A COVID-19 Visual Diagnosis Model Based on Deep Learning and GradCAM

Omar S. Hemied*, Non-Member
Mohammed S. Gadelrab***, Non-Member
Elsayed A. Sharara**, Non-Member
Taysir Hassan A. Soliman*, Non-Member
Akinori Tsuji****, Non-Member
Kenji Terada****, Senior Member

Recently, the whole world was hit by COVID-19 pandemic that led to health emergency everywhere. During the peak of the early waves of the pandemic, medical and healthcare departments were overwhelmed by the number of COVID-19 cases that exceeds their capacity. Therefore, new rules and techniques are urgently required to help in receiving, filtering and diagnosing patients. One of the decisive steps in the fight against COVID-19 is the ability to detect patients early enough and selectively put them under special care. Symptoms of this disease can be observed in chest X-rays. However, it is sometimes difficult and tricky to differentiate “only” pneumonia patients from COVID-19 patients. Machine-learning can be very helpful in carrying out this task. In this paper, we tackle the problem of COVID-19 diagnostics following a data-centric approach. For this purpose, we construct a diversified dataset of chest X-ray images from publicly available datasets and by applying data augmentation techniques. Then, we employ a transfer learning approach based on a pre-trained convolutional neural network (DenseNet-169) to detect COVID-19 in chest X-ray images. In addition to that, we employ Gradient-weighted Class Activation Mapping (GradCAM) to provide visual inspection and explanation of the predictions made by our deep learning model. The results were evaluated against various metrics such as sensitivity, specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV) and the confusion matrix. The resulting models has achieved an average detection accuracy close to 98.82%. © 2022 Institute of Electrical Engineers of Japan. Published by Wiley Periodicals LLC.

Keywords: machine learning; deep learning; data-centric ML; GradCAM; COVID-19 detection; diagnosis

Received 24 September 2021; Revised 14 January 2022

1. Introduction

Artificial Intelligence (AI) is playing an important role in medical applications and especially in diagnostics and disease detection. The critical situation and the challenges that were introduced by COVID-19 pandemic invites researchers to seek help in AI and Deep Learning (DL). As a branch of AI, DL is increasingly applied where it shows significant improvements. In many countries, during the first wave, health care systems have been overwhelmed because of the limited diagnostic groups, limited hospital beds for admission of these patients, lack of personal protection equipment (PPE) for healthcare professionals and limited ventilation equipment. The second and the third waves showed a sharp rise in the number of COVID-19 patients. It is expected that subsequent waves could add unprecedented burden on health care systems worldwide, especially with the arrival of new, possibly vaccination-resisting variants of the virus. Several measures should be considered to alleviate this stressful situation. For example, it is important to differentiate patients with acute respiratory illnesses such as pneumonia who may have a COVID-19 infection in order to efficiently utilize the limited resources. In this work we suggest deep learning application on chest X-rays to detect COVID-19 infection in pneumonia patients. With our approach we can classify X-rays into one of three categories: Normal, Pneumonia, COVID-19. The use of X-rays has several advantages over conventional diagnostic tests. For example: (a) The X-ray imaging is widespread and provides good means in diagnostic tests, besides its relative low cost. (b) Transmitting digital X-ray images does not require expensive means. It can be transmitted from the acquisition point to the point of analysis via the Internet, which makes the diagnosis process very fast. Sometimes, it may be difficult for some radiologists to notice the very subtle differences between pneumonia and COVID-19. This could be due to the lack of experience or because of the work stress. A computerized diagnosis system can provide a great help in the fight against COVID-19 by supporting medical teams to differentiate accurately COVID-19 patients from others. Such
AI-based tool can serve as a guide for radiologists to clarify the nuances. However, the proposed model is not supposed to completely replace the conventional diagnostic tests for COVID-19 infection. Rather, it can be considered as an initial screening tool for triaging patients with severe acute respiratory infection to undergo further infection tests or start treatment preparation.

It is worth to note that scientists do not have yet a full understanding of the continuously evolving COVID-19 and its dynamics where researchers continue to discover new surprising information. Consequently, new datasets should be collected from various regions and the DL model should be updated regularly. This why we carry out this work despite similar research work has been already published. In general, such research is most likely valid in the light of the current information and the symptoms captured in the available X-ray image datasets. New research should be conducted to keep updating the ML models as we get newer information, new virus variants or new datasets in order to maintain model accuracy [1].

In this research we present a visualized diagnosis assistance model based on deep learning and GradCAM to detect COVID-19 infection from chest X-ray images. The model can differentiate between COVID-19 and pneumonia infections with ability to highlight the input regions that guided the classification model to the prediction decision. That can support medical teams to accelerate and increase the accuracy of the COVID-19 diagnosis. Our contributions in this paper can be summarized as follows:

1. Developing a visualized diagnosis assistance method for COVID-19 through a data-centric ML approach.
2. Creating models that can differentiate not only between the normal and COVID-19 but also between pneumonia and COVID-19 that sometimes represent a difficult task even for specialists.
3. Constructing a new dataset from two different datasets that consolidates the diversity of the data and its resilience to bias. Also using dataset augmentation techniques to enhance and increase dataset size.
4. Modifying DenseNet-169 neural network architecture as a pre-trained model to enhance COVID-19 detection and to add the ability for visualization. Besides that, tuning the hyper-parameters to enhance the accuracy.

The paper is organized as follows: Section 2 presents some relevant research work. Section 3 briefly describes our proposed method and explains how we constructed our dataset. In Section 4, we describe the details of the various stages of our deep learning approach and how we designed and trained our learning model. Moreover, it also introduces the basis of Class Activation Maps (CAM). Then, we present in Section 5 the evaluation metrics, the results, the GradCAM output and a comparison with results from previous work. Finally, we conclude the paper and discuss some directions for the future work in Section 6.

2. Related Work

Several AI techniques were applied for detecting pneumonia caused by COVID-19 from chest X-ray images. Among the challenges AI is facing is how the system can differentiate whether the detected pneumonia in the chest X-ray is caused by COVID-19. In common practice, many researchers believe that most deaths from COVID-19 are owing to pneumonia in the lungs of vulnerable patients. AI algorithms can be trained to detect pneumonia in real time. AI used to help the community in different ways such as early warnings and alerts; diagnosis and prognosis; tracking and prediction; treatments and cures; data dashboards; and social control. Blue Dot [2], a Canadian AI model informed users about the COVID-19 outbreak on 31st December 2019 before World Health Organization (WHO) announced on 9th January 2020. Not only that, but researchers also working on Blue Dot published an article in the Journal of Travel Medicine in January 2020 [3] that predicts the spread of the virus in 20 different cities in the world from Wuhan, China by travelers. Another AI-based model is HealthMap [4], which also gave an early alert of significant outbreak of COVID-19 back on December 2019 at Boston Children’s Hospital in USA. These examples, show that AI-based models can accelerate response to save many lives.

AI can also help in predicting and controlling the spread of the disease. Scientists working with UN Global Pulse, reviewed several AI-based applications for the detection of COVID-19. They mentioned that both X-ray and Computed Tomography (CT) scans can be used for the detection of COVID-19 based on AI models. They also proposed the idea of using a mobile phone to scan CT images for the detection of COVID-19 [5]. Following the previous Zika-virus pandemic in 2015, AI-based system was developed to predict the spread of the disease [6]. Such existing models can be utilized for COVID-19 after re-training with data related to COVID-19. The algorithm trained for the prediction of seasonal flu can also be re-trained with new data from COVID-19 [7]. MIT technologies provided several lists to the AI-based dashboard for tracking and forecasting the outbreak of COVID-19 such that HealthMap [4], Next Train [8], and Upcode [9]. These dashboards provide a global view of COVID-19 outbreak in all the countries. Lock-down and social distancing are considered the preliminary and basic precautionary measures by healthcare department. AI can be used to control the scanning of people in crowded or potentially affected areas such as railway stations. In China, infrared camera has been used to detect the human body temperature for the prediction of fever against COVID-19 [1].

On the treatment side, AI was contributing long time before COVID-19 outbreak in discovering potentially emerging drugs. Upon the outbreak of COVID-19, several research lab indicated the need of AI to search the vaccine for the treatment. Scientists believe that the use of AI model can accelerate the process of searching cure and developing vaccines for COVID-19. Some research has already investigated the structure of the protein of COVID-19 that was predicted by Google’s DeepMind and suggested that it probably can provide useful information for the discovery of the vaccine [10].

Another paper works on the early detection and treatment of pneumonia to reduce mortality rates among children [11]. It presents Convolutional Neural Network models (CNN) to detect pneumonia using X-ray images. Several CNNs were trained to classify X-ray images into two classes: pneumonia and non-pneumonia, by tuning the hyper-parameters and the number of convolutional layers. Six models have been mentioned in the paper and the dataset used is available on Kaggle under the name “Chest X-Ray Images (Pneumonia)” (dataset URL is provided in Table II, 1st row). This 1.16 GB dataset contains 5216 images for training and 624 images for testing. Images in this dataset are grayscale with the dimension of $64 \times 64$. The dataset consists of three types of images: Normal, Bacterial Pneumonia, and Viral Pneumonia. The first and second models achieve a validation accuracy of...
Another interesting work applies deep learning to detect COVID-19 patients from their chest radiography images [12]. Its dataset contains about 5000 images that are a mix of chest X-ray and CT images from the publicly available datasets. Images exhibiting COVID-19 disease presence were identified by board-certified radiologist. Transfer learning on a subset of 2000 radiograms was used to train four popular convolutional neural networks. The Deep learning methods used as transfer learning approach include ResNet18, ResNet50, SqueezeNet, and DenseNet-121. They were evaluated on the remaining 3000 images where mostly achieved a sensitivity rate of 98% (±3%), while having a specificity rate of around 90%. Besides sensitivity and specificity rates, the results present the receiver operating characteristic curve (ROC), precision-recall curve, average prediction, and confusion matrix of each model. Moreover, heat-maps of lung regions potentially infected by COVID-19 were generated and their results show that the generated heatmaps contain most of the infected areas annotated by the board-certified radiologist. Their best performing model achieved a sensitivity rate of 98%, while having a specificity of 92% [12].

In Ref. [13], the authors present their approach for solving COVID-19 diagnosis problem based on two AI-based methods for classification and diagnosis from lung MRI images. The first technique consists of classification and feature extraction based on ANN and fractal methods. Moreover, they presented CNN-based segmentation methods that can determine infected lung tissues in the MRI images. Besides that, the authors presented the results of two algorithms for the diagnosis: Deep Neural Network (DNN) on the fractal feature of images and Convolutional Neural Network (CNN) on lung images and segmentation. CNN architecture showed higher accuracy (93.2%) and sensitivity (96.1%) than the DNN method with an accuracy of 83.4% and sensitivity of 86%. Results show that the presented method can almost detect infected regions with high accuracy of 83.84%. This finding also can be used to monitor and control patients from infected region growth [13]. Shibly et al. [14] applied faster R-CNN to detect COVID-19 from chest X-ray images using a dataset similar to [18].

Chowdhury et al. has proposed their own approach for COVID-19 detection [15]. They have trained eight models: SqueezeNet, MobileNetV2, ResNet18, InceptionV3, ResNet101, CheXNet, DenseNet-201 and VGG19. All the models were trained four times: first, to detect two-classes (COVID and Non-COVID) and secondly, to detect three classes (COVID, viral Pneumonia and Normal); both with and without image augmentation. They managed to achieve detection accuracy up to 97.94% using DenseNet-201 and 96.38% average accuracy of all the eight models on the three-class problem with augmented images. In Section 5, we compare our results with the results from these previous works and further explain the similarities and investigate the differences with our work.

### 3. Proposed Method

In this research, we aim to develop a diagnosis system with visual assistance that can classify a given frontal-view chest X-ray image into one of the following classes: Normal, Pneumonia and COVID-19. The motivation behind the three-class configuration is to provide better understanding in case of any confusion between regular pneumonia and COVID-19 that could be due to the similarity of pathological characteristics between COVID-19 and pneumonia.

To achieve that we follow a data-centric approach with significant focus on dataset construction. We start by analyzing available datasets and compare their features in order to select relevant images to be included in our dataset. Then, we generate more images by applying data augmentation technique to extend the dataset and to avoid over-fitting. Once the dataset has been constructed, we start applying a transfer deep learning process on DenseNet-169. The output of the training stage is a trained model that can classify X-ray images into one of three classes: Normal, Pneumonia and COVID-19. We treat each class as a categorical classification problem of the input that consists of a frontal-view chest X-ray image and the output represents a binary that results from the activation of the output layer ‘SoftMax’. Besides that, we use categorical cross entropy as a loss function. The results will be evaluated against various metrics such as sensitivity, specificity, PPV and NPV, and the confusion matrix. On top of that, we add GradCAM layers to visualize the parts of the image that led to the classification (i.e., the diagnosis) decision.

#### 3.1. Dataset

Unfortunately, few datasets are publicly available on COVID-19. After comparing several datasets, we decided to select images from (COVID-19 and Pneumonia) [16] (as Sub- Dataset1) with data augmentation [19] and COVID-19 Radiography Database [17] (as Sub- Dataset2). Both has a reasonable and diverse amount of the three types of chest X-ray images: COVID-19, pneumonia and normal. Moreover, the two sub-datasets have collected images from different sources. Hence, it will be more likely that there is no overlapping or duplicates when we combine the two sub-datasets. We have constructed our dataset as illustrated in Fig. 1. We combined each type of the X-ray images from the two sub-datasets together. We used all the available images for COVID-19 and Pneumonia while we selected only 85.26% and 92.31% respectively. The accuracy of the remaining models VGG16, VGG19, ResNet50 and Inception-v3 are 87.28%, 88.46%, 77.56% and 70.99% respectively.

![Fig. 1. Construction of our composed dataset](image-url)
4. Deep Learning

Transfer learning emerged as a popular method in computer vision because it allows accurate models to be built [20,21]. With transfer learning, a model learned from a domain can be re-used on different domains. It can be performed with or without a pre-trained model. A pre-trained model is a model developed to solve a similar task. Instead of creating a model from scratch to solve the new task, the model that was trained on another problem is utilized as a starting point. Even though a pre-trained model is trained on a task which is different from the original task, the features learned, in most cases, found to be useful for the new task. The objective of training a deep learning model is to find the correct weights for the network by numerous forward and backward iterations. By using pre-trained models that have been previously trained on large datasets, the weights and architecture obtained can be used and applied to the current problem [22]. One of the advantages of pre-trained models is the reduced cost and the accelerated time of training for the new model [23] thanks to the pre-trained weights. The new model only has to learn the weights of the last few layers. Many CNN architectures are pre-trained on ImageNet [24].

Dense Convolutional Network (DenseNet) is another popular architecture [5], which was the winner of the 2017 ImageNet competition. In DenseNet, each layer obtains additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers. Each layer is receiving a ‘collective knowledge’ from all preceding layers [12]. Since each layer receives feature maps from all preceding layers, network can be thinner and compact (i.e., the number of channels can be fewer). This why it has higher computational efficiency and memory efficiency [25].

4.1. Design of DL model

As shown in Fig. 3, we designed the learning model such that it contains pre-trained (transfer learning) with a 169-layer Dense Convolutional Network (DenseNet-169) backbone, followed by our customized, fully connected layers. We replaced DenseNet top layer with 3 layers: Convolution layer with 512 filter, (3×3) kernel size and ReLU as activation function + Global average Pooling + final layer with SoftMax. The final classifier produces one of three classes (COVID-19, Normal, Pneumonia). Each of these classes with a SoftMax activation to produce the final output as [0,1]. Fig. 4 shows the detailed structure of DenseNet-169.

4.2. Training

We initialize our DL model, that was described in the previous section, with pre-trained weights for DenseNet169 implementation by xhlulu on Kaggle [20]. Then, the training process goes through two stages as described hereafter. Firstly, DenseNet backbone weights of the first 295 layer from 595 DenseNet layers are frozen and only the remaining DenseNet layers and our layers are trained. Then, training is performed using RMSprop optimizer with learning rate parameter set to 10−4. We use categorical cross entropy as a loss function and mini batches of size 50, and train for about 50 epochs. All images are down sampled to 224 × 224 before being fed to the neural network (as these pre-trained models are usually trained with a specific image resolution). All our implementations are done in TensorFlow and Keras framework.

4.3. Class activation map

To further improve our approach, we use Class Activation Map (CAM) to check and
explain the prediction of the Convolutional Neural Networks. In image classification problems, deep learning is considered a black box because it does not give explanations about what the network has learned or which part of the input of the network was responsible for the prediction (decision) of the network. By applying GradCAM [26] and [27] which is a CAM variation, on our CNN models they become more transparent as we can visualize input regions that are significant in making a prediction or a decision. Class activation mapping helps to identify the bias in the training set and thus increases model accuracy. For example, if we discover that the network bases its predictions on wrong features, then we can make the network more robust by removing bad-quality data, collecting better data or modifying the feature set. CAM can be produced from image classification CNNs where the Global average pooled convolutional feature maps are fed directly into the output layer with ‘SoftMax’ as an activation function. Specifically, if the last CNN layer produces \( K \) feature maps \( A^k \in \mathbb{R}^{u \times v} \) of width \( u \) and height \( v \), these feature maps are then spatially pooled using Global Average Pooling 2D (GAP) and linearly transformed to produce a score \( y_c \) for each class \( c \), as shown by Equation (1):

\[
y_c = \sum_k w_k^c \frac{1}{Z} \sum_i \sum_j A^k_{ij}
\]  
(1)

GradCAM computes the gradient of a differentiable output, for example class score, with respect to the convolutional features in the chosen layer. The gradients are spatially pooled to find the neuron importance weights. These weights are then used to linearly combine the activation maps and determine which features are most important to the prediction. Suppose that we have an image classification network with output \( y^c \), representing the score for class \( c \), to compute the GradCAM map for a convolutional layer with \( k \) feature maps (channels), \( A^k_{ij} \), where \( i, j \) indexes the pixels. The neuron importance weight becomes:

\[
\alpha^c_k = \frac{1}{N} \sum_i \sum_j \frac{\partial y_c}{\partial A^k_{ij}}
\]  
(2)

To ensure that we get only the positive features of the class of interest we apply the rectified linear activation function (ReLU):

\[
M = RLU\left(\sum_k \alpha^c_k A^k\right)
\]  
(3)

The output is therefore a heatmap for the specified class that has the same size as the feature map.

5. Results and Comparison

5.1. Evaluation metrics

Four metrics are used to evaluate our results: sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) [28,29].

5.1.1. Sensitivity

Sensitivity measures how often a test generates correctly a positive result for people who have the condition that has being tested for (also known as the ‘true positive’ rate). A test that is highly sensitive will flag almost everyone who has the disease and not generate many false-negative results.
5.1.2. Specificity  Specificity measures a test ability to correctly generate a negative result for people who do not have the condition that has being tested for (also known as the “true negative” rate). A high-specificity test will correctly rule out almost everyone who does not have the disease and will not generate many false-positive results.

5.1.3. PPV and NPV  PPV and NPV are directly related to prevalence and allow to clinically say how likely it is that a patient has a specific disease.

In particular, the Positive Predictive Value (PPV) is the probability that following a positive test result, that individual will truly have that specific disease. The Negative Predictive Value (NPV) is the probability that following a negative test result, that individual will truly not have that specific disease.

5.2. Results  We trained our model for 50 epochs that finished training with approximately 99% accuracy. Fig. 5 shows the accuracy and the losses of the resulting model during the training and the validation stages. When tested against the Test Dataset, that was constructed as explained in Section 3.1, the model achieved an average accuracy 98.824%. Table I presents the results in terms of the evaluation metrics that were defined in the previous sub-section: Sensitivity, Specificity, PPV and NPV for each category. Regarding COVID-19, it has the following evaluation metrics: Sensitivity = 99.64%, Specificity = 99.50%, PPV = 98.69% and NPV = 99.86%. Besides that, Fig. 6 shows the corresponding confusion matrix of the model.

To evaluate the validity of our results and the robustness of our model vis-a-vis the selected dataset and how it was split, we have used stratified K-fold cross-validation method. In this case, we set K = 5 folds and repeated the experiment on the same dataset for 50 epochs for each fold. Fortunately, the 5-fold cross-validation results are very close to the results of the original experiment (that was 98.824%). with an average accuracy 98.4006%, minimum accuracy 98.171% and maximum accuracy 98.661%.

Table I. Evaluation against the test dataset

|          | Sensitivity | PPV  | Specificity | NPV  |
|----------|-------------|------|-------------|------|
| COVID-19 | 99.64%      | 98.69% | 99.50%      | 99.86%|
| Normal   | 97.72%      | 99.16% | 99.54%      | 98.73%|
| Pneumonia| 99.28%      | 98.67% | 99.22%      | 99.58%|
| Accuracy | 98.824%     |       |             |      |

5.3. GradCAM output  We show in this section the results of the visualization model by Gradient-weighted class activation mapping (GradCAM). Firstly, the original X-ray scan image and the GradCAM are shown separately. Then, the GradCAM and the original scan are superimposed to produce a new image that displays the localization of the image part (features) that the CNN model used to make its decision. In other words, to some extent it can identify the localization of the visual symptoms of COVID-19 or pneumonia. The visualization output was validated through a simple comparison with a separate and blind identification of the disease spots on same image samples by two physicians and fortunately there was a promising match with our GradCAM output. All these results were confirmed by a specialist of a chest radiologist. Fig. 7 displays sample images with the visual symptom spots highlighted.

With the aid of GradCAM, we tried to understand the causes of the classification errors of our model. When we examined the miss-classified images, we noticed that some images have either low resolution, poor quality (i.e., have some noise, undesirable shadows or blurry contents) or both, as shown in Fig. 8. This could be inherent to the original images or could be introduced later due to poor scanning devices during the digitalization process of paper-based images. The GradCAM output for most miss-classified images seems to be meaningless. In most of these
5.4. Comparison completeness, in this section we compare our work with the relevant works that were discussed in Section 2. Not only we compare the results, but we also compare the ML method, the categories, the utilized datasets, dataset types and the evaluation metrics. We noticed that all the discussed papers share the common goal to diagnose Pneumonia and/or COVID-19 and sometimes have common datasets. However, they differ in other several aspects. For example, they apply different ML methods and even if they apply same ML methods they are applied differently. Most are using variations of CNN pre-trained models. Some work is capable of distinguishing COVID from Non-COVID [12,13,15] while others (such as [14,15] and ours) offer to differentiate between COVID-19, Normal and Pneumonia. One work is focusing on detecting Normal, Bacterial Pneumonia and Viral Pneumonia [11]. Ref. [15] deals with both the 2-class and the 3-class detection. Dataset types vary between X-ray, CT and MRI images. The results of the compared models range from 70.99% to approximately 99% accuracy. Our model has an average accuracy of 98.824%. While it has achieved an Avg Sensitivity: 98.88%, Avg PPV: 98.84%, Avg Specificity: 99.42% and Avg NPV: 99.39%.

Table II summarizes the results and broadly compares our work with some relevant previous work to position our work with respect to others and to provide readers with an overview of similar research. It provides a broad idea about each paper such as: the applied ML method, output type, datasets and the resulting metrics. It is worth to note that our work mainly differs from the others in the way the dataset is constructed and split. Moreover, we follow a data-centric approach to train a single model rather than training several models on the same data as most other papers did. Having a detailed comparison is out-of-the-scope of the current paper due to space limitations and because of the difficulty to reproduce the same research work with exactly the same resulting metrics shown in the table.

6. Conclusion and Future Work

During the peak of its early waves, as COVID-19 pandemic spreads around the world, the number of cases was growing exponentially. While waiting for the upcoming waves, finding a method that can help in diagnosing COVID-19 through a cheap and fast method is fundamental to avoid overwhelming healthcare systems again. In this context, we propose a data-centric machine learning techniques to identify and differentiate between pneumonia and COVID-19 from X-ray images. Moreover, we apply GradCAM to highlight the input regions that guided the classification model to the prediction decision. In this paper, we have developed a deep learning diagnosis system for COVID-19 from chest X-ray images. We used transfer learning by DenseNet-169 Convolutional Neural Network while applying data augmentation techniques to generate more data and to avoid over-fitting. The resulting model employs GradCAM to provide highlights for the regions containing potential infections.

Our paper is distinguished from the previous work by two main aspects. First, by following a data-centric approach. It encompasses all the stages starting from fine-tuning a pre-trained convolutional model DenseNet-169 on the training dataset to the generation of the diagnosis results. Second, by adding the ability to generate the GradCAM that assists doctors visually to make accurate diagnosis for the diseases. For training and testing, we employed two publicly available datasets to construct our own dataset. When
| Topic                                                                 | ML Method                                                                 | Output Categories                        | Dataset Type     | Dataset                                                                 | Metrics                                                                 |
|----------------------------------------------------------------------|----------------------------------------------------------------------------|------------------------------------------|------------------|------------------------------------------------------------------------|------------------------------------------------------------------------|
| [11] Pneumonia detection in chest X-ray images using convolutional   | Six CNN models: two from scratch + four transfer learning: VGG16, VGG19,   | Normal, Bacterial Pneumonia, and Viral   | Chest X-ray      | https://www.kaggle.com/paultimothymooney/chest-xraypneumonia 5216       | Accuracy for the six models respectively: 85.26, 92.31, 87.28, 88.46, 77.56 and 70.99 |
| neural networks and transfer learning                                 | ResNet50, Inception-v3                                                     | Pneumonia                                | 624 testing      | [19]                                                                     |                                                                        |
| [12] Deep-COVID: Predicting COVID-19 from chest X-ray images using   | Four CNN Models (transfer learning): ResNet18, ResNet50, SqueezeNet,       | COVID-19, Non-COVID-19                   | Chest X-ray and CT | https://github.com/ieee8023/covid-chestxray-dataset. ~5 k dataset      | Most of these networks achieved a sensitivity rate of 98% (±3%), while   |
| deep transfer learning                                               | DenseNet-121, VGG19, ResNet101, Inception-v3, + heatmaps                 |                                          | images           | images constructed from two datasets. 2084 training 3100 test images    | having a specificity rate of around 90%                                 |
| [13] Diagnosis and detection of infected tissue of COVID-19 patients | Feature extraction based on ANN, fractal, CNN & segmentation               | COVID-19, Non-COVID-19                   | X-Ray image, lung MRI images | https://arxiv.org/abs/2003.13865 https://medicalsegmentation.com/COVID-19/ | - For classification: Accuracy: 93.2% Sensitivity: 96.1% - For segmentation: Accuracy: 83.84% Sensitivity: 97.36% |}
| based on lung x-ray image using CNN                                  |                                                                           |                                          |                  |                                                                        |                                                                        |
| [14] COVID faster R–CNN: A novel framework to Diagnose Novel         | Faster R–CNN                                                               | Normal, Pneumonia, COVID-19              | Chest X-ray      | https://kaggle.com/c/rsna-pneumonia-detection-challenge/data             |                                                                        |
| Coronavirus Disease (COVID-19) in X-Ray images                       |                                                                           |                                          |                  |                                                                        |                                                                        |
| [15] Can AI help in screening Viral and COVID-19 pneumonia?          | Eight Models: SqueezeNet, MobileNetv2, ResNet18, InceptionV3, ResNet101,  | - 2-classes: (COVID-19, Non-COVID-19)    | Chest X-ray      | https://kaggle.com/tawsifurrahman/covid19-radiography-database          | Avg results of the 8 models on the 3 classes problem: Avg Accuracy: 96.38% Avg Sensitivity: 96.37% Avg PPV: 96.50% Avg Specificity: 97.72% Avg F1-Score: 96.425% Avg Accuracy: 98.824% Avg Sensitivity: 98.88% Avg PPV: 98.84% Avg Specificity: 99.42% Avg NPV: 99.39% |
|                                                                     | CheXNet, DenseNet-201, VGG19                                              | - 3-classes: Normal, Viral Pneumonia,    |                  |                                                                        |                                                                        |
|                                                                     |                                                                           | COVID-19                                 |                  |                                                                        |                                                                        |
| Ours                                                                 | DenseNet-169 transfer learning + GradCAM activation mapping               | Normal, Pneumonia, COVID-19              | Chest X-ray      | (COVID-19 & Pneumonia) [17], (COVID-19 Radiography Database) [19]      |                                                                        |
compared to the results from previous work, our model managed to achieve 98.824% average accuracy. It is worth noting that the purpose of this work is to assist in the diagnosis during the triage of patients to reach an accurate separation in emergency medical support services.

As a future work, we plan to build a larger and more diverse dataset so we can apply more sophisticated deep learning techniques with proper depth in the samples. Moreover, we need to work on the newly discovered variants of COVID-19 such as delta, lambda and Omicron variants to ensure that our models remain resilient to variations in X-ray images that could occur due to the new generations of the virus. Moreover, we are planning to study the effects of various datasets on a fixed model as well as the behavior of various models against the same dataset, which deserves a separate paper.

One of the drawbacks of using open datasets on chest radiographs is the often lack of information related to different positioning methods while taking X-ray images. In general, many X-rays taken in a standing position are healthy, and those taken in a sitting position might reflect serious problems.

Acknowledgment

The authors would like to thank Dr. Mohamed Abdel-Tawab, Lecturer of radiology at Assiut University Hospitals, who provided medical insights and expertise that greatly assisted the research. We would like also to show our gratitude to the “anonymous” reviewers for their insightful comments on an earlier version of this manuscript to be further improved.

References

(1) Mangal A, Kalia S, Rajgopal K, Rangarajan K, Namboodiri VP, Banerjee S, Arora C. CovidAID: COVID-19 detection using chest X-ray. ArXiv200409803 Cs Eess, 2020. http://arxiv.org/abs/2004.09803 Accessed July 16, 2021.
(2) BlueDot: BlueDot: Outbreak Risk Software. https://bluedot.global/ Accessed July 16, 2021.
(3) Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of unknown etiology in Wuhan, China: Potential for international spread via commercial air travel. Journal of Travel Medicine 2020; 27(2). https://doi.org/10.1093/jtm/taa008.
(4) HealthMap. http://www.diseasedaily.com/about. Accessed July 16, 2021.
(5) Maghded HS, Ghafoor KZ, Sadiq AS, Curran K, Rawat DB, Rabie K. A novel AI-enabled framework to diagnose coronavirus COVID 19 using smartphone embedded sensors: Design study, ArXiv 200307434 Cs Q-Bio, 2020. http://arxiv.org/abs/2003.07434 Accessed July 16, 2021.
(6) Akhtar M, Kraemer MUG, Gardner LM. A dynamic neural network model for predicting risk of Zika in real time. Journal of BMC Medicine 2019; 17(1):1. https://doi.org/10.1186/s12916-019-1389-3.
(7) MIT Technology Review. https://www.technologyreview.com/s/615360/cdc-cmu-forecasts-coronavirus-spread/ Accessed July 16, 2021.
(8) Nextstrain/ncov/global. https://nextstrain.org/ncov/global Accessed July 16, 2021.
(9) ZP and U. Academy: COVID-19 Singapore Dashboard. https://www.againstCOVID19.com/singapore/dashboard Accessed July 16, 2021.
(10) Deepmind: Computational predictions of protein structures associated with COVID-19 DeepMind. https://deepmind.com/research/opensource/computational-predictions-of-protein-structures-associated-with-COVID-19 Accessed July 16, 2021.
(11) Jain R, Nugraph P, Gaurav Kataria V, Kaushik S, Jude Hemanth D. Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. Journal of Measurement 2020; 165:108046. https://doi.org/10.1016/j.measurement.2020.108046.
(12) Minaee S, Kafieh R, Sonka M, Yazdani S. And Ghazaleh Jamalipour Soufi: “Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning”, Journal of Medical Image Analysis 2020; 65:101794.
(13) Hassantabar S, Ahmadi M, Sharifi A. Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural network approaches. Journal of Chaos, Solitons and Fractals 2020; 140:110170. https://doi.org/10.1016/j.chaos.2020.110170.
(14) Shibly KH, Dey SK, Islam MT-U, Rahman MM. COVID faster R–CNN: A novel framework to diagnose novel coronavirus disease (COVID-19) in X-ray images. Journal of Informatics in Medicine Unlocked 2020; 20:100405. https://doi.org/10.1016/j.imu.2020.100405.
(15) Timothy P. Chest X-ray images (pneumonia). https://kaggle.com/paulimothymooney/chest-xray-pneumonia Accessed July 16, 2021.
(16) Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR, Khan MS, Iqbal A, Al-Emadi N, Reaz MBI, Islam TI. Can AI help in screening viral and COVID-19 pneumonia? IEEE Access 2020; 8:132665–132676.
(17) Patel P. Chest X-ray (Covid-19 & Pneumonia). https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia Accessed July 16, 2021.
(18) Deis A. Data Augmentation for Deep Learning, Medium, 2019. https://towardsdatascience.com/data-augmentation-for-deep-learning-4fe21da4eb9 Accessed July 16, 2021.
(19) Rahman T. COVID-19 Radiography Database. https://kaggle.com/tawisfurrahman/covid19-radiography-database Accessed April 13, 2021.
(20) Xhlulu: Keras DenseNet Weights. https://kaggle.com/xhlulu/densenet-keras Accessed August 24, 2021.
(21) Cao X, Wipf D, Wen F, Duan G, Sun J. A practical transfer learning algorithm for face verification. In 2013 IEEE International Conference on Computer Vision, Sydney, Australia: IEEE, 2013; 3208–3215. https://doi.org/10.1109/ICCV.2013.398.
(22) Kieu STH, Bade A, Hijazi MHA, Kolivand H. A survey of deep learning algorithm for face verification. In 2013 IEEE International Conference on Computer Vision, Venice, 2017:618–626. https://doi.org/10.1109/ICCV.2017.74.
(23) Wang C, Chen D, Hao L, Xuebo Liu Y, Zeng JC, Zhang G. Pulmonary image classification based on inception-v3 transfer learning model. IEEE Access 2019; 7:146533–146541. https://doi.org/10.1109/ACCESS.2019.2946000.
(24) Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Journal of ACM Communications 2017; 60(6):84–90. https://dl.acm.org/doi/10.1145/3065386 Accessed January 10, 2021.
(25) Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. ArXiv160806993 Cs, (2018). http://arxiv.org/abs/1608.06993 Accessed July 16, 2021.
(26) Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. GradCAM: Why did you say that?, ArXiv161107450 Cs Stat, (2017). http://arxiv.org/abs/1611.07450 Accessed April 20, 2021.
(27) Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. GradCAM: visual explanations from deep networks via gradient-based localization, 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017:618–626. https://doi.org/10.1109/ICCV.2017.74.
(28) Health News Review: Sensitivity, specificity and understanding medical tests. https://www.healthnewsreview.org/toolkit/tips-for-understanding-studies/understanding-medical-tests-sensitivity-specificity-and-positive-predictive-value/ Accessed July 16, 2021.
(29) Prarikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values.
Indian Journal of Ophthalmology 2008; 56(1):45–50. https://doi.org/10.4103/0301-4738.37595.

Omar S. Hemied (Non-Member) Omar has obtained a B.Sc degree in Software engineering (2021) with grade V. Good with honor from Computer Science Department, Faculty of Computers and Information, Assiut University, Egypt. He is currently working as machine learning and computer vision engineer at Aiactive Tech.

Mohammed S. Gadelrab (Non-Member) He is currently working as an assistant professor researcher at the National Institute of Standards, Egypt. His research interests span various areas including dataset engineering, security, SW testing, computer networks, applied ML. He earned a Master and PhD. in 2004 and 2008, respectively, from Toulouse University, France. Moreover, he has occupied several R&D positions in both academia and industry in France, Canada, Egypt where he gained solid multi-disciplinary experience.

Elsayed A. Sharara (Non-Member) He received his B.S. degree in communication and electronics from Al-Azhar University in Egypt, a Master and PhD. degrees in System Innovation Engineering from Tokushima University in Japan in 2018 and 2021 respectively. His current research interests include computer vision, deep learning, and fatigue detection.

Taysir Hassan. A. Soliman (Non-Member) She obtained her B.Sc. in Computer Science from Marymount University, USA in 1990. She got her M.Sc. in computational sciences from Cairo University in 1996 and her Ph.D. in Bioinformatics in 2004 from Ain Shams University, Egypt on a joint supervision with University of Connecticut, USA. She is currently working as the Dean of Faculty of Computers & Information, Assiut University, Egypt. Here research interests include data mining, data science, semantic data integration, biodiversity informatics, bioinformatics, big data analytics. Besides that, she was a visiting scientist at IBM, CAS Center, Cairo Division. She is also a member of IEEE, ACM SIGKDD, and ACM SIGBIO. Moreover, she supervised and currently leading and participating in several national and international research projects.

Akinori Tsuji (Non-Member) He received his B.E. degree from University of Tottori in 1998. He received his M.E. and Ph.D. degrees from University of Tokushima in 2002 and 2013, respectively. He has been an engineer in the Department of Information Science and Intelligent Systems at Tokushima University since 1998. His research interests include image processing and embedded systems. He is a member of the IEICE and IEEE.

Kenji Terada (Senior Member) He received a doctoral degree from Keio University in 1995. In 2009, he became a Professor in the Department of Information Science and Intelligent Systems, University of Tokushima. His research interests are in computer vision and image processing. He is a member of the IEICE, SICE, SCIE, and JSPE.