Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/110538

Detecting COVID-19 in X-ray Images using Transfer Learning

Jamal Alsakran[1,*], *Loai Alnemer*[2], *Nouh Alhindawi*[3] *and Omayya Muard*[4]

[1]Computer Information Science, Higher Colleges of Technology, Fujairah, UAE
[2]Department of Computer Information Science, King Abulllah II School for Information Technology, The University of Jordan, Amman, Jordan
[3]Department of Software Engineering, Faculty of Sciences and Information Technology, Jadara University, Irbid, Jordan
[4]Department of Research and Development, Association of Arab Universities, Amman, Jordan

Received: 2 Mar. 2022, Revised: 2 May 2022, Accepted: 21 Jun. 2022
Published online: 1 Sep. 2022

Abstract: Accurate and speedy detection of COVID-19 is essential to curb the spread of the disease and avoid overwhelming the health care system. COVID-19 detection using X-ray images is commonly practiced at medical centers; however, it requires the intervention of medical professionals trained in diagnosing and interpreting medical imagining. In this paper, we employ deep transfer learning models to detect COVID-19 on a dataset of over 20,000 X-ray images. Our results on 5 pretrained models (VGG19, InceptionV3, MobileNetV2, DenseNet121, and ResNet101V2) show high performance of 99% without image augmentation, and 93% when image augmentation is used.

Keywords: COVID-19, Deep learning, Transfer learning, Image augmentation, X-ray images

1 Introduction

The COVID-19 pandemic has claimed the lives of millions and continues to have a devastating effect on the global population's health and well-being. Therefore, advancing the technology of screening and detecting COVID-19 is critically essential for immediate treatment as well as mitigating and combating the disease. Reverse transcriptase-polymerase chain reaction (RT-PCR) testing remains the gold standard; however, it is time consuming and needs to be conducted in medical laboratories by trained professionals [1]. Furthermore, early studies have shown that RT-PCR has a relatively poor sensitivity [2] and a highly false positive rate [3].

An alternative screening method used for COVID-19 screening is radiography examination. In radiography screening, chest radiography imaging (e.g., chest X-ray (CXR) or computed tomography (CT) imaging) is conducted and analyzed by radiologists to look for visual characteristics associated with COVID-19. Radiography screening has shown quite promising results [4]. Nonetheless, successful screening of radiography imaging requires medical professionals specializing in diagnosing and interpreting medical imagining. Furthermore, acquiring chest X-rays in conventional rooms may spread the disease to medical staff, and uninfected patients through droplet-contaminated surfaces [5]. Therefore, it is recommended that COVID-19 screening is carried out in a dedicated radiology room or that the room is disinfected after each use.

Convolutional neural network (CNN) is a class of deep learning that has attracted lots of attention due to its remarkable success in computer vision and image analysis [6]. It has significantly improved the accuracy of image recognition and classification tasks and has become an integral part of medical diagnosis [7]. Transfer learning has remarkably facilitated deep learning to overcome the lack of large datasets that had been previously necessary for an effective deep learning model. With transfer learning, pretrained CNNs can be exploited to generalize to new tasks.

This research is motivated by the urgent need to advance the technology of screening and detecting COVID-19, the increasingly available open source radiography image datasets, and the advent of artificial intelligence systems. We aim to utilize deep convolutional

* Corresponding author e-mail: jalsakran@hct.ac.ae

neural network models to detect COVID-19 from chest X-ray (CXR) images.

2 Related Work

Since the outbreak of COVID-19, extensive research has been devoted to advancing the technology of screening and detecting infected patients via radiography imaging [8]. Early studies have shown that patients infected with COVID-19 are likely to present abnormalities in chest radiography images [4,9,10]. Another research concludes that chest CT has a high sensitivity for diagnosing COVID-19, and it may be considered a primary tool for the current COVID-19 detection in epidemic areas [8].

Huang et al. [4] identify that 98% of all patients (41 patients) involved in their study present bilateral abnormalities in chest CT images, and Guan et al. [4] identify that 86.2% of COVID-19 positive cases (975 patients) in their study present abnormal results in CXR and CT images The most common patterns on chest CT were ground-glass opacity (56.4%) and patchy bilateral shadowing (51.8%).

Several artificial intelligence systems based for detecting patients infected with COVID-19 via radiography imaging have been proposed, and results have shown quite promising accuracy [11,12]. Most studies have focused on exploring deep convolutional neural networks, given their overall successes in a variety of computer vision tasks. Zhan et al. have proposed a CNN model that can detect 96% of COVID-19 cases [13]. The dataset used in their experiments is small (less than 1500 images).

Wang et al. [14] have introduced COVID-Net, a deep convolutional neural network design tailored for detecting COVID-19 cases from CXR images that is open source and available to the general public. They have also introduced COVIDx, an open access benchmark dataset that contains 13,975 CXR images across 13,870 patient cases. Recently, three convolutional neural network models (ResNet50, InceptionV3, and ResNetV2) have been compared for detecting coronavirus pneumonia infected patients using chest CXR radiographs [15]. They report 98% accuracy on the pretrained ResNet50 model.

Transfer learning has been utilized in several recent studies for COVID-19 screening. DenseNet-121 has been employed on ChestX-ray8 dataset of 108,948 lung disease images [16]. It achieves AUROC of 0.973 when discriminating COVID-19 cases against mixed pneumonia and normal cases. Ioannis et al. [17] employ a transfer learning approach using AlexNet and GoogLeNet for classifying a dataset of 1427 X-ray images containing 224 COVID-19, 700 Bacterial Pneumonia, and 504 Normal X-ray images. Their best best-performing classifier had an AUC of 0.99.

In a recent effort [18], Chowdhury et al. have achieved a remarkable accuracy result of 99.7%. They have conducted experiments on many pretrained CNNs such as ResNet18, InceptionV3, CheXNet, and VGG19. Experiments included two versus three class settings and with and without data augmentation. The dataset that they have collected contains a mixture of 423 COVID-19, 1485 viral pneumonia, and 1579 normal chest X-ray images. They have constantly been updating the dataset which they have made publicly available, and it is used in our experiments.

3 Visualizing X-ray Images

COVID-19 screening through radiology imaging has been in practice since the pandemic's beginning, mainly due to the readily available radiological imaging at medical centers. However, successful screening requires the intervention of medical specialists, and it remains difficult for radiologists to distinguish COVID-19 from other viral pneumonia because they have similar symptoms.

Chest X-rays of mild cases or early-stage disease may be normal, but patients with moderate to severe symptoms are likely to have features associated with pneumonia. Common features associated with COVID-19 include a reticular pattern, ground-glass opacities, and consolidations, with rounded morphology and a confluent or patchy multifocal distribution. The distribution is usually bilateral and peripheral, predominating in the lower fields [19]. Fig. 1 shows an example of normal X-ray Fig. 1 (a) and COVID-19 infected X-ray Fig. 1 (b) that exhibits hazy gray areas indicating that the air spaces becoming partially filled with fluid or pus.

To interpret how the visual features associated with COVID-19 activate class discrimination in classification, we employ Grad-CAM [20] visualization. Grad-CAM uses class-specific gradient information to localize important regions and produces discriminative visualizations for the class of interest. It helps users establish appropriate trust in CNN predictions [21]. Fig 2 shows Grad-CAM visualizations of the X-ray images presented in Fig. 1. The figure emphasizes the area of interest to our model at the time of classification which helps to interpret the result. For instance Fig. 1 (b) highlights COVID-19 feature associated with lungs filled with gas.

4 Transfer Learning

Deep learning techniques are often described as *data hungry* techniques due to their dependence on the availability of extensive data. Compared to traditional machine learning, deep learning requires a large amount of data to learn latent patterns [22]. Nowadays, the amount of data generated and collected is growing at unprecedented levels, facilitating deep learning. Nonetheless, in some domains, such as medical imaging, the collection of high quality data is laborious and extremely difficult.

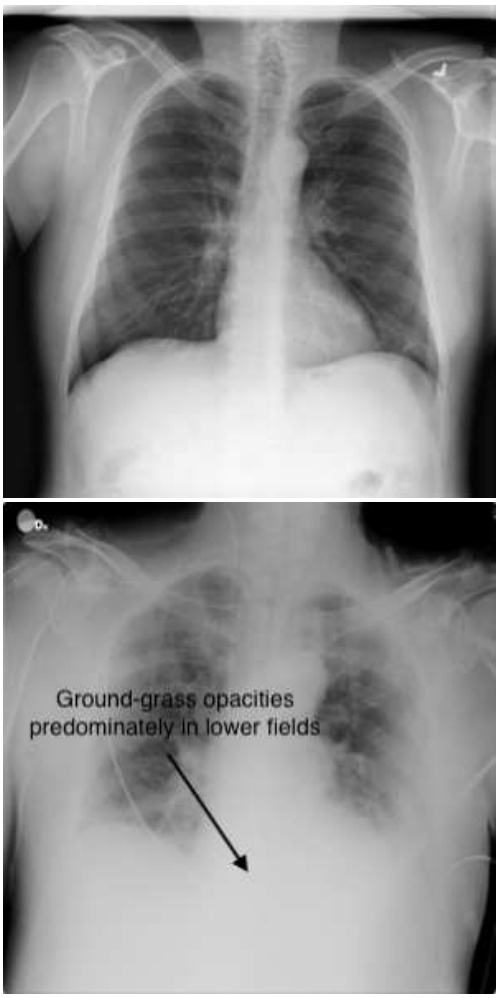

Fig. 1: X-ray images of (a) normal and (b) COVID-19 infected patient exhibits common features such as ground-glass opacities in lower fields

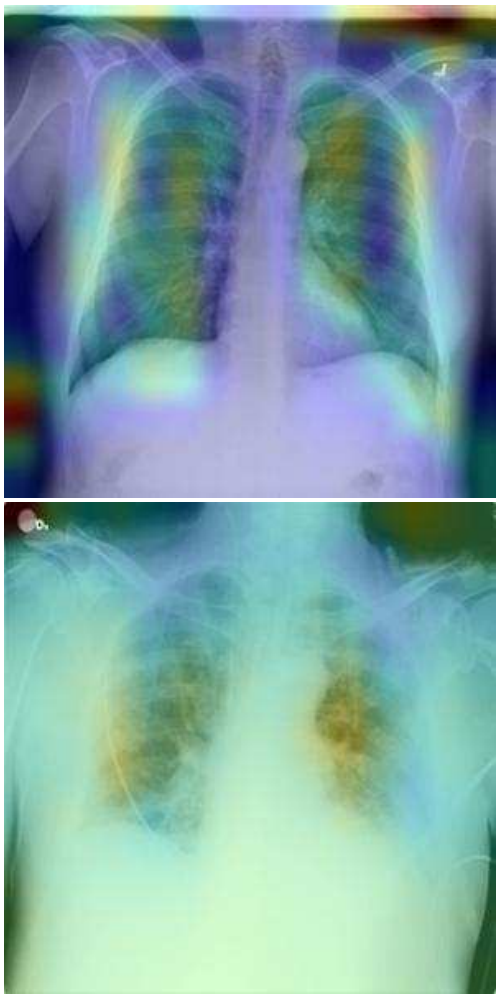

Fig. 2: Grad-CAM visualizations of (a) normal and (b) COVID-19 infected X-rays. High intensity visuals reflect areas of interest to the model

Transfer learning can enable learning when insufficient data is all that we have. It allows models trained in one setting to generalize to another setting [23]. Fig. 3 shows X-ray images are being fed to a pretrained CNN with nontrainable weights. The CNN weights must not be updated, so the previously learned patterns are not lost. The output of the pretrained CNN is then fed to a newly constructed fully connected NN that suits the new task. In contrast, the fully connected NN weights must be trainable, so it can learn to classify the new task.

Many famous and proven effective pretrained CNN, such as VGG, Inception, and ResNet are trained on more than a million images and 1000 classes of the ImageNet dataset (https://www.image-net.org/). These CNNs facilitate transfer learning because they are trained on a large corpus of images and require the model to classify a relatively large number of classes. They require the model to efficiently extract features from a broad range of image classes to perform well on the problem and generalize on unseen images.

5 Image Augmentation

Image augmentation is yet another technique that can tackle the insufficient data issue. It introduces minor alterations to the dataset to trick the model into thinking that new images are introduced. Image augmentation is encouraged in the medical imaging domain because it generates several variants of an X-ray image that realistically mimic variations in image acquisition collected in radiography.

Another advantage of image augmentation is that the model sees no image more than once as new alternations are generated in each epoch. This dramatically helps

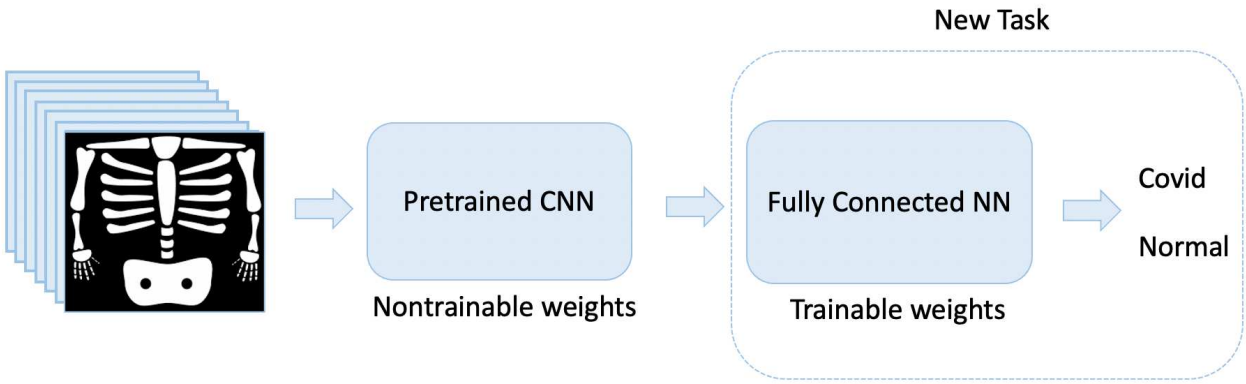

Fig. 3: Transfer learning with pretrained CNN that feeds new fully connected CNN

reduce overfitting because the model cannot overfit all the samples and is forced to generalize.

Fig. 4 shows a sample of 3 images that are generated from the original image Fig. 4 (a). The images present a rotated by 10 degrees image Fig. 4 (b), horizontally flipped image Fig. 4 (c), and width shifted image Fig. 4 (d). In this example, one image has produced three more images that will be seen by the model as new images.

6 Dataset

The dataset used in our experiments has been collected by a team of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh, along with their collaborators from Pakistan and Malaysia in collaboration with medical doctors, and is publicly available on Kaggle [24]. The dataset consists of 3,616 COVID-19 positive cases, 10,192 Normal, 6,012 Lung Opacity, and 1,345 Viral Pneumonia images. An earlier release of a smaller dataset has been used in the team's publications [18,25].

In our experiments, the dataset is split 70% for training, 20% for validation, and 10% for testing. The images contain three color channels and have been resized to 150x150x3.

7 Experiment Results

Our experiments are conducted using Keras deep learning framework. Keras provides out of the box support for pretrained models with configurable API to fine tune the models, for instance, freezing the layers of the pretrained model during training. For training, we have used batch size 64 and enabled *Early Stopping* to prevent overfitting. *Early Stopping* is set to monitor *validation loss* with *patience* set to 3.

Image augmentation in Keras provides a wide range of arguments that can be widely adjusted to alter the images. We have experimented with many of those arguments and chosen a few that are well suited for medical imaging. Table 1 shows the arguments and their respective values used in our experiments.

Table 1: Image augmentation arguments used in experiments

Argument	Value
Rotation range	10
Width shift range	0.1
Height shift range	0.1
Shear range	0.1
Zoom range	0.1
Horizontal flip	True
Fill mode	nearest

There are over 30 pretrained models available in Keras, and we have selected VGG19, InceptionV3, MobileNetV2, DenseNet121, and ResNet101V2 because of their high reported accuracy and fast training. During training, the pretrained CNN weights are frozen, and the output is fed to a fully connected network with 256 neurons and an output layer of 4 neurons (2 neurons in two-class experiments).

Table 2 shows the results of our experiments with and without image augmentation. The results report model *accuracy* and *precision*, *recall* and *fi-score* of COVID-19 class. When training is done without image augmentation, all models achieve above 93% accuracy, with ResNet101V2 outperforming others and achieving 99.6% accuracy. However, when image augmentation is used, there is a noticeable degradation in performance, with ResNet101V2 reaching 93.1% accuracy. The lower performance is attributed to the nature of image alterations that is generated in image augmentation, which does not suit medical images. For instance, rotating or scaling a full field X-ray image would generate images

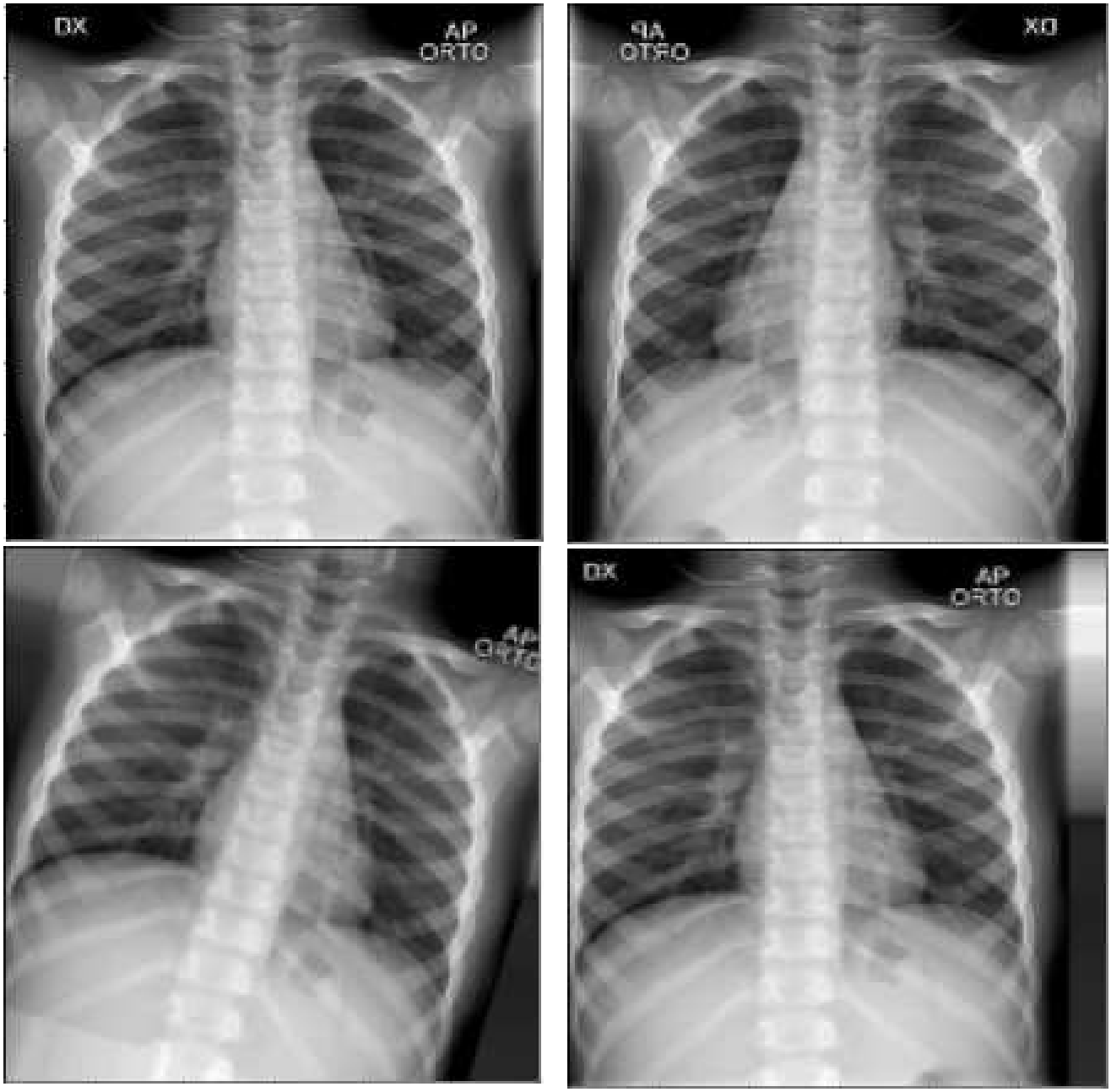

Fig. 4: Example of image augmentation. (a) original image, (b) image rotated, (c) image horizontally flipped, and (d) image shifted

that are not typically acquired by scanning equipment. For example, Fig 4 (b) shows a rotated image generated by data augmentation; however, in practice, such an image would be rejected if it were acquired by scanning equipment at a medical facility. Furthermore, another factor that has negatively affected performance with image augmentation is that the training was *Early Stopped* with fewer than 30 epochs due to increased validation loss. The increased validation loss suggests that the model suffers from overfitting; therefore, a larger dataset is needed to improve performance.

We have further conducted experiments on two-class classification: COVID versus Non-COVID. The dataset consists of 3,616 COVID-19 cases and 17,549 non-COVID-19 cases (Normal, Lung Opacity, and Viral Pneumonia). Table 3 reports our experiment results which show that MobileNetV2, DenseNet121, and ResNet101V2 achieve above 99% accuracy when

Table 2: Experiment results with and without image augmentation on four classes (COVID-19, Normal, Lung Opacity, and Viral Pneumonia)

Scheme	Model	Accuracy	Precision	Recall	F1-score
Without augmentation	VGG19	0.939	0.96	0.95	0.96
	InceptionV3	0.952	0.95	0.95	0.94
	MobileNetV2	0.992	0.99	0.99	0.99
	DenseNet121	0.9881	0.98	0.97	0.97
	ResNet101V2	0.996	0.99	0.99	0.99
With augmentation	VGG19	0.857	0.94	0.67	0.78
	InceptionV3	0.8751	0.78	0.87	0.81
	MobileNetV2	0.915	0.91	0.90	0.91
	DenseNet121	0.883	0.97	0.69	0.81
	ResNet101V2	0.931	0.98	0.88	0.93

Table 3: Experiment results with and without image augmentation on two classes (COVID-19 and non-COVID-19)

Scheme	Model	Accuracy	Precision	Recall	F1-score
Without augmentation	VGG19	0.938	0.95	0.67	0.78
	InceptionV3	0.972	0.97	0.97	0.97
	MobileNetV2	0.998	0.99	0.99	0.99
	DenseNet121	0.995	0.99	0.99	0.99
	ResNet101V2	0.998	0.99	0.99	0.99
With augmentation	VGG19	0.923	0.95	0.58	0.72
	InceptionV3	0.910	0.69	0.91	0.78
	MobileNetV2	0.964	0.97	0.91	0.89
	DenseNet121	0.947	0.95	0.73	0.82
	ResNet101V2	0.970	0.98	0.84	0.91

classification is done without image augmentation. The results show a slight degradation in performance when image augmentation is used with ResNet101V2 outperforming other classifiers. Compared to the four-class classification reported in table 2, the accuracy is slightly better, especially when augmentation is used. Nonetheless, f1-score results worsen because of the imbalanced dataset.

8 Conclusion and Future Work

Lately, COVID-19 detection using radiography images has received extensive attention and achieved remarkable results. In this effort, we have employed deep transfer learning to detect COVID-19 on a dataset containing over 20,000 X-ray images. Our results show robust performance that reaches 99% accuracy for ResNet101V2 pretrained model. This study has conducted experiments on a publicly available radiology dataset of over 20,000 images. In the future, we intend to expand our experiments by collecting more and larger datasets. Large datasets would help overcome the overfitting problem; however, it would pose other challenges such as data preprocessing and integration.

Additionally, we have experimented with 5 pretrained models, and we plan to expand our experiments to include several others and conduct a comprehensive comparative analysis of their performance. Furthermore, we plan to investigate further the impact of image augmentation on X-ray image classification. For instance, in this study, we have employed a handful of data augmentation arguments; however, many more arguments deserve to be explored, such as *brightness range* and *fill mode*.

Acknowledgement

This research was supported by grant number #98121 from Higher Colleges of Technology (HCT). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of HCT.

Conflicts of Interests

The authors declare that they have no conflicts of interests

References

[1] W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, and W. Tan, *Detection of SARS-CoV-2 in Different Types of Clinical Specimens*, *JAMA*, **323**, 1843–1844 (2020).

[2] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, *Sensitivity of chest ct for covid-19: comparison to rt-pcr*, *Radiology*, **296**, E115–E117 (2020).

[3] P. Wikramaratna, R. S. Paton, M. Ghafari, and J. Lourenco, *Estimating false-negative detection rate of sars-cov-2 by rt-pcr*, *MedRxiv*, **2020** (2020).

[4] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, et al., *Clinical features of patients infected with 2019 novel coronavirus in wuhan, china*, *The lancet*, **395**, 497–506 (2020).

[5] G. D. Rubin, C. J. Ryerson, L. B. Haramati, N. Sverzellati, J. P. Kanne, S. Raoof, N. W. Schluger, A. Volpi, J.-J. Yim, I. B. Martin, et al., *The role of chest imaging in patient management during the covid-19 pandemic: a multinational consensus statement from the fleischner society*, *Radiology*, **296**, 172–180 (2020).

[6] F. Chollet, *Deep learning with Python*, Simon and Schuster (2021).

[7] K. Kallianos, J. Mongan, S. Antani, T. Henry, A. Taylor, J. Abuya, and M. Kohli, *How far have we come? artificial intelligence for chest radiograph interpretation*, *Clinical radiology*, **74**, 338–345 (2019).

[8] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, *Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: a report of 1014 cases*, *Radiology*, **296**, E32–E40 (2020).

[9] M.-Y. Ng, E. Y. Lee, J. Yang, F. Yang, X. Li, H. Wang, M. M.-s. Lui, C. S.-Y. Lo, B. Leung, P.-L. Khong, et al., *Imaging profile of the covid-19 infection: radiologic findings and literature review*, *Radiology: Cardiothoracic Imaging*, **2**, e200034 (2020).

[10] W.-j. Guan, Z.-y. Ni, Y. Hu, W.-h. Liang, C.-q. Ou, J.-x. He, L. Liu, H. Shan, C.-l. Lei, D. S. Hui, et al., *Clinical characteristics of coronavirus disease 2019 in china*, *New England journal of medicine*, **382**, 1708–1720 (2020).

[11] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. Siegel, *Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis*, *arXiv preprint arXiv:2003.05037* (2020).

[12] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su, et al., *A deep learning system to screen novel coronavirus disease 2019 pneumonia*, *Engineering*, **6**, 1122–1129 (2020).

[13] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, *Covid-19 screening on chest x-ray images using deep learning based anomaly detection*, *arXiv preprint arXiv:2003.12338*, **27** (2020).

[14] L. Wang, Z. Q. Lin, and A. Wong, *Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images*, *Scientific Reports*, **10**, 1–12 (2020).

[15] A. Narin, C. Kaya, and Z. Pamuk, *Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks*, *Pattern Analysis and Applications*, **24**, 1207–1220 (2021).

[16] X. Li and D. Zhu, *Covid-xpert: An ai powered population screening of covid-19 cases using chest radiography images* (2020).

[17] P. Lakhani and B. Sundaram, *Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks*, *Radiology*, **284**, 574–582 (2017).

[18] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, and M. T. Islam, *Can ai help in screening viral and covid-19 pneumonia?*, *IEEE Access*, **8**, 132665–132676 (2020).

[19] E. M. Chamorro, A. D. Tascón, L. I. Sanz, S. O. Vélez, and S. B. Nacenta, *Radiologic diagnosis of patients with covid-19*, *RadiologíA (English Edition)*, **63**, 56–73 (2021).

[20] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, *Grad-cam: Why did you say that?*, *arXiv preprint arXiv:1611.07450* (2016).

[21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, *Grad-cam: Visual explanations from deep networks via gradient-based localization*, in *Proceedings of the IEEE international conference on computer vision*, 618–626, (2017).

[22] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, *A survey on deep transfer learning*, in *International conference on artificial neural networks*, 270–279, (2018).

[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT press (2016).

[24] T. Rahman, https://doi.org/10.34740/kaggle/dsv/3122958 (2020).

[25] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan, et al., *Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images*, *Computers in biology and medicine*, **132**, 104319 (2021).