Performance Comparison for COVID-19 Chest X-ray Images Classification based on Different CNNs

WESSAM S. ELARABY[1], AHMED H. MADIAN[1, 2]
[1]Radiation Engineering Department
NCRRT, Egyptian Atomic Energy Authority
Nasr City
CAIRO, EGYPT

[2]Nanoelectronics Integrated System Center (NISC)
Nile University
6th October
CAIRO, EGYPT

Abstract: - Nowadays, the detection of the disease that is called Coronavirus or COVID-19 is essential for the whole world. Scientific researchers have spent significant efforts on better understanding the characteristics of the virus and possible means to prevent diagnose and treat COVID-19. Convolutional neural networks (CNNs), have obtained remarkable results in numerous applications. One of these applications is image classification. Chest radiograph (X-ray) images can be requested for early COVID-19 classification of patients. Hence, this paper makes demonstrates different CNN architectures utilizing Chest radiograph database images for COVID-19 detection (Kaggle's X-ray chest images). It contains three different classes of images: 1) COVID-19, 2) normal, and 3) viral pneumonia Chest radiograph images. Therefore, three alternative CNN architectures like SqueezeNet, GoogleNet, and ResNet 50 have been realized using Matlab 2019a and numerical simulation has been performed. GoogleNet has achieved good performance based on the accuracy obtained with a value of 97.02% and it saves time-consuming. A performance comparison between different techniques has been carried out and this comparison shows that the detection is accurate enough for the non-uniform structure of the chest radiograph images.

Key-Words: - Transfer Learning, Deep Learning, Image Classification, Medical Images, COVID-19

Received: September 22, 2021. Revised: October 8, 2022. Accepted: November 10, 2022. Published: December 6, 2022.

1 Introduction

By the end of 2019, a new virus has arisen in the Wuhan region of China that, at that point, spread from one side of the planet to the other, [1]. The World Health Organization (WHO) gave a name to this new virus "COVID-19",[2]. According to the World Meter website, on 10 August 2022, 584,065,952 cases of COVID-19 have been reported all around the world, including 6,418,958 deaths. Covid-19 (Coronavirus) emerges from the very family that's the Middle East Respiratory Syndrome (MERS-CoV) virus that has been within 2015 and Severe Acute Respiratory Syndrome (SARS-CoV) that has been within 2003, [3]. The initial side effects habitually in patients are fever , cough, weariness, and pain in a muscle or group of muscles, [4], [5]. From the patients, the side effects found by the specialists for COVID-19 show up in pneumonia and Chest radiograph tests. Chest radiograph images play a crucial part in the diagnosis of COVID-19. The chest radiograph is like Chest-Computed Tomography (CT), particularly in intensely influenced regions where the detection delay is critical and the resources are limited. CT scans have a few restrictions: 1) the time for image acquisition, 2) the related expense, and 3) the accessibility of CT devices. On the other hand, radiograph images have

numerous benefits such as availability and faster diagnostic instruments. X-ray imaging is likewise less expensive, and patients are less harmed than CTs because of the radiation during the acquisition process. Also, there are portable X-ray gadgets; X-rays can diminish the danger of tainting contrasted with CT for suspects where the person can spread the disease in the transport path, [6]. Therefore, Chest Radiograph (CXR) images can be beneficial to reveal COVID-19 cases. COVID-19-positive cases don't differ too much from the other cases of viral pneumonia and it may overlap with other kinds of lung diseases. So, distinguishing COVID-19 from another type of disease that affects the lungs is difficult and it is necessary to make a diagnosis as soon as possible. A radiologist can make a wrong diagnosis due to working nonstop and the wrong diagnosis can expose a positive COVID-19 case to healthy people. It can increase the number of people infected and overload the hospitals. Also, the radiologist can be infected by the patients, so they need some more intelligent techniques without interfering with patients. Due to these reasons, it is possible to use Artificial Intelligence (AI), to make a diagnosis with high accuracy, and help with: (1) better and faster diagnosis, and (2) help unload the hospitals, [7].

The use of customary machine learning methods for medical image classification started quite a while ago. As they have disadvantages: 1) the performance is far from the practical standard and the improvement of them is highly slow. 2) The feature extraction and selection can vary according to different objects and it is time-consuming [8]. Deep Learning (DL) is really a subfield of machine learning, as shown in Fig.1; it can perform undertakings such as image classification and recognition, by exposing multilayered neural networks to a vast amount of data. On DL, CNNs are a deep Artificial Neural Network (ANN) category. It may be utilized for many applications, like image classification and object recognition. It makes an improvement in image classification since 2012, [9]. With the quick advancement of CNNs, numerous robust architectures have risen up, like AlexNet, [10], GoogleNet, [11], Inception V3, [12], Residual Network (ResNet) 50, and ResNet 18 [13]. Deep learning is one of the successful algorithms that demonstrated its capability in diagnostic imaging with high accuracy.

Recent Studies have used deep learning for chest radiograph images to reveal COVID-19. In [14] used a dataset of 5071 chest radiograph images, 71 images for then Covid-19, and 5000 images for non-COVID-19. It trained four rife DL models including ResNet 18, ResNet 50, DenseNet -121, and SqueezeNet , to reveal Coronavirus. The foremost performing model carried out sensitivity rate of 97.5% and specificity rate of 95%.

The performance evaluation in [15] was done on several cutting-edge CNNs architectures utilizing two datasets. The very first dataset encompasses 1427 chest radiograph images comprising 224 images for Covid-19, 504 images of normal, and 700 images for bacterial pneumonia cases. The next dataset comprises 224 images for Covid-19, 714 images for viral pneumonia and bacterial, besides 504 images of normal cases. As per their outcomes, the best achievement for the second dataset using Mobile Net v2 with sensitivity, accuracy, and specificity: is 98.66%, 96.78%, and 96.46% respectively.

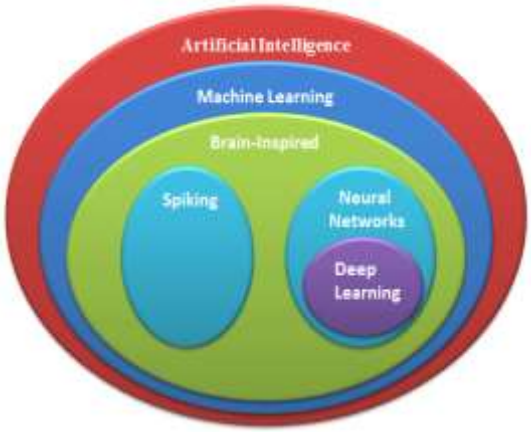

Figure 1: Deep learning is a subfield of machine learning

The DarkCovidNet Model was proposed in [7] for distinguishing proof of COVID-19 from chest radiograph images. The dataset used included 127 chest radiograph images of COVID-19 besides 500 Normal and 500 pneumonia chest radiograph images. Their model created a classification accuracy of 98.08% for COVID-19 versus Normal (binary classes) and 87.02% for COVID-19 versus Normal versus Pneumonia (multi-class cases).

In [16], different CNN architectures were compared: MobileNetV2, ResNet50V2, ResNet152V2, Xception, VGG16, and DenseNet12 to reveal ground glass opacities (GGO) present in radiograph images of COVID-19 patients centered on accuracy. The dataset was collected from IEEE 8023 and Kaggle CXR Pneumonia Dataset. The best result is MobileNetV2 which provides 91.28% accuracy.

Moreover in [17], the performance of various CNN architectures was measured on Chest

radiograph images for COVID-19 recognition with the general public dataset Kaggle's COVID-19 Radiography dataset. The CNN architectures used were InceptionResNetV2, VGG16, DesNet-201, Xception, ResNet 50, and ResNet-101. The best result is Xception Net which provides an accuracy value of 97.34%.

The main contribution of this paper is to analyze the performance of various CNN architectures on Chest radiograph images for COVID-19 recognition to get the best degree of diagnostic accuracy. The dataset utilized in this paper encompasses three different classes of images: Coronavirus, normal , and viral pneumonia Chest radiograph images. The CNN architectures used are SqueezeNet, GoogleNet, and ResNet 50. The performance evaluation is measured by Area Under the Curve (AUC), precision, specificity, precision, accuracy, recall, F-score, sensitivity, and G-mean. The best result was obtained from GoogleNet with an accuracy value of 97.02%. Fig. 2 shows samples of the dataset utilized.

(a) (b) (c)

Figure 2: (a) COVID-19, (b) Normal, and (c) Viral Pneumonia are samples from the dataset

2 Methods

Analysts in the image domain faced many challenges, one of these challenges is named training data that may possibly not be accessible from a specific application. In classification applications, to do classification on a dataset with a specific group of labels could be bounded in availability and distinctive from enormous resources like ImageNet . It could potentially cause issues because neural networks require great training data to construct from scratch without any preparation. A central issue about image data is the extricated features from a specific dataset. For instance, the ImageNet dataset contains in excess of 1,000,000 images drawn from 1000 classes experienced in regular day-to-day existence. The extracted features from the ImageNet dataset are utilized to represent an entirely different image dataset by going it via a pre-trained CNN and taking out the multidimensional features from the fully connected layers, [18]. Some of the convolution architectures are well-known models that can be utilized for image classification by designing training algorithms. The subsequent features can be utilized for applications like classification. This is called transfer learning. It is a machine-learning technique that depends on the idea of reusability. It is executed utilizing two steps: feature extraction and parameter tuning (optimization strategy). Throughout the feature extraction, the pre-trained model hold the new features from the dataset using the training data. Secondly, to improve the performance of a model in the applied domain, the model architecture needs to be rebuilt and updated along with the parameter tuning. This technique could be particularly ideal for medical applications since it doesn't need as much training data, which may be difficult to obtain in medical situations. In the analysis of medical data, the limited number of available datasets is among the biggest difficulties faced by researchers. Deep learning models regularly need plenty of data. Experts label this data, but it is costly and time-consuming. The greatest benefit of utilizing the transfer learning method is that it allows the training of data with fewer datasets and needs fewer calculation costs, [19].

First of all, the preprocessed procedures used are explained as follows: the data is enhanced to add more images to the latter blocks and emphasize specific aspects to be detected. This is accomplished by rotating, shearing, and translating images. The images may have different sizes, so the augmented image is used to automatically resize the training images. Data augmentation is used to extract more data from the dataset to make perturbed duplicates of the existing photos and helps prevent the network from overfitting. The objective is to strengthen the neural network with different diversities. Several ways may be used to enhance images. Our approach combines a variety of strategies to support a sizable dataset for various situations, as follows:

- Randomly flip the training images along the vertical axis and randomly translate them up to 30 pixels.
- Scale them up to 10% horizontally and vertically.

Deep learning (CNN) has gained success in computer vision for medical imaging. CNN has been applied to many medical image classification problems due to its successful extraction of image features. There are several types of deep CNNs including Visual Geometry Group Network (VGG-

Net); SqueezeNet; Residual Network (ResNet); Dense Convolutional Network (DenseNet); GoogleNet; Inception, and Xception.

SqueezeNet was proposed by [20]. It is a little CNN architecture. In the fire module, as shown in Fig. 3 (a), the incoming data squeezes a 1x1 layer alternately in the vertical dimension followed up by two parallel 1x1 and 3x3 convolutional layers that "expand" the profundity of the data once more. It starts 18 layers deep by having an autonomous convolution layer, followed closely by 8 Fire modules, finishing with the last Conv layer as shown in Fig. 3 (b). It progressively increases how many filters per fire module from the start to the termination of the network.

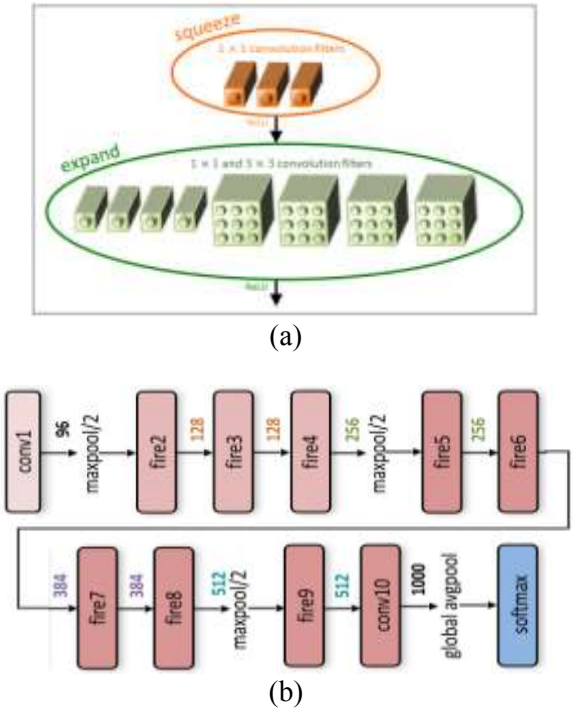

Figure 3: (a) Fire module in SqueezeNet, (b) SqueezeNet architecture [20]

GoogleNet was presented by [11] and was the victor of ILSVRC in 2014. GoogleNet has 7,000,000 parameters; it encompasses four convolutional layers, nine inception modules, three average pooling layers, five fully-connected layers, four max-pooling layers, and three softmax layers for the fundamental classifiers. It likewise utilizes dropout regularization in the fully-connected layers and applies Rectified Linear Unit (ReLU) activation in the whole of the convolutional layers. This network is a lot more profound and wider with 22 all-out layers.

The Residual Network (ResNet) models were developed by [13]. It relies upon deep architectures that have shown good coming-together behaviors and convincing accuracy. In 2015, they triumphed first place in the ILSVRC and Common Objects in Context (COCO) classification challenge . ResNet comprises a number of stacked residual units and was developed with a large variety of layers: 18, 34, 50, 101, 152, and 1202. The residual units are convolutional, pooling, and layers. ResNet and Visual Geometry Group (VGG) net are similar, but ResNet is about eight times more profound than VGG. The ResNet 18 presents a great compensation between the profundity and performance, and it is made out of one average pooling, five convolutional layers, and a fully connected layer with a softmax. The network of ResNet 50 consists of forty-nine convolutional layers and a fully connected layer. At last, ResNet 50 was picked for the development benefit of saving computing resources and training time, and its architecture is shown in Fig. 4.

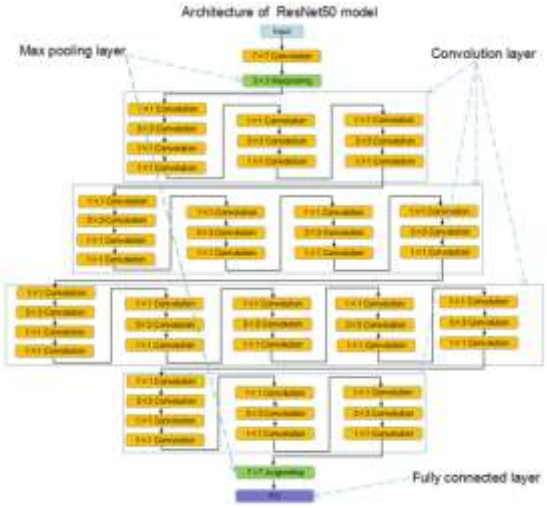

Figure 4: ResNet50 architecture, [21]

3 Results and Discussion

In this paper, the COVID-19 chest radiograph database is used, [22], [23]. This database evolved from the Italian Society of Medical and Interventional Radiology (SIRM) COVID-19 Database . It incorporates 219 COVID-19-positive images, 1341 normal images, and 1345 viral pneumonia images. Portable Network Graphics (PNG) with a resolution of 1024*1024 pixels, is the image file format.

The training process was done by Matlab 2019a. The laptop specifications are: Processor is Intel core i7-4510U CPU @ 2.60 GHz, memory is 8GB, and operating system (OS) is Windows 8.1, 64

bits. Fig. 5 shows the flowchart for Images Classification based on different CNN architectures. First of all, the preprocessed procedures used are explained as follows: the data is enhanced to add more images to the latter blocks and emphasize specific aspects to be detected. This is accomplished by rotating, shearing, and translating images. The images may have different sizes, so the augmented image is used to automatically resize the training images. Data augmentation is used to extract more data from the dataset to make perturbed duplicates of the existing photos and helps prevent the network from overfitting. Then, prepare the pre-trained network for transfer learning. Finally, train the network and measure the performance.

The performance measures were assessed by assembling a comparison of the pre-trained models as shown in Table 1 and Fig. 6. Receiver Operating Characteristic Curve (ROC) is famed as Area Under the Curve (AUC), which is vastly utilized for performance measurement in the supervised classification. The AUC is used for classifying diagnostic test accuracy, which simplifies the comparison between diagnostic tests. It is based on the relation between the false positive rate (FPR) (X-axis) and the true positive rate (TPR) (Y-axis), [24]. Fig. 7 shows the confusion matrices.

The accuracy of positive examples can be measured by sensitivity/recall. It refers to the positive classes labeled correctly. This is shown in equation 1, where true positive (TP) is the number of instances that are effectively identified, and false negative (FN) is the number of positive cases which is negative since it doesn't classify correctly [9].

$$Sensitivity\ (Recall\ ) = \frac{TP}{TP+FN} \quad (1)$$

Specificity is the probability of true negative given in the second class. The probability of the negative label approximates to be true as in equation 2. True negative (TN) is negative and classified as negative, and false positive (FP) is the negative instances classified incorrectly as positive cases. In general, the effectiveness of the algorithm on a single class is assessed by sensitivity and specificity, positive and negative respectively, [9].

$$Specificity = \frac{TN}{TN+FP} \quad (2)$$

The performance of the classification was assessed by calculating the accuracy. It calculates the percentage of correct samples as shown in equation 3, [9].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Precision assesses the predictive power of the algorithm. Precision is the number of the predicted positive that is actually correct and how to "precise" the model is out of them as in equation 4, [9].

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

F-score concentrates on the analysis of positive class. It combines precision and recall and gets a higher value as shown in equation 5. The model performs better on the positive classes by a high value [9].

$$F_{score} = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (5)$$

G-mean is defined as the square root of the accuracy of both classes as shown in equation 6. As the G-mean value is higher, the learning algorithm is better in both classes.

$$G_{mean} = \sqrt{Sensitivity * Specificity} \quad (6)$$

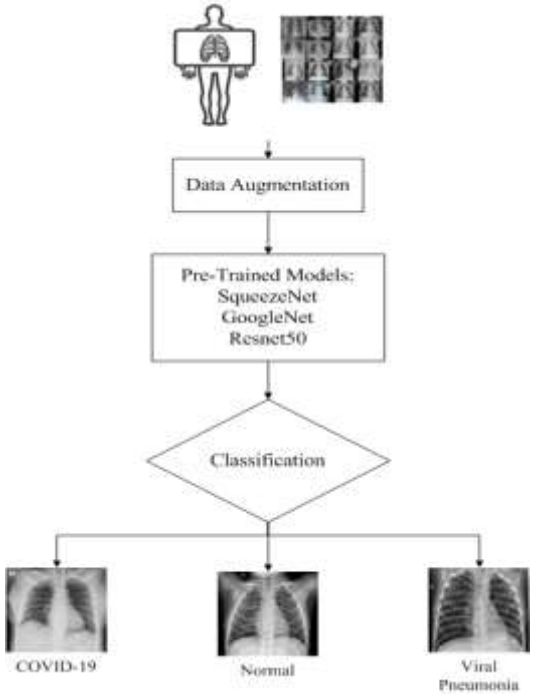

Figure 5: Images classification based on different CNN architectures

The dataset was split into two groups; 70% of it is for training the model and 30% for validation. In Table 1, the best results from two pre-trained models, GoogleNet and ResNet 50. The ResNet 50 takes more time than GoogleNet, but the result obtained from the GoogleNet with the highest accuracy value of 97.02%. It is preferred to use GoogleNet to save computational time and get the best results for COVID-19 detection.

Table 1: Performance measures for every pre-trained model

Performance Measures	SQUEEZENET	GOOGLENET	RESNET50
Area under the curve (AUC)	99.9%	99.87%	99.97%
Accuracy	95.76%	97.02%	96.22%
Sensitivity	92.42%	96.97%	98.48%
Specificity	96.03%	97.02%	96.03%
Precision	65.59%	72.73%	67.01%
Recall	92.42%	96.97%	98.48%
F-score	76.73%	83.12%	79.75%
G-mean	94.21%	97%	97.25%
Time	224 min 53sec	572 min 29 sec	1422 min 38 sec

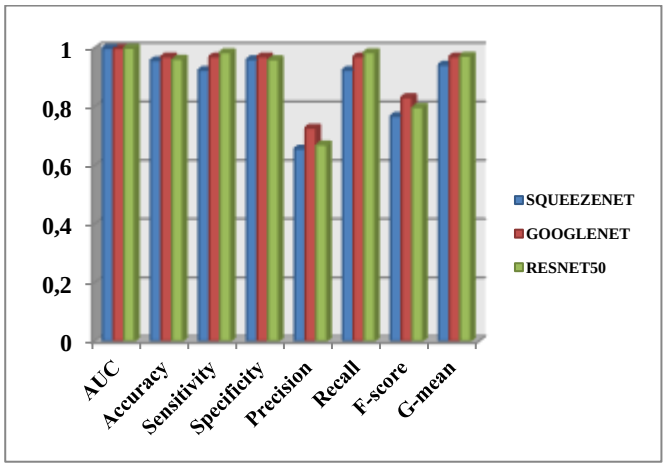

Figure 6: Performance evaluation measures

COVID-19	61	0	1
NORMAL	2	393	22
VIRAL PNEUMONIA	3	9	381
Output Class/Target Class	**COVID-19**	**NORMAL**	**VIRAL PNEUMONIA**

(a)

COVID-19	64	0	5
NORMAL	1	397	14
VIRAL PNEUMONIA	1	5	385
Output Class/Target Class	**COVID-19**	**NORMAL**	**VIRAL PNEUMONIA**

(b)

COVID-19	65	1	1
NORMAL	0	389	18
VIRAL PNEUMONIA	1	12	385
Output Class/Target Class	**COVID-19**	**NORMAL**	**VIRAL PNEUMONIA**

(c)

Figure 7: Confusion matrix for the three pre-trained models used: (a) for the SqueezeNet, (b) for the GoogleNet, and (c) for the ResNet50

4 Conclusion and Future Work

Since there isn't a particular treatment for COVID-19 now, then the early detection of the disease is necessary for forestalling the spread of the pandemic. The DL with CNNs may have effectively affected the automatic detection and extraction of fundamental features from radiograph images, linked to the diagnosis of Covid-19. This paper makes a comparison between three different pre-trained models to detect COVID-19 by utilizing chest radiograph images to get more reliable diagnostic performance. The performance evaluation is measured by AUC, specificity, accuracy, precision, sensitivity, recall, F-score, and

G-mean. The best result was obtained from GoogleNet with an accuracy value of 97.02% and also it saves more time than ResNet 50.

In the Future, research can include adding more data, but not be restricted to x-ray images. Moreover, to increase the detection power, COVID - 19 diagnosis with sonography (lung ultrasound) combined with radiography can be used as ultrasound frequency analysis using acoustic models would be good enough in identifying COVID-19 presence. It is additionally important to develop models that can distinguish Covid-19 cases from other comparable viral cases, like SARS.

References:

[1] Moscoviz, L., and Evans, D. K. Learning loss and student dropouts during the covid-19 pandemic: A review of the evidence two years after schools shut down. *Center for Global Development, Working Paper*, *609*, 2022.

[2] Abbas, A., Abdelsamea, M. M., and Gaber, M. M., Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network, *arXiv preprint arXiv:2003.13815,* 2020.

[3] Narin, A., Kaya, C., and Pamuk, Z., Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks, *arXiv preprint arXiv:2003.10849,*2020.

[4] Kaklauskas, A., Abraham, A., and Milevicius, V., Diurnal emotions, valence and the coronavirus lockdown analysis in public spaces, *Engineering Applications of Artificial Intelligence*, Vol. 98, 2021, pp. 104122.

[5] Ren, L. L., et al., Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study, *Chinese Medical Journal,* Vol. 133, No. 9, 2020, pp. 1015.

[6] Ahishali, M., Degerli, A., Yamac, M., Kiranyaz, S., Chowdhury, M. E., Hameed, K., Hamid, T., Mazhar, R., and Gabbouj, M., A Comparative Study on Early Detection of COVID-19 from Chest X-Ray Images, *arXiv preprint arXiv:2006.05332,* 2020.

[7] Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., and Acharya, U. R., Automated detection of COVID-19 cases using deep neural networks with X-ray images, *Computers in Biology and Medicine*, Vol. 121, 2020, pp. 103792.

[8] Yadav, S. S., and Jadhav, S. M., Deep convolutional neural network based medical image classification for disease diagnosis, *Journal of Big Data*, Vol. 6, No. 1, 2019.

[9] Maeda-Gutiérrez, V., Galván-Tejada, C. E., Zanella-Calzada, L. A., Celaya-Padilla, J. M., Galván-Tejada, J. I., Gamboa-Rosales, H., Luna-García, H., Magallanes-Quintanar, R., Méndez, C. A. G., and Olvera-Olvera, C. A., Comparison of Convolutional Neural Network Architectures for Classification of Tomato Plant Diseases, *Applied Sciences*, Vol. 10, No. 4, 2020, pp. 1245.

[10] Krizhevsky, A., Sutskever, I., and Hinton, G. E., Imagenet classification with deep convolutional neural networks, *In Advances in Neural Information Processing Systems; Curran Associates,* Inc.: New York, NY, USA, 2012, pp. 1097–1105.

[11] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., Going deeper with convolutions, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston,* MA, USA, 2015, pp. 1–9.

[12] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., Rethinking the inception architecture for computer vision, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* Las Vegas, NV, USA, 2016, pp. 2818–2826.

[13] He, K., Zhang, X., Ren, S., and Sun, J., Deep residual learning for image recognition, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* Las Vegas, NV, USA, 2016, pp. 770–778.

[14] Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., and Soufi, G. J., Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning, *arXiv preprint arXiv:2004.09363*, 2020.

[15] Apostolopoulos, I. D., and Mpesiana, T. A., Covid-19 : automatic detection from X-ray images utilizing transfer learning with convolutional neural networks, *Physical and Engineering Sciences in Medicine*, Vol. 43, No. 2, 2020, pp. 635–640.

[16] Ammar, C., Purbojo, S., Umitaibatin, R., and Rudiyanto, D., Comparative Study of Convolutional Neural Network Feature Extractors Used for COVID-19 Detection From Chest X-Ray Images, 2020. Doi:10.13140/RG.2.2.15462.24642.

[17] Sisamon, F. L., and Vidarte, R. L., Analysis of Deep Learning Models for COVID-19 Diagnosis from X-Ray Chest Images, 2020. Doi:10.13140/RG.2.2.16482.35520.

[18] Aggarwal, C. C., *Neural Networks and Deep Learning*. Springer, NY, USA, 2018.

[19] Asif, S., and Wenhui, Y., Automatic Detection of COVID-19 Using X-ray Images with Deep Convolutional Neural Networks and Machine Learning, *medRxiv*, 2020. https://doi.org/10.1101/2020.05.01.20088211

[20] Landola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K., SqueezeNet: AlexNet-level Accuracy with 50x fewer parameters and <0.5MB model size. *arXiv preprint arXiv:1602.07360*, 2016.

[21] Peng, J., Kang, S., Ning, Z., Deng, H., Shen, J., Xu, Y., Zhang, J., Zhao, W., Li, X., Gong, W., Huang, J., and Liu, L., Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging, *European Radiology* Vol. 30, No. 1, 2020, pp. 413-424.

[22] Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K. R., Khan, M. S., Iqbal, A., Al-Emadi, N., and Reaz, M. B. I., Can AI help in screening Viral and COVID-19 pneumonia?,. *IEEE Access*, Vol. 8, 2020, pp. 132665-132676. https://ieeexplore.ieee.org/document/9144185

[23] Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S. B. A., Islam, M. T., Maadeed, S. A., Zughaier, S. M., Khan, M. S., and Chowdhury, M. E., Exploring, the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-ray Images, *Computers in Biology and Medicine*, Vol. 132, 2021, pp. 104319. https://doi.org/10.1016/j.compbiomed.2021.104319

[24] Hanley, J. A., and McNeil, B. J., The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, Vol. 143, 1982, pp. 29–36.