	70%	70%

IV. RESULTS

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.

TABLE 4. Study Two Model Performance on Train Set.

Model	Accuracy	Precision	Recall	F-score
VGG16	99%	99%	99%	99%
InceptionResNetV2	99%	100%	99%	99%
ResNet50	93%	96%	93%	93%
MobileNetV2	100%	100%	100%	100%
ResNet101	96%	97%	88%	92%
VGG19	99%	98%	96%	97%

TABLE 5. Study Two Model Performance on Test Set.

Model	Accuracy	Precision	Recall	F-score
VGG16	99%	100%	97%	98%
InceptionResNetV2	99%	99%	98%	98%
ResNet50	93%	94%	93%	92%
MobileNetV2	99%	99%	99%	99%
ResNet101	96%	97%	88%	92%
VGG19	97%	97%	91%	94%

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

Figure 6 shows that 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.

Table 5 presents the performance results for all models on the test set. Models VGG16, InceptionResNetV2, and MobileNetV2 showed 99% accuracy; however, the precision, recall, and f score were distinct for each model, yet all above 97%. On the lower end, ResNet50 demonstrated relatively lower performance across all measures.

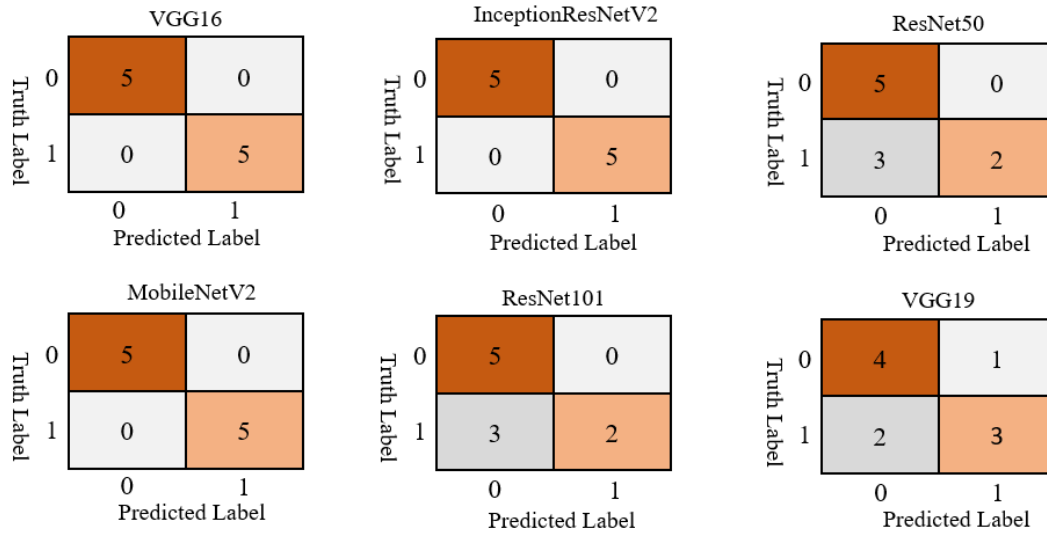


FIGURE 4. Study one confusion matrices for six different deep learning models applied on the test set.

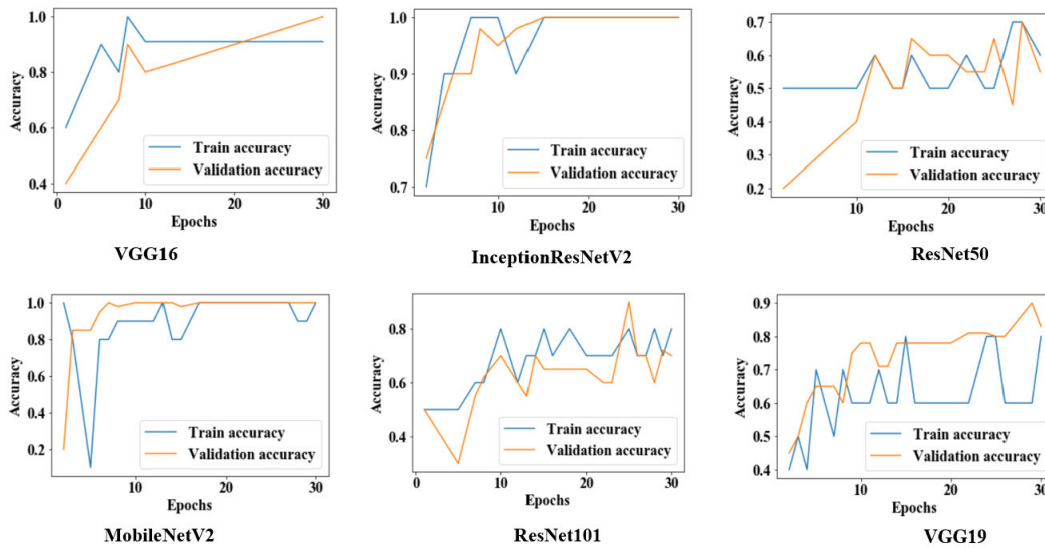


FIGURE 5. Training and validation accuracy throughout the execution of each model in study one.

1) CONFUSION MATRIX

Figure 7 shows that most of the models performance is satisfactory on the test set. In Study Two, classification accuracy for ResNet50 and ResNet101 is significantly better compared to Study One, possibly as an effect of the models being trained with more data and more epochs. In general, MobileNetV 2 performed better among all the models and misclassified only 2 images out of 369 images, while ResNet50 showed lower performance and misinterpreted 25 images out of 369 images.

2) MODEL ACCURACY

Figure 8 suggests that the overall training and validation accuracy were more steady during Study Two than Study One. The performance of ResNet50 and ResNet101 significantly

improved once trained with more data (1845 images) and more epochs (50 epochs).

3) MODEL LOSS

Figure 9 provides evidence that both training and validation losses were minimized following each epoch for all models, potentially as an effect of the increased batch size, number of epochs, and data amount.

C. STUDY THREE

As means of highlighting the potential of our proposed models with more complex classifications, we executed a small-scale pilot study to assess the performance of the VGG16 model on a multi-class dataset. The performance outcomes for the train and test runs are presented in Table 6. The accuracy remained above 90% on both runs, which suggests

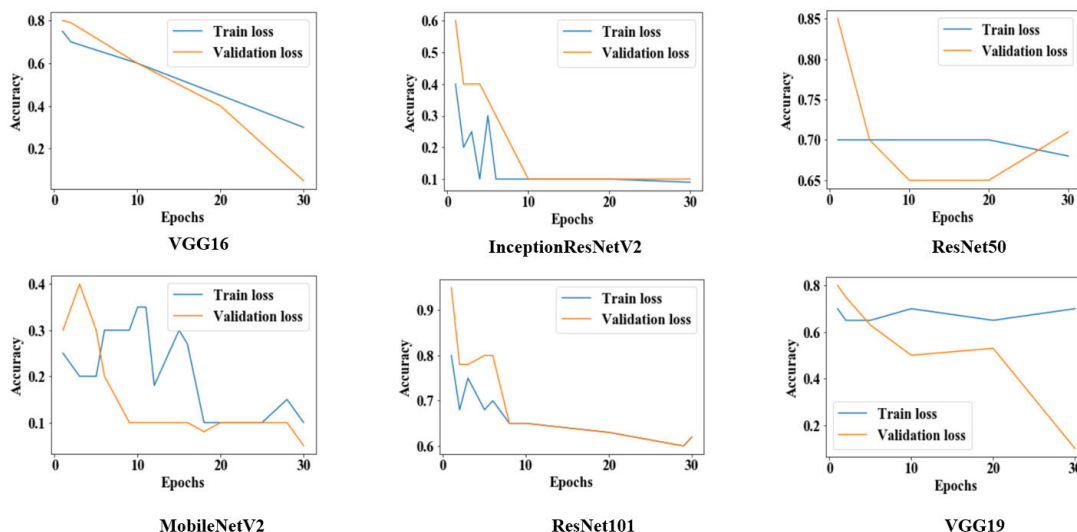


FIGURE 6. Training and validation loss throughout the execution of each model in study one.

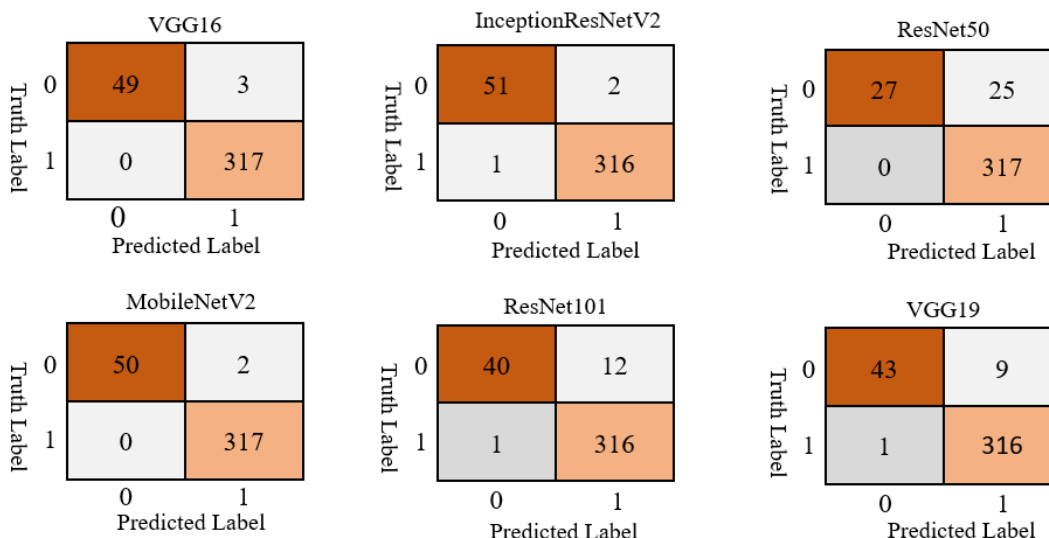


FIGURE 7. Study two confusion matrices for six different deep learning models applied on the test set.

TABLE 6. VGG16 Model Performance on Train and Test Datasets of Study Three.

VGG16	Accuracy	Precision	Recall	F-score
Train	93%	93%	88%	91%
Test	91%	90%	91%	86%

a notably high performance of our model with either binary or multi-class datasets.

D. TEST RESULTS WITH CONFIDENCE INTERVALS

Table 7 presents 95% confidence intervals for model accuracy on the test sets for Studies One and Two. For instance, in Study One, the average accuracies for VGG16 and MobileNetV2 were found to be 100%; however, the

Wilson score and Bayesian interval show that the estimated accuracies lie between 72.2% to 100% and 78.3% to 100%, respectively. On the other hand, Study Two reported relatively narrower interval ranges.

A paired t-test was conducted to compare model accuracies on both studies as shown in Table 8. There was no significant difference identified within the scores for Study One ($M = 84.50$, $SD = 15.922$) and Study Two ($M = 97.39$, $SD = 2.38$); $t(5) = -2.251$, $p = .074$. These results suggest that model accuracy is competent on both datasets and makes no statistically significant differences ($p > 0.05$).

V. DISCUSSION

As a means of comparing our results with those available in the literature, Table 9 contrasts the accuracies of our three best

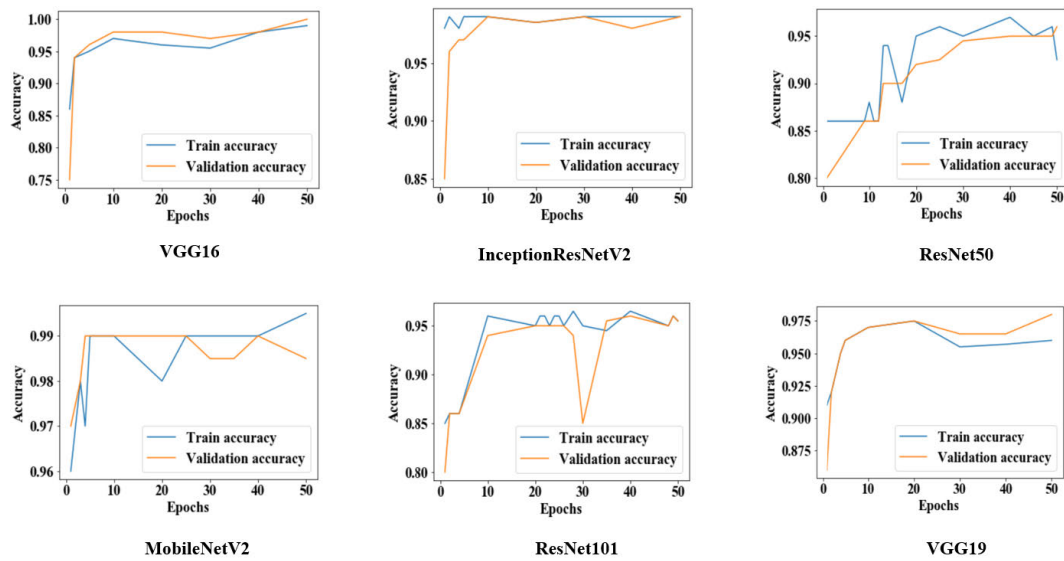


FIGURE 8. Training and validation accuracy throughout the execution of each model in study two.

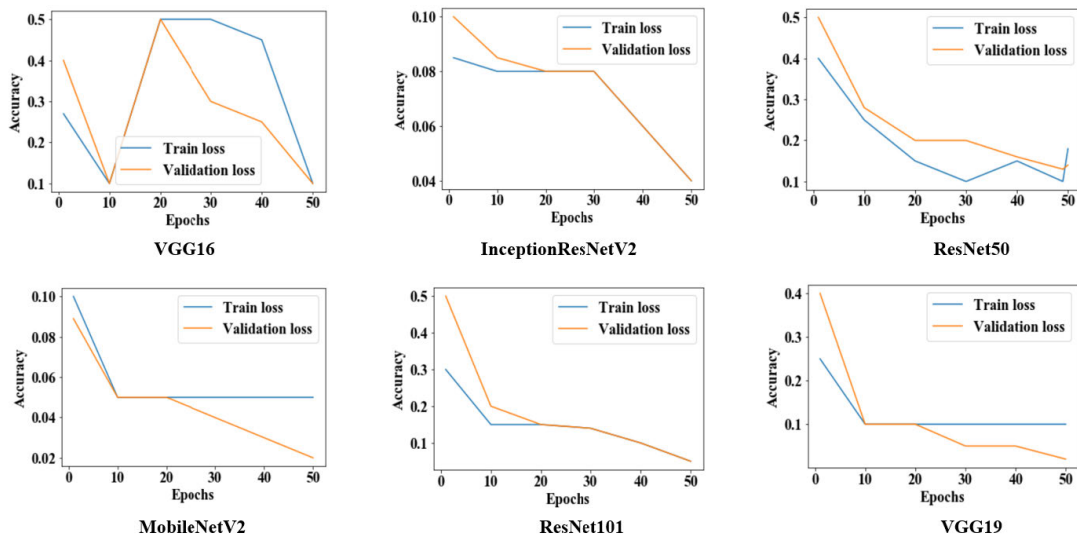


FIGURE 9. Training and validation loss throughout the execution of each model in study two.

performing CNN models on small datasets as part of Study One. It is relevant to emphasize that none of the referenced studies presents their results as confidence intervals, which hinders a direct comparison, but still allows for a higher-level assessment of the reported performance measures.

Using 50 chest X-ray images, we have achieved accuracy ranges from 68.1% to 99.8% using InceptionResNetV2, while Narin *et al.* (2020) used 100 images and obtained 86% accuracy [12]. Hemdan *et al.* (2020) and Sethy & Behera (2020) used small datasets of 50 images and acquired 90% and 9% accuracy using VGG19 and ResNet50 + SVM, respectively [32], [33].

Additionally, In Study Two, some of our models—VGG16, InceptionResNetV2, MobileNetV2, VGG19— demonstrated

almost similar accuracy while considering a highly imbalanced dataset than referenced literature [37], [38] that also used imbalanced datasets (Table 10). For the imbalanced dataset, we used 262 COVID-19 and 1583 non-COVID-19 patients’ (1 : 6.04) chest X-ray images. Apostolopoulos and Mpesiana (2020) used 1428 chest X-ray images where the data ratio was 1 : 5.4 (224 COVID-19: 1208 others) and achieved 98% accuracy [36]. Similarly, Khan *et al.* (2020) used 1251 chest X-ray images, data proportion 1 : 3.4 (284 COVID-19:967 others), and acquired 89.6% accuracy [38]. In Study Two, some of the best models we acquired were VGG16, VGG19, InceptionResNetV2, and MobileNetV2 and accuracy lies between 97% to around 100%.

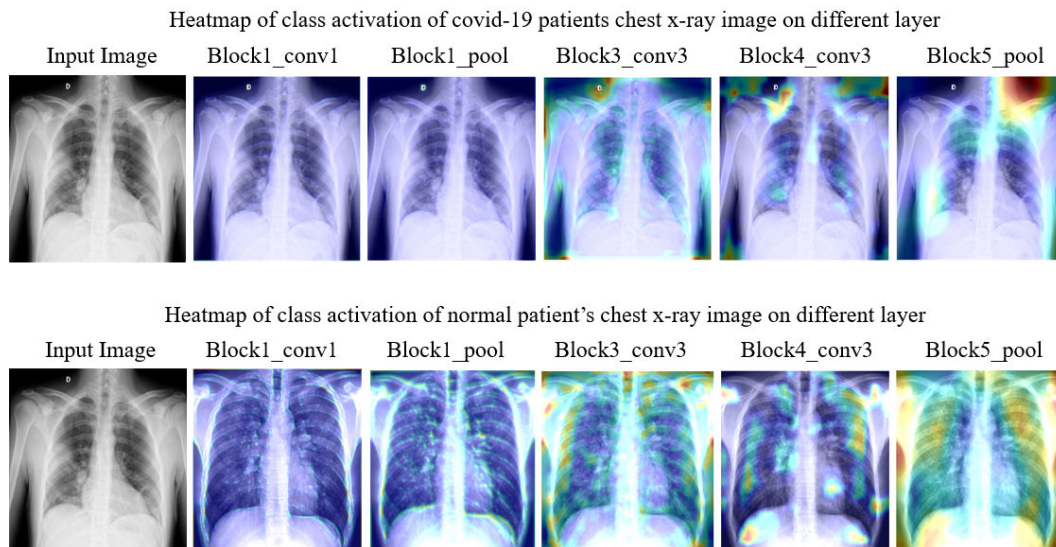


FIGURE 10. Heatmap of class activation on different layers.

TABLE 7. Confidence Interval ($\alpha = 0.05$) for Studies One and Two on Test Accuracy.

Study	Model	Actual accuracy	Methods	
			Wilson Score	Bayesian Interval
One	VGG16	100%	72.2% – 100%	78.3% – 100%
	InceptionResNetV2	97%	68.1% – 99.8%	72.5% – 99.9%
	ResNet50	70%	39.7% – 89.2%	39.4% – 90.7%
	MobileNetV2	100%	72.2% – 100%	78.3% – 100%
	ResNet101	70%	39.7% – 89.2%	39.4% – 90.7%
	VGG19	70%	39.7% – 89.2%	39.4% – 90.7%
Two	VGG16	99%	97.6% – 99.7%	97.8% – 99.8%
	InceptionResNetV2	99%	98.0% – 99.9%	98.3% – 99.9%
	ResNet50	93%	90% – 95.4%	90.3% – 95.5%
	MobileNetV2	99%	97.6% – 99.7%	97.8% – 99.8%
	ResNet101	96%	94.1% – 97.9%	94.2% – 98.0%
	VGG19	97%	95.1% – 98.5%	95.2% – 98.6%

A. FEATURE SELECTION

Figure 10 highlights extracted features as an effect of different CNN layers of VGG16 models applied on chest X-ray images from Study One. For instance, in block1_conv1 and block1_pool1, the extracted features were slightly fuzzy, while in block4_conv3 and block5_pool, those features become more visible/prominent. The heatmap also demonstrates a considerable difference in both COVID-19 and other patient images corresponding to each layer. For instance, as shown in Figure 11 (left), two specific regions were highlighted by heatmap for the COVID-19 patient’s X-ray image, whereas for other patients’ images, the areas were found to be haphazard and small.

During the experiment, each layer plays a significant role in identifying essential features from the images in the training phase. As a result, it is also deemed possible to see which features are learned and play a crucial role in differentiating between the two classes. In Figure 11, the left frame represents a chest X-ray image of a COVID-19 patient, and the right one highlights infectious regions of that same image, as spotted by the VGG16 model during Study One. The highlighted

TABLE 8. Descriptive Statistics of Paired t-Test for Study One and Study Two. M – Mean; SD – Standard Deviation; SEM – Standard Error Mean; DF – Degree of Freedom.

Study	M	SD	SEM
One	84.50	15.922	6.50
Two	97.3967	2.38362	.9731
Paired Difference	-12.89	14.03	5.72
Results			
t		-2.251	
DF		5	
$\alpha = .05$.074	

TABLE 9. Different Deep CNN Models Performance on Small Chest X-Ray Image Dataset.

Reference	Model	Datasize	Accuracy
Hemdan et al. [32]	VGG19	50	90%
Sethy & Behera [33]	ResNet50+SVM	50	95%
Narin et al. [12]	InceptionResNetV2	100	86%
Best model from	VGG16		72.2% – 100%
Study One with	InceptionResNetV2	50	68.1% – 99.8%
95% CI	MobileNetV2		72.2% – 100%

region on the upper right shoulder, which resulted from the individual layer of the VGG16 model (Study One), can be considered an irrelevant and therefore unnecessary feature identified by the network. The following topics extend the discussion on this issue:

- 1) The models attained unnecessary details from the images since the dataset is small compared to the model architecture (contains multiple CNN layers).
- 2) The models extracted features beyond the center of the images, which might not be essential to differentiate the COVID-19 patients from the non-COVID-19 patients.
- 3) The average age of COVID-19 patients in the first case study is 55.76 years. Therefore it is possible that individual patients might have age-related illnesses (i.e. weak/damaged lungs, shoulder disorder), apart from

TABLE 10. Comparison of Models Performance on Imbalanced Datasets.

Reference	Model	Datasize	Data ratio	Accuracy
Brunese et al. [42]	VGG16	6505	3003 : 3520 \approx 1:1.17	99%
Apostolopoulos & Mpesiana [36]	CNN	1428	224 : 1208 \approx 1:5.4	98%
Ozturk et al. [37]	DarkNet	1750	250 : 1500 = 1 : 6	98%
Khan et al. [38]	Xception	1251	284 : 967 \approx 1:3.4	89.6%
Lee et al. [67]	VGG-16	1821	607 : 1214 = 1 : 2	95.9%
Shelke et al. [68]	DenseNet-VGG16	2271	500 : 1771 \approx 1:3.54	95%
Best model from Study Two with 95% CI	VGG16 InceptionResNetV2 MobileNetV2 VGG19	1845	262 : 1583 \approx 1:6.04	97.6% – 99.7% 97.6% – 99.7% 97.6% – 99.7% 95.2% – 98.6%

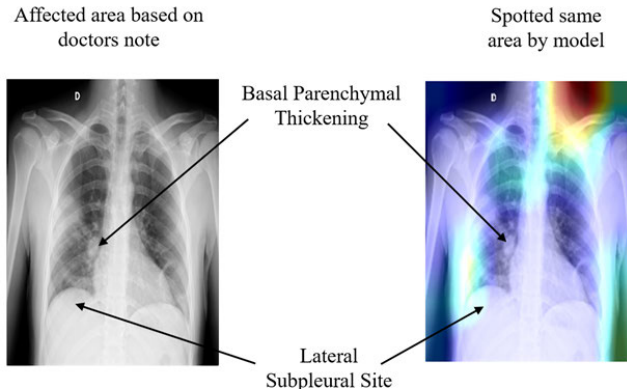


FIGURE 11. Model’s ability to identify important features on chest X-ray using VGG16.

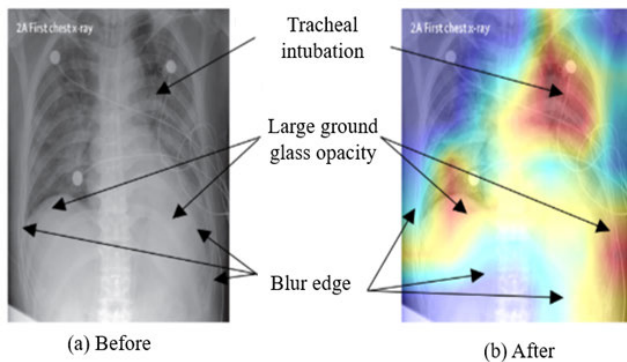


FIGURE 12. Model’s competency to identify essential features on chest X-ray using MobileNetV2.

complications related to COVID-19, which are not necessarily considered by the doctor’s notes.

Interestingly, these irrelevant regions spotted by our models decreased significantly when trained with larger datasets (1845 images) and increased epochs (50 epochs). For instance, Figure 12, presents the heatmap of the Conv-1 layer of MobileNetV2, acquired during the Study Two. The heatmap verifies that the spotted regions are very similar and match closely with the doctor’s findings.

VI. LIMITATIONS OF THE STUDY

We present the following items as limitations of our study, which shall be addressed in future works that consider our choice of tools and methods:

- At the time of writing, the limited availability of data represented a challenge to confidently assess the performance of our models. Open databases of COVID-19 patient records, especially those containing chest X-ray images, are rapidly expanding and should be considered in ongoing and future studies.
- We did not consider categorical patient data such as age, gender, body temperature, and other associated health conditions that are often available in medical datasets. More robust classification models that use those variables as inputs should be investigated as a means of achieving higher performance levels.
- We were limited to assessing the classification performance of our models against the gold standards of COVID-19 testing. However, those gold standards themselves are imperfect and often present false positives/negatives. It is imperative to ensure that the training sets of AI models like those herein presented are classified to the highest standards.
- Lastly, our study did not explore the compatibility of our proposed models with existing computer-aided diagnosis (CAD) systems. From a translational perspective, future works should explore the opportunity to bridge that gap with higher priority.

VII. CONCLUSION AND FUTURE WORKS

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 cross-checked 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.

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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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 multititudinal/multimodal data," *J. Med. Syst.*, vol. 44, no. 5, pp. 1–5, Mar. 2020, doi: [10.1007/s10916-020-01562-1](https://doi.org/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](https://doi.org/10.1016/s2213-2600(20)30132-6).
- [14] M. Barstugan, U. Ozkaya, and S. Ozturk, "Coronavirus (COVID-19) classification using CT images by machine learning methods," 2020, *arXiv:2003.09424*. [Online]. Available: <http://arxiv.org/abs/2003.09424>
- [15] F. Shi, L. Xia, F. Shan, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. Gao, H. Sui, and D. Shen, "Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification," 2020, *arXiv:2003.09860*. [Online]. Available: <http://arxiv.org/abs/2003.09860>
- [16] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. Siegel, "Rapid ai development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," 2020, *arXiv:2003.05037*. [Online]. Available: <http://arxiv.org/abs/2003.05037>
- [17] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases," *Radiology*, vol. 296, no. 2, pp. E32–E40, Aug. 2020, doi: [10.1148/radiol.2020200642](https://doi.org/10.1148/radiol.2020200642).
- [18] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest CT for COVID-19: Comparison to RT-PCR," *Radiology*, vol. 296, no. 2, pp. E115–E117, Aug. 2020, doi: [10.1148/radiol.2020200432](https://doi.org/10.1148/radiol.2020200432).
- [19] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, Q. Chen, S. Huang, M. Yang, X. Yang, and S. Hu, "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography," *Sci. Rep.*, vol. 10, no. 1, Nov. 2020, Art. no. 19196, doi: [10.1038/s41598-020-76282-0](https://doi.org/10.1038/s41598-020-76282-0).
- [20] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103795, doi: [10.1016/j.combiomed.2020.103795](https://doi.org/10.1016/j.combiomed.2020.103795).
- [21] X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and C. Zheng, "A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2615–2625, Aug. 2020, doi: [10.1109/tmi.2020.2995965](https://doi.org/10.1109/tmi.2020.2995965).
- [22] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, and B. Xu, "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)," *MedRxiv*, Apr. 2020, doi: [10.1101/2020.02.14.20023028](https://doi.org/10.1101/2020.02.14.20023028).
- [23] J. Brownle. (Sep. 2014). *Machinelearningmastery*. [Online]. Available: <http://machinelearningmastery.com/discover-feature-engineering-howtoengineer-features-and-how-to-getgood-at-it>
- [24] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, and Q. Ni, "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, Oct. 2020, doi: [10.1016/j.eng.2020.04.010](https://doi.org/10.1016/j.eng.2020.04.010).
- [25] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia, and J. Xia, "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy," *Radiology*, vol. 296, no. 2, pp. E65–E71, Aug. 2020, doi: [10.1148/radiol.2020200905](https://doi.org/10.1148/radiol.2020200905).
- [26] I. D. Apostolopoulos, S. I. Aznaouridis, and M. A. Tzani, "Extracting possibly representative COVID-19 biomarkers from X-ray images with deep learning approach and image data related to pulmonary diseases," *J. Med. Biol. Eng.*, vol. 40, no. 3, pp. 462–469, May 2020, doi: [10.1007/s40846-020-00529-4](https://doi.org/10.1007/s40846-020-00529-4).
- [27] C. Rachna. *Difference Between X-ray and CT Scan*. Accessed: Apr. 5, 2020. [Online]. Available: <https://bioldifferences.com/difference-between-x-ray-and-ct-scan.html>
- [28] N. Eldeen M. Khalifa, M. Hamed N. Taha, A. Ella Hassanien, and S. Elghamrawy, "Detection of coronavirus (COVID-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest X-ray dataset," 2020, *arXiv:2004.01184*. [Online]. Available: <http://arxiv.org/abs/2004.01184>
- [29] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Jamalipour Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101794, doi: [10.1016/j.media.2020.101794](https://doi.org/10.1016/j.media.2020.101794).
- [30] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images," *Pattern Recognit. Lett.*, vol. 138, pp. 638–643, Oct. 2020, doi: [10.1016/j.patrec.2020.09.010](https://doi.org/10.1016/j.patrec.2020.09.010).
- [31] M. M. Ahsan, K. D. Gupta, M. M. Islam, S. Sen, M. L. Rahman, and M. Shakhawat Hossain, "COVID-19 symptoms detection based on Nas-NetMobile with explainable AI using various imaging modalities," *Mach. Learn. Knowl. Extraction*, vol. 2, no. 4, pp. 490–504, Oct. 2020, doi: [10.3390/make2040027](https://doi.org/10.3390/make2040027).
- [32] E. El-Din Hemdan, M. A. Shouman, and M. Esmail Karar, "COVIDX-net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images," 2020, *arXiv:2003.11055*. [Online]. Available: <http://arxiv.org/abs/2003.11055>
- [33] P. K. Sathy and S. K. Behera. (2020). *Detection of Coronavirus Disease (COVID-19) Based on Deep Features and Support Vector Machine*. [Online]. Available: <https://www.preprints.org/manuscript/202003.0300/v2>
- [34] M. M. Ahsan, T. E. Alam, T. Trafalis, and P. Huebner, "Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and Non-COVID-19 patients," *Symmetry*, vol. 12, no. 9, p. 1526, Sep. 2020, doi: [10.3390/sym12091526](https://doi.org/10.3390/sym12091526).
- [35] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "An ensemble learning approach for brain cancer detection exploiting radiomic features," *Comput. Methods Programs Biomed.*, vol. 185, Mar. 2020, Art. no. 105134, doi: [10.1016/j.cmpb.2019.105134](https://doi.org/10.1016/j.cmpb.2019.105134).
- [36] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, Apr. 2020, doi: [10.1007/s13246-020-00865-4](https://doi.org/10.1007/s13246-020-00865-4).

- [37] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103792, doi: [10.1016/j.compbiomed.2020.103792](https://doi.org/10.1016/j.compbiomed.2020.103792).
- [38] A. I. Khan, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105581, doi: [10.1016/j.cmpb.2020.105581](https://doi.org/10.1016/j.cmpb.2020.105581).
- [39] T. B. Chandra, K. Verma, B. K. Singh, D. Jain, and S. S. Netam, "Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble," *Expert Syst. Appl.*, vol. 165, Mar. 2021, Art. no. 113909, doi: [10.1016/j.eswa.2020.113909](https://doi.org/10.1016/j.eswa.2020.113909).
- [40] B. Sekeroglu and I. Ozsahin, "Detection of COVID-19 from chest X-ray images using convolutional neural networks," *SLAS Technol., Translating Life Sci. Innov.*, vol. 25, no. 6, pp. 553–565, Sep. 2020, doi: [10.1177/2472630320958376](https://doi.org/10.1177/2472630320958376).
- [41] M. K. Pandit, S. A. Banday, R. Naaz, and M. A. Chisht, "Automatic detection of COVID-19 from chest radiographs using deep learning," *Radiography*, Nov. 2020, doi: [10.1016/j.radi.2020.10.018](https://doi.org/10.1016/j.radi.2020.10.018).
- [42] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105608, doi: [10.1016/j.cmpb.2020.105608](https://doi.org/10.1016/j.cmpb.2020.105608).
- [43] J. Paul Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "COVID-19 image data collection: Prospective predictions are the future," 2020, *arXiv:2006.11988*. [Online]. Available: <http://arxiv.org/abs/2006.11988>
- [44] Kaggle. *COVID-19 Chest X-Ray*. Accessed: Mar. 14, 2020. [Online]. Available: <https://www.kaggle.com/search?q=covid-19-datasetFileTypes%3Apng>
- [45] Kaggle. *COVID-19 chest X-Ray*. Accessed: Mar. 14, 2020. [Online]. Available: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>
- [46] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [47] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 2017, vol. 31, no. 1. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/11231>
- [48] T. Akiba, S. Suzuki, and K. Fukuda, "Extremely large minibatch SGD: Training ResNet-50 on ImageNet in 15 minutes," 2017, *arXiv:1711.04325*. [Online]. Available: <http://arxiv.org/abs/1711.04325>
- [49] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520, doi: [10.1109/cvpr.2018.00474](https://doi.org/10.1109/cvpr.2018.00474).
- [50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [51] F. Chollet, "Deep learning for computer vision," in *Deep Learning with Python and Keras: The Practical Manual from the Developer of the Keras Library*, 2nd ed. Nordrhein-Westfalen, German: MITP-Verlags GmbH, 2018, pp. 119–177, ch. 5.
- [52] M. Denil, B. Shakibi, L. Dinh, M. Ranzato, and N. de Freitas, "Predicting parameters in deep learning," 2013, *arXiv:1306.0543*. [Online]. Available: <http://arxiv.org/abs/1306.0543>
- [53] F. Hutter, J. Lücke, and L. Schmidt-Thieme, "Beyond manual tuning of hyperparameters," *KI-Künstliche Intell.*, vol. 29, no. 4, pp. 329–337, Jul. 2015, doi: [10.1007/s13218-015-0381-0](https://doi.org/10.1007/s13218-015-0381-0).
- [54] B. Qolomany, M. Maabreh, A. Al-Fuqaha, A. Gupta, and F. Benhaddou, "Parameters optimization of deep learning models using Particle swarm optimization," in *Proc. 13th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2017, pp. 1285–1290, doi: [10.1109/IWCMC.2017.7986470](https://doi.org/10.1109/IWCMC.2017.7986470).
- [55] J. Brownlee. (2018). *What is the Difference Between a Batch and an Epoch in a Neural Network*. [Online]. Available: <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch>
- [56] B. Chandra and R. K. Sharma, "Deep learning with adaptive learning rate using Laplacian score," *Expert Syst. Appl.*, vol. 63, pp. 1–7, Nov. 2016, doi: [10.1016/j.eswa.2016.05.022](https://doi.org/10.1016/j.eswa.2016.05.022).
- [57] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay," 2018, *arXiv:1803.09820*. [Online]. Available: <http://arxiv.org/abs/1803.09820>
- [58] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le, "Don't decay the learning rate, increase the batch size," 2017, *arXiv:1711.00489*. [Online]. Available: <http://arxiv.org/abs/1711.00489>
- [59] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.
- [60] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*. [Online]. Available: <http://arxiv.org/abs/1712.04621>
- [61] P. Filipczuk, T. Fevens, A. Krzyzak, and R. Monczak, "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies," *IEEE Trans. Med. Imag.*, vol. 32, no. 12, pp. 2169–2178, Dec. 2013, doi: [10.1109/tmi.2013.2275151](https://doi.org/10.1109/tmi.2013.2275151).
- [62] S. P. Mohanty, D. P. Hughes, and M. Salathā, "Using deep learning for image-based plant disease detection," *Frontiers Plant Sci.*, vol. 7, Art. no. 1419, Sep. 2016, doi: [10.3389/fpls.2016.01419](https://doi.org/10.3389/fpls.2016.01419).
- [63] T. Menzies, J. Greenwald, and A. Frank, "Data mining static code attributes to learn defect predictors," *IEEE Trans. Softw. Eng.*, vol. 33, no. 1, pp. 2–13, Jan. 2007, doi: [10.1109/tse.2007.256941](https://doi.org/10.1109/tse.2007.256941).
- [64] S. J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. K. Chan, "Cost-based modeling for fraud and intrusion detection: Results from the JAM project," *Proc. DARPA Inf. Survivability Conf. Expo.*, vol. 2, Jan. 2000, pp. 130–144, doi: [10.1109/DISCEX.2000.821515](https://doi.org/10.1109/DISCEX.2000.821515).
- [65] J. Zhang, Y. Xie, G. Pang, Z. Liao, J. Verjans, W. Li, Z. Sun, J. He, Y. Li, C. Shen, and Y. Xia, "Viral pneumonia screening on chest X-ray images using confidence-aware anomaly detection," 2020, *arXiv:2003.12338*. [Online]. Available: <http://arxiv.org/abs/2003.12338>
- [66] M. M. Ahsan, Y. Li, J. Zhang, M. T. Ahad, and M. M. S. Yazdan, "Face recognition in an unconstrained and real-time environment using novel BMC-LBPH methods incorporates with DJI vision sensor," *J. Sensor Actuator Netw.*, vol. 9, no. 4, p. 54, Nov. 2020, doi: [10.3390/jsan9040054](https://doi.org/10.3390/jsan9040054).
- [67] K.-S. Lee, J. Y. Kim, E.-T. Jeon, W. S. Choi, N. H. Kim, and K. Y. Lee, "Evaluation of scalability and degree of fine-tuning of deep convolutional neural networks for COVID-19 screening on chest X-ray images using explainable deep-learning algorithm," *J. Personalized Med.*, vol. 10, no. 4, p. 213, Nov. 2020, doi: [10.3390/jpm10040213](https://doi.org/10.3390/jpm10040213).
- [68] A. Shelke, M. Inamdar, V. Shah, A. Tiwari, A. Hussain, T. Chafekar, and N. and Mehendale, "Viral pneumonia screening on chest X-ray images using confidence-aware anomaly detection," *MedRxiv*, Jun. 2020, doi: [10.1101/2020.06.21.20136598](https://doi.org/10.1101/2020.06.21.20136598).



MD MANJURUL AHSAN received the M.S. degree in industrial engineering from Lamar University, USA, in 2018. He is currently pursuing the Ph.D. degree in industrial and systems engineering from The University of Oklahoma, Norman, USA. He is also a Graduate Research Assistant with The University of Oklahoma. His research interests include image processing, computer vision, deep learning, machine learning, and model optimization.



MD TANVIR AHAD (Graduate Student Member, IEEE) received the B.S. degree in electrical and electronic engineering from the Ahsanullah University of Science and Technology, Dhaka, Bangladesh, in 2013, and the master's degree in electrical engineering in 2018. He is currently pursuing the Ph.D. degree in aerospace and mechanical engineering with The University of Oklahoma. From January 2015 to July 2015, he was with the Dana Engineering International Ltd. (GE—Waukesha), Dhaka, Bangladesh as an Assistant Engineer (GE—Waukesha) in gas engine. From 2015 to 2018, he was with the Applied DSP Laboratory, Lamar University, Beaumont, TX, USA. He is also a Graduate Research Assistant with The University of Oklahoma. He has published several IEEE conferences (ICAEE, ISAECT, and IEEEGreenTech). His current research interests include signal processing, neural engineering, process design, AI and machine learning, and energy.



FARZANA AKTER SOMA received the Medical degree Bachelor of Medicine, Bachelor of Surgery (M.B.B.S.) degree from the Holy Family Red Crescent Medical College & Hospital, Dhaka, Bangladesh, in 2015. She has been an Intern Doctor with Holy Family Red Crescent Medical College & Hospital, since 2015. Her interests include infectious disease and internal medicine.



SHUVA PAUL (Member, IEEE) received the B.S. degree in electrical and electronics engineering and the M.S. degree from American International University-Bangladesh, Dhaka, Bangladesh, in 2013 and 2015, respectively, and the Ph.D. degree in electrical engineering and computer science from South Dakota State University, Brookings, SD, USA, in 2019. He is currently a Postdoctoral Fellow with the Georgia Institute of Technology. His current research interests include

power system cybersecurity, computational intelligence, reinforcement learning, game theory, and smart grid security for power transmission and distribution systems, events and anomaly detection, vulnerability and resilience assessment, and big data analytics.

Dr. Paul has been actively involved in numerous conference, including the Session Chair of the IEEE EnergyTech in 2013 and the IEEE EIT in 2019. He also serves as a reviewer for many reputed conference including IEEE PES General Meeting, SSCI, IJCNN, PECC, and ECCI, and journals including the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON SMART GRID, *Neurocomputing*, IEEE ACCESS, *IET The Journal of Engineering*, and *IET Cyber-Physical Systems: Theory & Applications*.



ANANNA CHOWDHURY received the Bachelor of Medicine, Bachelor of Surgery (M.B.B.S.) degree from Z. H. Sikder Womens Medical College & Hospital, Dhaka, Bangladesh, in 2018. From 2018 to 2019, she served as an Intern Doctor for Z. H. Sikder Womens Medical College & Hospital. Her research interests include infectious disease, epidemiology, and internal medicine.



SHAHANA AKTER LUNA received the medical (M.B.B.S.) degree from Dhaka Medical College & Hospital, Dhaka, Bangladesh, in 2019. She was an Active Internee Doctor for COVID-19 patients with Dhaka Medical College & Hospital.



MUNSHI MD. SHAFWAT YAZDAN received the B.S. degree in civil engineering from the Bangladesh University of Engineering and Technology (BUET), in 2012, and the M.S. degree in environmental engineering from Lamar University, TX, USA, in 2017. He is currently pursuing the Ph.D. degree with the Department of Civil and Environmental Engineering, Idaho State University. Since 2017, he has been an Adjunct Faculty with Idaho State University, ID, USA. His current

research interests include hydraulic fracturing, nanotechnology, hydrology, climate change, water and wastewater management, and treatment technology. He is a member of the ASCE and a GATE scholar of ISU.



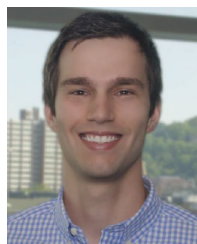
AKHLAQR RAHMAN (Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from American International University-Bangladesh, in 2012, and the Ph.D. degree in electrical engineering from the Swinburne University of Technology, Melbourne, Australia, in 2019. He has been teaching at a higher education level for 8 years since 2013. He has also been the Department Coordinator and a Lecturer with Uttara University, Dhaka, Bangladesh. He is currently the Course Coordinator and a Lecturer with the Engineering Institute of Technology (EIT).

He has published several Q1 (IEEE TII and Computer Networks) and Q2 (MDPI JSAN) Journals, while also attending several IEEE conferences (GLOBECOM, ISAECT, ICAMEchS, and ISGT-Asia). His research interests include machine learning, artificial intelligence, cloud networked robotics, optimal decision-making, industrial automation, and the IoT. Besides that, he is the Guest Editor for MDPI JSAN special issue on "Industrial Sensor Networks", and an Official Reviewer for *Robotics* (MDPI). He has also been part of the TPC for several conference such as CCB, IEEE Greentech, and IEEE ICAMEchS.



ZAHED SIDDIQUE received the Ph.D. degree in mechanical engineering from the Georgia Institute of Technology, in 2000. He is currently a Professor of mechanical engineering with the School of Aerospace and Mechanical Engineering and also the Associate Dean of research and graduate studies with the Gallogly College of Engineering, The University of Oklahoma. He is the coordinator of the industry sponsored capstone from at his school and is the advisor of OU's FSAE team.

He has published more than 163 research articles in journals, conference proceedings, and book chapters. He has also conducted workshops on developing competencies to support innovation, using experiential learning for engineering educators. His research interests include product family design, advanced material, engineering education, the motivation of engineering students, peer-to-peer learning, flat learning environments, technology assisted engineering education, and experiential learning.



PEDRO HUEBNER received the B.S. degree in production engineering from the Pontifical Catholic University of Parana, Brazil, in 2013, the master's degree in industrial engineering in 2015, and the Ph.D. degree in 2018. He joined the graduate program in industrial and systems engineering with the Edward P. Fitts Department of Industrial and Systems Engineering, NC State. He is currently an Assistant Professor with the School of Industrial and Systems Engineering, The

University of Oklahoma.

As a Researcher, he pursues his interests include bridging the gap between engineering and modern medicine, while investigating the applications of current manufacturing technologies in the field of Tissue Engineering and Regenerative Medicine. His latest research explores the design and fabrication of 3D printed bioresorbable implants assisting in the regeneration of musculoskeletal tissues such as the knee meniscus and osteochondral.

...