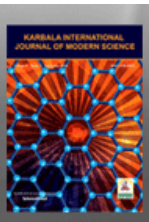

Karbala International Journal of Modern Science

Volume 9 | Issue 3 Article 13

Deep Learning-based CAD System for Predicting the COVID-19 X-ray Images

Aqeel R. Talib
Department of Mathematics, College of Science, University of Basrah, Basrah, Iraq, aqeelraed1987@gmail.com

Hana' M. Ali
Department of Mathematics, College of Science, University of Basrah, Basrah, Iraq

Recommended Citation

Talib, Aqeel R. and Ali, Hana' M. (2023) "Deep Learning-based CAD System for Predicting the COVID-19 X-ray Images," *Karbala International Journal of Modern Science*: Vol. 9 : Iss. 3 , Article 13.
Available at: https://doi.org/10.33640/2405-609X.3316

Deep Learning-based CAD System for Predicting the COVID-19 X-ray Images

Abstract

According to World Health Organization data, Coronavirus (COVID-19) has infected about 660, 378, 145 patients around the world. It is nonetheless difficult for physicians to detect COVID-19 infections out of CT or X-ray radiographs. Thus, several computer-aided diagnosis (CAD) systems based on deep learning and radiographs were developed to detect COVID-19 infections. However, the majority of approaches considered small datasets, which is ineligible to provide diverse COVID-19 radiographs. This work utilizes a massive number of X-ray radiographs, and compared standard CNN, DenseNet-121, and GoogLeNet for isolating COVID-19 infections out from normal and other pneumonia radiographs. The dataset in this work is large enough to evaluate the realistic performance of those models in labeling COVID-19 infections. Considering the time complexity, accuracy, precision, recall, and F1 score, the experimental results shows that the DenseNet-121 is not only the optimal model, but also there is superior for standard CNN compared to the second output of GoogLeNet, which is an unexplained phenomenon.

Keywords

Deep Learning, Convolutional Neural Network, GoogLeNet, DenseNet, Covid-19, Pneumonia.

RESEARCH PAPER

Deep Learning-based CAD System for Predicting the COVID-19 X-ray Images

Aqeel R. Talib*, Hana' M. Ali

Department of Mathematics, College of Science, University of Basrah, Basrah, Iraq

Abstract

According to World Health Organization data, Coronavirus (COVID-19) has infected about 660, 378, 145 patients around the world. It is nonetheless difficult for physicians to detect COVID-19 infections out of CT or X-ray radiographs. Thus, several computer-aided diagnosis (CAD) systems based on deep learning and radiographs were developed to detect COVID-19 infections. However, the majority of approaches considered small datasets, which is ineligible to provide diverse COVID-19 radiographs. This work utilizes a massive number of X-ray radiographs, and compared standard CNN, DenseNet-121, and GoogLeNet for isolating COVID-19 infections out from normal and other pneumonia radiographs. The dataset in this work is large enough to evaluate the realistic performance of those models in labeling COVID-19 infections. Considering the time complexity, accuracy, precision, recall, and F1 score, the experimental results shows that the DenseNet-121 is not only the optimal model, but also there is superior for standard CNN compared to the second output of GoogLeNet, which is an unexplained phenomenon.

Keywords: Deep learning, Convolutional neural network, GoogLeNet, DenseNet, Covid-19, Pneumonia

1. Introduction

Coronavirus (COVID-19) is a disease caused by severe acute respiratory syndrome coronavirus-2, which was spread in 2019 and occupied the attention of the whole world. Most countries were infected and the healthcare agreement was under evaluation by the World Health Organization [1]. COVID-19 is characterized by a rapid spread and infection among individuals due to the lack of intensive treatment, and further, the majority of diagnosis is taken after the appearance of symptoms, which most of the time leads to a lessened ability to confront the virus due to the weak immunity of the patient and the exacerbation of the disease [2]. According to recent data, COVID-19 causes fewer symptoms in around 99% of patients, whereas the remaining are severe cases. As of January 1, 2023, there have been 665, 003, 425 COVID-19 cases reported globally, with 6,697,442 deaths, and there are 21, 381, 993 active patients out of these [3].

A crucial, simple, cheap, and fast technique for diagnosing pneumonia is chest radiography (X-ray) [4]. In this regard, in comparison with computed tomography (CT) and magnetic resonance imaging (MRI), the chest X-ray exposes the patient to less radiation. However, an accurate X-ray picture diagnosis requires specialized knowledge and expertise. Consequently, compared to other imaging modalities like CT or MRI, a chest X-ray is far more challenging to diagnose.

Only specialist physicians can precisely diagnose the X-ray, and decide whether it is embedded with a COVID-19 virus. Unfortunately, the number of those specialist physicians capable of making such diagnosis is less than the number of non-specialist doctors. Even during normal times, in most nations around the world, there aren't enough doctors per population. According to some data, Greece tops the list with 607 doctors per 100,000 people; this figure is considerably lower in other nations.

For a rapid and precise diagnosis of pneumonia from the chest X-ray, doctors can resort to exploiting

Received 22 February 2023; revised 10 July 2023; accepted 12 July 2023.
Available online 21 August 2023

* Corresponding author.
E-mail address: aqeelraed1987@gmail.com (A.R. Talib).

https://doi.org/10.33640/2405-609X.3316

artificial intelligence-based diagnosis methods, especially CADs. Artificial intelligence (AI) is one technique that is increasingly being used in the realm of medical services since it can handle massive datasets better than humans can [5]. To reduce the workload of doctors, the CAD systems can be integrated with the radiologist diagnostic methods, and this would also improve the reliability and the quantitative analysis. Such a combination of deep learning-based medical imaging CAD system has revolutionized the state of the art [6–8]. Deep learning (DL) paradigms have been effectively applied in many fields such as skin cancer classification [9], breast cancer detection [10], brain disease classification [11], fundus image segmentation [12], arrhythmia detection [13], pneumonia detection from chest X-ray images [14], and lung segmentation [15].

The rapid spread of the COVID-19 outbreak has made it necessary to cover this area. This has raised interest in creating AI-based automated detection systems. It is a challenge to extend expert clinicians to all hospitals owing to the limited number of radiologists. Additionally, AI methods can also help remove drawbacks including the lack of readily available RT-PCR test kits, test fees, and long wait times for test results.

Recently, many X-ray images have been widely exploited for COVID-19 classification. In [16], the authors utilize the standard CNN to classify normal versus pneumonia X-rays. P. Saha et al. [17] utilized a DL paradigm to diagnose COVID-19 through X-ray images and proposed a COVIDX-Net model, which is composed of 7 CNN models. Wang and Wong [2] scored a 92.4% accuracy in classifying COVID-19 versus normal and other pneumonia classes by proposing a deep model of COVID-19 detection (COVID-Net). Ioannis et al. [18] constructed a deep-learning model using only 224 confirmed COVID-19 images. Their paradigms scored 98.75% and 93.48% success rates for two and three classes, respectively. Finally, there have been many recent studies on the detection of COVID-19 that used CT scans and various deep-learning models [19–21].

Our contribution to this work is to examine multiple automatic CAD predictions of COVID-19 using a deep CNN based on chest X-ray images. In this regard, we checked three models of CNN to investigate their role in distinguishing the X-rays of three classes of images, which are COVID-19, normal, and other pneumonia diseases. The first model is a standard CNN. The second model is GoogLeNet [22], and the third model is the Dense-121 which is proposed in [23]. In contrast to some previous works in the state of art, which have been performed on limited disorganized, a huge number of chest X-ray images of three classes have been collected from multiple resources to fairly evaluate those models. Further, more performance analytics on different CNN outputs were discussed.

The rest of the paper is organized as follows: In Section 2, research on using DL techniques on chest X-ray and CT images used to detect COVID-19 disease is discussed. Section 3 discusses the material, dataset, and the three models. In section 4, the experimental results were discussed along with the performance metrics. Finally, section 5 summarizes the conclusion of this work.

2. Related work

Recent studies that aim to build diagnosis models for COVID-19 detections have relied on chest X-rays and CT radiography, and they were either binary or multiple classifications models. While several research works depend on feature extraction, some others utilized raw data. Other studies also used a variety of data set sizes to construct a well-generalized model. However, most of those techniques have been trained with tiny or high-variance data set that produce high-accuracy models in the training stage, but lack model generalization, which as a result makes the model not applicable to the real-world problem.

Convolutional neural networks are the technique most often adopted in the majority of CAD-COVID-19 systems. M. Mijwil suggested implementing DL paradigms such as Random Forest, Logistic Regression, Naïve Bayes, and Support Vector Machine to detect COVID-19 out of the X-ray images. The work shows that the DL technique is the best regarding disease detection [24], however, the author trained all models with only 389 COVID-19 images, which would make the model inappropriate to be considered as a real-world CAD system. Mukherjee et al. [25] proposed a CNN architecture that is appropriate to be a CAD system for both CT and chest X-ray radiographs. Their CNN can impeccably convolve both CT and X-ray images with the same effectiveness; hence, they achieved an overall accuracy of 96.28%, but with raw data consisting of 336 COVID-19 CT and CXR images only. The author in [3] proposed a CT images of the CAD system to construct a CNN tailored for COVID-19 CT image segmentation by introducing the feature variation block (FV) which modifies the global information of the features. The FV block enhances the ability of feature representation in diverse cases. They also utilized pyramid pooling to deal with the sophisticated infection position in the images.

Considering the difficulty of obtaining and cleaning the CT image data, their data were medium in size to show a good performance, hence the model scored nearly 98%. However, the chest CT images in their work was taken from only 861 patients with confirmed COVID-19 by RT-PCR.

Apostolopoulos and Bessiana utilized an evolutionary ANN for healthy differentiation on automatic prediction of COVID-19-induced pneumonia, and other pneumonia classes. They utilized transfer learning techniques to detect various abnormalities in datasets. Although their model achieved remarkable results, the dataset was embedded with only 224 confirmed Covid-19 images [18].

Zhang et al. [26] developed a model based on DL that can detect COVID-19 aiming to produce high sensitivity and provide reliable and rapid scanning. Although their work employed the idea of rapid scanning with thousands of viral images, their data consisted of merely 106 confirmed COVID-19 cases as they stated that the majority of data were collected before the pandemic, and thus the model would not be considered as a CAD system for COVID-19 detection.

In [27], Narin et al. utilized a transfer learning CNN that uses shifted cross-validation to detect COVID-19 in chest X-ray radiographs from other pneumonia diseases. They have been testing their 341 conformed COVID-19 cases dataset on five models, which are (ResNet50, ResNet101, ResNet152, InceptionV3, and Inception-ResNetV2).

Singh et al., in [28] employed a multi-objective differential evolution (MODE)-based CNN that supplied CT images to detect the COVID-19 infected and non-infected CT images. Their model was built with a dataset embedded with a total of 3883 images of depicted Pneumonia, out of which 2538 images belong to Bacterial Pneumonia and 1345 images depicted virus Pneumonia, hence the COVID-19 images in their data are within 1345 virus Pneumonia, which makes their CAD system suitable for the prediction of all Pneumonia cases rather than concentrated at COVID-19 detection system.

Adhikari, in [29] proposed a network named Auto Diagnostic Medical Analysis by which the infected areas in the radiographs might be detected to assist the physicians in better finding the diseased part if any. The study made use of both X-ray and CT scans. It has been suggested that the DenseNet network (DN) be used to eliminate and mark contaminated lung regions, however, there are in total 152 images utilized for training and testing the CAD system, which consists of different cases of the infections that include ARDS, SARS, Pneumocystis, Streptococcus along the COVID-19. Distributing only 152 images of several diseases to a CNN model is not an appropriate application for building a COVID-19 CAD system.

In [30], the authors trained their novel CNN model to detect COVID-19 infections by applying three types of filters to get and locate the edges from the images, which would help to get the desired segmented target. However, the model receives only nearly 200 X-ray images containing only up to 70 COVID-19 cases. With limited data, it can be very hard to evaluate the robustness and accuracy of the proposed approach. Further, with a small number of images, it is hard to generalize the result.

Another earlier deep COVID-19 detection model is proposed by Wang and Wong [2] using CNN and X-ray images. The model was fair enough to be a perfect CAD system, as it has been tested over 1203 normal chests, 45 COVID-19, 660 viral pneumonia, and 931 bacterial pneumonia patients. Their model accomplished a very good testing accuracy of 83.50%, taking into consideration the data diversity and size, which contains 45 confirmed COVID-19 cases. Ozturka et al. [31] suggested one of the earliest COVID-19 CAD systems that use other than the standard CNN, which is the DarkNet model. This model has 17 convolutional layers and numerous filtering stages in each layer, as classifiers for the YOLO real-time object identification system. However, the dataset was too small and taken from only 43 female and 82 male cases.

The author [32] proposed a CNN-based CAD system called COVIDX-Net to assist radiologists in automatically diagnosing COVID-19 in 50 X-ray images, which contains only 25 confirmed positive COVID-19 cases.

From the above works, one conclusion could be drawn, which is the dataset set size itself, and the number of COVID-19 images, which should be considered when multiple models are compared. The dataset of COVID-19 in early 2021 and 2022 was very rare, which makes the accuracies of the earlier models very high, and sometimes reach 98% or even 99%. Some of the above works have trained models even to deal with data that contain less than 500 COVID-19-labeled radio images. As a result, in this work, the dataset that has been used to train three CNN models is very large compared with the above-mentioned works. The dataset of one type of radio-images, i.e., X-rays, has 36,240 images with 16,194 COVID-19 radiographs, a matter that makes this dataset large enough COVID-19 data compared with earlier works.

3. Material and models

This section describes the X-ray radiographs dataset that has been used to develop the CNN models. The section also gives preliminaries of the three CNN models, which are the standard CNN, GoogLeNet, and DN-121 networks.

3.1. Dataset

The chest X-ray image dataset, visualized in Fig. 1, has been collected from two different authorized resources. Cohen JP developed a COVID-19 chest X-ray image dataset extracted from various open-access sources and it is available at https://github.com/ieee8023/covid-chestxray-dataset. The second source is the COVID-Net Open Initiative databases, available at, https://www.kaggle.com/code/chaitanya99/viral-pneumonia-classification-googlenet/data, which is an initiative committed to speeding up machine learning usage to support front-line healthcare professionals and clinical institutions around the world in their fight against the ongoing

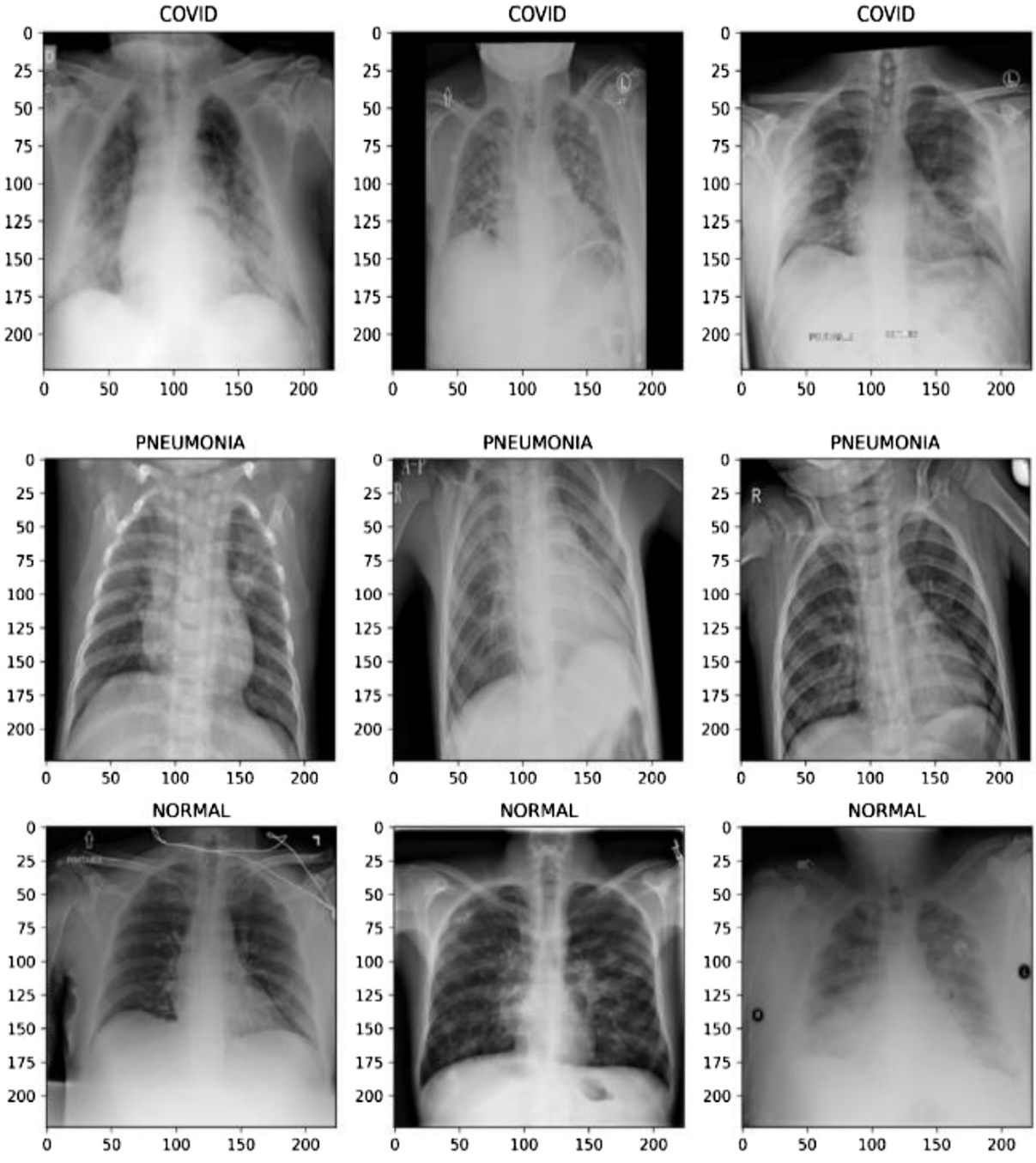

Fig. 1. Samples of chest X-ray radiographs of three labeled classes.

pandemic. The dataset from the second source is updated constantly throughout the course, as new cases with chest X-rays of COVID-19 and other pneumonia types are added every period. The obtained dataset, which contains a pre-diagnosed COVID-19, Pneumonia, and normal cases, is packed with 36,240 grayscale chest X-ray images taken for infected and normal females and males with an average age of approximately 47 years old. In Fig. 1. COVID-19 shows a shadow on the Cardio-vassal within the limits, infiltrate, consolidation, basilar opacity, and vaguely defined opacities in the bilateral lungs. The general picture of those samples shows that COVID-19 X-ray images reveal a whiter shade than pneumonia and normal ones.

3.2. Models

This work builds a CAD system based on CNN, which is one of the common DL architectures applied to many detection, prediction, recognition, classification, and regression applications.

CNN is inspired by the natural visual perception mechanism of living organisms. Generally, CNN is formulated from three basic layers, which are convolutional, pooling, and a fully connected network. While the convolutional layer aims to gain information about the feature representation, the pooling layer performs a shift-invariance by lessening the feature map resolution. The fully-connected layer generates a high global level of semantic information.

CNN has been subjected to several developments to overcome the difficulties encountered in deep training. The most straightforward method of enhancing the performance of CNN is by rising their size (i.e., depth and width). However, there would be two major disadvantages of such a procedure. The first disadvantage is that the bigger the size, the larger number of trainable parameters, which enforce the CNN to be more disposed to overfitting. The second disadvantage relates to the high computations due to the large number of layers and required-to-adjust parameters.

The forthcoming subsections briefly describe the different three CNN models, GoogLeNet, DN-121, and the standard CNN, that have been compared to reveal the efficiency of each in terms of computational complexity, and performance matrices in predicting the COVID-19 diagnosis.

3.2.1. Standard CNN

The basic components of the CNN are ordinarily very similar. This paper considers the LeNET-5 network as a representation of the standard CNN. In such architecture, the convolutional layer is employed to formulate a new summarized feature map by applying a learned kernel on all spatial locations of the input along with a nonlinear activation function. Mathematically, the feature value at the position (i,j) in the $m-th$ feature map of the l-layer , $V^l_{i,j,m}$, can be extracted by:

$$V^l_{i,j,m} = C^l_m \; x^l_{i,j} + b^l_m \tag{1}$$

where C^l_m and b^l_m are the weighted vector (kernel) and bias of the $m-th$ filter of the l-layer respectively. $x^l_{i,j}$ is the patch input which is centered at the position (i, j). On the other hand, the nonlinear activation function is utilized to extract the nonlinear features. Let the activation function be f, thus the activation function $f^l_{i,j,m}$ of $V^l_{i,j,m}$ the convolutional feature can be calculated as:

$$f^l_{i,j,m} = f\left(V^l_{i,j,m}\right) \tag{2}$$

To implement a shift-invariance, a pooling layer is often located between two successive convolutional layers. In the pooling layer, each feature map is linked to its corresponding feature map of the previous convolutional layer. Let P denote the pooling function of each feature map, then for each $V^l_{i,j,m}$ there is:

$$y^l_{i,j,m} = P\left(V^l_{n,k,m}\right), \forall\,(n,k) \in \varphi_{i,j} \tag{3}$$

where $\varphi_{i,j}$ is the local neighboring around the position of (i,j). It is worth mentioning that the most used pooling function in CNN is the average pooling method.

After several layers of convolution and pooling, the high-level reasoning is achieved by one or more fully connected layers, which linked all neurons of the proceeding layer with each neuron of the current layer to generate the semantic information.

Finally, the output layer is placed at the end of CNN architecture to execute the classification task, which is usually done by a function called softmax operator. CNN is a higher class of the basic multilayer perceptron in which a global optimization technique should be backpropagated to gain the optimal solution for the CNN parameters that reduce the loss function τ. Let ξ refer to all CNN parameters that must be optimized. We assume K is the desired relation between input and output to be $\{(\boldsymbol{x}^{(k)}, \boldsymbol{y}^{(k)}); k \in [1,2,\ldots,N]\}$, where $\boldsymbol{x}^{(k)}$ and $\boldsymbol{y}^{(k)}$ $k-th$ is the input data and its desired target data respectively.

Let $t^{(n)}$ be the output of the network, then τ can be computed as follow:

Table 1. Standard CNN Architecture.

Layers	Description
Convolution	32 × 3 × 3
Pooling	2 × 2 max pool
Convolution	64 × 3 × 3
Pooling	2 × 2 max pool
Convolution	128 × 3 × 3
Pooling	2 × 2 max pool
Convolution	32 × 3 × 3
Pooling	4 × 4 max pool
Flatten	1568

$$\tau = \frac{1}{K}\sum_{k=1}^{N} l\left(\xi; y^{(n)}, t^{(n)}\right) \quad (4)$$

By optimizing the trainable parameters of CNN, the loss function can be reduced which would enhance the classification label and approximate the matching of input-desired relations. The figure below demonstrates the basic architecture of standard CNN that is being used in this work. Table 1 shows the CNN architecture that has been used to classify the chest X-ray images.

The standard CNN architecture which was utilized in this work is demonstrated in Fig. 2, and Table 1.

3.2.2. GoogLeNet

GoogLeNet was proposed by a collaboration between Google and various universities around the world in 2014 [22]. This architecture employs some techniques such as 1 × 1 convolutions and a global average pooling which enables the architecture to go deep.

Fully connected layers have been used at the end of the network in the previous architectures of CNN such as AlexNet, and this would cause high

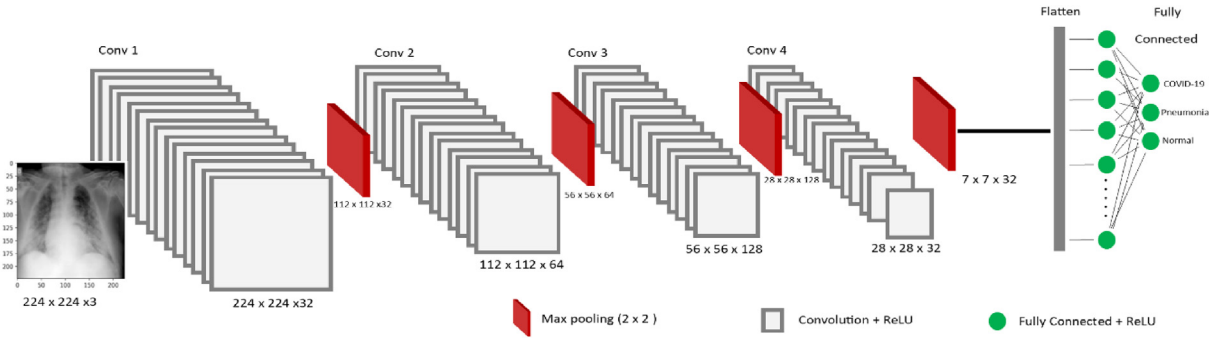

Fig. 2. The Standard CNN architecture.

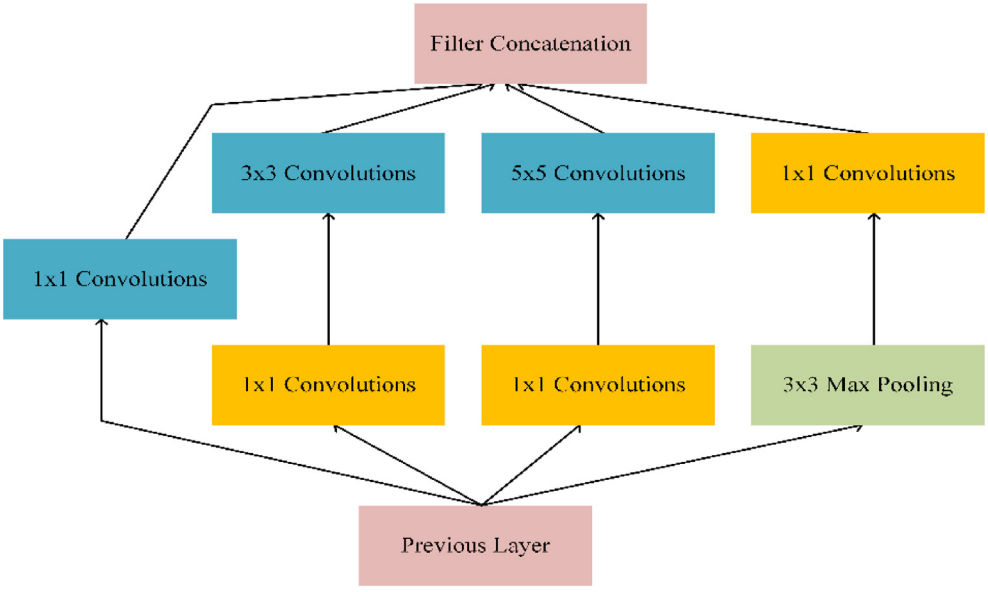

Fig. 3. Inception module with dimension reductions.

computation complexity because more parameters are embedded in the fully connected layers. In contrast, GoogLeNet uses a method called global average pooling at the end of the network, which decreases the number of the trainable parameters due to the feature map of 7×7 and the average 1×1 layer containing the global average pooling.

This network adopts the inception module, which is a new architecture different from the previous models. The inception module, illustrated in Fig. 3, consists of 1×1, $3 \times 3, 5 \times 5$ convolutions, and 3×3 max pooling that operate in a parallel course to the input and output of these which are packed together to generate the final output.

The final building block in such a network is the auxiliary classifier, which branches at the heart of the architecture. Its branches perform only during the training stage. Each branch is embedded in i) 5×5 average pooling layer with a stride 3, ii) 128 filtered 1×1 convolutions, iii) fully connected layers of 1024 units with RELU activation function, iv) a 70% ratio of drooped layers, v) Softmax classifier.

The overall architecture comprises 22 layers when one counts only the parameterized layers and 27 layers when pooling layers are counted. Fig. 4 shows the overall architecture of the GoogLeNet that has been used in this work.

Fig. 4. GoogLeNet network with all the bells and whistles [22].

3.2.3. Dense-121

DNs are proposed to be the new models that resolve the problem of the vanishing gradient existing in the standard CNN, which is a problem that occurs due to the increasing path of information in the routes from input to output. DNs is a modified standard CNN that is subjected to simplified connectivity patterns among layers. The main idea in the DNs architecture is called connectivity according to which every layer is linked directly with all other layers; hence, it gains the name of Dense. For the L layer, there are $\frac{L(L+1)}{2}$ connections. The

Table 2. DN-121 architecture.

Layers	Output Size	DenseNet-121
Convolution	112×112	7×7 conv, stride 2
Pooling	56×56	3×3 max pool, stride 2
Dense Block (1)	56×56	$6 \times \begin{bmatrix} 1 \times 1 \ conv \\ 3 \times 3 \ conv \end{bmatrix}$
Transition Layer (1)	56×56	1×1 conv
	28×28	2×2 average pool, stride 2
Dense Block (2)	28×28	$12 \times \begin{bmatrix} 1 \times 1 \ conv \\ 3 \times 3 \ conv \end{bmatrix}$
Transition Layer (2)	28×28	1×1 conv
	14×14	2×2 average pool, stride 2
Dense Block (3)	14×14	$24 \times \begin{bmatrix} 1 \times 1 \ conv \\ 3 \times 3 \ conv \end{bmatrix}$
Transition Layer (3)	14×14	1×1 conv
	7×7	2×2 average pool, stride 2
Dense Block (4)	7×7	$16 \times \begin{bmatrix} 1 \times 1 \ conv \\ 3 \times 3 \ conv \end{bmatrix}$
Classification Layer	1×1	7×7 global average pool 1000 fully connected, Softmax

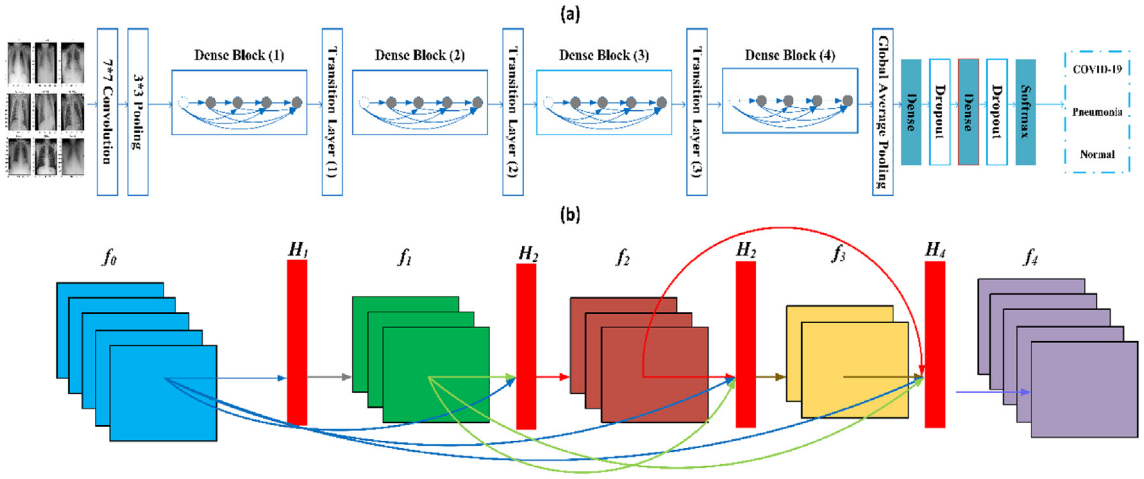

Fig. 5. (a) DN-121 Architecture and (b) the connectivity pattern of dense blocks.

feature maps (FM) of all former layers are concatenated and utilized as inputs to the current and hence allowing the idea of discarded feature maps.

The i^{th} layer collects the FM of all former layers, f_0, $f_1,\ldots,f_{i-1}$ as input:

$$f_i = U_i\left(\left[f_0, f_1,\ldots,f_{i-1}\right]\right) \tag{5}$$

where: $f_0, f_1,\ldots,f_{i-1}$ are the FM concatenations, which are the outputs, delivered by all former layers. The inputs of U_i are combined into a single tensor to facilitate the implementation.

DNs architecture is alienated into dense blocks in which the dimension of the FM remains fixed within one block, nevertheless, the number of filters between those blocks is altered. The transition layer between blocks reduces the number of channels to half of that existing once. In Equation (5), U_i is a composite function that carries out three successive processes i) batch normalization, ii) RELU, and iii) convolution.

DNs are also characterized by their growth rate, which regulates the amount of information linked to each layer in DNs. The dense block of DNs is attached to one 1×1 bottleneck convolutional layer before splitting into the 3×3 convolutions.

The DN-121, as illustrated in Table 2, gains the overmentioned main characterizations, which include four dense blocks and three transition layers. Every dense block in the architecture of DN-121 has a different number of layers characterized by two 1×1 convolutions as bottleneck layers and a 3×3 kernel to carry out the convolution operation. The transition layers in DN-121 are embedded in the 1×1 convolutional layer and a 2×2 average pooling layer with stride 2.

The rest of the architecture contains a 7×7 convolutional layer, a 3×3 basic pooling layer, a global average pooling layer, and an output layer. Fig. 5 illustrates the basic dense blocks and the structure of the entire model.

Table 3. Dataset sizes.

Data type	Size
Dataset	36,240
Training set	28,992
Testing set	3624
Validation set	3624

4. Experimental setup & evaluation

4.1. Data and model preparation

The dataset is a collection of different extensions (i.e., JPEG, JPG, and PNG) of 36,240 chest X-ray radiologic images, and it had been labeled as COVID-19, pneumonia, normal of 16,194, 15,773, and 4273 images, respectively. All the labeled images are in the form of a grayscale. The form of the images was changed to 224 × 224 pixels with 3 channels RGB with the help of the PIL library, hence providing more contrast between COVID-19 and other pneumonia diseases.

The images thereafter have been splitted into three packets, which are the training, testing, and validation samples. The training samples were about 80% of the entire data, and the testing and validation samples make up the remaining 20%. Table 3 shows the volume of the datasets.

The models, which are GoogLeNet, Standard CNN, and DN-121 were subjected to identical

Table 4. Training time.

Model	Time (sec)	No. of Epochs
Standrad CNN	58683.68	81
GoogleNet	114966.87	76
DN-121	509946.4	20

parameter values. Python programming language along with anaconda environment and several libraries such as Numpy, Keras, Scipy, PIL, Pandas, Seaborn, and Matplotlib was utilized to implement all frameworks. All experiments were carried out using NVIDIA graphical processing unit (GPU) RTX, 16 GB random access memory.

The standard CNN consists of successive 32, 64, and 128 filters in 3 × 3 kernel size convolutional layers and terminates with 32 filters convolutional layers of 3× 3 kernel and 4× 4 max pooling. A pooling layer with a pooling size of 3× 3 is embedded between every two consecutive layers. After building the structure, the flattened vector has 1568 elements, which represent the array image of 7× 7× 32. This architecture is supposed to own the superiority to gain the best accuracy, since the flattened vector is larger than that of the other model, and this size will be the same as the neurons in the fully connected layer. This would help in comparing the two fairly. Table 1 shows the entire architecture for this model. After building the structure, the trainable parameters are about 2,604,291. We have constructed several models of standard CNN; however, this architecture has the best accuracy so far.

The DN-121 network has been constructed with the same architecture that is described in the previous section.

DN-121 contains four dense blocks with 6, 12, 24, and 16 (1 ×1) and (3 × 3) convolutional layers respectively, along with three transitional layers each with a (1× 1) convolutional layer and (2× 2) pooling layer. The flattened vector has 1024 elements passed to a 1024 dense network. The total trainable parameters are about 7,040,579.

GoogleNet starts with two convolution-pooling combinations followed by nine inception blocks with a pooling layer after blocks 3 and blocks 7,

Table 5. Accuracy values.

Model	Training Accuracy	Testing Accuracy	Best epoch
Standard CNN	0.9436	0.9451	70
DN-121	0.9969	0.9818	17
GoogLeNet output 1	0.9413	0.9392	65
GoogLeNet output 2	0.9587	0.9536	
GoogLeNet output 3	0.9638	0.9588	

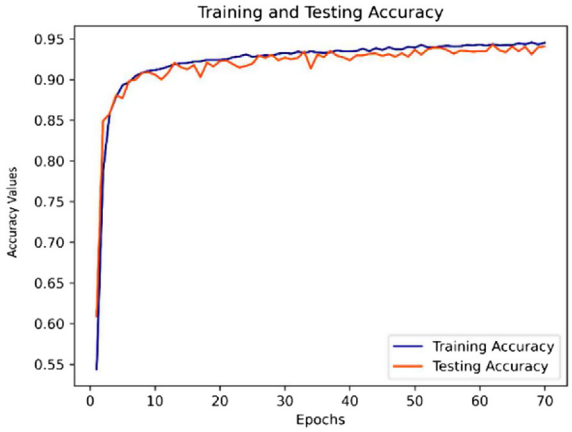

Fig. 6. Training versus testing accuracy in Standard CNN.

thereafter an average pooling layer and a dropout rate of 60%. The total trainable parameters for this model are 8,477,497. See Fig. 4.

For all models, Softmax function is used to extract the classed output, and all were trained in the initialized random weights and bias terms by employing the adaptive moment estimation (ADAM) to optimize the decision variables of the cross-entropy function. For ADAM, the $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon =$ 1e-07. All models were trained by a maximum of 300 epochs, however, they were all subjected to a patience technique that aims to hold the training process if the loss function value does not improve in 10 concussive epochs.

4.2. Evaluation metrics

In order to measure the performance of the deep transfer, four criteria were considered along with the confusion matrix, they are as follows:

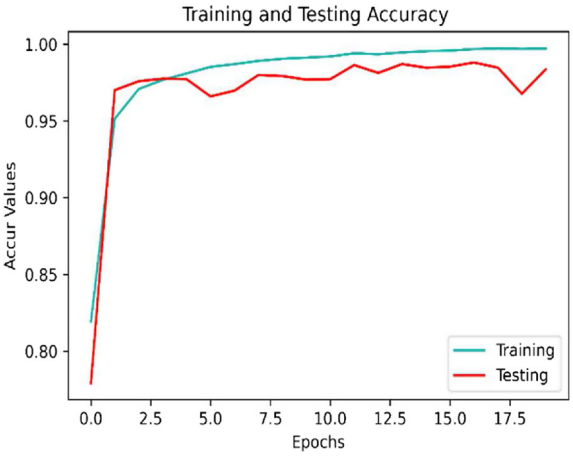

Fig. 7. Training versus testing accuracy in DN-121.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{F1 Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

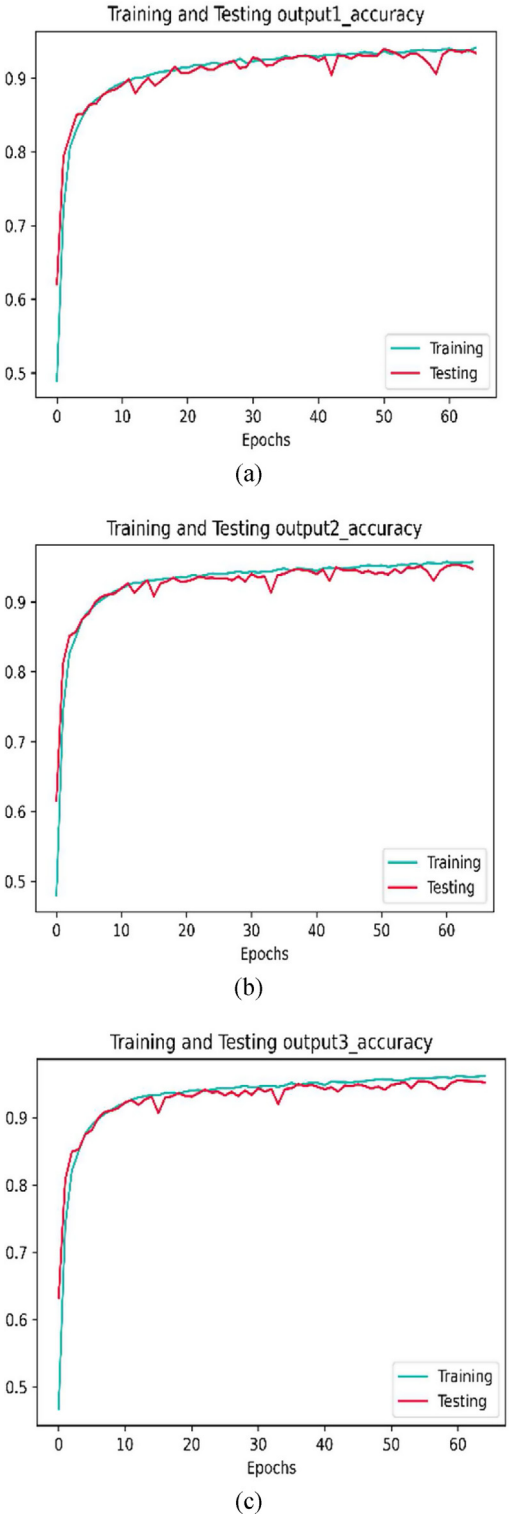

Fig. 8. (a,b,c) Training versus testing accuracy in GoogLeNet.

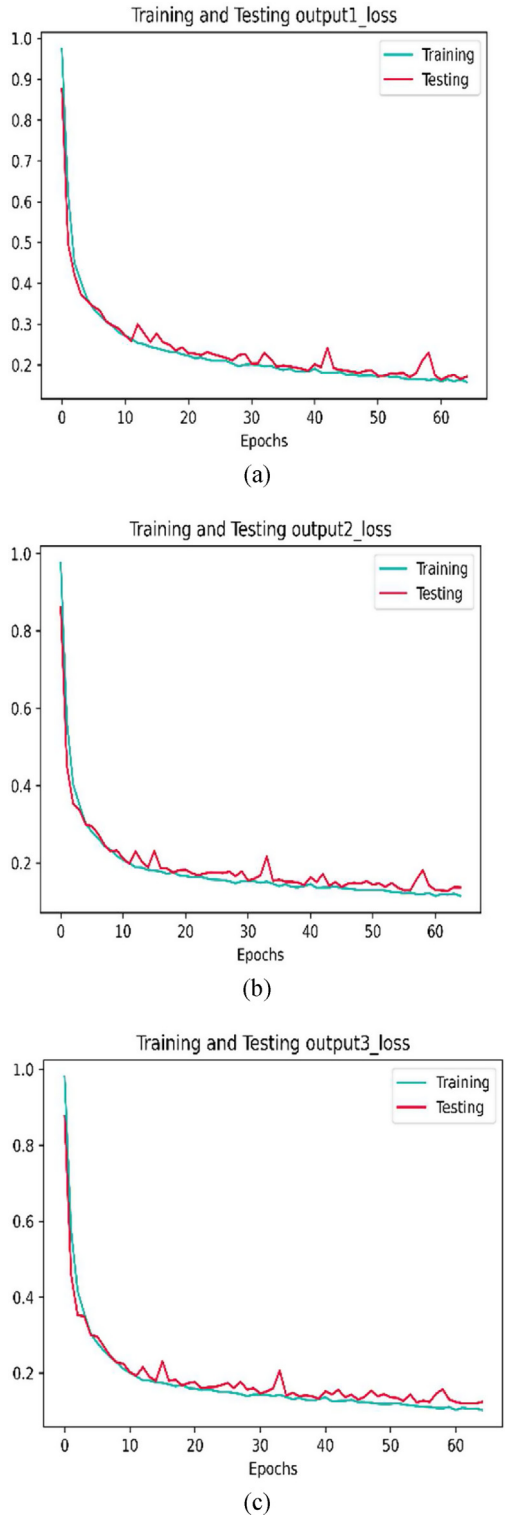

Fig. 9. (a,b,c) Training versus testing Loss in GoogLeNet.

Where TN, FP, TP, and FN in equations (6)–(9) refer to the number of true negative, false positive, true positive, and false negative cases, respectively. Given a model and test dataset, when considering the COVID-19 images as a target for evaluation, TN is the number of negative cases (normal or pneumonia) that are labeled clearly as either normal or pneumonia; FP is the number of negative cases (normal or pneumonia) that are incorrectly labeled as positive (COVID-19); TP refers to the number of positive (COVID-19) cases that are clearly labeled as COVID-19; and FN is the number of positive (COVID-19) cases that are incorrectly labeled as normal or pneumonia.

4.3. Experimental result

In this work, we constructed three different CNN networks for a classification CAD system that attempt to predict the class of X-ray images, which are COVID-19, pneumonia, and normal. While 80% of dataset is packed for training, the other remaining 20% is packed to be testing.

Table 4 shows the computational complexity for all three models. The training time is defined as the time from the starting epoch till the time at which the loss function is stable at a range of values, and stops improving.

Table 5 shows the training and testing accuracy of the best epochs obtained so far. DN-121 owns the best training accuracy. Output 3 in GoogleNet also performs well and has better training and testing accuracy than standard CNN and the other two outputs. The standard CNN trains better than output 1 in GoogLeNet.

Figs. 6–8, which are plotted using Spyder IDE, show the training and testing accuracy versus the number of epochs. Moreover, Figs. 9–11 show the loss function versus the number of epochs for the standard CNN, DN-121, and GoogLeNet. GoogLeNet has three dense branches embedded in its structure; hence, Fig. 8 shows all the output of those branches.

It's noticeable that the performance of DN-121 is better and has been adequately trained more than the other models in terms of the accuracy and the number of epochs required to reach the accuracy peak when epochs are considered as a limit of the patience factor. Further, the loss function in DN-121 is intuitively optimized and stabilized quickly.

The term standardization refers to the ability of a model to be familiar with various datasets that have never been used to train or test that model. A model gains standardization if it has a fair difference between its training and testing accuracy values. Considering the figures, it is clear that the standardization in DN-121 is also better as the difference between the two lines of training and testing shows enough contrast, which is not the case in the standard CNN or the GoogLeNet as the two lines match for a very long period of an epoch.

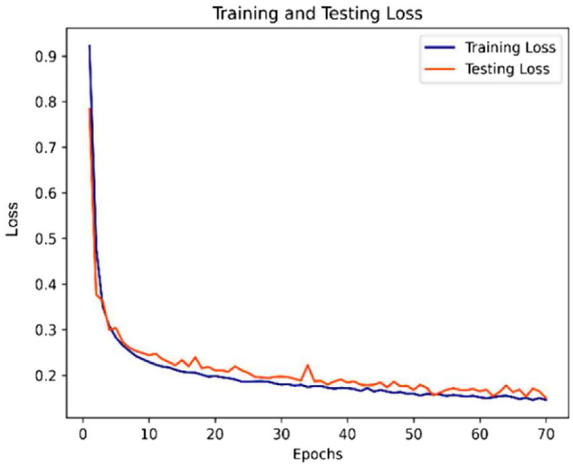

Fig. 10. Training versus testing Loss in Standard CNN.

Fig. 12 shows the confusion matrix (CM) of the standard CNN, DN-121, and GoogLeNet respectively. Those CMs are generated based on newly shuffled data different from the original testing data set. It can be seen that DN-121 with no more them 20 epochs performs a better prediction of COVID-19 images than the 81 epochs of standard CNN and the 76 epochs of GoogLeNet.

Among those new datasets, DN-121 is still the best model for COVID-19 detection so far. Nevertheless, the standard CNN predicts the COVID-19 X-rays better than output 2 of the GoogLeNet, and further,

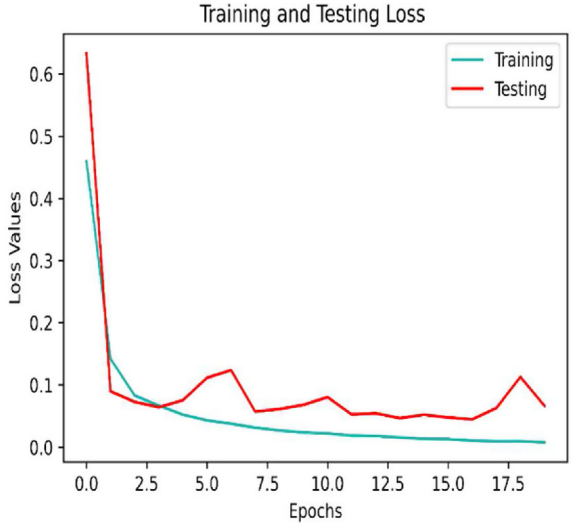

Fig. 11. Training versus testing Loss in DN-121.

output1 performs better at predicting the COVID-19 X-rays, which would lead to a standardization problem. Tables 6–8 list all the classification reports of each model and the first output branch of GoogLeNet, which is the best GoogLeNet output for newly shuffled data, respectively. Those tables also contained the Macro and weighted average. The Macro average gives equal weight to all classes

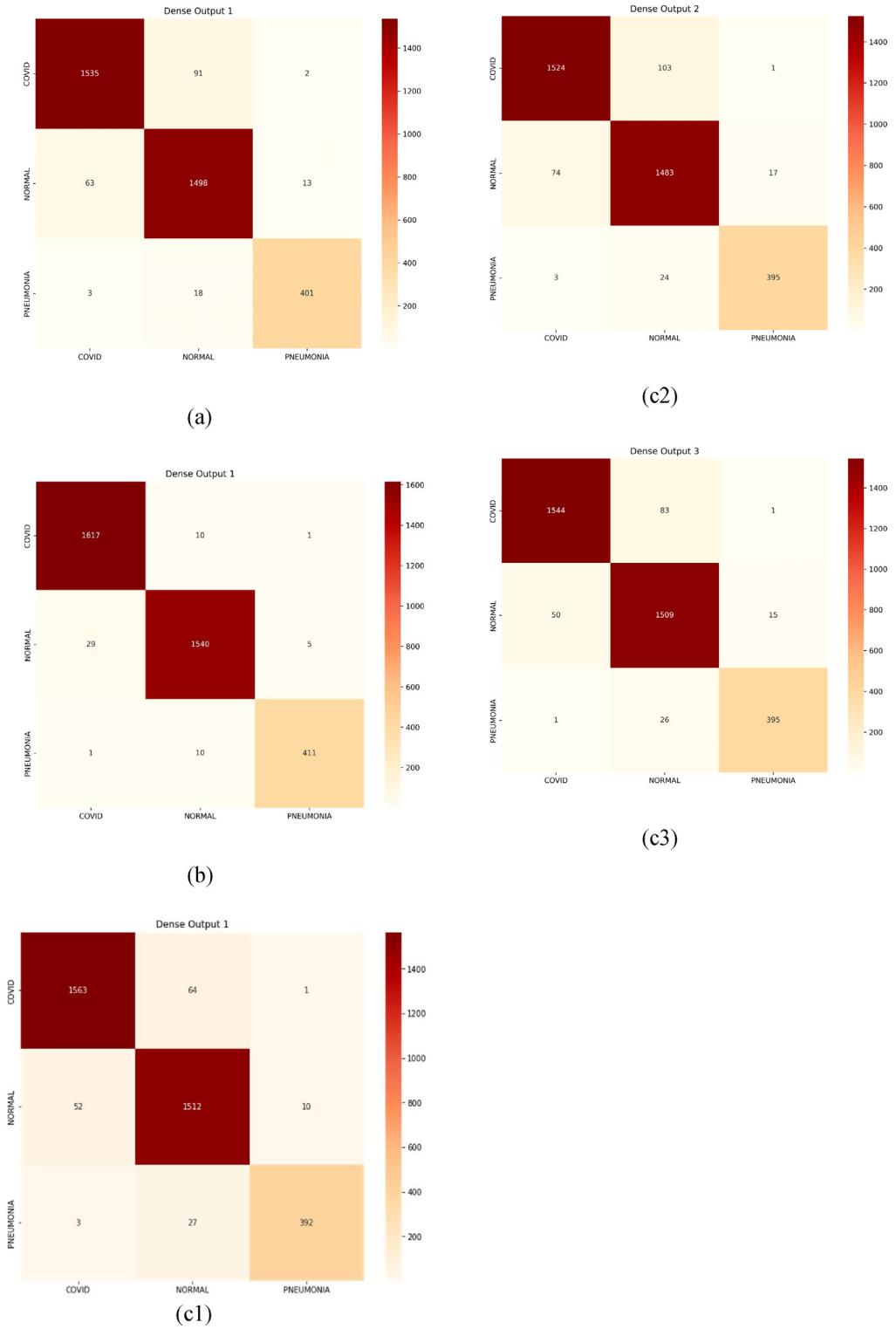

Fig. 12. The Confusion matrix of (a): Standard CNN, (b): the DN- 121, and (c1), (c2), and (c3): the three outputs of GoogLeNet.

Table 6. Classification report of CNN.

Averages	Category	precision	Recall	f1-score
–	COVID-19	0.96	0.94	0.95
–	Normal	0.93	0.95	0.94
–	PNEUMONIA	0.96	0.95	0.96
macro avg		0.95	0.95	0.95
weighted avg		0.95	0.95	0.95

Table 7. Classification report of DN-121.

Averages	Category	precision	Recall	f1-score
	COVID-19	0.98	0.99	0.99
	Normal	0.99	0.98	0.98
	PNEUMONIA	0.99	0.97	0.98
macro avg		0.98	0.98	0.98
weighted avg		0.98	0.98	0.98

Table 8. Classification report of output 1 GoogLeNet.

Averages	Category	precision	Recall	f1-score
	COVID-19	0.97	0.96	0.96
	Normal	0.94	0.96	0.95
	PNEUMONIA	0.97	0.93	0.95
macro avg		0.96	0.95	0.96
weighted avg		0.96	0.96	0.96

contributing to the final average metric, while the weighted average considers each class by its size.

5. Conclusion

DL techniques have revolutionized the state of the art through developing CAD radiography systems for detecting the infection of diseases. Unlike the previous works, where a small amount of datasets were used, this work contributes to training three models of CNN, which are the standard CNN, DN-121, and GoogLeNet to work on a large number of COVID-19 X-rays for a realistic performance evaluation of those three models. The works show that the DN-121 network gives the best performance as it detects COVID-19 in a less number of epochs than the other two CNN models. Further, the standard CNN, which is the simplest basic class of CNN, gains the superiority to detect COVID-19 infections over the second output of GoogLeNet are compared.

Tow limitations can be listed in our work. In recent times, many works claim that GoogLeNet outperformed the standard CNN in all aspects, however, this work shows that the standard CNN outperformed the second outputs of GoogLeNet in predicting the COVID-19 radiographies, even with a same amount of maximum of 300 epochs, but we still don't provide a clear explanation on such superiority, and thus this phenomenon will be discussed in future work.

Conflict of interest

There is no conflict of interest. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgment

The authors would like to thank Mohammed Aljubaily, and Kaggle, for their support in data collection.

References

[1] A.H. Abed, D.A. Abdulwahid, H.A. Jassim, National health systems response to COVID-19 outbreak, Iraq an example, The Medical Journal of Basrah University 39 (2021) 1–6, https://doi.org/10.33762/mjbu.2021.130218.1071.

[2] L. Wang, Z.Q. Lin, A. Wong, Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, Sci. Rep. 10 (2020) 1–12, https://doi.org/10.1038/s41598-020-76550-z.

[3] Q. Yan, B. Wang, D. Gong, C. Luo, W. Zhao, J. Shen, J. Ai, Q. Shi, Y. Zhang, S. Jin, L. Zhang, Z. You, COVID-19 chest CT image segmentation network by multi-scale fusion and enhancement operations, IEEE Transactions on Big Data 7 (2021) 13–24, https://doi.org/10.1109/TBDATA.2021.3056564.

[4] E.T. Khalid, E.B. Talal, M.K. Faraj, A.A. Yassin, Sentiment analysis system for COVID-19 vaccinations using data of Twitter, Indones. J. Electr. Eng. Comput. Sci. 26 (2022) 1156–1164, https://doi.org/10.11591/ijeecs.v26.i2.pp1156-1164.

[5] G. Liang, L. Zheng, A transfer learning method with deep residual network for pediatric pneumonia diagnosis, Comput. Methods Progr. Biomed. 187 (2020) 1–9, https://doi.org/10.1016/j.cmpb.2019.06.023.

[6] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, Deep learning for healthcare applications based on physiological signals: a review, Comput. Methods Progr. Biomed. 161 (2018) 1–13, https://doi.org/10.1016/j.cmpb.2018.04.005.

[7] J. Ker, L. Wang, J. Rao, T. Lim, Deep learning applications in medical image analysis, IEEE Access 6 (2017) 9375–9389, https://doi.org/10.1109/ACCESS.2017.2788044".

[8] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88, https://doi.org/10.1016/j.media.2017.07.005.

[9] N.C.F. Codella, Q.B. Nguyen, S. Pankanti, D. Gutman, B. Helba, A. Halpern, J.R. Smith, Deep learning ensembles for melanoma recognition in dermoscopy images, IBM J. Res. Dev. 61 (2017) 1–28, https://doi.org/10.1147/JRD.2017.2708299.

[10] M.J.J. Ghrabat, Z.A. Hussien, M.S. Khalefa, Z.A. Abduljabba, V.O. Nyangaresi, M.A. Al Sibahee, E.W. Abood, Fully automated model on breast cancer classification using deep learning classifiers, Indones. J. Electr. Eng. Comput. Sci. 28 (2022) 183–191, https://doi.org/10.11591/ijeecs.v28.i1.pp183-191.

[11] A.Z. Atiyah, K.H. Ali, Brain MRI images segmentation based on U-net architecture, IJEEE Journal 18 (2022) 21–27, https://doi.org/10.37917/ijeee.18.1.3.

[12] A. Ghani, C.H. See, V. Sudhakaran, J. Ahmad, R. Abd-Alhameed, Accelerating retinal fundus image classification using artificial neural networks (ANNs) and reconfigurable hardware (FPGA), Electronics 8 (2019) 1–17, https://doi.org/10.3390/electronics8121522.

[13] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nat. Med. 25 (2019) 65–69, https://doi.org/10.1038/s41591-018-0268-3.

[14] V. Chouhan, S.K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damaševicius, V.H.C. de Albuquerque, A novel transfer learning based approach for pneumonia detection in chest X-ray images, Appl. Sci. 10 (2020) 1–17, https://doi.org/10.3390/app10020559.
[15] J.C. Souza, J.O.B. Diniz, J.L. Ferreira, G.L.F. da Silva, A.C. Silva, A.C. de Paiva, An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks, Comput. Methods Progr. Biomed. 177 (2019) 285–296, https://doi.org/10.1016/j.cmpb.2019.06.005.
[16] R. Pramanik, S. Sarkar, R. Sarkar, An adaptive and altruistic PSO-based deep feature selection method for Pneumonia detection from Chest X-rays, Appl. Soft Comput. 128 (2022) 109464, https://doi.org/10.1016/j.asoc.2022.109464..
[17] P. Saha, M.S. Sadi, M.M. Islam, EMCNet: automated COVID-19 diagnosis from X-ray images using convolutional neural network and ensemble of machine learning classifiers, Inform. Med. Unlocked 22 (2021) 10–1016, https://doi.org/10.1016/j.imu.2020.100505.
[18] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, Physical and engineering sciences in medicine 43 (2020) 635–640, https://doi.org/10.1007/s13246-020-00865-4.
[19] Y. Song, S. Zheng, L. Li, X. Zhang, Z. Huang, J. Chen, R. Wang, H. Zhao, Y. Chong, J. Shen, Y. Zha, Y. Yang, Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images, IEEE ACM Trans. Comput. Biol. Bioinf 18 (2021) 2775–2780, https://doi.org/10.1109/TCBB.2021.3065361.
[20] X. Chen, L. Yao, Y. Zhang, Residual Attention U-Net for Automated Multi-Class Segmentation of Covid-19 Chest Ct Images, arXiv preprint arXiv:2004.05645 vol. 14 (2020) 1–7, https://doi.org/10.48550/arXiv.2004.05645.
[21] M. Turkoglu, COVID-19 detection system using chest CT images and multiple kernels-extreme learning machine based on deep neural network, Irbm 42 (2021) 207–214, https://doi.org/10.1016/j.irbm.2021.01.004.
[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, 1-9, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 7–12, https://doi.org/10.1109/CVPR.2015.7298594.
[23] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017-Janua, Honolulu, HI, USA. 2017, pp. 2261–2269, https://doi.org/10.1109/CVPR.2017.243.
[24] M.M. Mijwil, Implementation of machine learning techniques for the classification of lung X-ray images used to detect COVID-19 in humans, Iraqi J. Sci. 62 (2021) 2099–2109, https://doi.org/10.24996/ijs.2021.62.6.35.
[25] H. Mukherjee, S. Ghosh, A. Dhar, S.M. Obaidullah, K. Santosh, K. Roy, Deep neural network to detect COVID-19: one architecture for both CT Scans and Chest X-rays, Appl. Intell. 51 (2021) 2777–2789, https://doi.org/10.1007/s10489-020-01943-6.
[26] J. Zhang, Y. Xie, G. Pang, Z. Liao, J. Verjans, W. Li, Z. Sun, J. He, Y. Li, C. Shen, Y. Xia, Viral pneumonia screening on chest X-rays using confidence-aware anomaly detection, IEEE Trans. Med. Imag. 40 (2021) 879–890, https://doi.org/10.1109/TMI.2020.3040950.
[27] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks, Pattern Anal. Appl. 24 (2021) 1207–1220, https://doi.org/10.1007/s10044-021-00984-y.
[28] D. Singh, V. Kumar, M. Kaur, Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks, Eur. J. Clin. Microbiol. Infect. Dis. 39 (2020) 1379–1389, https://doi.org/10.1007/s10096-020-03901-z.
[29] N.C.D. Adhikari, Infection severity detection of CoVID19 from X-Rays and CT scans using artificial intelligence, Int. J. Comput. 38 (2020) 73–92.
[30] M. Umer, I. Ashraf, S. Ullah, A. Mehmood, G.S. Choi, COVINet: a convolutional neural network approach for predicting COVID-19 from chest X-ray images, J. Ambient Intell. Hum. Comput. 13 (2022) 535–547, https://doi.org/10.1007/s12652-021-02917-3.
[31] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med. 121 (2020) 1–11, https://doi.org/10.1016/j.compbiomed.2020.103792.
[32] E. Hemdan, M.A. Shouman, M.E. Karar, COVIDX-net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images, arXiv preprint arXiv:2003.11055 (2020) 1–14, https://doi.org/10.48550/arXiv.2003.11055.