Article

Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning

Mohamed Loey [1*], Florentin Smarandache [2], Nour Eldeen M. Khalifa [3]

[1] Department of Computer Science, Faculty of Computers Artificial Intelligence, Benha University, Benha 13511, Egypt; mloey@fci.bu.edu.eg (M.L.)

[2] Department of Mathematics, University of New Mexico, Gallup Campus, NM 87301, USA; smarand@unm.edu (F.L.)

[3] Department of Information Technology, Faculty of Computers & Artificial Intelligence, Cairo University, Cairo 12613, Egypt; nourmahmoud@cu.edu.eg (N.E.M.K.)

*Correspondence Author: mloey@fci.bu.edu.eg.

Abstract: The coronavirus (COVID-19) pandemic is putting healthcare systems across the world under unprecedented and increasing pressure according to the World Health Organization (WHO). With the advances in computer algorithms and especially Artificial Intelligence, the detection of this type of virus in the early stages will help in fast recovery and help in releasing the pressure off healthcare systems. In this paper, a GAN with deep transfer learning for coronavirus detection in chest x-ray images is presented. The lack of datasets for COVID-19 especially in chest x-rays images is the main motivation of this scientific study. The main idea is to collect all the possible images for COVID-19 that exists until the writing of this research and use the GAN network to generate more images to help in the detection of this virus from the available x-rays images with the highest accuracy possible. The dataset used in this research was collected from different sources and it is available for researchers to download and use it. The number of images in the collected dataset is 307 images for four different types of classes. The classes are the COVID-19, normal, pneumonia bacterial, and pneumonia virus. Three deep transfer models are selected in this research for investigation. The models are the Alexnet, Googlenet, and Restnet18. Those models are selected for investigation through this research as it contains a small number of layers on their architectures, this will result in reducing the complexity, the consumed memory and the execution time for the proposed model. Three case scenarios are tested through the paper, the first scenario includes 4 classes from the dataset, while the second scenario includes 3 classes and the third scenario includes 2 classes. All the scenarios include the COVID-19 class as it is the main target of this research to be detected. In the first scenario, the Googlenet is selected to be the main deep transfer model as it achieves 80.6% in testing accuracy. In the second scenario, the Alexnet is selected to be the main deep transfer model as it achieves 85.2% in testing accuracy, while in the third scenario which includes two classes(COVID-19, and normal), Googlenet is selected to be the main deep transfer model as it achieves 100% in testing accuracy and 99.9% in the validation accuracy. All the performance measurement strengthens the obtained results through the research.

Keywords: 2019 novel coronavirus, Deep Transfer Learning, Machine Learning, COVID-19, SARS-CoV-2, Convolutional Neural Network, GAN.

1. Introduction

In 2019, Wuhan is a commercial center of Hubei province in China that faced a flare-up of a novel 2019 coronavirus that killed more than hundreds and infected over thousands of individuals within the initial days of the novel coronavirus pestilence. The Chinese researchers named The novel virus

as 2019 novel coronavirus (2019-nCov) or Wuhan-virus [1]. The International Committee of Viruses titled the virus of 2019 as Severe Acute Respiratory Syndrome CoronaVirus-2 (SARS-CoV-2) and the malady as Coronavirus disease 2019 (COVID-19) [2–4]. The subgroups of coronaviruses family are alpha-CoV (α), beta-CoV (β), gamma-CoV (δ), and delta-CoV (γ) coronavirus. SARS-CoV-2 was announced to be an organ of the beta-CoV (β) group of coronaviruses. In 2003, the Kwangtung people were infected with a 2013 virus lead to Severe Acute Respiratory Syndrome (SARS-CoV). SARS-CoV was assured as a family of the beta-CoV (β) subgroup and was title as SARS-CoV [5]. In the history, SRAS-CoV across 26 countries in the world infected more than 8000 individuals with death-rate of 9%, moreover, SARS-CoV-2 infected more than 750,000 individuals with death-rate of 4%, across 150 states, till the date of this lettering. It demonstrates that the broadcast rate of SARS-CoV-2 is higher than SRAS-CoV. The transmission ability enhanced because of authentic recombination of S protein in the RBD region[6].
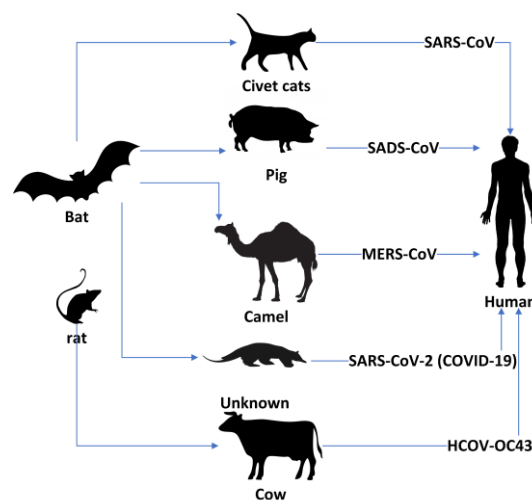

Fig. 1. Coronavirus transmission from animals to human

Beta-coronaviruses have caused malady to people that have had wild animals generally either in bats or rats [7,8]. SARS-CoV-1 and MERS-CoV (camel flu) were transmitted to people from some wild cats and Arabian camels respectively as shown in Figure 1. The sale and buy of unknown animals may be the provenance of coronavirus infection. The invention of the various progeny of pangolin coronavirus and their propinquity to SARS-CoV-2 suggests that pangolins should be a thinker as possible hosts of novel 2019 coronaviruses. Wild animals must be taken away from wild animal markets to stop animal corona-virus transmission [9]. Coronavirus transmission has been assured by World Health Organization (WHO) and by The Centers for Diseases of the US, with evidence of human-to-human conveyance from five different cases outside China, namely in Italy [10], US [11], Nepal [12], Germany [13], and Vietnam [14]. On 31 March 2020, SARS-CoV-2 Confirmed more than 750000 cases, 150000 recovered cases, and 35000 death cases. Table 1 show some statistics about SARS-CoV-2 [15].

Table 1. SARS-CoV-2 statistics in some countries

Location	Confirmed	Recovered	Deaths
United States	164,345	5,945	3,171
Italy	101,739	14,620	11,591
Spain	94,417	19,259	8,269
China	81,518	76,052	3,305
Germany	67,051	7,635	682
Iran	44,606	14,656	2,898
France	43,973	7,202	3,018
United Kingdom	22,141	135	1,408

Table 2. Major Contributions in the history of the neural network to deep learning [21,22]

Milestone/contribution	Year
McCulloch-Pitts Neuron	1943
Perceptron	1958
Backpropagation	1974
Neocognitron	1980
Boltzmann Machine	1985
Restricted Boltzmann Machine	1986
Recurrent Neural Networks	1986
Autoencoders	1987
LeNet	1990
LSTM	1997
Deep Belief Networks	2006
Deep Boltzmann Machine	2009

1.1 Deep learning

Nowadays, Deep Learning (DL) is a subfield of machine learning concerned with techniques inspired by neurons of the brain [16]. Today, DL is quickly becoming a crucial technology in image/video classification and detection. DL depends on algorithms for reasoning process simulation and data mining, or for developing abstractions [17]. Hidden deep layers on DL maps inputs data to labels to analyze hidden patterns in complicated data [18]. Besides their use in medical x-ray recognition, DL architectures are also used in other areas in the application of image processing and computer vision in medical. DL improves such a medical system to realize higher outcomes, widen illness scope, and implementing applicable real-time medical image [19,20] disease detection systems. Table 2 shows a series of major contributions in the field of the neural network to deep learning [21].

1.2 Generative Adversarial Network

Generative Adversarial Network (GAN) is a class of deep learning models invented by Ian Goodfellow in 2014 [23]. GAN models have two main networks called the generative network and

discriminative network. The first neural network is the generator network, responsible for generating new fake data instances that look like training data. The discriminator tries to distinguish between real data and fake (artificially generated) data generated by the generator network as shown in Figure 2. The mission GANs models that generator network is to try fooling the discriminator network and the discriminator network tries to fight from being fooled [24–27].
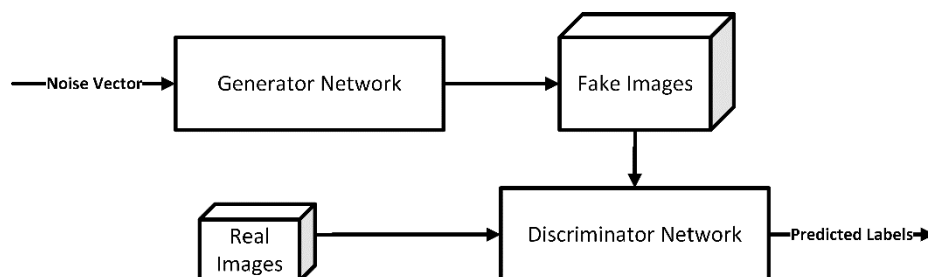

Fig. 2. Generative Adversarial Network model

1.3   Convolution Neural Networks

Convolutional Neural Networks (ConvNets or CNNs) are a category of deep learning techniques used primarily to recognize and classify the image. Convolutional Neural Networks have accomplished extraordinary success for medical image/video classification and detection. In 2012, Ciregan et al. and Krizhevsky and et al. [28,29] showed how CNNs based on Graphics Processing Unit (GPU) can enhance many vision benchmark records such as MNIST [30], Chinese characters [31], Arabic digits recognition [32], Arabic handwritten characters recognition [33], NORB (jittered, cluttered) [34], traffic signs [35], and large-scale ImageNet [36] benchmarks. In the following years, various advances in ConvNets further increased the accuracy rate on the image detection/classification competition tasks. ConvNets pre-trained models introduced significant improvements in succeeding in the annual challenges of ImageNet Large Scale Visual Recognition Competition (ILSVRC). Deep Transfer Learning (DTL) is a deep learning (DL) model that focuses on storing weights gained while solving one image classification and applying it to a related problem. Many DTL models were introduced like VGGNet [37], GoogleNet [38], ResNet [39], Xception [40], Inception-V3 [41] and DenseNet [42].

The novelty of this paper is conducted as follows: i) The introduced ConvNet models have end-to-end structure without classical feature extraction and selection methods. ii) We show that GAN is an effective technique to generate x-ray images. iii) Chest X-ray images are one of the best tools for the classification of SARS-CoV-2. iv) The deep transfer learning models have been shown to yield very high outcomes in the small dataset COVID-19. The rest of the paper is organized as follows. In section 2, explores related work and determines the scope of this works. In section 3, discusses the dataset used in our paper. In section 4, presents the proposed models, while section 5 illustrates the achieved outcomes and its discussion. Finally, section 6 provides conclusions and directions for further research.

2.   Related works

This part conducts a survey on the recent scientific researches for applying machine learning and deep learning in the field of medical pneumonia and coronavirus x-ray classification. Classical image classification stages can be divided into three main stages: image preprocessing, feature extraction, and feature classification.   Stephen et al. [43] proposed a new study of classifying and detect the presence of pneumonia from a collection of chest X-ray image samples based on a ConvNet model trained from scratch based on dataset [44]. The outcomes obtained were training loss = 12.88%, training accuracy = 95.31%, validation loss= 18.35%, and validation accuracy= 93.73%.

In [45], the Authors introduced an early diagnosis system from Pneumonia chest X-ray images based on Xception and VGG16. In this study, a database containing approximately 5800 frontal chest X-ray images introduced by Kermany et al [44] 1600 normal case, 4200 up-normal pneumonia case in the Kermany x-ray database. The trial outcomes showed that VGG-16 network better than X-ception network with a classification rate of 87 %. Forasmuch X-ception network better than VGG-16 network by sensitivity 85%, precision 86% and recall 94%. X-ception network is more felicitous for classifying x-ray images than VGG-16 network. Varshni et al. [46] proposed pre-trained ConvNet models (VGG-16, Xception, Res50, Dense-121, and Dense-169) as feature-extractors followed by different classifiers (SVM, Random Forest, k-nearest neighbors, Naïve Bayes) for the detection of normal and abnormal pneumonia X-Rays images. The prosaists used ChestX-ray14 introduced by Wang et al. [47].

Chouhan et al. [48] introduced an ensemble deep model that combines outputs from all transfer deep models for the classification of pneumonia using the connotation of deep learning. The Guangzhou Medical Center [44] database introduced a total of approximately 5200 x-ray images, divided to 1300 x-ray normal, 3900 x-rays abnormal. The proposed model reached a miss-classification error of 3.6% with a sensitivity of 99.6% on test data from the database. In paper [49]   proposed a Compressed Sensing (CS) with a deep transfer learning model for automatic classification of pneumonia on X-ray images to assist the medical physicians. The dataset used for this work contained approximately 5850 X-ray data of two categories (abnormal /normal) obtained from Kaggle. Comprehensive simulation outcomes have shown that the proposed approach detects the classification of pneumonia (abnormal /normal) with 2.66% miss-classification.

In this research, we introduced a transfer of deep learning models to classify COVID-19 x-ray images. To input adopting x-ray images of the chest to the convolutional neural network, we embedded the medical x-ray images using GAN to generate x-ray images. After that, a classifier is used to ensemble the outputs of the classification outcomes. The proposed transfer model was evaluated on the proposed dataset.

3.   Dataset

The COVID-19 dataset [50] utilized in this research [51] was created by Dr. Joseph Cohen, a postdoctoral fellow at the University of Montreal. The Pneumonia [44] dataset Chest X-Ray Images was used to build the proposed dataset. The dataset [52] is organized into 2 folders (train, test) and contains sub-folders for each image category (COVID-19/normal/pneumonia bacterial/ pneumonia virus). There are 306 X-Ray images (JPEG) and 4 categories (COVID-19/normal/pneumonia bacterial/ pneumonia virus). The number of images for each class is presented in Table 3. Figure 3 illustrates

samples of images used for this research. Figure 4 also illustrates that there is a lot of variation of image sizes and features that may reflect on the accuracy of the proposed model which will be presented in the next section.

Table 3. Number of images for each class in COVID-19 dataset

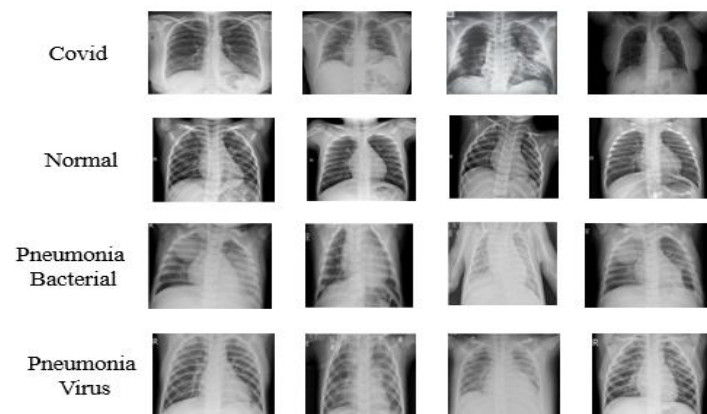| Dataset/ Class	covid	normal	pneumonia_bac	pneumonia_vir	total
Train	60	70	70	70	**270**
Test	9	9	9	9	**36**
Total	**69**	**79**	**79**	**79**	**306**

Fig. 3. Samples of the used images in this research

4.   The Proposed Model

The proposed model includes two main deep learning components, the first component is GAN and the second component is the deep transfer model. Figure 4 illustrates the proposed GAN/Deep transfer learning model. Mainly, the GAN used for the preprocessing phase while the deep transfer model used in the training, validation and testing phase.
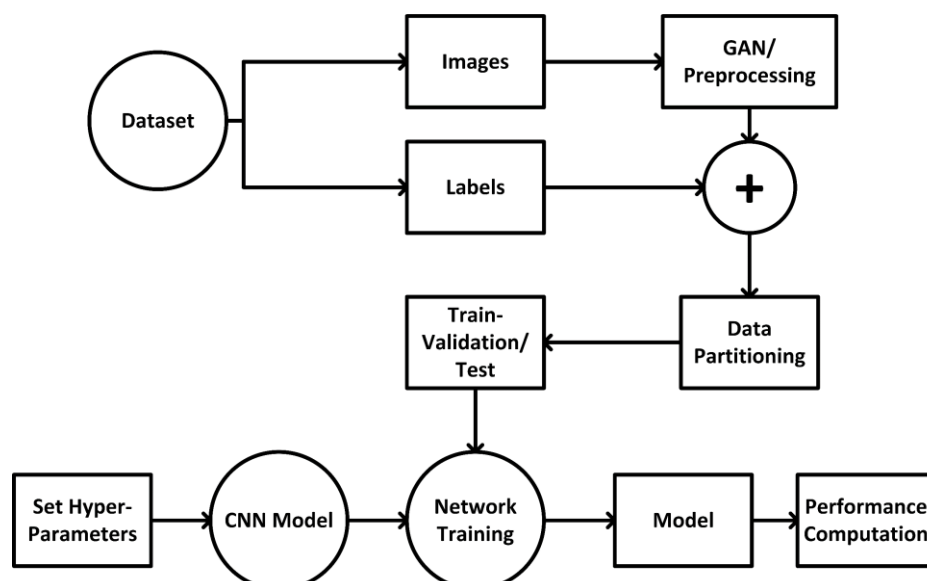
Fig. 4. The proposed GAN/Deep transfer learning mode

Algorithm 1 introduces the proposed transfer model in detail. Let $D = \{$Alexnet, Googlenet, Resnet18$\}$ be the set of transfer models. Each deep transfer model is fine-tuned with the COVID-19 x-ray Images dataset (X, Y); where X the set of N input data, each of size, 512 lengths × 512 widths, and Y have the identical class, $Y = \{y/y \in \{$ COVID-19; normal; pneumonia bacterial; pneumonia virus $\}\}$. The dataset divided to train and test, training set $(X_{\text{train}}; Y_{\text{train}})$ for 90% percent for the training and then validation while 10% percent for the testing. The 90% percent was divided into 80% for training and 20% for the validation. The selection of 80% for the training and 20% in the validation proved it is efficient in many types of research such as [53–57]. The training data then divided into mini-batches, each of size $n = 64$, such that $(X_q; Y_q) \in (X_{\text{train}}; Y_{\text{train}})$; $q = 1, 2, \dots, \frac{N}{n}$ and iteratively optimizes the DCNN model $d \in D$ to reduce the functional loss as illustrated in equation (1).

$$C(w, X_i) = \frac{1}{n} \sum_{x \in X_i, y \in Y_i} c(d(x, w), y) \qquad (1)$$

where $d(x, w)$ is the ConvNet model that true label y for input x given w is a weight and $c(.)$ is the multi-class entropy loss function.

This research relied on the deep transfer learning CNN architectures to transfer the learning weights to reduce the training time, mathematical calculations and the consumption of the available hardware resources. There are several types of research in [53,58,59] tried to build their architecture, but those architecture are problem-specific and can't fit the data presented in this paper. The used deep transfer learning CNN models investigated in this research are Alexnet [29], Resnet18 [39], Googlenet [60], The mentioned CNN models had a few numbers of layers if it is compared to large CNN models such as Xception [40], Densenet [42], and Inceptionresnet [61] which consist of 71, 201 and 164 layers accordingly. The choice of these models will reflect on reducing the training time and the complexity of the calculations.

Algorithm 1 Introduced Algorithm

Input data: COVID-19 Chest x-ray Images (X, Y); where $Y = \{y/y \in \{$COVID-19; normal; pneumonia bacterial; pneumonia virus $\}\}$

Output data: The transfer model that detected the COVID-19 Chest x-ray image $x \in X$

Pre-processing steps:

- modify the x-ray input to dimension 512 height × 512 width
- Generate x-ray images using GAN
- Mean normalize each x-ray data input

download and reuse transfer models D= $\{$Alexnet, Googlenet, Resnet18$\}$

Replace the last layer of each transfer model by (4×1) layer dimension.

foreach $\forall d \in D$ **do**

    $\mu = 0.01$

    for epochs = 1 to 20 **do**

        foreach mini-batch $(X_i; Y_i) \in (X_{\text{train}}; Y_{\text{train}})$ do

Modify the coefficients of the transfer $d(\cdot)$

if the error rate is increased for five epochs **then**

$\mu = \mu \times 0.01$

end

end

end

end

foreach $\forall x \in X_{test}$ **do**

the outcome of all transfer architectures, $d \in D$

end

4.1 Generative Adversarial Network

GANs consist of two different types of networks. Those networks are trained simultaneously. The first network is trained on image generation while the other is used for discrimination. GANs are considered a special type of deep learning models. The first network is the generator, while the second network is the discriminator. The generator network in this research consists of 5 transposed convolutional layers, 4 ReLU layers, 4 batch normalization layers, and Tanh Layer at the end of the model, while the discriminator network consists of 5 convolutional layers, 4 leaky ReLU, and 3 batch normalization layers. All the convolutional and transposed convolutional layers used the same window size of 4*4* pixel with 64 filters for each layer. Figure 5 presents the structure and the sequence of layers of the GAN network proposed in this research.

Generator Network		Discriminator Network	
	Input		Input
	Transposed Convolution 1		Convolution 1
	Batch Normalization 1		Leaky ReLU 1
	ReLU 1		Convolution 2
	Transposed Convolution 2		Batch Normalization 2
	Batch Normalization 3		Leaky ReLU 2
	ReLU 2		Convolution 3
	Transposed Convolution 3		Batch Normalization 3
	Batch Normalization 3		Leaky ReLU 3
	ReLU 3		Convolution 4
	Transposed Convolution 4		Batch Normalization 4
	Batch Normalization 4		Leaky ReLU 4
	ReLU 4		Convolution 5
	Transposed Convolution 5		
	Tanh		

Fig. 5. The structure and the sequence of layers for the proposed GAN network.

The GAN network helped in overcoming the overfitting problem caused by the limited number of images in the dataset. Moreover, it increased the dataset images to be 30 times larger than the original dataset. The dataset number of images reached 8,100 images after using the GAN network for 4 classes. This will help in achieving a remarkable testing accuracy and performance matrices. The achieved results will be deliberated in detail in the experimental outcomes section. Figure 6 presents samples of the output of the GAN network for the COVID-19 class.
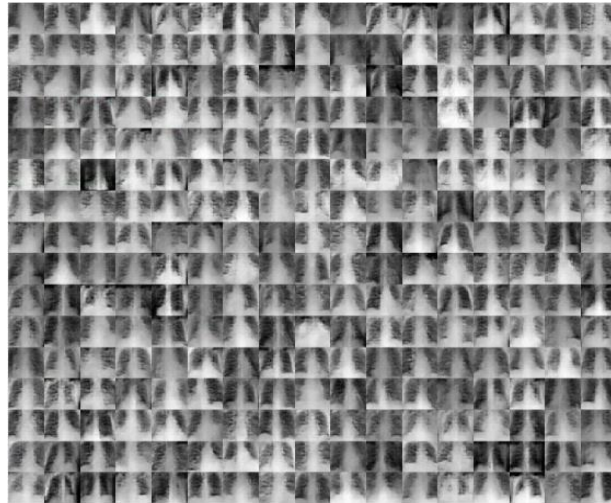
Fig. 6. Samples of images generated using the proposed GAN structure.

4.2 Deep Transfer Learning

Convolutional Neural Networks (ConvNet) is the most successful type of model for image classification and detection. A single ConvNet model contains many different layers of neural networks that work on labeling edges and simple/complex features on neural network layers and more complex deep features in deeper network layers. An image is convolved with filters (kernels) and then max pooling is applied, this process may go on for some layers and at last recognizable features are obtained. Take the size of $W^{l-1} \times H^{l-1} \times C^{l-1}$ (where $W \times H$ is width $\times$ height) feature map and a filterbank in layer $l-1$ for example within C^l kernels at the size of $f^l \times C^{l-1}$, augmenting the other two coefficients stride s^l and padding p^l, the outcome feature box in layer l is $W^l \times H^l \times C^l$ as shown in equation (2):

$$(W^l, H^l) = \left\lfloor \frac{(W^{l-1} \times H^{l-1}) + 2p^l - f^l}{s^l} + 1 \right\rfloor \quad (2)$$

where [·] indicate to floor math. Kernels must be equal to that of the input map. as in equation (3):

$$x_j^l = \sigma\left(\sum_{i \in V_j} x_j^{l-1} \times f_{ij}^l + b_j^l\right) \quad (3)$$

where i and j are indexes of input/output network maps at a range of $W^l \times H^l$ and $W^{l-1} \times H^{l-1}$ respectively. V_j here indicates the receptive field of kernel and b_j^l is the bias term. In equation (3), $\sigma(.)$ is a non-linearity function applied to get non-linearity in deep transfer learning. In our transfer method, we used ReLU in equation (4) as the non-linearity function for rapid training process:

$$\sigma(x_{input}) = \max(0, x_{input}) \quad (4)$$

Our cost function in equation (5):

$$L(s, t) = L_{cls}(s_{c^*}) + \lambda[p^* > 0]L_{reg}(g, g^*) \quad (5)$$

where s_{c^*} is output label c^* while g and g^* denote $[g_x, g_y, g_w, g_h]$ of bounding boxes. $\lambda[p^* > 0]$ consider the boxes of non-background (if $p^* = 0$ is background). This cost function have detection loss L_{cls} and regression loss L_{reg}, in equation (6-8):

$$L_{cls}(s_{c^*}) = -\log(s_{c^*}) \tag{6}$$

and

$$L_{reg}(g, g^*) = \sum_{i \in (x,y,w,h)} R_{L1}(g_i - g_i^*) \tag{7}$$

where:

$$R_{L1}(x) = \begin{cases} 0.5x^2, & if \ |x| < 0 \\ |x| - 0.5, & otherwise \end{cases} \tag{8}$$

In terms of optimizer technique, the momentum Stochastic Gradient Descent (SGD) [62] with momentum 0.9 is chosen as our optimizer technique, which updates weights parameters. This optimizer technique updates the weights of the gradient at the previous iteration and fine-tuning of the gradient. To bypass deep learning network overfitting problems, we utilize this problem by using the dropout technique [63] and the early-stopping technique [64] to select the best training steps. As to the learning rate policy, the step size technique is performed in SGD. We introduced the learning rate (μ) to 0.01 and the number of iterations to be 2000. Also, the mini-batch size is set to 64 and early-stopping to be 5 epochs if the accuracy didn't improve.

5. Experimental Results

The introduced model was coded using a software package (MATLAB). The development was CPU specific. All outcomes were conducted on a computer server equipped by an Intel Xeon processor (2 GHz), 96 GB of RAM. The proposed model has been tested under three different scenarios, the first scenario is to test the proposed model for 4 classes, the second scenario for 3 classes and the third one for 2 classes. All the test experiment scenarios included the COVID-19 class. Every scenario consists of the validation phase and the testing phase. In the validation phase, 20% of total generated images will be used while in the testing phase consists of around 10% from the original dataset will be used.

The main difference between the validation phase and testing phase accuracy is in the validation phase, the data used to validate the generalization ability of the model or for the early stopping, during the training process. In the testing phase, the data used for other purposes other than training and validating. The data used in training, validation, and testing never overlap with each other to build a concrete result about the proposed model.

Before listing the major results of this research, Table 4 presents the validation and the testing accuracy for 4 classes before using GAN as an image augmenter. The presented results in table 4 show that the validation and testing accuracy is quite low and not acceptable as a model for the detection of coronavirus.

Table 4. Validation and testing accuracy for 4 classes according to 3 deep transfer learning models without using GAN.

Model/ Validation- Testing Accuracy	ALexnet	Googlenet	Resnet18
Validation Accuracy	73.1%	76.9%	67.3%
Testing Accuracy	52.0%	52.8%	50.0%

5.1   Verification and Testing Accuracy Measurement

Testing accuracy is one of the estimations which demonstrates the precision and the accuracy of any proposed models. The confusion matrix also is one of the accurate measurements which give more insight into the achieved validation and testing accuracy. First, the 4 classes scenario will be investigated with the three types of deep transfer learning which include Alexnet, Googlenet, and Resnet18. Figure 7,8, and 9 illustrates the confusion matrices for the validation and testing phases for 4 classes in the dataset.
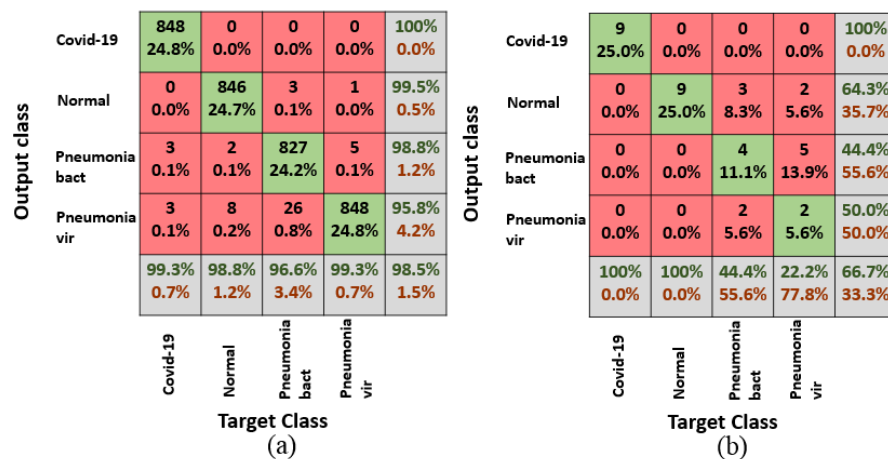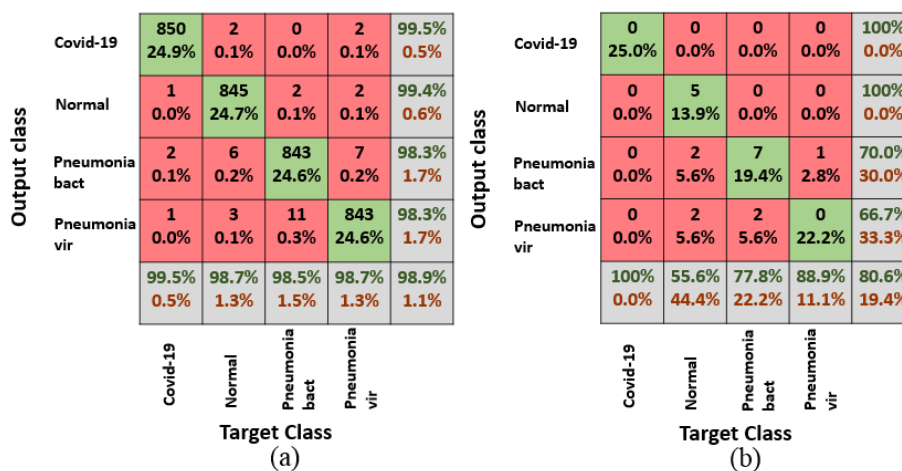

Fig. 7. Confusion matrices of Alexnet for 4 classes (a) validation accuracy, and (b) testing accuracy

Fig. 8. Confusion matrices of Googlenet for 4 classes (a) validation accuracy, and (b) testing accuracy

Table 5 summarizes the validation and the testing accuracy for the different deep transfer models for 4 classes. The table illustrates according to validation accuracy, the Resnet18 achieved the highest

accuracy with 99.6%, this is due to the large number of parameters in the Resnet18 architecture which contains 11.7 million parameters which are not larger than Alexnet but the Alexnet only include 8 layers while the Resnet18 includes 18 layers. According to testing accuracy, the Googlenet achieved the highest accuracy with 80.6%, this is due to a large number of layers if it is compared to other models as it contains about 22 layers.
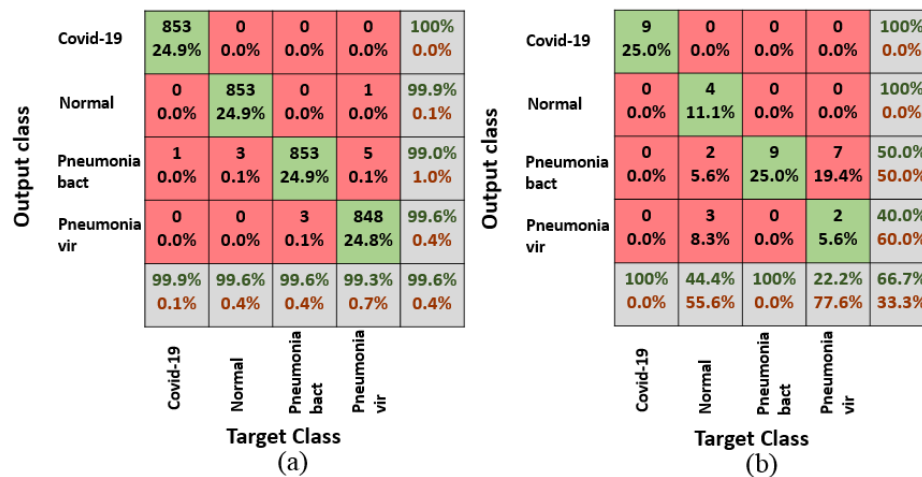

Fig. 9. Confusion matrices of Resnet18 for 4 classes (a) validation accuracy, and (b) testing accuracy.

Table5. Validation and testing accuracy for 4 classes according to 3 deep transfer learning models

Model/ Validation- Testing Accuracy	ALexnet	Googlenet	Resnet18
Validation Accuracy	98.5%	98.9%	99.6%
Testing Accuracy	66.7%	80.6%	66.7%

The second scenario to be tested in this research when the dataset only contains 3 classes. Figure 10,11, and 12 illustrates the confusion matrices for the validation and testing phases for 3 classes in the dataset including the covid class.
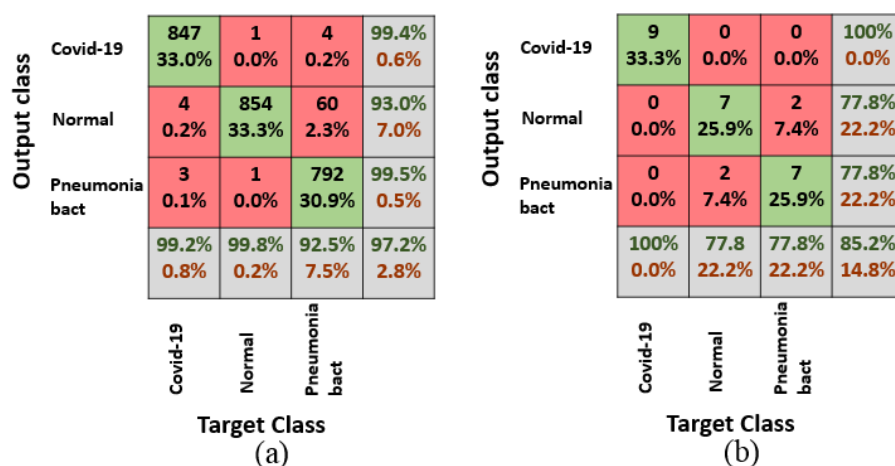

Fig. 10. Confusion matrices of Alexnet for 3 classes (a) validation accuracy, and (b) testing accuracy
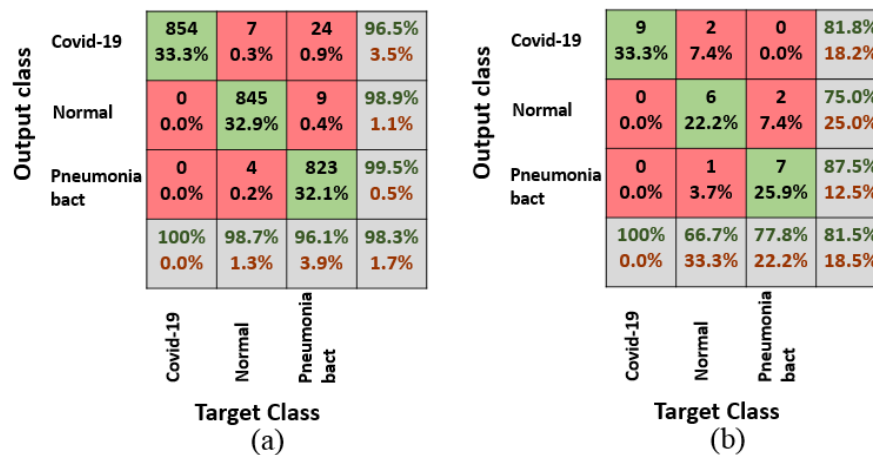
Fig. 11. Confusion matrices of Googlenet for 3 classes (a) validation accuracy, and (b) testing accuracy
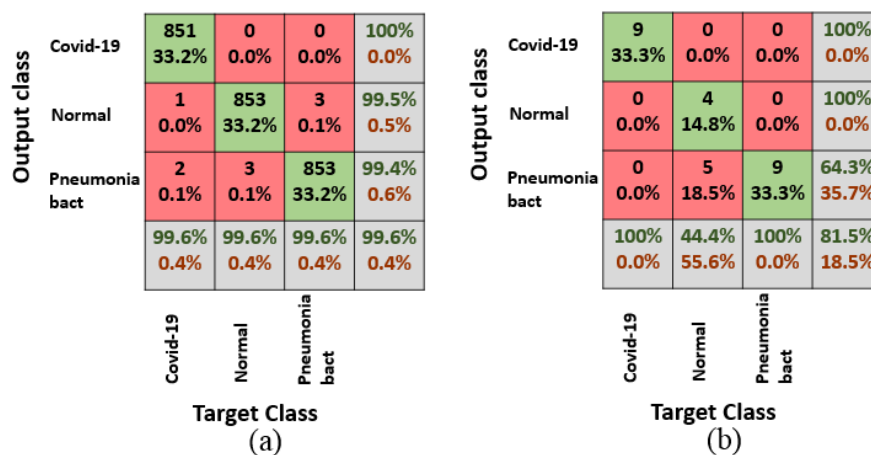

Fig. 3. Confusion matrices of Resnet18 for 3 classes (a) validation accuracy, and (b) testing accuracy

Table 6 summarizes the validation and the testing accuracy for the different deep transfer models for 3 classes. The table illustrates according to validation accuracy, the Resnet18 achieved the highest accuracy with 99.6%. According to testing accuracy, the Alexnet achieved the highest accuracy with 85.2%, this is maybe due to the large number of parameters in the Alexnet architecture which include 61 million parameters and also due to the elimination of the fourth class which include the pneumonia virus which has similar features if it is compared to COVID-19 which is also considered a type of pneumonia virus. The elimination of the pneumonia virus helps in achieving better testing accuracy for the all deep transfer model than when it is trained over 4 classes as mentioned before as COVID-19 is a special type of pneumonia virus.

Table 6. Validation and testing accuracy for 3 classes according to 3 deep transfer learning models

Model/ Validation- Testing Accuracy	ALexnet	Googlenet	Resnet18
Validation Accuracy	97.2%	98.3%	99.6%
Testing Accuracy	85.2%	81.5%	81.5%

The third scenario to be tested when the dataset only includes 2 classes, the covid class, and the normal class. Figure 13 illustrates the confusion matrix for the 3 different transfer models for validation accuracy, While the confusion matrix for testing accuracy is presented in Figure 14 which is the same for all the deep transfer models selected in this research.
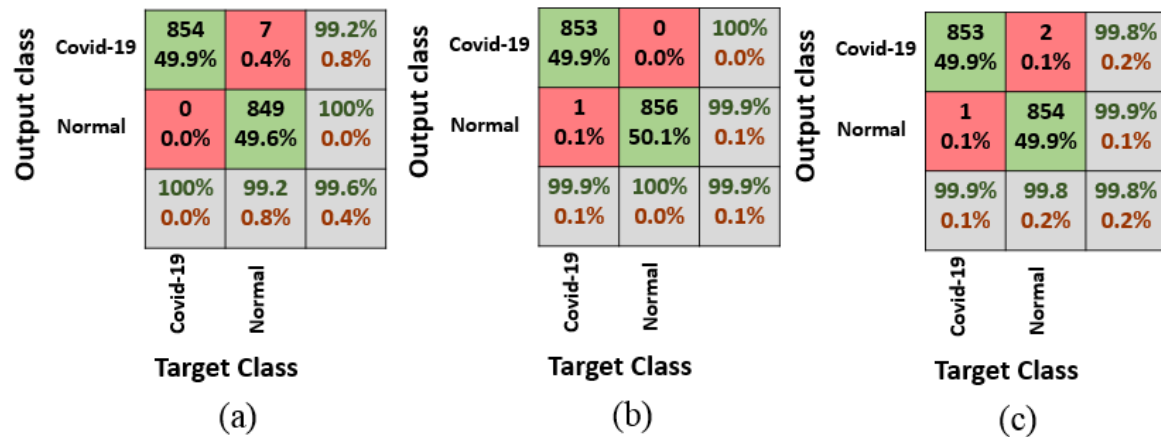

Fig. 4. Confusion matrices the verification accuracy for (a) Alexnet, (b) Googlenet, and (c) Resnet.
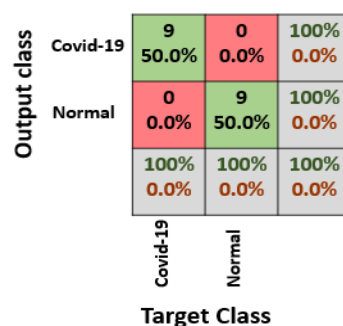

Fig. 5. Confusion matrix for testing accuracy for Alexnet, Googlenet, and Resnet18

Table 7 summarizes the validation and the testing accuracy for the different deep transfer models for 2 classes. The table illustrates according to validation accuracy, the Googlenet achieved the highest accuracy with 99.9%. According to testing accuracy, all the pretrained model Alexnet, Goolgenet, and Resnet18 achieved the highest accuracy with 100%, This due to the elimination of the third and the fourth class which includes pneumonia bacterial and pneumonia virus which has similar features if it is compared to COVID-19. This leads to a noteworthy enhancement in the testing accuracy which reflects on whatever the deep transfer model will be used the testing accuracy will reach 100%. The choice of the best model here will be according to validation accuracy which achieved 99.9%. So the Googlenet will be the selected deep transfer model in the third scenario.

Table 7. Validation and testing accuracy for 2 classes according to 3 deep transfer learning models

Model/ Validation- Testing Accuracy	ALexnet	Googlenet	Resnet18
Validation Accuracy	99.6%	99.9%	99.8%
Testing Accuracy	100%	100%	100%

Table 8. Testing accuracy for every class for the different 3 scenarios

# of classes	Class Name	Alexnet	Googlenet	Resnet18
4 classes	Covid	100%	100%	100%
	Normal	64.3%	100%	100%
	Pneumonia _bac	44.4%	70%	50%
	Pneumonia _vir	50%	66.7%	40%
3 classes	Covid	100%	81.8%	100%
	Normal	77.7%	75.0%	100%
	Pneumonia _bac	77.8%	87.5%	64.3%
2 classes	Covid	100%	100%	100%
	Normal	100%	100%	100%

To conclude this part, every scenario has it is own deep transfer model. In the first scenario, Googlenet was selected, while the second scenario, Alexnet was selected, and finally, in the third scenario, Googlenet was selected as a deep transfer model. To draw a full conclusion for the selected deep transfer learning that fit the dataset and all scenarios, testing accuracy for every class is required for the different deep transfer model. Table 7 presents the testing accuracy for every class for the different three scenarios. Table 8 doesn't help much to determine the deep transfer model that fits all scenarios but for the distinction of COVID-19 class among the other classes, Alexnet and Resent18 will be the selected as deep transfer model as they achieved 100% testing accuracy for COVID-19 class whatever the number of classes is 2,3 or 4.

5.2   Performance Evaluation and Discussion

To estimate the performance of the proposed model, extra performance matrices are required to be explored through this study. The most widespread performance measures in the field of deep learning are Precision, Sensitivity (recall), F1 Score [65] and they are presented from equation (9) to equation (11).

$$\text{Precision} = \frac{\text{TrueP}}{(\text{TrueP}+\text{FalseP})} \tag{9}$$

$$\text{Sensitivity} = \frac{\text{TrueP}}{(\text{TrueP}+\text{FrueN})} \tag{10}$$

$$\text{F1 Score} = 2.\frac{\text{Precision.Sensitivity}}{(\text{Precision}+\text{Sensitivity})} \tag{11}$$

Where TrueP is the count of True Positive samples, TrueN is the count of True Negative samples, FalseP is the count of False Positive samples, and FalseN is the count of False Negative samples from a confusion matrix.

Table 9 presents the performance metrics for different scenarios and deep transfer models for the testing accuracy. The table illustrates that in the first scenario which contains 4 classes, Googlenet achieved the highest percentage for precision, sensitivity and F1 score metrics which strengthen the

research decision for choosing Googlenet as a deep transfer model. The table also illustrates that in the second scenario which contains 3 classes, Alexnet achieved the highest percentage for precision and recall score metrics while Resnet achieved the highest score in F1 with 88.10% but overall the Alexnet had the highest testing accuracy which Also strengthens the research decision for choosing Alexnet as deep transfer model.

The table 9 also illustrates that in the third scenario which contains 2 classes, All deep transfer learning models achieved similar the highest percentage for precision, recall and F1 score metrics which strengthen the research decision for choosing Googlenet as it achieved the highest validation accuracy with 99.9% as illustrated in Table 6.

Table 9. Performance measurements for different scenarios.

# of classes	Class Name	Alexnet	Googlenet	Resnet18
4 classes	Precision	64.68%	84.17%	72.50%
	Recall	66.67%	80.56%	66.67%
	F1 Score	65.66%	82.32%	69.46%
	Testing Accuracy	66.67%	80.56%	69.46%
3 classes	Precision	85.19%	81.44%	88.10%
	Recall	85.19%	81.48%	81.48%
	F1 Score	85.19%	81.46%	84.66%
	Testing Accuracy	85.19%	81.48%	81.48%
2 classes	Precision	100%	100%	100%
	Recall	100%	100%	100%
	F1 Score	100%	100%	100%
	Testing Accuracy	100%	100%	100%

6.   Conclusion and future works

2019 novel Coronaviruses (COVID-19) are a family of viruses that leads to illnesses ranging from the common cold to more severe diseases and may lead to death according to World Health Organization (WHO), with the advances in computer algorithms and especially Artificial Intelligence, the detection of this type of virus in early stages will help in fast recovery. In this paper, a GAN with deep transfer learning for COVID-19 detection in limited chest x-ray images is presented. The lack of benchmark datasets for COVID-19 especially in chest x-rays images was the main motivation of this research. The main idea is to collect all the possible images for COVID-19 and use the GAN network to generate more images to help in the detection of the virus from the available x-ray's images. The dataset in this research was collected from different sources. The number of images of the collected dataset was 307 images for four types of classes. The classes are the covid, normal, pneumonia bacterial, and pneumonia virus.

Three deep transfer models were selected in this research for investigation. Those models are selected for investigation through this research as it contains small number of layers on their architectures, this will result in reducing the complexity and the consumed memory and time for the proposed model. A three-case scenario was tested through the paper, the first scenario which included the 4 classes from the dataset, while the second scenario included 3 classes and the third

scenario included 2 classes. All the scenarios included the COVID-19 class as it was the main target of this research to be detected. In the first scenario, the Googlenet was selected to be the main deep transfer model as it achieved 80.6% in testing accuracy. In the second scenario, the Alexnet was selected to be the main deep transfer model as it achieved 85.2% in testing accuracy while in the third scenario which included 2 classes(COVID-19, and normal), Googlenet was selected to be the main deep transfer model as it achieved 100% in testing accuracy and 99.9% in the validation accuracy. One open door for future works is to apply the deep models with a larger dataset benchmark.

References

1. Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). *The Indian Journal of Pediatrics* **2020**, *87*, 281–286.

2. Lai, C.-C.; Shih, T.-P.; Ko, W.-C.; Tang, H.-J.; Hsueh, P.-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents* **2020**, *55*, 105924.

3. Li, J.; Li, J. (Justin); Xie, X.; Cai, X.; Huang, J.; Tian, X.; Zhu, H. Game consumption and the 2019 novel coronavirus. *The Lancet Infectious Diseases* **2020**, *20*, 275–276.

4. Sharfstein, J.M.; Becker, S.J.; Mello, M.M. Diagnostic Testing for the Novel Coronavirus. *JAMA* **2020**.

5. Chang, L.; Yan, Y.; Wang, L. Coronavirus Disease 2019: Coronaviruses and Blood Safety. *Transfusion Medicine Reviews* **2020**.

6. Shereen, M.A.; Khan, S.; Kazmi, A.; Bashir, N.; Siddique, R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research* **2020**, *24*, 91–98.

7. Rabi, F.A.; Al Zoubi, M.S.; Kasasbeh, G.A.; Salameh, D.M.; Al-Nasser, A.D. SARS-CoV-2 and Coronavirus Disease 2019: What We Know So Far. *Pathogens* **2020**, *9*, 231.

8. York, A. Novel coronavirus takes flight from bats? *Nature Reviews Microbiology* **2020**, *18*, 191.

9. Lam, T.T.-Y.; Shum, M.H.-H.; Zhu, H.-C.; Tong, Y.-G.; Ni, X.-B.; Liao, Y.-S.; Wei, W.; Cheung, W.Y.-M.; Li, W.-J.; Li, L.-F.; et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* **2020**.

10. Giovanetti, M.; Benvenuto, D.; Angeletti, S.; Ciccozzi, M. The first two cases of 2019-nCoV in Italy: Where they come from? *Journal of Medical Virology* **2020**, *92*, 518–521.

11. Holshue, M.L.; DeBolt, C.; Lindquist, S.; Lofy, K.H.; Wiesman, J.; Bruce, H.; Spitters, C.; Ericson, K.; Wilkerson, S.; Tural, A.; et al. First Case of 2019 Novel Coronavirus in the United States. *New England Journal of Medicine* **2020**, *382*, 929–936.

12. Bastola, A.; Sah, R.; Rodriguez-Morales, A.J.; Lal, B.K.; Jha, R.; Ojha, H.C.; Shrestha, B.; Chu, D.K.W.; Poon, L.L.M.; Costello, A.; et al. The first 2019 novel coronavirus case in Nepal. *The Lancet Infectious Diseases* **2020**, *20*, 279–280.

13. Rothe, C.; Schunk, M.; Sothmann, P.; Bretzel, G.; Froeschl, G.; Wallrauch, C.; Zimmer, T.; Thiel, V.; Janke, C.; Guggemos, W.; et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *New England Journal of Medicine* **2020**, *382*, 970–971.

14. Phan, L.T.; Nguyen, T. V; Luong, Q.C.; Nguyen, T. V; Nguyen, H.T.; Le, H.Q.; Nguyen, T.T.; Cao, T.M.; Pham, Q.D. Importation and Human-to-Human Transmission of a Novel Coronavirus in Vietnam. *New England Journal of Medicine* **2020**, *382*, 872–874.

15. Coronavirus (COVID-19) map Available online: https://www.google.com/COVID-19-map/ (accessed on Mar 31, 2020).

16. Rong, D.; Xie, L.; Ying, Y. Computer vision detection of foreign objects in walnuts using deep learning. *Computers and Electronics in Agriculture* **2019**, *162*, 1001–1010.

17. Eraslan, G.; Avsec, Ž.; Gagneur, J.; Theis, F.J. Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics* **2019**.

18. Riordon, J.; Sovilj, D.; Sanner, S.; Sinton, D.; Young, E.W.K. Deep Learning with Microfluidics for Biotechnology. *Trends in Biotechnology* **2019**, *37*, 310–324.

19. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik* **2019**, *29*, 102–127.

20. Maier, A.; Syben, C.; Lasser, T.; Riess, C. A gentle introduction to deep learning in medical image processing. *Zeitschrift für Medizinische Physik* **2019**, *29*, 86–101.

21. Shrestha, A.; Mahmood, A. Review of Deep Learning Algorithms and Architectures. *IEEE Access* **2019**, *7*, 53040–53065.

22. Pouyanfar, S.; Sadiq, S.; Yan, Y.; Tian, H.; Tao, Y.; Reyes, M.P.; Shyu, M.-L.; Chen, S.-C.; Iyengar, S.S. A Survey on Deep Learning: Algorithms, Techniques, and Applications. *ACM Comput. Surv.* **2018**, *51*, 92:1--92:36.

23. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2; MIT Press: Cambridge, MA, USA, 2014; pp. 2672–2680.

24. Cao, Y.; Jia, L.; Chen, Y.; Lin, N.; Yang, C.; Zhang, B.; Liu, Z.; Li, X.; Dai, H. Recent Advances of Generative Adversarial Networks in Computer Vision. *IEEE Access* **2019**, *7*, 14985–15006.

25. Gonog, L.; Zhou, Y. A Review: Generative Adversarial Networks. In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA); 2019; pp. 505–510.

26. Lee, M.; Seok, J. Controllable Generative Adversarial Network. *IEEE Access* **2019**, *7*, 28158–28169.

27. Caramihale, T.; Popescu, D.; Ichim, L. Emotion Classification Using a Tensorflow Generative Adversarial Network Implementation. *Symmetry* **2018**, *10*, 414.

28. Ciregan, D.; Meier, U.; Schmidhuber, J. Multi-column deep neural networks for image classification. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012; pp. 3642–3649.

29. Krizhevsky Ilya and Hinton Alex and Sutskever, G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **2012**, 1097–1105.

30. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **1998**, *86*, 2278–2324.

31. Yin, F.; Wang, Q.; Zhang, X.; Liu, C. ICDAR 2013 Chinese Handwriting Recognition Competition. In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition; 2013; pp. 1464–1470.

32. El-Sawy, A.; EL-Bakry, H.; Loey, M. CNN for Handwritten Arabic Digits Recognition Based on LeNet-5 BT - Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016.; Hassanien, A.E., Shaalan, K., Gaber, T., Azar, A.T., Tolba, M.F., Eds.; Springer International Publishing: Cham, 2017; pp. 566–575.

33. El-Sawy, A.; Loey, M.; EL-Bakry, H. Arabic Handwritten Characters Recognition Using Convolutional Neural Network. *WSEAS Transactions on Computer Research* **2017**, *5*.

34. LeCun, Y.; Huang, F.J.; Bottou, L. Learning methods for generic object recognition with invariance to pose and lighting. In Proceedings of the Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.; 2004; Vol. 2, pp. II-104 Vol.2.

35. Stallkamp, J.; Schlipsing, M.; Salmen, J.; Igel, C. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In Proceedings of the The 2011 International Joint Conference on Neural Networks; 2011; pp. 1453–1460.

36. Deng, J.; Dong, W.; Socher, R.; Li, L.; Kai, L.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009; pp. 248–255.

37. Liu, S.; Deng, W. Very deep convolutional neural network based image classification using small training sample size. In Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR); 2015; pp. 730–734.

38. Szegedy, C.; Wei, L.; Yangqing, J.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015; pp. 1–9.

39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; pp. 770–778.

40. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; pp. 1800–1807.

41. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition; 2016; pp. 2818–2826.

42. Huang, G.; Liu, Z.; Maaten, L. v. d.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; pp. 2261–2269.

43. Stephen, O.; Sain, M.; Maduh, U.J.; Jeong, D.-U. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. *Journal of Healthcare Engineering* **2019**, *2019*, 4180949.

44. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122-1131.e9.

45. Ayan, E.; Ünver, H.M. Diagnosis of Pneumonia from Chest X-Ray Images Using Deep Learning. In Proceedings of the 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT); 2019; pp. 1–5.

46. Varshni, D.; Thakral, K.; Agarwal, L.; Nijhawan, R.; Mittal, A. Pneumonia Detection Using CNN based Feature Extraction. In Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT); 2019; pp. 1–7.

47. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; pp. 3462–3471.

48. Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damaševičius, R.; de Albuquerque, V.H.C. A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images. *Applied Sciences* **2020**, *10*, 559.

49. Islam, S.R.; Maity, S.P.; Ray, A.K.; Mandal, M. Automatic Detection of Pneumonia on Compressed Sensing Images using Deep Learning. In Proceedings of the 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE); 2019; pp. 1–4.

50. Cohen, J.P.; Morrison, P.; Dao, L. COVID-19 Image Data Collection. *arXiv:2003.11597 [cs, eess, q-bio]* **2020**.

51. Dao, J.P.C. and P.M. and L. COVID-19 image data collection Available online: https://github.com/ieee8023/covid-chestxray-dataset (accessed on Mar 31, 2020).

52. Dataset Available online: https://drive.google.com/uc?id=1coM7x3378f-Ou2l6Pg2wldaOI7Dntu1a (accessed on Mar 31, 2020).

53. Khalifa, N.; Taha, M.; Hassanien, A.; Mohamed, H. Deep Iris: Deep Learning for Gender Classification Through Iris Patterns. *Acta Informatica Medica* **2019**, *27*, 96.

54. Khalifa, N.E.M.; Taha, M.H.N.; Hassanien, A.E.; Hemedan, A.A. Deep bacteria: robust deep learning data augmentation design for limited bacterial colony dataset. *International Journal of Reasoning-based Intelligent Systems* **2019**.

55. Lemley, J.; Bazrafkan, S.; Corcoran, P. Smart Augmentation Learning an Optimal Data Augmentation Strategy. *IEEE Access* **2017**.

56. Khalifa, N.E.M.; Taha, M.H.N.; Ezzat Ali, D.; Slowik, A.; Hassanien, A.E. Artificial Intelligence Technique for Gene Expression by Tumor RNA-Seq Data: A Novel Optimized Deep Learning Approach. *IEEE Access* **2020**.

57. Khalifa, N.; Loey, M.; Taha, M.; Mohamed, H. Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection. *Acta Informatica Medica* **2019**, *27*, 327.

58. Khalifa, N.E.; Hamed Taha, M.; Hassanien, A.E.; Selim, I. Deep galaxy V2: Robust deep convolutional neural networks for galaxy morphology classifications. In Proceedings of the 2018 International Conference on Computing Sciences and Engineering, ICCSE 2018 - Proceedings; KW, 2018; pp. 1–6.

59. Khalifa, N.E.M.; Taha, M.H.N.; Hassanien, A.E. Aquarium Family Fish Species Identification System Using Deep Neural Networks. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018*; 2018; pp. 347–356.

60. Shah, R.; Yang, Y. GoogLeNet. *Population Health Management* **2015**.

61. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017; 2017.

62. Bottou, L. Stochastic Gradient Descent Tricks. In *Neural Networks: Tricks of the Trade: Second Edition*; Montavon, G., Orr, G.B., Müller, K.-R., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp. 421–436 ISBN 978-3-642-35289-8.

63. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* **2014**, *15*, 1929–1958.

64. Caruana, R.; Lawrence, S.; Giles, L. Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping. In Proceedings of the Proceedings of the 13th International Conference on Neural Information Processing Systems; MIT Press: Cambridge, MA, USA, 2000; pp. 381–387.

65. Goutte, C.; Gaussier, E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In; 2010.