

COVID-19 Diagnosis from Medical Images Using Transfer Learning

Elaf Alshehri^a Manal Kalkatawi^b Felwa Abukhodair^b Khalid Khashoggi^c
Reem Alotaibi^b

^aKing Abdulaziz and His Companions Foundation for Giftedness and Creativity (Mawhiba), Riyadh, Saudi Arabia;

^bFaculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia; ^cFaculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia

Keywords

Coronavirus disease 2019 · Deep learning · CT scan · X-ray · Convolutional neural network

Abstract

Introduction: The novel coronavirus (COVID-19) originated in Wuhan, China, in December 2019. To date, the virus has infected more than 110 million people worldwide and claimed 2.5 million lives. With the rapid increase in the number of infected cases, some countries face a shortage of testing resources. Computational methods such as deep learning algorithms can help in such a situation to expedite and automate the diagnosis of COVID-19. **Methods:** In this research, we trained eight convolutional neural network models to automatically detect and diagnose COVID-19 from medical imaging, including X-ray and CT scan images. Those deep learning networks have a predefined structure in which we re-train on medical images to serve our purpose, which is called transfer learning. **Results:** We used two different medical images known as X-ray and CT scan. The experimental results show that CT scan achieved better performance than X-ray. Specifically, the Xception network model has achieved an overall performance on CT scan of 84%, 91%,

and 77% for accuracy, sensitivity, and specificity, respectively. That was the highest in all models that we trained. On the other hand, the same network model (Xception) was applied on X-ray and performed 69%, 83%, and 55% for accuracy, sensitivity, and specificity, respectively. **Conclusion:** The performance of our proposed model to detect COVID-19 from CT scan is acceptable and promising to start in the field. We target the medical sectors to help them by providing rapid and accurate diagnosis of COVID-19 cases using an alternative detection approach to the traditional ones.

© 2022 The Author(s).

Published by S. Karger AG, Basel

Introduction

Coronavirus disease 2019 (COVID-19) was first reported in Wuhan city, Hubei Province of China, in December 2019. COVID-19 emerged in China, and soon after, it was reported in many other countries worldwide [1, 2]. With the rapid spread and increase in the number of cases worldwide, on March 11, 2020, the WHO announced the novel COVID-19 outbreak as a pandemic with more than 118,000 cases and more than 4,000 deaths globally [3]. The most common symptoms of COVID-19,

as declared by the WHO, are fever, dyspnea, cough, myalgia, and headache [4]. COVID-19 can be defined as an infectious disease that has affected around 109,426,406 individuals worldwide and caused 2,419,363 deaths as of January 18, 2021 [5].

The current diagnostic test is based on reverse transcription polymerase chain reaction (RT-PCR). The test takes 6–8 h to show results [6]. Such a time is long compared with the continuously growing spread rate of COVID-19. Given the widespread and the lack of fast accurate tests, medical experts are motivated to study and research alternative testing methods that can be readily available, faster, and cheaper. Besides RT-PCR, medical images, such as chest X-ray and chest computed tomography (CT), have become a primary method to diagnose and manage patients with COVID-19 [7].

With the advancement of modern technology and artificial intelligence, there have been many contributions to leveraging machine learning to produce other diagnostic methods that can help detect COVID-19 automatically and rapidly. The literature shows that many research contributions focused on the automatic diagnosis of COVID-19 from X-ray images using deep learning models. However, chest CT scan images have higher sensitivity for diagnosing COVID-19 [8]. Therefore, there have been recent efforts studying the effectiveness of using deep learning algorithms using CT scan imaging for screening and testing for COVID-19, and the results are promising [9–11].

To support the research and development of computational methods for diagnosing COVID-19 from medical imaging, Yang et al. [12] built a dataset of CT scan images of COVID-19 findings and made it available for the researchers to use. The authors used this dataset for the experiments, but training deep learning models on a small dataset can easily lead to overfitting. This means that the model would perform well on the training data but generalizes badly on testing data. For this reason, they adopted two approaches: transfer learning and data augmentation. Transfer learning is used to leverage a large collection of data from a relevant domain to help with the learning in the domain of interest [13]. Specifically, they used a large collection of chest X-ray images to pre-train CNN and then used the pre-trained network, with a learned set of weights in the deep network, on the COVID-19 CT dataset. Data augmentation can be expressed as creating a new image label by adding some changes to the original images, such as random crop, affine transformation, and flip. It mathematically generates new images similar to the existing images with a minor difference to

increase the number of data samples [14]. The authors collected 471 CT scans for both COVID and non-COVID to build a model that binary classifies the patients' images. They took about 25% of the data for testing and the other 75% for training and resized all the images to 224×224 . The accuracy was 84.7%, precision was 97%, and recall was 76.2%. The results are relatively good, but the recall is not satisfactory. Using more advanced methods can enhance the recall.

Another study by Khan et al. [15] proposed a CoroNet model to automatically classify X-ray images with COVID-19, normal, bacterial pneumonia, and viral pneumonia. This model was built using a pre-trained Xception network using the ImageNet dataset [16]. To achieve high accuracy results, they trained the proposed model using three classification scenarios: CoroNet of 4-class, 3-class, and binary-class classification. Cross-validation with $k = 4$ was used to evaluate the performance of the 4-class basic model. The CoroNet model reached an average accuracy of 89.6%. The precision was 93.17%, and the recall was 98.25%.

Afshar et al. [17] proposed an alternative modeling framework based on Capsule networks (CapsNet), named as COVID-CAPS, to detect COVID-19 from X-ray images. The inputs to the network are three-dimensional X-ray images. The proposed model consisted of four convolutional layers and three capsule layers. The first layer is followed by batch normalization, and the second layer is followed by an average pooling layer, wherein the fourth layer is reconstituted to form the first capsule layer. Hence, the last three capsule layers are included in the COVID-CAPS to perform routing by agreement process. To train the model, they used 295,488 trainable parameters. Therefore, COVID-CAPS can be trained and deployed in a timely manner, and, hence, there is no need for robust computational resources. For the experiments, two settings were performed, with and without pre-training parameters. COVID-CAPS without pre-training achieved 95.7% accuracy, 90% sensitivity, and 95.8% specificity. Pre-trained COVID-CAPS reached 98.3% accuracy, 80% sensitivity, and 98.6% specificity.

Correspondingly, Toraman et al. [18] proposed the CapsNet convolutional network for diagnosing COVID-19 from chest X-ray images with capsule networks. The proposed model has a convolution, base, and digit layer followed by three fully connected layers. The image size for the proposed model is 128×128 . Due to the limited samples of COVID-19 X-ray images, data augmentation techniques have been used to avoid overfitting issues. Therefore, COVID-19 X-ray images have been increased from 231 to

1,050. In [18], two different approaches were used to detect COVID-19 from X-ray images: binary-class and multiclass classification. The categorization in the former method was either COVID-19 or normal X-ray. In the latter approach, COVID-19 versus normal versus pneumonia categories were considered. COVID-19 was recognized with an accuracy of 97.04% in binary classification, and an accuracy of 94.57% in multiclass classification, as the variation in images that the model needs to learn.

Another model was proposed by Tabik et al. [19]. They built a balanced dataset, COVIDGR-1.0, that includes all levels of severity, from normal cases with positive RT-PCR, mild, and moderate to severe cases. The authors proposed the COVID Smart Data-based Network (COVID-SDNet) approach. In their proposed model, one of the components of COVID-SDNet is the CNN-based classifier. Using a transfer learning approach, they have adopted Resnet-50 initialized with ImageNet parameters. To make this CNN applicable to their problem, they deleted the last layer of the network and included 512 neuron layers with ReLU activation and two or four neuron layers with SoftMax activation. Each experiment uses a data split of 80/20, for training and testing, respectively. The model achieved accuracy up to 97.37%.

More recent studies were published about using prediction models for early detection of COVID-19 from X-ray images. In [20], four datasets were applied with five-fold cross-validation on CNN-based models. Out of 16 experiments, two proposed models, ensemble deep transfer learning CNN model and hybrid LSTMCNN, perform the best. The accuracy of the ensemble CNN was 96.51% average-wise, while the accuracy of LSTMCNN was 96.46% average-wise. Using the proposed models may contribute to delivering correct results to patients rapidly.

The study by Saygılı [21] is another exciting paper that uses X-ray and CT scan images without deep learning modeling. This study used three datasets, two CT scan images and one X-ray image. The authors used the same dataset we intended to use in our research. The proposed model performs differently for the datasets; it achieves in dataset-1 (CT), dataset-2 (X-ray), and dataset-3 (CT) an accuracy of 89.41%, 99.02%, and 98.11%, respectively. The same author published another paper [22] in the same area but using only CT scan images. The best accuracy values obtained using the kernel support vector machines method for Dataset-1, Dataset-2, and Mixed Dataset are 98.5%, 86.3%, and 94.5%, respectively.

An additional study used chest X-ray images to diagnose COVID-19 [23]. ANN-based segmentation is applied, so that only the lung area of interest is evaluated for

COVID-19 detection. Also, the authors used augmentation for the images in the COVID-19 class. Two deep learning architectures were proposed; both use AlexNet architecture, one as transfer learning and the other is a hybrid structure as it contains a Bidirectional Long Short-Term Memories (BiLSTM) layer. The classification accuracy of the first architecture is 98.14% and 98.70% in the second hybrid architecture.

The main contribution of this study is as follows:

1. We trained several deep learning models, specifically convolutional neural network (CNN), on publicly available datasets.
2. We used different types of medical images, CT scan and X-ray, to train deep learning models.
3. We provided experimental comparative study to the different deep learning models that we trained and selected the best performing one as the suggested alternative method to diagnose COVID-19 using computational techniques. Also, we ran a comparison between the results of the models seeking the best performing model to diagnose COVID-19.

The remaining parts of this research paper are organized as follows: the Materials and Methods section, which describes the dataset, methodology, and experiments. In the Results section, we present detailed results of all the experiments that we conducted using different evaluation metrics. Finally, the Discussion/Conclusion section of the study is presented.

Materials and Methods

Deep learning is about self-learning from large amounts of raw data. The model learns the features that discriminate the input images automatically. Here, we will describe the dataset that we adopted in our experimental. After that, we will describe the methodology that we followed.

Dataset

In this research, we used two types of medical images: CT scan images and X-ray images.

CT Scan Dataset¹

We used the COVID-CT dataset [12]. There are a total of 746 CT scans. The dataset consists of 349 COVID-19-positive CT scan images collected from 216 patients and confirmed by a senior radiologist in Tongji hospital. Also, the dataset has 397 COVID-19-negative CT scan images collected from other published datasets as described in [12]. The COVID-19-negative images are mixed between healthy patients and others with lung diseases other than COVID-19. Figure 1 shows some examples from the dataset.

¹ <https://github.com/UCSD-AI4H/COVID-CT>.

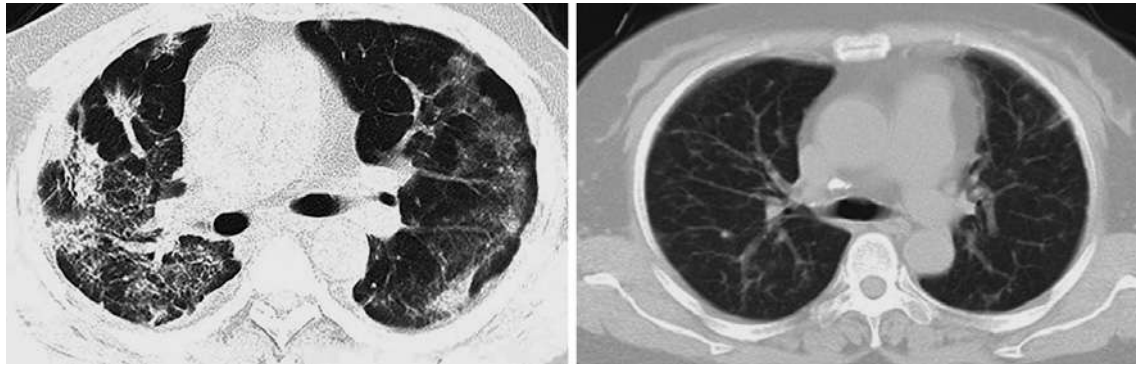


Fig. 1. Examples of CT scans that are negative and positive for COVID-19.

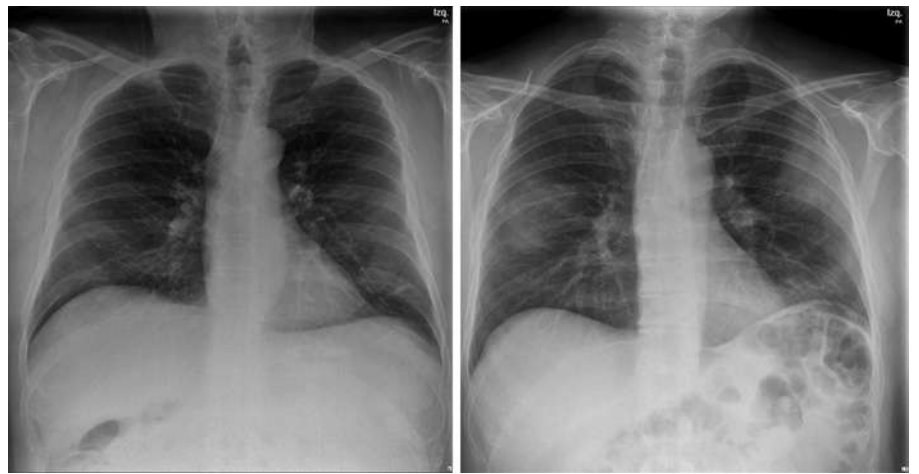


Fig. 2. Examples of X-rays that are negative and positive for COVID-19.

X-Ray Dataset²

Here, we used the COVIDGR dataset [19]. There are 852 X-rays, containing 426 X-rays that are positive for COVID-19 and 426 X-rays that are negative. The positive images are for patients who were positive for the RT-PCR test, and the X-ray images were taken in a period of at most 24 h between the X-ray image and the test. The negative images were collected from other datasets as explained in [19] and may contain other lung diseases. Figure 2 shows some examples from the dataset.

The Datasets

COVID-CT and COVIDGR are available online, and ethical approval is not required.

Methodology

We applied the COVID-19 CT dataset on eight different pre-trained CNN models (CoroNet, InceptionV3, Inception-ResNetV2, MobileNetV2, NASNetMobile, VGG16, VGG19, and Xception). In every model, we conducted many experiments by changing many parameters, such as data split ratios, batch

size, and epochs by callbacks functions to choose the best accuracy for each model; those parameters are shown in Table 1. Then, we tested the COVIDGR X-ray dataset with the same parameters, to compare the results between the different kinds of medical images: CT scan and X-ray. At the end, we evaluated the methods using three most common metrics: accuracy, sensitivity, and specificity, which can be calculated by the following equations:

$$\text{Accuracy} = \frac{Tp + Tn}{Tp + Fn + Tn + Fp},$$

$$\text{Sensitivity} = \frac{Tp}{Tp + Fn},$$

$$\text{Specificity} = \frac{Tn}{Tn + Fp},$$

where Tp is the number of true positives, Tn is the number of true negatives, Fp is the number of false positives, and Fn is the number of false negatives. The operations performed in this study are indicated in Figure 3.

² <https://github.com/ari-dasci/OD-covidgr>.

Table 1. Hyperparameter ranges

Experiment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Batch size	10	10	10	10	12	12	12	12	16	16	16	16	32	32	32	32
Test size	0.3	0.3	0.2	0.2	0.3	0.3	0.2	0.2	0.3	0.3	0.2	0.2	0.3	0.3	0.2	0.2
Validation	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1

Table 2. Statistics of data split for different medical images

Image type	Split		COVID	Non-COVID	Total
CT scans	80/20	Train	279	317	596
		Test	70	80	150
	70/30	Train	244	277	521
		Test	105	120	225
X-rays	80/20	Train	341	341	682
		Test	85	85	170
	70/30	Train	298	298	596
		Test	128	128	256

Experiments

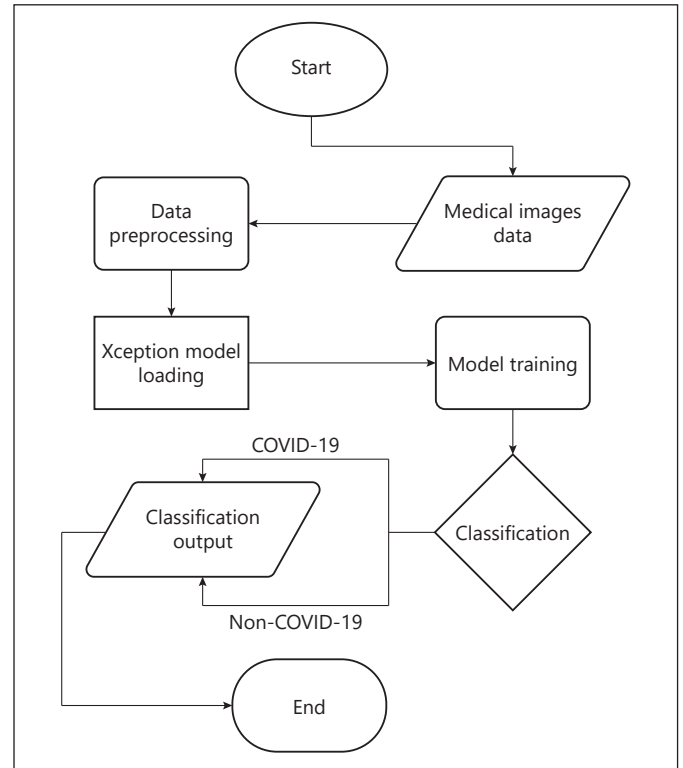
We split the dataset into three sets: training, validation, and testing sets. Table 2 summarizes the number of COVID and non-COVID images for CT scan and X-ray in each experimental setup (without validation).

All images were resized to 244×244 . We changed the data split for training, validation, and testing each time and changed the batch size. So, the total number of experiments with the combination of parameters that we changed is 16 experiments.

Results

The overall accuracy, sensitivity, and specificity computed for each model are summarized in Tables 3 and 4 and Figure 4. CT scan achieved better accuracy than X-ray which is expected because CT scan contains more details than X-ray. The Xception model achieved the highest accuracy among all models – shown in Tables 3 and 4. The overall performance is accuracy 84%, sensitivity 91%, and specificity 77% for CT scan, while the overall performance for X-ray is accuracy 69%, sensitivity 83%, and specificity 55%. An example of misclassified CT scan images is shown in Figure 5.

Our proposed model has some limitations, such as the number of medical images used for training and testing the model. It is well known that the larger the number of samples, especially in deep learning, the better the results.

**Fig. 3.** Flowchart of our proposed method.

Additionally, using transfer learning might speed up building a model and start with pre-trained weights for the network and overcome overfitting due to limited samples, yet that would give different results than training a model from scratch to work on the problem under study.

Discussion/Conclusion

With the rapid increase of daily cases of COVID-19, some countries are struggling to provide adequate testing methods. Advancements in technology and research in developing computational methods for rapid and accurate testing methods have yielded promising results.

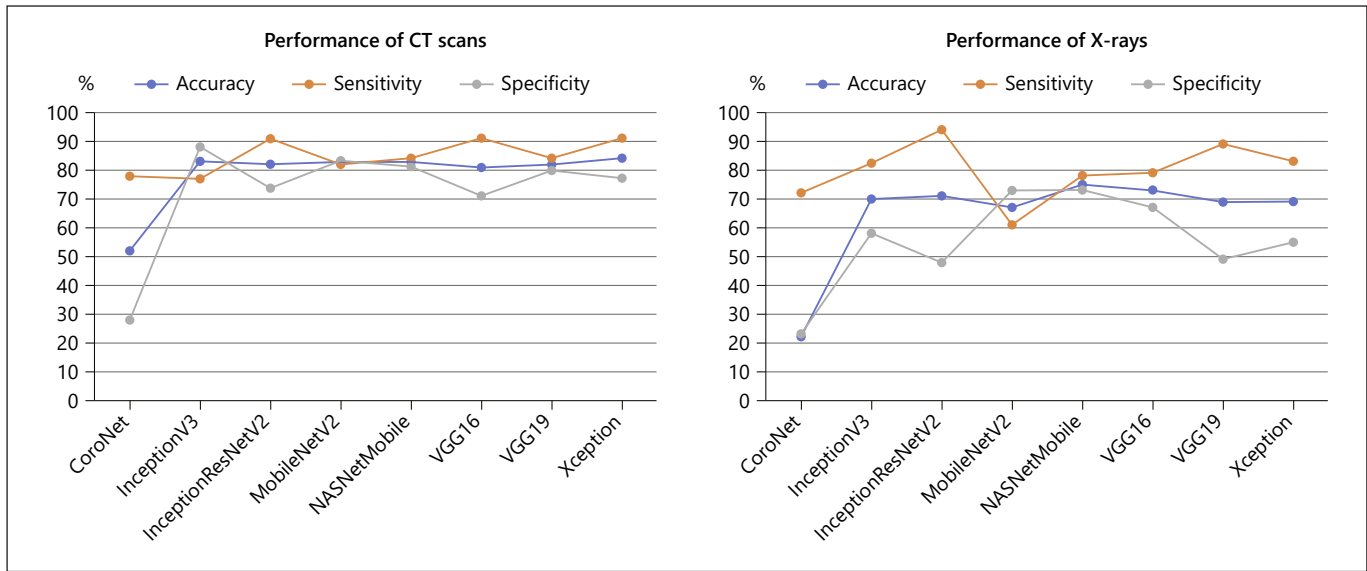


Fig. 4. Plots of accuracy, sensitivity, and specificity.

Table 3. Performance on CT scans for each model

Model	Batch size	Epochs	Test size	Validation	Accuracy	Sensitivity	Specificity
CoroNet	10	22	0.2	0.2	0.52	0.78	0.28
InceptionV3	32	45	0.2	0.2	0.83	0.77	0.88
InceptionResNetV2	32	100	0.2	0.2	0.82	0.91	0.74
MobileNetV2	16	57	0.3	0.2	0.83	0.82	0.83
NASNetMobile	16	100	0.2	0.2	0.83	0.84	0.81
VGG16	16	100	0.2	0.1	0.81	0.91	0.71
VGG19	10	39	0.2	0.2	0.82	0.84	0.80
Xception	32	77	0.3	0.2	0.84	0.91	0.77

Table 4. Performance on X-rays for each model

Model	Batch size	Epochs	Test size	Validation	Accuracy	Sensitivity	Specificity
CoroNet	10	22	0.2	0.2	0.22	0.72	0.23
InceptionV3	32	45	0.2	0.2	0.70	0.82	0.58
InceptionResNetV2	32	100	0.2	0.2	0.71	0.94	0.48
MobileNetV2	16	57	0.3	0.2	0.67	0.61	0.73
NASNetMobile	16	100	0.2	0.2	0.75	0.78	0.73
VGG16	16	100	0.2	0.1	0.73	0.79	0.67
VGG19	10	39	0.2	0.2	0.69	0.89	0.49
Xception	32	77	0.3	0.2	0.69	0.83	0.55



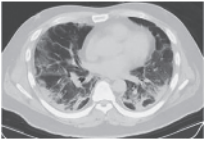

Image	True label	Predicted label
	P	P
	N	N
	P	N
	N	P

Fig. 5. Example of correctly classified and misclassified CT scan images.

In this work, we trained eight different deep learning (CNN) models and achieved up to 84% accuracy, which is close to the performance of similar studies using both CNN and CT scan images for prediction, such as in [14] they achieved 84.7% accuracy. However, using another medical image (X-ray) and the CapsNet model results in higher performance up to ~97% accuracy. Our proposed solution found that using CT scan medical images would result in more accurate performance than X-ray images. Additionally, our work proved that using deep learning algorithms with predefined structure known as transfer learning might save some time and affect the results variably. Although the performance of the proposed models is somehow less than the performance of other published work, experimenting with transfer learning with a limited number of data samples and predicting only two classes can be contributing factors. Thus, research results show that deep learning methods can yield promising results in detecting and diagnosing COVID-19 from medical imaging faster and cheaper than conventional testing methods, especially in case of the availability of CT scan.

As future work, it is recommended that we obtain a larger dataset along with multiclass images, resulting in a

multiclass classification problem rather than binary classification. Moreover, we plan to improve the method and optimize the hyperparameters to achieve higher and more reliable accuracy. If that could be achieved, we might launch our model to be tested and used for a more enclosed population, like Saudis, to improve the diagnosis of COVID-19, therefore decreasing the infection rate.

Acknowledgments

This research was supported by the King Abdulaziz and His Companions Foundation for Giftedness and Creativity (Mawhiba). The first author also thanks the research supervisors who provided insight and expertise that greatly advanced the research, although they may not agree with all of the interpretation of this paper.

Statement of Ethics

This study does not require statement of ethics. The ethical approval for this study was not required. The written informed consent was not required for this study.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

This research was supported by the King Abdulaziz and His Companions Foundation for Giftedness and Creativity (Mawhiba).

Author Contributions

Elaf Alshehri is a student at Mawhiba, and she did the training and testing of deep learning models. Dr. Manal, Dr. Felwa, and Dr. Reem are the direct supervisors for Elaf and contributed equally to the directing, advising, and guiding Elaf; also they contributed equally in paper writing. Dr. Khalid covered the medical part in regards of explaining the medical information and double checked its correctness.

Data Availability Statement

Data are available publicly online at <https://github.com/UCSD-AI4H/COVID-CT> and <https://github.com/ari-dasci/OD-co-vidgr>.

References

- 1 Chen J, Wu L, Zhang J, Zhang L, Gong D, Zhao Y, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. *medRxiv*. 2020.
- 2 Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med*. 2020;382(10):929–36.
- 3 Gumbrecht J, Howard J. WHO declares novel coronavirus outbreak a pandemic. Cable News Network; 2020 [cited 2021 Dec 18]. Available from: <https://edition.cnn.com/2020/03/11/health/coronavirus-pandemic-world-health-organization/index.html>.
- 4 Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497–506.
- 5 World Health Organization. WHO coronavirus (COVID-19) dashboard. World Health Organization; [cited 2021 Dec 18]. Available from: <https://covid19.who.int/>.
- 6 Jawerth N. How is the COVID-19 virus detected using real time RT-PCR? *IAEA Bulletin*. 2020;61(2):8–11.
- 7 Sekeroglu B, Ozsahin I. Detection of COVID-19 from chest X-ray images using Convolutional Neural Networks. *SLAS Technol*. 2020;25(6):553–65.
- 8 Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1,014 cases. *Radiology*. 2020;296(2):E32–40.
- 9 Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology*. 2020;296(2):E65–71.
- 10 Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *medRxiv*. 2020.
- 11 Butt C, Gill J, Chun D, Babu BA. Deep learning system to screen coronavirus disease 2019 pneumonia. *Appl Intell*. 2020:1.
- 12 Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P. COVID-CT-dataset: a CT scan dataset about COVID-19 [preprint]. *arXiv*. 2020. arXiv:200313865.
- 13 Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *J Big Data*. 2016;3(1):9.
- 14 Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data*. 2019;6(1):60.
- 15 Khan AI, Shah JL, Bhat MM. CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput Methods Programs Biomed*. 2020;196:105581.
- 16 Khan AI, Shah JL, Bhat MM. CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput Methods Programs Biomed*. 2020;196:105581.
- 17 Afshar P, Heidarian S, Naderkhani F, Oikonomou A, Plataniotis KN, Mohammadi A. COVID-CAPS: a capsule network-based framework for identification of COVID-19 cases from X-ray images. *Pattern Recognit Lett*. 2020;138:638–43.
- 18 Toraman S, Alakus TB, Turkoglu I. Convolutional capsnet: a novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. *Chaos Solitons Fractals*. 2020;140:110122.
- 19 Tabik S, Gómez-Ríos A, Martín-Rodríguez JL, Sevillano-García I, Rey-Area M, Chartre D, et al. COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. *IEEE J Biomed Health Inform*. 2020;24(12):3595–605.
- 20 Khanna M, Agarwal A, Singh LK, Thawkar S, Khanna A, Gupta D. Radiologist-level two novel and robust automated computer-aided prediction models for early detection of COVID-19 infection from chest X-ray images. *Arab J Sci Eng*. 2021:1–33.
- 21 Saygılı A. A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods. *Appl Soft Comput*. 2021;105:107323.
- 22 Saygılı A. Computer-aided detection of COVID-19 from CT images based on Gaussian mixture model and Kernel support vector machines classifier. *Arab J Sci Eng*. 2021:1–19.
- 23 Aslan MF, Unlarsen MF, Sabanci K, Durdu A. CNN-based transfer learning–BiLSTM network: a novel approach for COVID-19 infection detection. *Applied Soft Computing*. 2021;98:106912.