COVID-19 Identification from Chest X-Rays

Iosif Mporas
School of Physics Engineering and Computer Science,
University of Hertfordshire,
Hatfield AL10 9AB, United Kingdom
i.mporas@herts.ac.uk

Prasitthichai Naronglerdrit
Department of Computer Engineering,
Faculty of Engineering at Sriracha,
Kasetsart University Sriracha Campus,
Chonburi, Thailand
prasitthichai@eng.src.ku.ac.th

***Abstract*— Artificial Intelligence and Data Science community has contributed to the global response against the new coronavirus, COVID-19. Significant attention has been given to detection and diagnosis tools with rapid diagnostic tools based on X-rays using deep learning being proposed. In this paper we present an evaluation of several well-known pretrained deep CNN models in a transfer learning setup for COVID-19 detection from chest X-ray images. Two different publicly available datasets were employed and different setups were tested using each of them separately of mixing them. The best performing models among the evaluated ones were the DenseNet, ResNet and Xception models, with the results indicating the possibility of identifying COVID-19 positive cases from chest X-ray images.**

Keywords*— *COVID-19, X-rays, transfer learning, convolutional neural networks.

I. Introduction

COVID-19 is a highly infectious disease, which is caused by the SARS-CoV-2 virus [1]. In March 2020 and after spreading to more than 100 countries and leading to several thousands of cases, the World Health Organization (WHO) officially declared the outbreak of the new coronavirus as a pandemic [2]. Although COVID-19 affects the entire population, young people that have been affected by COVID-19 are in most cases either asymptomatic or present mild symptoms like cough, headache, fatigue and fever. For non-young ages and especially for elders and/or for patients with chronic conditions COVID-19 positive cases may progress to more serious symptoms like diarrhea, dyspnea, pneumonia and death [3]. Young and middle-aged patients being diagnosed with COVID-19 are having significantly lower mortality rates comparing to elder patients with COVID-19 which are more likely to progress to severe disease [4, 5]. With COVID-19 being highly infectious it can easily be spread from asymptomatic to vulnerable population. To stop the spread of COVID-19 virus and to protect vulnerable people many countries worldwide have introduced isolation measures like social distancing and lockdown and in parallel they perform diagnostic tests to key staff and the general population to detect COVID-19 positive cases.

The diagnosis of COVID-19 is performed by the reverse-transcription polymerase chain reaction (RT-PCR) test after collection of proper respiratory tract specimen, which is a laboratory-based test for detection and quantification of a targeted DNA molecule [6]. The RT-PCR test can be done only in laboratories that are equipped with the needed infrastructure. Moreover, in some cases the COVID-19 test may need be repeated after one or two days while the cost of the equipment and the required PCR reagents is not low, thus making this diagnostic test expensive and sometimes time consuming, without counting the need for specialized microbiologists to do the tests analyses and the appropriate safety measures (personal protective equipment) that are required to keep laboratory staff safe [6]. Due to these difficulties and restrictions many countries are restricting the performed diagnostic tests for COVID-19 to only suspicious cases and/or vulnerable groups of population as it is not possible to do massive testing of the general population. At the same time governments introduce isolation measures, which are causing socioeconomical problems, such as increase of the number of domestic abuse cases [7], reduction of the economic growth [8] and the global trade [9]. Based on the above-mentioned facts, the development of alternative, complementary and low-cost tools for detection of COVID-19 positive cases and for decision making support is essential.

The recent development of technology in deep learning and medical image processing in combination with big data repositories for COVID-19 could offer support in the global effort against the new coronavirus. Several studies have investigated the potential of using X-ray and/or CT images identify COVID-19. An automatic X-ray COVID-19 lung image classification system was presented in [10], which first increases the contrast of the image by applying a median filter on it and using threshold based segmentation and support vector machines (SVM) it classifies infected from non-infected lung images. A decompose, transfer and compose convolutional neural networks (CNNs) based method for classification of COVID-19 chest X-ray images was presented in [11], extracting deep local features of each image and a class-composition layer to refine the classification of the images. Pre-trained CNN models together with SVM classifier to detect the COVID-19 from chest X-ray images were reported in [12]. In [13] a deep convolutional neural network design was proposed tailored for detection of COVID-19 cases from chest X-ray images, named COVID-Net, using a human-machine collaborative design strategy to design it. In [14] a 3D deep learning framework to detect COVID-19 cases from chest X-ray images called COVNet was presented, first extracting the lung region of interest using U-net [15]. An automated CT image analysis tools for detection, quantification, and tracking of COVID-19 cases was presented in [16], with the method consisting of two sub-systems and analysing the CT case at two distinct levels. A deep CNN based transfer learning approach for automatic detection of COVID-19 pneumonia was presented in [17], using four different popular CNN based deep learning algorithms.

In this paper we present an evaluation of several pre-trained deep convolutional neural network (CNN) models on the detection of positive cases the new COVID-19 virus from chest X-ray images. Different experimental setups are tested using two X-ray datasets either separately to each other or by mixing them. The remainder of this paper is organized as follows: In Section II the evaluation architecture followed is briefly described. In Section III the experimental setup and in

© IEEE 2020. This article is free to access and download, along with rights for full text and data mining, re-use and analysis

Section IV the experimental results are presented. Conclusions are provided in Section V.

II. COVID-19 DETECTION FROM X-RAY IMAGES

COVID-19 positive cases detection is performed using convolutional neural networks. In particular, we used well known deep CNN models for classification of images which have are pre-trained from large image databases and retrained them to learn COVID-19 positive vs negative cases. Chest X-ray images which have been clinically diagnosed as COVID-19 positive are preprocessed and afterwards are used to retrain existing deep CNN models for image classification. X-ray images preprocessing consists of image resizing and pixel values normalization to meet the input specifications of each pretrained deep CNN model. The CNN models are retrained as binary classifiers to identify positive COVID-19 against non-COVID-19 chest X-ray images.

The retrained deep CNN models and used for testing, receiving as input new chest X-rays with unknown clinical diagnosis in order to automatically label them as positive COVID-19 cases or not, i.e. providing a binary decision per chest scan. The block diagram of the evaluated architecture for detection of COVID-19 positive cases from chest X-ray images is illustrated in Fig 1.

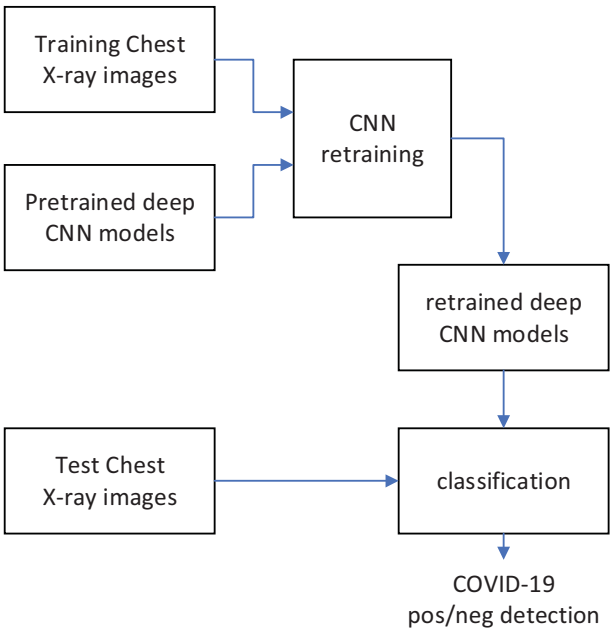

Fig. 1. Block diagram of the evaluated architecture for COVID-19 positive/negative cases detection from chest X-ray images.

III. EXPERIMENTAL SETUP

For the retraining of the deep CNN models and the evaluation of the new/retrained ones two X-ray datasets were used which are available online. The first dataset [17] (Dataset-A) consists of grayscale chest X-ray images of size equal to 1024×1024 pixels. The dataset has three classes and each of the X-ray images has been labeled as 'COVID-19', 'normal' or 'viral pneumonia'. The number of X-ray images per class of [17] is tabulated in Table I.

The second dataset [18] (Dataset-B) consists of grayscale chest X-ray images of size equal to 300×400 pixels. The dataset has four classes and each of the X-ray images has been labeled as 'COVID-19', 'normal', 'viral pneumonia' or 'bacterial pneumonia'. The number of X-ray images per class of [18] is tabulated in Table II.

TABLE I. NUMBER OF X-RAY IMAGES PER CLASS IN THE [17] DATASET (DATASET-A).

Class Name	Number of X-Ray Images
COVID-19	219
Normal	1341
Viral Pneumonia	1345

TABLE II. NUMBER OF X-RAY IMAGES PER CLASS IN THE [18] DATASET (DATASET-B).

Class Name	Number of X-Ray Images
COVID-19	60
Normal	880
Viral Pneumonia	412
Bacterial Pneumonia	650

During preprocessing of the X-ray images, they were resized to 224×224 pixels, using bilinear interpolation, in order to fit to the pretrained deep models input size. The resized X-ray images' pixel values were then normalized to the range [0, 1] in order the retraining of the deep CNN models to converge faster. For the preprocessing of the X-ray images the computer vision and image processing library OpenCV [19] was used.

To apply transfer learning on the X-ray datasets described above and develop models for COVID-19 detection we relied on several well known and widely used deep CNN models for image classification. The pre-trained deep CNN models used are DenseNet-121, DenseNet-169 and DenseNet-201 [20]; Inception-ResNet-V2 [21]; Inception-V3 [22]; MobileNet [23]; MobileNet-V2 [24]; NASNet-Large and NASNet-Mobile [25]; ResNet-50, ResNet-101 and ResNet-152 [26]; ResNet-50-V2, ResNet-101-V2 and ResNet-152-V2 [27]; VGG-16 and VGG-19 [28]; and Xception [29].

IV. EXPERIMENTAL RESULTS

The evaluation architecture presented in Section II was evaluated using the experimental setup described in Section III. The performance of the evaluated deep CNN models was measured in terms of classification accuracy, precision and recall (or sensitivity), i.e.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

$$recall = \frac{TP}{TP + FN} \quad (3)$$

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 of the classified dermatoscopic images. To avoid overlap between the training and testing subsets a 10-fold cross validation protocol was used.

The X-ray image classification results for Dataset-A [25] and for Dataset-B [31] for all evaluated deep CNN models after retraining them are tabulated in Table III. The best performance is indicated in bold.

TABLE III. CLASSIFICATION ACCURACY, PRECISION AND RECALL (IN PERCENTAGES) FOR DIFFERENT RETRAINED DEEP CNN MODELS IN BINARY COVID-19 CLASSIFICATION FOR DATASETS A AND B USING 10-FOLD CROSS VALIDATION.

	Dataset A			Dataset B		
Model Name	**Acc**	**Prec**	**Rec**	**Acc**	**Prec**	**Rec**
DenseNet-121	99.38	94.67	97.26	**99.90**	98.33	**98.33**
DenseNet-169	99.72	97.31	99.09	99.85	98.31	96.67
DenseNet-201	99.31	98.07	92.69	99.85	96.72	**98.33**
Incept.-ResNet-V2	**99.90**	99.09	**99.54**	**99.90**	98.33	**98.33**
Inception-V3	97.15	73.29	97.72	99.75	95.08	96.67
MobileNet	99.69	96.88	99.09	99.55	96.36	88.33
MobileNet-V2	99.45	99.51	93.15	99.75	98.25	93.33
NASNet-Large	95.21	68.02	68.95	98.20	96.15	41.67
NASNet-Mobile	88.04	37.48	88.13	99.05	85.96	81.67
ResNet-50	99.52	95.96	97.72	99.75	95.08	96.67
ResNet-101	99.66	96.44	99.09	99.70	95.00	95.00
ResNet-152	**99.90**	99.54	99.09	99.55	89.23	96.67
ResNet-50-V2	99.52	**100**	93.61	99.85	**100.00**	95.00
ResNet-101-V2	90.24	43.52	98.17	99.35	84.06	96.67
ResNet-152-V2	99.69	98.61	97.26	99.30	87.10	90.00
VGG-16	92.67	**100**	2.74	97.00	0.00	0.00
VGG-19	92.46	0.00	0.00	97.00	0.00	0.00
Xception	**99.90**	99.09	**99.54**	**99.90**	98.33	**98.33**

As can be seen in Table III the best performing models in terms of accuracy and recall are the Inception-ResNet-V2 and Xception, in both datasets. ResNet-50-V2 achieved high precision score, however, recall scores were significantly worse in both datasets when compared with the previously mentioned two best performing models. VGG models failed to classify COVID-19 vs non COVID-19 chest X-ray images.

In addition, we performed binary classification by using each of datasets A/B as training/test and vice versa, with the results being tabulated in Table IV. In specific, the first set of three columns tabulate the accuracy, precision and recall scores when dataset A was used to train CNN models and test them on dataset B, while in the second set of three columns the same performance metrics were used with dataset B being used to train the COVID-19 identification models and dataset B to test them. The results presented in Table IV demonstrate the transferability of the COVID-19 models as models that were trained on one dataset preserved high accuracy, precision and recall scores when tested on another dataset.

Finally, in Table V we evaluated COVID-19 identification performance in terms of accuracy, precision and recall when merging the two datasets A and B. The evaluation results in Table V show that with the increase of the available data to retrain the deep CNN models the performance of them is in general improved. Data augmentation with artificially generated chest X-ray images was not evaluated as considered outside the scope of the present study.

V. CONCLUSION

The new COVID-19 virus has caused thousands of deaths, especially in elders and patients with health conditions. The standard method for detection and diagnosis of COVID-19 is the reverse-transcription polymerase chain reaction (RT-PCR) test after collection of proper respiratory tract specimen, which is time-consuming and in many cases not affordable thus the development of new low-cost rapid tests of diagnostic tools to support clinical assessment is needed.

We presented an evaluation of transfer learning using pretrained deep convolutional neural network models for COVID-19 identification using chest X-ray images. Two publicly available datasets were used in different experimental setups. In specific, we tested the binary COVID-19 identification performance of several convolutional neural network models using 10-fold cross validation on each dataset separately, then we tested the transferability of the models by using one dataset for training and the other for testing and vice versa. Finally, we merged the two datasets and performed 10-fold cross validation to investigate the effect of the size of available data in accuracy, precision and recall. The experimental evaluation demonstrated the potential of building diagnostic tools for automatic detection of COVID-19 positive cases from chest X-ray images and deep convolutional neural networks and the development of larger and clinically standardized datasets would further help in this direction.

TABLE IV. CLASSIFICATION ACCURACY, PRECISION AND RECALL (IN PERCENTAGES) FOR DIFFERENT RETRAINED DEEP CNN MODELS IN BINARY COVID-19 CLASSIFICATION FOR DATASETS A AND B USED AS TRAINING/TEST SUBSETS.

	Train: A/ Test: B			Train: B/ Test: A		
Model Name	**Acc**	**Prec**	**Rec**	**Acc**	**Prec**	**Rec**
DenseNet-121	**100.00**	**100.00**	**100.00**	99.00	95.67	90.87
DenseNet-169	**100.00**	**100.00**	**100.00**	98.38	90.95	87.21
DenseNet-201	**100.00**	**100.00**	**100.00**	98.42	93.47	84.93
Incept.-ResNet-V2	99.85	96.77	**100.00**	98.97	97.49	88.58
Inception-V3	99.78	95.24	**100.00**	98.55	89.33	91.78
MobileNet	**100.00**	**100.00**	**100.00**	**99.24**	94.57	**95.43**
MobileNet-V2	85.72	23.51	98.33	98.73	88.56	**95.43**
NASNet-Large	96.15	**100.00**	13.33	92.56	80.00	1.83
NASNet-Mobile	95.56	0.00	0.00	92.77	**100.00**	4.11
ResNet-50	**100.00**	**100.00**	**100.00**	96.97	90.68	66.67
ResNet-101	17.01	5.08	**100.00**	98.66	90.54	91.78
ResNet-152	**100.00**	**100.00**	**100.00**	95.25	85.84	44.29
ResNet-50-V2	99.48	93.44	95.00	98.00	88.15	84.93
ResNet-101-V2	**100.00**	**100.00**	**100.00**	93.70	55.14	88.13
ResNet-152-V2	97.26	61.86	**100.00**	97.01	95.21	63.47
VGG-16	95.56	0.00	0.00	92.46	0.00	0.00
VGG-19	95.56	0.00	0.00	92.46	0.00	0.00
Xception	**100.00**	**100.00**	**100.00**	98.90	93.49	91.78

TABLE V. CLASSIFICATION ACCURACY, PRECISION AND RECALL (IN PERCENTAGES) FOR DIFFERENT RETRAINED DEEP CNN MODELS IN BINARY COVID-19 CLASSIFICATION FOR MERGED DATASETS A AND B USING 10-FOLD CROSS VALIDATION.

Model Name	Acc	Prec	Rec
DenseNet-121	**99.98**	**100.00**	99.64
DenseNet-169	99.88	**100.00**	97.85
DenseNet-201	96.29	60.52	**100.00**
Incept.-ResNet-V2	99.94	**100.00**	98.92
Inception-V3	99.92	99.64	98.92
MobileNet	99.86	98.23	99.28
MobileNet-V2	99.23	99.59	86.74
NASNet-Large	98.08	84.13	81.72
NASNet-Mobile	97.78	81.25	79.21
ResNet-50	96.88	64.58	**100.00**
ResNet-101	99.90	99.28	98.92
ResNet-152	99.63	94.54	99.28
ResNet-50-V2	99.82	98.56	98.21
ResNet-101-V2	99.78	97.18	98.92
ResNet-152-V2	99.74	97.50	97.85
VGG-16	94.31	0.00	0.00
VGG-19	94.31	0.00	0.00
Xception	**99.98**	**100.00**	99.64

REFERENCES

[1] Andersen, K.G., Rambaut, A., Lipkin, W.I. et al. The proximal origin of SARS-CoV-2. Nat Med 26, 450–452 (2020). https://doi.org/10.1038/s41591-020-0820-9.

[2] Catrin Sohrabia, Zaid Alsafi, Niamh O'Neill, Mehdi Khan, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, Riaz Agha, World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19), International Journal of Surgery, Volume 76, April 2020, Pages 71-76.

[3] Hussin A.Rothan and Siddappa N.Byrareddy, The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak, Journal of Autoimmunity, Volume 109, May 2020, 102433, https://doi.org/10.1016/j.jaut.2020.102433.

[4] Kai Liu, Ying Chen, Ruzheng Lin, Kunyuan Han, Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients, Journal of Infection, Available online 27 March 2020, In Press, https://doi.org/10.1016/j.jinf.2020.03.005.

[5] John P. A. Ioannidis, Cathrine Axfors, Despina G. Contopoulos-Ioannidis, Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters, medRxiv 2020.04.05.20054361; doi: https://doi.org/10.1101/2020.04.05.20054361.

[6] Yi-Wei Tang, Jonathan E. Schmitz, David H. Persing, Charles W. Stratton, The Laboratory Diagnosis of COVID-19 Infection: Current Issues and Challenges, Journal of Clinical Microbiology Apr 2020, JCM.00512-20; DOI: 10.1128/JCM.00512-20.

[7] Kim Usher, Navjot Bhullar, Joanne Durkin, Naomi Gyamfi, Debra Jackson, Family violence and COVID‐19: Increased vulnerability and reduced options for support, International Journal of Mental Health Nursing, First published: 20 April 2020, https://doi.org/10.1111/inm.12735.

[8] World Economic Outlook Reports, World Economic Outlook, April 2020: The Great Lockdown, available at: https://www.imf.org/en/Publications/WEO/Issues/2020/04/14/weo-april-2020.

[9] The impact of the Covid-19 pandemic on global and EU trade, Chief Economist Team, DG Trade, European Commission, 17 April 2020, available at: https://trade.ec.europa.eu/doclib/docs/2020/april/tradoc_158713.pdf.

[10] Aboul Ella Hassanien Sr., Lamia Nabil Mahdy Jr., Kadry Ali Ezzat Jr., Haytham H. Elmousalami Jr., Hassan Aboul Ella Jr. (2020), Automatic X-ray COVID-19 Lung Image Classification System based on Multi-Level Thresholding and Support Vector Machine, medRxiv, doi: https://doi.org/10.1101/2020.03.30.20047787.

[11] Abbas A , Abdelsamea M M, Medhat Gaber M, (2020) Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network, medRxiv, doi: https://doi.org/10.1101/2020.03.30.20047456.

[12] Sethy PK, Behera SK. (2020) Detection of Coronavirus Disease (COVID-19) Based on Deep Features, International Journal of Mathematical, Engineering and Management Sciences Vol. 5, No. 4, 643-651.

[13] Wang L, Wong A. (2020), COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images; arXiv: 2003.09871.

[14] Li et al. (2020), Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. RSNA Radiology, https://doi.org/10.1148/radiol.2020200905.

[15] Ronneberger, O., Fischer, P. and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). [online] Switzerland: Springer, Cham,159-168.

[16] Gozes O et. Al (2020), Rapid AI Development Cycle for the Coronavirus (COVID-19) Pandemic: Initial Results for Automated Detection & Patient Monitoring using Deep Learning CT Image Analysis, Radiology: Artificial Intelligence.

[17] Chowdhury, M. E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Al-Emadi, N. (2020). Can AI help in screening Viral and COVID-19 pneumonia? arXiv preprint arXiv:2003.13145.

[18] Darshan. (2020). COVID-19 Detection X-Ray Dataset. Retrieved from https://www.kaggle.com/darshan1504/covid19-detection-xray-dataset.

[19] Bradski, G. (2000). The opencv library. Dr Dobb's J. Software Tools, 25, 120-125.

[20] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition.

[21] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A.A., 2017, February. Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence.

[22] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[23] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[24] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[25] Zoph, B., Vasudevan, V., Shlens, J. and Le, Q.V., 2018. Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8697-8710).

[26] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

[27] He, K., Zhang, X., Ren, S. and Sun, J., 2016, October. Identity mappings in deep residual networks. In European conf. on computer vision (pp. 630-645).

[28] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[29] Chollet, F., Xception: Deep learning with depthwise separable convolutions. In 2017 IEEE CVPR (pp. 1251-1258).