CLASSIFICATION OF VIRAL, BACTERIAL, AND COVID-19 PNEUMONIA USING DEEP LEARNING FRAMEWORK FROM CHEST X-RAY IMAGES

Muhammad E. H. Chowdhury, Tawsifur Rahman,
Amith Khandakar and Sakib Mahmud

Department of Electrical Engineering, Qatar University, Doha, Qatar

ABSTRACT

The novel coronavirus disease (COVID-19) is a highly contagious infectious disease. Even though there is a large pool of articles that showed the potential of using chest X-ray images in COVID-19 detection, a detailed study using a wide range of pre-trained convolutional neural network (CNN) encoders-based deep learning framework in screening viral, bacterial, and COVID-19 pneumonia are still missing. Deep learning network training is challenging without a properly annotated huge database. Transfer learning is a crucial technique for transferring knowledge from real-world object classification tasks to domain-specific tasks, and it may offer a viable answer. Although COVID-19 infection on the lungs and bacterial and viral pneumonia shares many similarities, they are treated differently. Therefore, it is crucial to appropriately diagnose them. The authors have compiled a large X-ray dataset (QU-MLG-COV) consisting of 16,712 CXR images with 8851 normal, 3616 COVID-19, 1485 viral, and 2740 bacterial pneumonia CXR images. We employed image pre-processing methods and 21 deep pre-trained CNN encoders to extract features, which were then dimensionality reduced using principal component analysis (PCA) and classified into 4-classes. We trained and evaluated every cutting-edge pre-trained network to extract features to improve performance. CheXNet surpasses other networks for identifying COVID-19, Bacterial, Viral, and Normal, with an accuracy of 98.89 percent, 97.87 percent, 97.55 percent, and 99.09 percent, respectively. The deep layer network found significant overlaps between viral and bacterial images. The paper validates the network learning from the relevant area of the images by Score-CAM visualization. The performance of the various pre-trained networks is also thoroughly examined in the paper in terms of both inference time and well-known performance criteria.

KEYWORDS

Novel Coronavirus disease, COVID-19, viral pneumonia, bacterial pneumonia, deep learning, Convolutional neural network, Principal component analysis.

1. INTRODUCTION

The COVID-19 pandemic struck the world, and it has severely overrun healthcare systems worldwide. It had an impact on social, economic, and all facets of human life [1, 2]. As of August 2022, there were more than 6 million fatalities and more than 579 million active cases worldwide [3]. Reverse transcription-polymerase chain reaction (RT-PCR), which recognizes viral nucleic acid, is the gold standard for COVID-19 diagnosis. Low viral load and sample

David C. Wyld et al. (Eds): SIGEM, MLTEC, SEAPP, ITCON, NATL, FUZZY, CSEA - 2022
pp. 01-21, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.122001

mistakes might lead to inaccurate RT-PCR results [4, 5]. Antigen testing is quick but not particularly accurate [6-8].

Therefore, it has become necessary to search for more readily available, dependable, and easily accessible diagnostic equipment. The authors have used deep learning in a variety of areas, including food [9-11], renewable energy [12, 13], education [14], communication [15, 16], and others. The authors investigated the use of machine learning for biomedical solutions, such as reliable heart sound diagnosis in smart digital stethoscopes [17], real-time heart attack detection in reducing road accidents [18], for estimating blood pressure from Photoplethysmogram signal and demographic features [19], etc. The current advances in artificial intelligence have been a lifesaver in various biomedical abnormalities detection [20]. Radiological tests can be useful in the diagnosis and assessment of disease progression by assessing the severity of pneumonia because pneumonia has been found in the majority of COVID-19 patients. As a result, typical diagnosis methods for COVID-19 include Chest X-ray (CXR) and Chest computed tomography (CT)-scan [21-23]. The procedures used in CT scans can be costly and occasionally readily can contaminate the equipment due to the nature of this contagious disease, endangering the safety of the subsequent patients. The American College of Radiology does not endorse it either [24, 25]. However, X-ray machines are more readily accessible, more affordable, and portable (as opposed to CT machines), making them a more affordable option for treating lung-related conditions. The impact of Chest X-rays and artificial intelligence on lung-related disorders including tuberculosis [26], pneumonia [27], and even COVID-19 [22, 23, 28, 29] has been the subject of numerous investigations by the authors. Portable tools used in solitary spaces can also help to lower the risk of infection [30-32]. Therefore, if artificial intelligence (AI) on CXR can be made more dependable with the aid of more research, it would be a more cost-effective solution.

Convolutional neural networks (CNNs), one type of deep learning artificial intelligence network, need large training data. Unfortunately, in the initial phase of the pandemic, CXR images are scarce for deep neural network training [33-36]. It is challenging to gather enough information from the small number of CXR images. In several investigations, it was suggested that increasing synthetic images for training might lessen the drawbacks. CovidGAN, an Auxiliary Classifier Generative Adversarial Network (ACGAN) based model, was utilized by Wang et al. in [37] to generate artificial data. A dataset of 403 COVID-CXR images and 721 normal images was used for their investigation. The accuracy of CNN's COVID-19 detection (using CXRs) has increased from 85% to 95%; thanks to the use of CovidGAN to generate synthetic data [37]. To achieve an accuracy of 96.58 percent, Chowdhury et al. in [38] built a unique framework called PDCOVIDNet using a dilated convolution in the parallel stack. Decompose, transfer, and compose (DeTraC) is a network proposed by Abbas et al. in [39] that checked anomalies by examining class borders of the images and reported an accuracy of 93.1 percent and sensitivity of 100 percent in COVID-19 detection. In [35], Wang and Wong proposed a deep COVID19 detection (COVID-Net) model, which classified normal, non-COVID pneumonia, and COVID-19 groups with 92.4 percent accuracy.

Some studies looked into how well learning of CNN models can be transferred to CXRs related learning. Transfer learning using CNN models has significantly helped in the process by using pre-trained networks for the task of differentiating between CXR images with normal and COVID-19-affected CXR by varying parameters like weights and biases of the pre-trained model. Azemin et al. in [40] used the ResNet-101 model to stratify COVID-19 with an accuracy of only 71.9 percent while Khan et al. [41] investigated a couple of pre-trained deep learning models like ResNet50, VGG16, VGG19, and DensNet121 and found that VGG16 and VGG19 had the highest performance with 99.3 percent accuracy in COVID-19 detection. The pre-trained networks AlexNet, GoogLeNet, and ResNet18 were examined by Loey et al. in [42] using a

dataset of 307 images divided into four classes: COVID-19, normal, pneumonia bacterial, and pneumonia virus. According to the authors, GoogLeNet achieved 99.9% validation accuracy and 100% testing accuracy. However, the study was carried out on a very small dataset and therefore, the performance reported in this study cannot be generalized on a large dataset.

Using a modified version of VGG-16, Brunese et al. [43] analyzed a dataset of 6,523 chest X-rays from patients with COVID-19, other lung illnesses, and healthy individuals. A 97 percent accuracy rate has been reported. In [44], the authors utilized the Xception model to categorize pneumonia, positive COVID-19, and negative COVID-19 on a dataset acquired from [45]. Training accuracy was stated to be 99.5% while the testing accuracy was 97.4%. Using a COVID-19 dataset made up of 2,951 CXR images and annotated ground-truth infection segmentation masks, Degerli et al. developed a novel method for COVID-19 infection map development in [46]. On the generated dataset, several encode-decoder (E-D) CNNs were trained and tested, with the best network achieving an F1 score for infection localization of 85.81 percent.

Deep CNN models were employed by Chowdhury et al. in [28] to classify images of normal, viral pneumonia, and COVID-19 chest X-rays into binary and three-class categories. Transfer learning was investigated on the generated dataset using pre-trained Squeezenet, Mobilenetv2, Inceptionv3, CheXNet, ResNet, and Densenet201 models. Three-class classification tasks produced an accuracy score of 97.9 percent while binary classification had a score of 99.7 percent. However, the study did not include bacterial pneumonia which is often challenging to classify using CXR images. Moreover, the dataset size was much smaller compared to the current study. Additionally, few deep learning models were investigated in this and other similar transfer learning-based works.

To the best of the authors' knowledge, an in-depth analysis of the performance of a large pool of state-of-the-art transfer learning models has not yet been demonstrated, even though a lot of research has been done on the use of artificial intelligence for COVID-19 detection from CXR images. The research community will benefit from this paper's addition to the body of knowledge about transfer learning's role in COVID-19 detection because it will help them to decide whether to deploy a particular network given that networks vary not only in terms of performance but also in terms of size, parameters, and inference time. These will be covered in the paper's remaining section. The network's performance's dependability as a visualization tool is tested, and since it uses one of the largest datasets accessible, it can also be regarded as trustworthy. The next sections of the chapter are organized as follows: Section II explored the technique in detail and gave a description of each transfer learning model, Section III presented the results and analyses and discussed them, and Section IV reported the conclusion.

2. METHODOLOGY

Figure 1 depicts the details of the proposed methodology used in this investigation. As can be observed, the authors investigated how well state-of-the-art pre-trained CNN models perform in useful feature extraction as a CNN encoder and then principal component analysis (PCA) was used to reduce the dimensionality to avoid overfitting of the classifier.

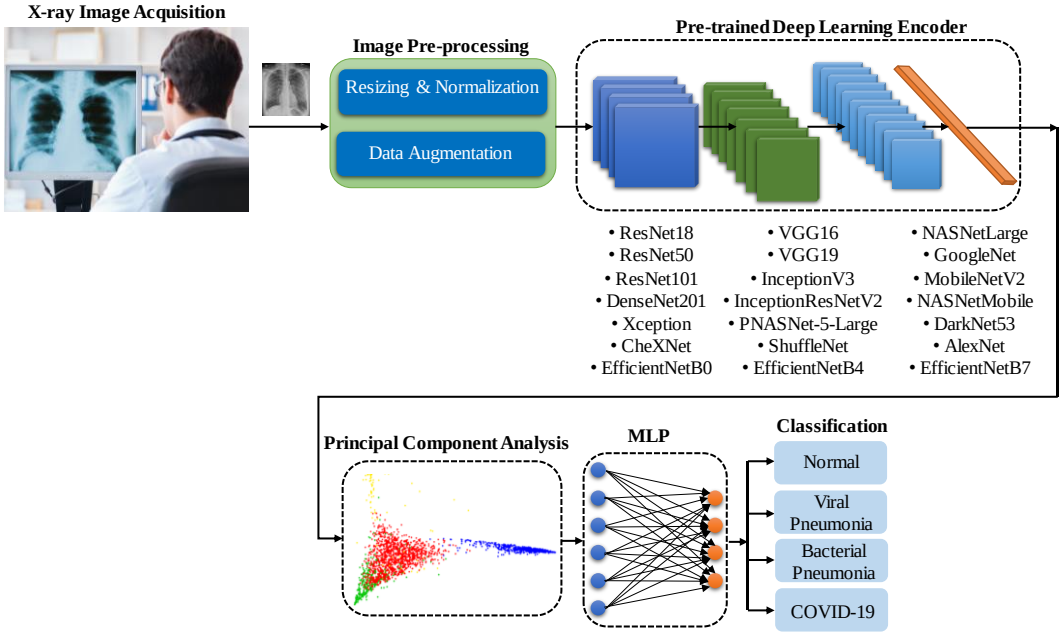

Figure 1. Framework of the proposed methodology.

Finally, a 3-layer multilayer-perceptron (MLP) classifier is used in identifying a dataset made up of COVID-19, Bacterial Pneumonia, Viral Pneumonia, and Normal CXR images. In MLP, one input, one hidden layer and one output layer, where the rectilinear unit (ReLU) activation function is used in the input and hidden layers and sigmoid function is used in the output and Interpretable maps are used afterwards to describe the usefulness and reliability of the proposed framework. This section will go into detail on the cutting-edge pre-trained model that was investigated in this study, the specifics of the dataset that was used, as well as the pre-processing and training aspects.

2.1. Deep Learning Pre-Trained Models

Several pre-trained CNN models were trained on a very large dataset, ImageNet, which has demonstrated cutting-edge performance [47]. These pre-trained networks have been shown to perform consistently well in different computer vision problems and can be trained on new databases to adjust their weights and biases for that dataset or application. The introduction section already showed examples of such applications. Network depth, which is the maximum number of consecutive convolutional layers of fully connected layers along the path from the input layer to the output layer, varies between these networks in terms of size and the number of parameters in millions. The parameters utilized in the investigation are described in detail in Table 1.

Table 1 compares the several cutting-edge models that were used in the study.

Network	*Network Depth*	*Parameters [Millions]*
ResNet18	72	11.7
ResNet50	107	25.6
ResNet101	209	44.7
VGG16	16	138.4
VGG19	19	143.7
InceptionV3	189	23.9
InceptionResNetV2	449	55.9
NASNetLarge	533	88.9
PNASNet-5-Large	*	86.1
Xception	81	22.9
CheXNet (DenseNet 121)	242	8.1
DenseNet201	402	20.2
ShuffleNet	50	1.4
GoogLeNet	22	7
MobileNetV2	105	3.5
NASNetMobile	132	5.3
DarkNet53	53	41
AlexNet	8	61.1
EfficientNetB0	132	5.3
EfficientNetB4	258	19.5
EfficientNetB7	438	66.7

*Values are not known

All the experiments were carried out in Python by importing all the models from the Pytorch library in Google ColabPro. Through transfer learning, specific characteristics from the X-ray images of the COVID-infected pneumonia patients were extracted using the rich set of features that these networks had learned from the ImageNet dataset. To categorize the X-ray images into one of the following classes: Normal, Bacterial Pneumonia, Viral Pneumonia, and COVID-19, the dense layers of each network were dropped and after flattening features were extracted from each CNN encoder. The dimensionality of the feature vector was reduced using principal component analysis (PCA) with 90% variance to avoid over-fitting of the models on a comparatively small dataset, then the multi-layer perceptron (MLP) model with SoftMax layer with four neurons was used as a classifier to classify the 4-class problem. Below are further specifics about the networks used in this study:

AlexNet - In the AlexNet, convolution processes were performed several times between max-pooling operations, which allows the network to acquire richer features at all spatial scales. AlexNet placed first with a top-5 test error rate of 15.3% [48] in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

VGG - VGG uses a standard CNN design and simply varies the depth: one network has 11 weight layers (8 convolutional and 3 fully connected layers), while another network has 19

weight layers (16 convolutional and 3 fully connected layers). Convolutional layers have a relatively small width (number of channels), starting at 64 in the first layer and rising by a factor of 2 after each max-pooling layer to reach 512 [49].

DarkNet 53 - Like the VGG models, it primarily makes use of filters and doubles the number of channels after each phase of pooling. Batch normalization is used to regularize the model batch, speed up convergence, and stabilize training. It serves as the foundation for the well-known localization network, YOLOv3 [50].

ResNet - Overfitting, a well-known paradigm for deep networks trained on small datasets, can significantly reduce the generalization performance. When a lot of training epochs are run, the "vanishing gradient" problem eventually leads to network saturation, especially at the initially hidden layers, making the problem worse. By incorporating the idea of shortcut connections, where the activations of one layer that are given to the next layer are fed to the deeper layers as well, which is the core concept of a residual network (ResNet). This solves the vanishing gradient problem with deep CNN networks. ResNet is made up of 8 residual blocks, each of which has two convolutional layers with three kernels on each layer [51] and it has a couple of variants: ResNet 18, 50, 101, and 152. Going farther into the network causes the layer depth to rise every two blocks, with layer sizes of 64, 128, 256, and 512 kernels, respectively. In addition, a 7×7 Conv layer is utilized in the network's beginning, followed by a pooling layer of stride 2, and a SoftMax classification layer at its end.

GoogleNet - The classification of many types of problems performed better with GoogLeNet-Inception networks. Smaller kernels are typically preferred for an area-specific feature that is dispersed over an image frame, whereas larger kernels are typically selected for global characteristics that are distributed over a vast region of images. This gave rise to the concept of inception layers, where kernels of various sizes - such as 1×1, 3×3, and 5×5 were combined within the same layer rather than moving further into the network [52, 53]. The Inception network begins with several traditional layers of 3×3 kernel, and 3 inception blocks, and culminates with an 8×8 global average pooling layer, followed by a SoftMax classifier. This architecture expands the network space where training can choose the best features.

InceptionV3 - This network suggested several improvements over version InceptionV1, which improved accuracy and decreased computational cost. High-quality networks can be trained on relatively small training sets thanks to the combination of a decreased parameter count, extra regularization, batch-normalized auxiliary classifiers, and label smoothing [54].

InceptionResNetV2 – It is the result of combining the most recent revision of the Inception architecture [54] with the residual connections, reported in [51]. It was claimed that the combination would keep the computational efficiency of the Inception network architecture while gaining all the advantages of the residual technique. Although it is a more expensive hybrid version of Inception, it has been demonstrated to have better performance [55].

Xception - In neural computer vision architectures, this network substitutes depthwise separable convolutions for Inception modules. The feature extraction base of the network in the Xception architecture is composed of 36 convolutional layers. The Xception architecture is a linear stack of residually connected depthwise separable convolution layers [56].

DenseNet - Contrary to residual networks, DenseNet concatenates all feature maps as opposed to simply adding up residuals [57]. All layers inside a thick block are closely connected to one another, allowing for more monitoring between levels. The four dense blocks that make up

DenseNet, each have numerous convolution layers with 1×1 and 3×3 filters. Transition layers made up of a batch normalization layer, a 1×1 convolutional layer, and a 2×2 average pooling layer are used to divide the dense blocks. The network begins with a 7×7 convolutional layer, then moves on to a 3×3 max-pooling layer, both with a stride of 2 and concludes with a 7×7 global average pooling layer, then a SoftMax layer. DenseNet has several variants, such as DenseNet121, DenseNet169, and DenseNet201.

ChexNet - One of the largest publicly accessible Chest X-ray datasets is ChestX-ray14 [58], which has over 100,000 frontal view X-ray images with 14 disease classes. CheXNet [59] is a DenseNet (DenseNet121) model which was re-trained on this dataset. Therefore, CheXNet is the only pre-trained network, which is already trained on a large X-ray dataset, unlike other pre-trained models.

MobileNetV2 – There are real-time applications like robots, self-driving cars, augmented reality, etc., which need compact networks (lightweight networks). MobileNet is built on a simplified architecture that uses depthwise separable convolutions to construct compact deep neural networks. It was created for mobile and embedded vision applications. The model builder can select the appropriate model size for their application based on the constraints of the problem using two straightforward global hyperparameters (width multiplier and resolution multiplier) [60]. For MobileNet, Depthwise Separable Convolutions are a crucial component. The MobileNetV2 design starts with a fully convolutional layer with 32 filters, followed by 19 bottleneck layers with residual connections between point wise convolutional layers. Except for the last fully connected layer, which has no nonlinearity and feeds into a SoftMax layer for classification, all Conv layers are followed by batch normalization and rectilinear unit (ReLU) nonlinear activation function [61].

ShuffleNet - The architecture of ShuffleNet makes use of two novel operations – point wise group convolution and channel shuffle - to significantly lower computation costs while retaining accuracy, it was also created with mobile phone deployment in mind. It has been improved to achieve lower complexity and is based on Residual network design [62].

NASNet - The NASNetLarge and NASNetMobile models' generalization has been greatly improved thanks to a new regularization method called ScheduledDropPath [63].

EfficientNet - Tan et al. presented EfficientNet in [64], which, in contrast to existing CNN scaling algorithms that use one-dimension scaling, balances the network's width, depth, and resolution.

2.2. Description of the Database

The authors have compiled a large dataset called, QU-MLG-COV in this study, which consists of 16,712 CXR images with 8851 normal, 3616 COVID-19, 1485 viral pneumonia, and 2740 bacterial pneumonia CXR images. The authors created this dataset by using and modifying various open-access databases for four different types of CXR images (COVID-19, normal (healthy), viral pneumonia, and bacterial pneumonia). The QU-MLG-COV dataset merged the COVID-19 dataset, the CXR dataset from the Radiological Society of North America (RSNA) [65], and the Chest X-Ray Images (Pneumonia) Kaggle dataset.

COVID-19 dataset

COVID-19 CXR images that make up the COVID-19 dataset were gathered from several publically accessible datasets,

have various lung abnormalities. We have taken 8851 healthy (normal) CXR images from the RSNA dataset for this study. Radiologists with the necessary training assessed the CXRs in the dataset, and clinical history, vital signs, and laboratory tests were used to confirm online sources and published studies. A total of 3616 X-ray images were collected; 2473 of them came from the BIMCV-COVID19+ dataset [66], 183 from a German medical school [67], 559 from the Italian Society of Medical Radiology (SIRM), GitHub, Kaggle, and Twitter [68-71], and 400 from another COVID-19 CXR repository [72]. The BIMCV-COVID19+ dataset, which includes 2473 CXR pictures of COVID-19 patients obtained from digital X-ray (DX) and computerized X-ray (CX) equipment, is the single largest available dataset.

RSNA Chest X-ray dataset

About 26,684 CXR DICOM images make up the RSNA pneumonia detection challenge dataset [65], of which 8851 images are normal, and 17842 images are the condition. To categorize the CXR images into healthy control (normal) and lung infections, they were connected with clinical symptoms and history.

Chest X-Ray Images (Pneumonia)

On Kaggle, 5824 chest X-ray images of bacterial, viral, and normal pneumonia were found with resolutions ranging from 400p to 2000p. Out of 5824 chest X-ray images, 2760 images with bacterial pneumonia and 1485 with viral pneumonia are used in this study. Figure 2 provides some examples of the Chest X-Ray images used in this investigation.

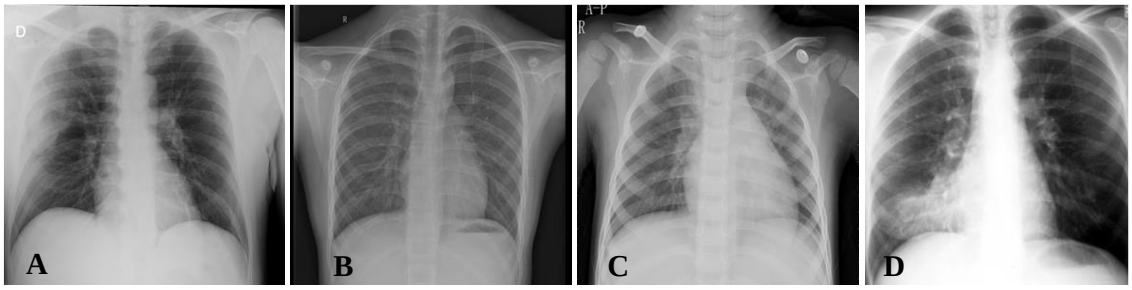

Figure 2. Sample CXR images from the dataset for COVID-19 (A), Normal (B), Viral Pneumonia (C), and Bacterial Pneumonia (D).

2.3. Experimental Setup

The dataset's distribution of labelled images across classes was unbalanced, which could have influenced training results. Data augmentation is a well-liked remedy for this unbalanced dataset [20, 27, 28], which may also be applied to expand the dataset because CNN models learn best from large databases. With 80% of the data used for training and 20% of the data being unseen for testing, the performance of the experiment conducted in this study was evaluated using five-fold cross-validation. Additionally, to prevent overfitting, 20% of the training data are used as a validation set.

Table 2. Before and after data augmentation, the number of photos per class and per fold Class

Class	*# of Samples*	*Training Samples*	*Augmented Training Samples*	*Validation Samples*	*Test Samples*
COVID-19	3616	2893	2314×2=4628	579	723
Normal	8851	7081	5665	1416	1770
Viral	1485	1188	950×6 = 5700	238	297
Bacterial	2760	2208	1766×3 = 5298	442	552

Data augmentation was used to balance the dataset by applying rotations of 5 and 10 degrees to avoid unrealistic rotations of the images. Additionally, image translations in both the horizontal and vertical directions were applied within the range [-0.15, +0.15], as shown in Figure 3. The number of CXR images per class utilized for training, validation, and testing at each fold is listed in Table 2.

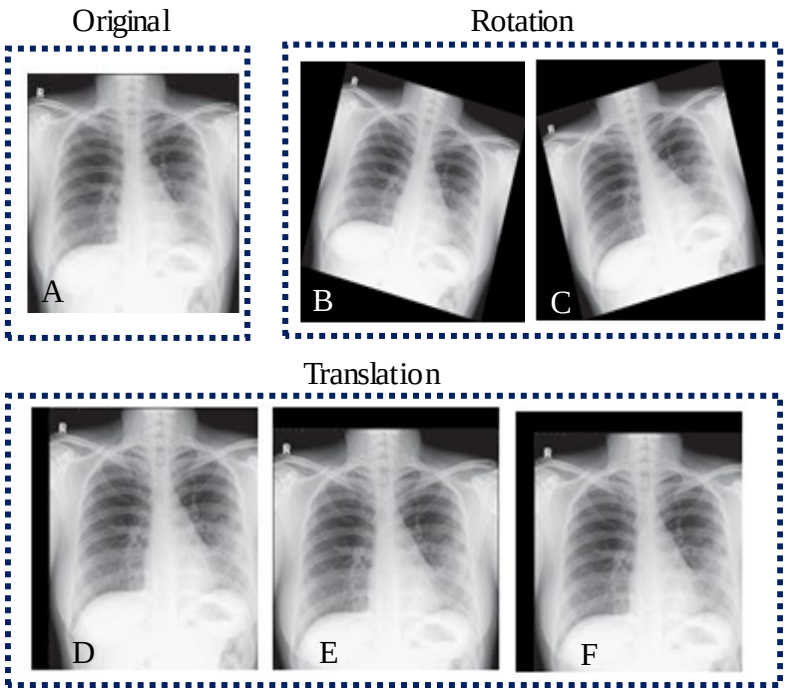

Figure 3. Image augmentation: clockwise and anticlockwise rotation and horizontal and vertical translation.

The experiment was conducted on ColabPro using the Pytorch library. Table 3 displays the training experiment's specifics. To create the final receiver operating characteristic (ROC) curve, confusion matrix, and evaluation matrices, a fivefold cross-validation result was averaged.

Table 3. Details of the experiment hyper-parameters

Training parameter	*Value*
batch size	16
learning rate	0.001
epochs	15
epochs patience	8
stopping criteria	8
Loss function	BCE
optimizer	ADAM

Four evaluation metrics - Accuracy, Precision, Sensitivity, and F1-score with 95 percent confidence intervals (CIs) - were used to evaluate the performance of the deep CNN-based encoders. The overall confusion matrix, which compiles all test fold results from the 5-fold cross-validation, was used to derive per-class values.

$$Accuracy_{class\_i} = \frac{TP_{class\_i} + TN_{class\_i}}{TP_{class\_i} + TN_{class\_i} + FP_{class\_i} + FN_{class\_i}} \quad (1)$$

$$Precision_{class\_i} = \frac{TP_{class\_i}}{TP_{class\_i} + FP_{class\_i}} \quad (2)$$

$$Sensitivity_{class_i} = \frac{TP_{class_i}}{TP_{class_i} + FN_{class_i}} \quad (3)$$

$$F1\_score_{class_i} = 2\frac{Precision_{class_i} \times Sensitivity_{class_i}}{Precision_{class_i} + Sensitivity_{class_i}} \quad (4)$$

where $class_i = COVID-19, normal, viral\ and\ bacterial.$

The weighted average values of each class were used to calculate the overall performance. Since class frequencies differ for the given task, the weighted average provides a better indication of overall performance.

$$Precision = \frac{n1(Precision_{COVID}) + n2(Precision_{normal}) + n3(Precision_{viral}) + n4(Precision_{bacterial})}{n1 + n2 + n3 + n4} \quad (5)$$

$$Sensitivity = \frac{n1(Sensitivity_{COVID}) + n2(Sensitivity_{normal}) + n3(Sensitivity_{viral}) + n4(Sensitivity_{bacterial})}{n1 + n2 + n3 + n4} \quad (6)$$

$$F1\_score = \frac{n1(F1\_score_{COVID}) + n2(F1\_score_{normal}) + n3(F1\_score_{viral}) + n4(F1\_score_{bacterial})}{n1 + n2 + n3 + n4} \quad (7)$$

$$Accuracy = \frac{n1(Accuracy_{COVID}) + n2(Accuracy_{normal}) + n3(Accuracy_{viral}) + n4(Accuracy_{bacterial})}{n1 + n2 + n3 + n4} \quad (8)$$

Where $n1, n2, n3\ and\ n4$ are the total number of COVID-19, normal, viral, and bacterial cases respectively.

The authors also used a different metric, known as inference time, $\delta(t)$, to measure how long it took the framework to classify the image. The trade-off between accuracy and decision-making time can also be better understood by plotting inference time against the F1 score. To ensure that the best-performing network is picking up knowledge from important areas of the image, the authors have additionally verified the validity of the best-trained network using the well-known Score-CAM visualization technique [73].

3. RESULTS AND DISCUSSION

Table 3 shows the overall performance of the various pre-trained network-based encoders in descending order of the F1-score. As can be observed, the performance of all different CNN encoder-based frameworks is comparable. It demonstrates that all networks, regardless of size, perform well in classifying the CXR images into various classes. This merely verifies how well pre-trained models on CXRs are, as described in numerous earlier literature. It is also noteworthy that EfficientNet B7 has the longest inference time—roughly 42.8 milliseconds—while AlexNet has the shortest—roughly 1.8 milliseconds. However, the best performing model is CheXNet, which is the only network pre-trained on a large CXR dataset along with already trained on ImageNet. There is a clear performance gap of more than 2% compared to other pre-trained models which are only trained on ImageNet. This highlights the importance of retraining the pre-trained model on domain data, which can improve the model's performance.

The best option initially can be appeared to be ResNet18 and MobileNetV2 because of their excellent performance and quick inference times. The greatest option for even deploying for mobile and smart devices would be MobileNetV2. CheXNet is DenseNet121 trained on large CXR images and performed better than the DenseNet201 variant of DenseNet. This also clearly reflects the importance of domain knowledge of the model during re-training. However, it is worth mentioning that the pre-trained model-based framework overall performs close to each other as they are already trained on a large image database and these models are very good at extracting useful image features.

Table 3. Overall network performance metrics in descending order of F1-Score.

Encoders	***Inference time $\delta(t)$(ms)***	***Accuracy (%)***	***Precision (%)***	***Sensitivity (%)***	***F1-score (%)***
CheXNet (DenseNet 121)	17.6	96.7	96.51	97.18	96.81
EfficientNet B0	10.7	94.24	94.32	94.42	94.18
ResNet 18	5.5	93.72	93.76	93.54	93.99
MobileNet V2	2.5	93.19	93.29	93.07	92.99
DenseNet201	25.8	93.02	92.67	93.07	92.99
InceptionV3	23.8	92.52	92.26	92.52	92.3
ResNet 101	10.8	92.81	91.45	91.85	91.45
ResNet 50	6.1	91.71	91.23	91.71	91.35
EfficientNet B7	42.8	91.47	91.47	91.47	91.47
VGG 16	11	91.42	91.12	91.42	91.22
AlexNet	1.8	91.16	91.06	91.37	91.16
PNASNET-5-Large	28.8	91.15	91.04	91.15	91.09
EfficientNet B4	28.7	90.96	90.86	90.96	90.91
GoogLeNet	3.3	90.03	90.79	90.03	90.89
Darknet53	7.2	90.88	90.56	90.88	90.66
Xception	4.6	90.96	90.52	90.96	90.62
VGG 19	12.1	90.91	90.47	90.92	90.61
NASNetLarge	30.1	89.66	89.58	89.58	89.57
ShuffleNet	6.6	89.77	89.46	89.77	89.51
InceptionResNet	29.8	88.07	88.78	88.07	88.89
NASNetMobile	4.8	87.24	87.11	87.25	87.17

Table 4 shows the class-wise performance comparison of the top performing 5 networks, namely, CheXNet, EfficientNet B0, ResNet18, MobileNetV2, and DenseNet201. It is further confirmed by the class-specific performance that CheXNet outperforms all other models as it performs well not only overall but also in classes. It is also evident from the table that the networks may become perplexed when attempting to differentiate between viral and bacterial pneumonia CXRs. This is supported by the literature as the signature of infection from community-acquired viral pneumonia and bacterial pneumonia has some overlapping features, which is confusing the networks. It is clear from Tables 3 and 4 that the performance of CheXNet is higher than the other four top-performing models. This explains that CheXNet can extract more useful CXR features from the CXR images compared to other top-performing models.

Table 4. Class-wise performance comparison for the Top-5 networks

Encoder	*Inference time $\delta(t)$(ms)*	*Class*	*Accuracy (%)*	*Precision (%)*	*Sensitivity (%)*	*F1-score (%)*
CheXNet (DenseNet 121)	17.6	Bacterial Pneumonia	97.87	93	95	94
		COVID-19	98.89	97	99	98
		Normal	99.09	99	99	99
		Viral Pneumonia	97.55	87	86	86
		Overall	96.7	96.51	97.18	96.81
EfficientNet B0	10.7	Bacterial Pneumonia	96.74	89	92	90
		COVID-19	97.19	94	93	93
		Normal	97.75	98	98	98
		Viral Pneumonia	96.81	83	81	82
		Overall	94.24	94.32	94.42	94.18
ResNet18	5.5	Bacterial Pneumonia	96.51	88	91	90
		COVID-19	96.91	93	93	93
		Normal	97.5	98	97	98
		Viral Pneumonia	96.52	81	79	80
		Overall	93.72	93.76	93.54	93.99
MobileNetV2	2.5	Bacterial Pneumonia	96.28	87	90	89
		COVID-19	96.61	92	92	92
		Normal	97.24	98	97	97
		Viral Pneumonia	96.25	80	78	79
		Overall	93.19	93.29	93.07	92.99
DenseNet201	25.8	Bacterial Pneumonia	96.17	87	90	89
		COVID-19	96.53	92	92	92
		Normal	97.09	97	97	97
		Viral Pneumonia	96.25	79	78	79
		Overall	93.02	92.67	93.07	92.91

The confusion matrix depicted in Figure 4 can be used to further confirm CheXNet's effectiveness. It is clear that the network does a good job of differentiating between normal and pathological behaviour (i.e. Viral Pneumonia, Bacterial Pneumonia, and COVID -19). It is doing an excellent job of determining the COVID-19 patients and healthy control. Although understandably, bacterial and viral pneumonia are both atypical forms of pneumonia, it can be difficult to distinguish them in the early stage as mentioned earlier. There is a good number miss-classification between bacterial and viral pneumonia. The framework is missing some of the control patients to the unhealthy group, which could be due to the early stage CXR images, where the signature of infection is not evident.

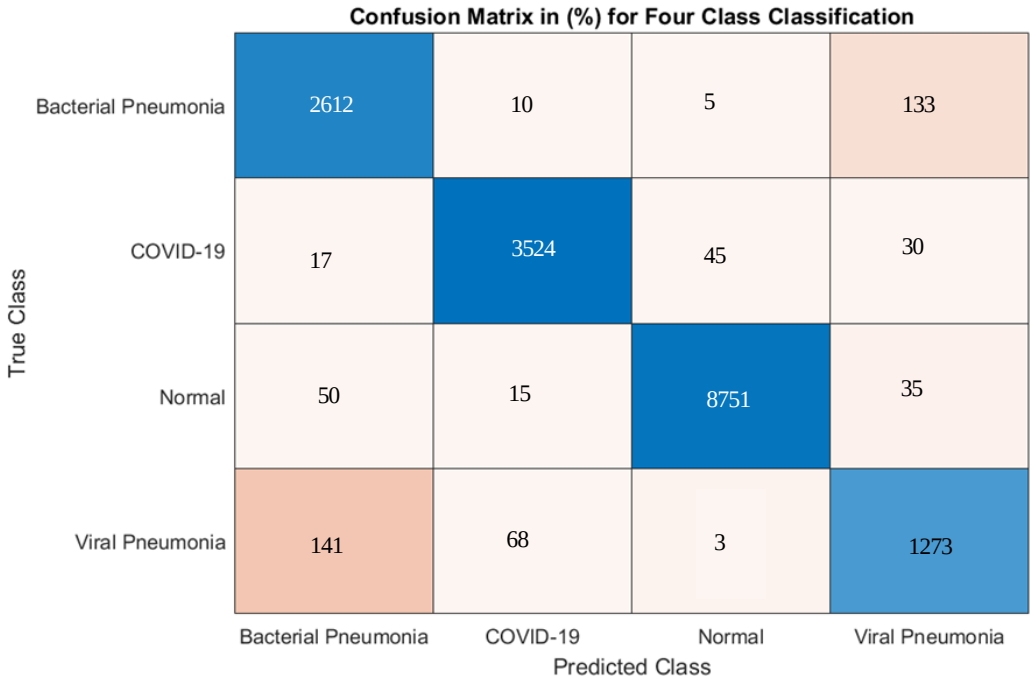

Figure 4. Confusion Matrix for the top-performing model, CheXNet.

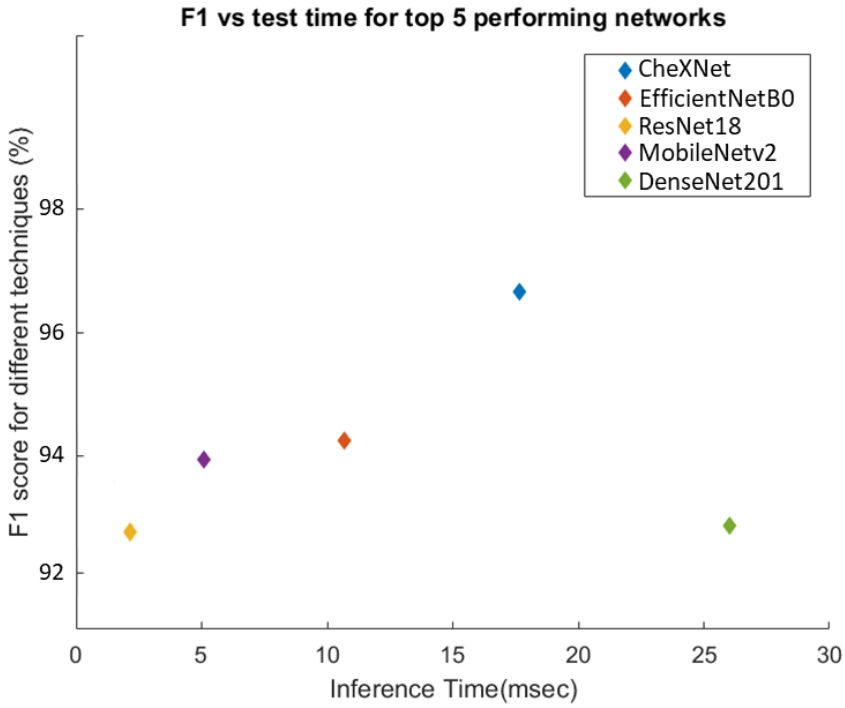

Figure 5. F1-Score versus Inference time for the top-performing 5 models.

Figure 5 shows a comparison of the Top-5 networks in terms of F1-Score and inference time. It is clear that while the networks' overall performances are comparable, the inference times vary significantly. The authors have further examined whether the networks are genuinely picking up

knowledge from the problem's "lungs," or region of interest. The Score-CAM visualization results displayed in Figure 6 can be used to confirm this. The network choosing the lung area for all classes can be verified by the Score-CAM visualization for the top-performing CheXNet network.

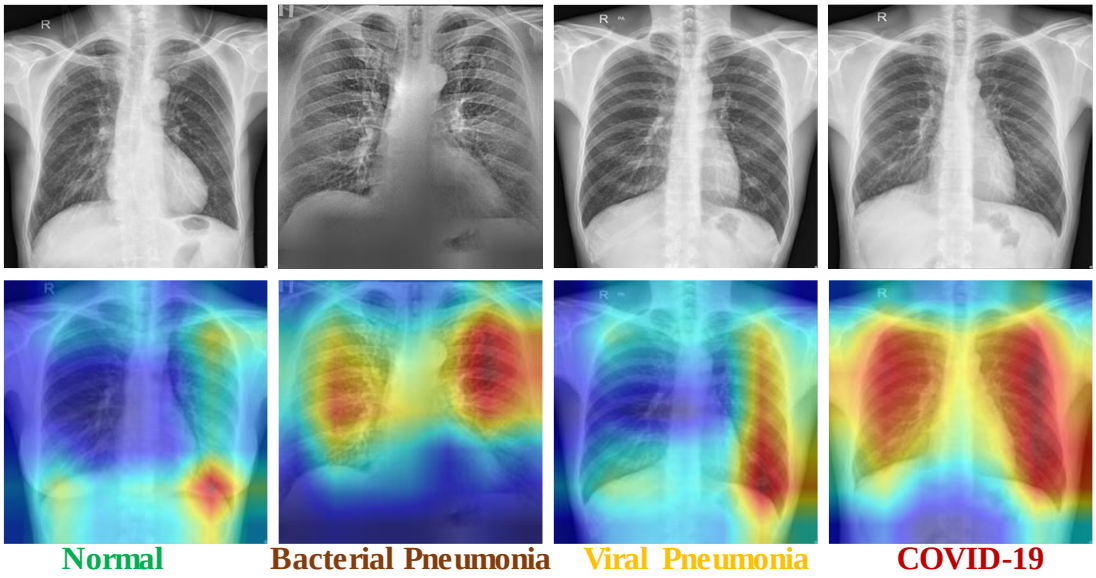

Figure 6. Score-CAM-based heat-map for the best performing CheXNet model, showing where the network is learning more to take the decision.

Table 5 summarizes the recent works on multi-class classification using deep learning algorithms. In all the cases, apart from the studies from our group, the size of the COVID dataset is small. However, the performance of this work is superior compared to the relevant literature.

Table 5. Comparison with the current state-of-art/relevant studies.

Articles	Techniques	Dataset	Performance
Tsung et al. [75]	CNN (ResNet50)	15478 chest X-ray images (473 COVID)	accuracy, sensitivity, and specificity obtained are 93%, 90.1%, and 89.6%
Abbas et al. [39]	CNN (DeTraC)	1768 chest X-ray images (949 COVID)	Accuracy-93.1%
Jain et al. [76]	CNN (Inception V3, Xception, and ResNet)	6432 chest X-ray images (490 COVID)	Accuracy-96% and Recall-92%
Ohata et al. [77]	Transfer learning + machine learning method (DenseNet201 + MLP)	388 chest X-ray images (194 COVID)	Acc: 95.641%, F1-score: 95.633%, FPR: 4.103%
Ioannis et al. [78]	CNN	1427 chest X-ray images (224 COVID)	accuracy, sensitivity, and specificity obtained are 96%, 96.66%, and 96.46%
Chowdhury et al. [28]	Seven different deep CNN networks for classification	423 COVID-19, 1485 viral pneumonia, and 1579 normal chest X-ray	The classification accuracy, precision, sensitivity, and specificity

Articles	Techniques	Dataset	Performance
		images	were 99.55% and 97.9%, 97.95%, 97.9%, and 98.8%, respectively
Rahman et al. [29]	Seven different deep CNN networks for classification and a modified Unet network for segmentation	18479 chest x-ray images (3616 COVID)	accuracy of 96.29%, sensitivity of 97.28%, and the F1-score of 96.28%. In segmentation, Accuracy of 98.63%, and Dice score of 96.94%
Proposed study	Many different deep CNN networks and PCA-based framework for classification	16,712 CXR images with 8851 normal, 3616 COVID-19, 1485 viral, and 2740 bacterial pneumonia	Overall accuracy of 96.7% for 4-class problem

4. CONCLUSION

To the best of the authors' knowledge, an in-depth analysis of the performance of the popular twenty-one state-of-the-art transfer learning model has not yet been demonstrated, even though a lot of research has been done on the use of artificial intelligence for COVID-19 detection from CXR images. The research community will benefit from this paper's addition to the body of knowledge about transfer learning's role in COVID-19 detection because it will help them decide whether to deploy a particular network given that networks vary not only in terms of performance but also in terms of size, parameters, and inference time. Overall performance measurements reveal that all the networks perform well, however, CheXNet is found to perform better than all the other networks. It is found that the CheXNet is picking up information from the lung areas in the reliability test conducted utilizing Score-CAM visualization. The results show that CheXNet outperforms with an overall accuracy of 96.7%, with an inference time of 17.6 ms, which is inline with the findings reported in the previous work of the authors [74]. In future, the authors will deploy such framework in cloud platform to carryout a multi-centre study and to improve the model generalizability.

ACKNOWLEDGEMENTS

This work was supported by the Qatar National Research Grant NPRP11S-0102-180178 and UREP28-144-3-046. The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] S. A. Harmon, T. H. Sanford, S. Xu, E. B. Turkbey, H. Roth, Z. Xu, et al., "Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets," Nature communications, vol. 11 (1), pp. 1-7(2020)

[2] N. Abughanam, S. S. M. Gaben, M. E. Chowdhury, and A. Khandakar, "Investigating the effect of materials and structures for negative pressure ventilators suitable for pandemic situation," Emergent Materials, vol. 4 (1), pp. 313-327(2021)

[3] M. Black, A. Lee, and J. Ford, "Vaccination against COVID-19 and inequalities–avoiding making a bad situation worse," Public Health in Practice (Oxford, England), 2021)

[4] I. Arevalo-Rodriguez, D. Buitrago-Garcia, D. Simancas-Racines, P. Zambrano-Achig, R. Del Campo, A. Ciapponi, et al., "False-negative results of initial RT-PCR assays for COVID-19: a systematic review," PloS one, vol. 15 (12), p. e0242958(2020)

[5] A. Tahamtan and A. Ardebili, "Real-time RT-PCR in COVID-19 detection: issues affecting the results," Expert review of molecular diagnostics, vol. 20 (5), pp. 453-454(2020)
[6] A. Afzal, "Molecular diagnostic technologies for COVID-19: Limitations and challenges," Journal of advanced research, 2020)
[7] W. H. Organization, "Use of chest imaging in COVID-19: a rapid advice guide, 11 June 2020," World Health Organization2020.
[8] W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, et al., "Detection of SARS-CoV-2 in different types of clinical specimens," Jama, vol. 323 (18), pp. 1843-1844(2020)
[9] M. E. Chowdhury, A. Khandakar, S. Ahmed, F. Al-Khuzaei, J. Hamdalla, F. Haque, et al., "Design, construction and testing of iot based automated indoor vertical hydroponics farming test-bed in qatar," Sensors, vol. 20 (19), p. 5637(2020)
[10] M. E. Chowdhury, T. Rahman, A. Khandakar, M. A. Ayari, A. U. Khan, M. S. Khan, et al., "Automatic and Reliable Leaf Disease Detection Using Deep Learning Techniques," AgriEngineering, vol. 3 (2), pp. 294-312(2021)
[11] M. E. Chowdhury, T. Rahman, A. Khandakar, N. Ibtehaz, A. U. Khan, M. S. Khan, et al., "Tomato Leaf Diseases Detection Using Deep Learning Technique," 2021)
[12] A. Khandakar, M. EH Chowdhury, M. Khoda Kazi, K. Benhmed, F. Touati, M. Al-Hitmi, et al., "Machine learning based photovoltaics (PV) power prediction using different environmental parameters of Qatar," energies, vol. 12 (14), p. 2782(2019)
[13] A. Khandakar, A. Rizqullah, A. Ashraf Abdou Berbar, M. Rafi Ahmed, A. Iqbal, M. E. Chowdhury, et al., "A Case Study to Identify the Hindrances to Widespread Adoption of Electric Vehicles in Qatar," Energies, vol. 13 (15), p. 3994(2020)
[14] A. Khandakar, M. E. H. Chowdhury, A. Gonzales Jr, S. Pedro, F. Touati, N. A. Emadi, et al., "Case study to analyze the impact of multi-course project-based learning approach on education for sustainable development," Sustainability, vol. 12 (2), p. 480(2020)
[15] A. Khandakar and A. Mahmoud Salem Mohamed, "Understanding probabilistic cognitive relaying communication with experimental implementation and performance analysis," Sensors, vol. 19 (1), p. 179(2019)
[16] A. Khandakar, A. Touati, F. Touati, A. Abdaoui, and A. Bouallegue, "Experimental setup to validate the effects of major environmental parameters on the performance of FSO communication link in Qatar," Applied Sciences, vol. 8 (12), p. 2599(2018)
[17] M. E. Chowdhury, A. Khandakar, K. Alzoubi, S. Mansoor, A. M Tahir, M. B. I. Reaz, et al., "Real-time smart-digital stethoscope system for heart diseases monitoring," Sensors, vol. 19 (12), p. 2781(2019)
[18] M. E. Chowdhury, K. Alzoubi, A. Khandakar, R. Khallifa, R. Abouhasera, S. Koubaa, et al., "Wearable real-time heart attack detection and warning system to reduce road accidents," Sensors, vol. 19 (12), p. 2780(2019)
[19] M. H. Chowdhury, M. N. I. Shuzan, M. E. Chowdhury, Z. B. Mahbub, M. M. Uddin, A. Khandakar, et al., "Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques," Sensors, vol. 20 (11), p. 3127(2020)
[20] M. E. Chowdhury, A. Khandakar, Y. Qiblawey, M. B. I. Reaz, M. T. Islam, and F. Touati, "Machine Learning in Wearable Biomedical Systems," in Sports Science and Human Health-Different Approaches, ed: IntechOpen, 2020.
[21] Y. Qiblawey, A. Tahir, M. E. Chowdhury, A. Khandakar, S. Kiranyaz, T. Rahman, et al., "Detection and severity classification of COVID-19 in CT images using deep learning," Diagnostics, vol. 11 (5), p. 893(2021)
[22] A. M. Tahir, M. E. Chowdhury, A. Khandakar, T. Rahman, Y. Qiblawey, U. Khurshid, et al., "COVID-19 Infection Localization and Severity Grading from Chest X-ray Images," arXiv preprint arXiv:2103.07985, 2021)
[23] A. Tahir, Y. Qiblawey, A. Khandakar, T. Rahman, U. Khurshid, F. Musharavati, et al., "Coronavirus: Comparing COVID-19, SARS and MERS in the eyes of AI," arXiv preprint arXiv:2005.11524, 2020)
[24] A. Jacobi, M. Chung, A. Bernheim, and C. Eber, "Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review," Clinical imaging, 2020)
[25] S. Kooraki, M. Hosseiny, L. Myers, and A. Gholamrezanezhad, "Coronavirus (COVID-19) outbreak: what the department of radiology should know," Journal of the American college of radiology, vol. 17 (4), pp. 447-451(2020)

[26] T. Rahman, A. Khandakar, M. A. Kadir, K. R. Islam, K. F. Islam, R. Mazhar, et al., "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," IEEE Access, vol. 8 pp. 191586-191601(2020)
[27] T. Rahman, M. E. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, et al., "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection using Chest X-ray," Applied Sciences, vol. 10 (9), p. 3233(2020)
[28] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, et al., "Can AI help in screening viral and COVID-19 pneumonia?," IEEE Access, vol. 8 pp. 132665-132676(2020)
[29] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, et al., "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," Computers in biology and medicine, vol. 132 p. 104319(2021)
[30] E. S. Amis Jr, P. F. Butler, K. E. Applegate, S. B. Birnbaum, L. F. Brateman, J. M. Hevezi, et al., "American College of Radiology white paper on radiation dose in medicine," Journal of the american college of radiology, vol. 4 (5), pp. 272-284(2007)
[31] E. Baratella, P. Crivelli, C. Marrocchio, A. M. Bozzato, A. D. Vito, G. Madeddu, et al., "Severity of lung involvement on chest X-rays in SARS-coronavirus-2 infected patients as a possible tool to predict clinical progression: an observational retrospective analysis of the relationship between radiological, clinical, and laboratory data," Jornal Brasileiro de Pneumologia, vol. 46 (5), 2020)
[32] G. D. Rubin, C. J. Ryerson, L. B. Haramati, N. Sverzellati, J. P. Kanne, S. Raoof, et al., "The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society," Chest, vol. 158 (1), pp. 106-116(2020)
[33] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, et al., "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," Radiology, vol. 296 (2), pp. E32-E40(2020)
[34] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, et al., "Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT," Radiology, 2020)
[35] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," Scientific Reports, vol. 10 (1), pp. 1-12(2020)
[36] P. R. Bassi and R. Attux, "A deep convolutional neural network for covid-19 detection using chest x-rays," Research on Biomedical Engineering, pp. 1-10(2021)
[37] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection," Ieee Access, vol. 8 pp. 91916-91923(2020)
[38] N. K. Chowdhury, M. M. Rahman, and M. A. Kabir, "PDCOVIDNet: a parallel-dilated convolutional neural network architecture for detecting COVID-19 from chest X-ray images," Health information science and systems, vol. 8 (1), pp. 1-14(2020)
[39] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," Applied Intelligence, vol. 51 (2), pp. 854-864(2021)
[40] M. Z. Che Azemin, R. Hassan, M. I. Mohd Tamrin, and M. A. Md Ali, "COVID-19 deep learning prediction model using publicly available radiologist-adjudicated chest X-ray images as training data: preliminary findings," International Journal of Biomedical Imaging, vol. 2020 2020)
[41] I. U. Khan and N. Aslam, "A deep-learning-based framework for automated diagnosis of COVID-19 using X-ray images," Information, vol. 11 (9), p. 419(2020)
[42] M. Loey, F. Smarandache, and N. E. M Khalifa, "Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning," Symmetry, vol. 12 (4), p. 651(2020)
[43] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays," Computer Methods and Programs in Biomedicine, vol. 196 p. 105608(2020)
[44] N. N. Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, "Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays," Irbm, 2020)
[45] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," Computers in biology and medicine, vol. 121 p. 103792(2020)

[46] A. Degerli, M. Ahishali, M. Yamac, S. Kiranyaz, M. E. Chowdhury, K. Hameed, et al., "COVID-19 Infection Map Generation and Detection from Chest X-Ray Images," arXiv preprint arXiv:2009.12698, 2020)
[47] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248-255.
[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, vol. 25 pp. 1097-1105(2012)
[49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014)
[50] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018)
[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
[52] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9.
[53] B. Raj, "A simple Guide to the versions of the Inception network," Retrieved from Towards Data Science, 2018)
[54] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818-2826.
[55] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
[56] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251-1258.
[57] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700-4708.
[58] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2097-2106.
[59] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, et al., "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017)
[60] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017)
[61] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510-4520.
[62] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6848-6856.
[63] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8697-8710.
[64] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in International Conference on Machine Learning, 2019, pp. 6105-6114.
[65] kaggle. RSNA Pneumonia Detection Challenge [Online]. Available: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data. [Accessed on 09-June-2020]
[66] (2020). BIMCV-COVID19, Datasets related to COVID19's pathology course [Online]. Available: https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/#1590858128006-9e640421-6711. [Accessed on: 06 August 2020]
[67] (2020). covid-19-image-repository [Online]. Available: https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png. [Accessed on: 06 August 2020]

[68] R. Chen, W. Liang, M. Jiang, W. Guan, C. Zhan, T. Wang, et al., "Risk factors of fatal outcome in hospitalized subjects with coronavirus disease 2019 from a nationwide analysis in China," Chest, 2020)
[69] Z. Weng, Q. Chen, S. Li, H. Li, Q. Zhang, S. Lu, et al., "ANDC: an early warning score to predict mortality risk for patients with Coronavirus Disease 2019," 2020)
[70] J. Liu, Y. Liu, P. Xiang, L. Pu, H. Xiong, C. Li, et al., "Neutrophil-to-lymphocyte ratio predicts severe illness patients with 2019 novel coronavirus in the early stage," MedRxiv, 2020)
[71] I. Huang and R. Pranata, "Lymphopenia in severe coronavirus disease-2019 (COVID-19): systematic review and meta-analysis," Journal of Intensive Care, vol. 8 (1), pp. 1-10(2020)
[72] (2020). COVID-CXNet [Online]. Available: https://github.com/armiro/COVID-CXNet. [Accessed on: 06 August 2020]
[73] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, et al., "Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 24-25.
[74] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, et al., "Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-rays Images," Computers in Biology and Medicine, p. 104319(2021).
[75] Lin, T. C., & Lee, H. C. (2020, August). Covid-19 chest radiography images analysis based on integration of image preprocess, guided grad-CAM, machine learning and risk management. In Proceedings of the 4th International Conference on Medical and Health Informatics (pp. 281-288).
[76] Jain, R., Gupta, M., Taneja, S., & Hemanth, D. J. (2021). Deep learning-based detection and analysis of COVID-19 on chest X-ray images. Applied Intelligence, 51(3), 1690-1700.
[77] Ohata, Elene Firmeza, et al. "Automatic detection of COVID-19 infection using chest X-ray images through transfer learning." IEEE/CAA Journal of Automatica Sinica 8.1 (2020): 239-248.
[78] Apostolopoulos, Ioannis D., and Tzani A. Mpesiana. "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks." Physical and engineering sciences in medicine 43.2 (2020): 635-640.

AUTHORS

Muhammad E. H. Chowdhury received his PhD degree from the University of Nottingham, U.K., in 2014. He worked as a Postdoctoral Research Fellow at the Sir Peter Mansfield Imaging Centre, University of Nottingham. He is currently working as an Assistant Professor with the Department of Electrical Engineering, Qatar University. He has filed several patents and published more than 100 peer-reviewed journal articles, conference papers, and several book chapters. His current research interests include biomedical instrumentation, signal processing, wearable sensors, medical image analysis, machine learning and computer vision, embedded system design, and simultaneous EEG/fMRI. He is currently running several NPRP, UREP, and HSREP grants from Qatar National Research Fund (QNRF) and internal grants (IRCC and HIG) from Qatar University along with academic projects from HBKU and HMC. He has been involved in EPSRC, ISIF, and EPSRC-ACC grants along with different national and international projects during his tenure at the University of Nottingham. He is a Senior Member of IEEE, and a member of British Radiology, ISMRM, and HBM. He is serving as an Associate Editor for IEEE Access and a Topic Editor and Review Editor for Frontiers in Neuroscience. He has recently won the COVID-19 Dataset Award, AHS Award from HMC and National AI Competition awards for his contribution to the fight against COVID-19.

Amith Khandakar received a B.Sc. degree in electronics and telecommunication engineering from North South University, Bangladesh, and a master's degree in computing (networking concentration) from Qatar University, in 2014. He graduated as the Valedictorian (President Gold Medal Recipient) of North South University. He is an IEEE Senior Member. He is also a certified Project Management Professional and the Cisco Certified Network Administrator. He has 2 patents and published around 60 peer-reviewed journal articles, conference papers, and four book chapters. His current research interests include biomedical instrumentation, wearable sensors, medical image analysis, machine learning and Engineering Education. He is also running UREP grants from QNRF and internal grants from Qatar University.

Tawsifur Rahman is currently working as a research assistant in the machine learning group, at Qatar University. He received his B.Sc. degree in Electrical and Electronic Engineering from the University of Chittagong (Bangladesh), and his M.Sc. degree in Biomedical Engineering from the University of Dhaka (Bangladesh). His research interests include medical image analysis, big clinical data processing, computer vision, and deep learning. He is the author and co-author of about 30+ research journal articles, and a few book chapters on medical imaging, and clinical bio-markers data using artificial intelligence. Thus far, his publications have been cited 1700+ times (Source: Google Scholar). He has received 'An ICT fellowship 2019-2020' from the Information Communication & Technology (ICT) ministry of Bangladesh for his research. Moreover, he and his research team have received the 'Kaggle COVID-19 dataset award' from the Kaggle community.

Sakib Mahmud received his Bachelor of Science (BSc.) degree in Electrical Engineering with Honors from Qatar University (QU) in June 2020 and currently pursuing his Master of Science (MSc.) in Electrical Engineering from the same department. As an undergraduate student, he received the Dean's Award for six semesters during 2016-2019. As an undergraduate student at QU, He has been a part of two UREP projects funded by the QNRF. Upon graduation, he joined QU's master's program to pursue a postgraduate degree in electrical engineering and work as a Graduate Research Assistant (GRA) at the same time, under the same department. As a GRA, and a member of QU Machine Learning Group, he was hired in two QNRF funder NPRP grants and three High Impact grants from QU. Currently, he is hired in a High Impact grant funded by QU. Apart from his research work, he has been an Electrical Engineering Intern in the Electrical Consultant Group (E.C.G.) Qatar Branch, Data Science Intern in Data Glacier and has been selected as an Artificial Intelligence (AI) Intern in AI Ready Academy, a joint training program for Machine Learning and Data Science by Microsoft and ZAKA, among 2000 competitive applicants around the Middle East and North Africa (MENA) region, He has expertise in Electronics Circuitry design, 3D Modeling, Biomedical Signal and Image Processing, Computer Vision, Machine and Deep Learning, and Data Analytics and Visualization in various platforms. Currently, so far, he has published 10 articles in peer reviewed Journals, one conference paper and one patent, many under review, thus accumulating around 60 citations in Google Scholar.