Shallow Convolutional Neural Network for COVID-19 Outbreak Screening using Chest X-rays

SUBMISSION DATE / POSTED DATE

20-04-2020 / 22-04-2020

CITATION

Mukherjee, Himadri; Ghosh, Subhankar; Dhar, Ankita; Obaidullah, Sk. Md.; Santosh, KC; Roy, Kaushik (2020): Shallow Convolutional Neural Network for COVID-19 Outbreak Screening using Chest X-rays. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.12156522.v1

DOI

10.36227/techrxiv.12156522.v1

Shallow Convolutional Neural Network for COVID-19 Outbreak Screening using Chest X-rays

Himadri Mukherjee[1], Subhankar Ghosh[2], Ankita Dhar[1], Sk. Md. Obaidullah[3], K.C. Santosh[4,*] and Kaushik Roy[1]

[1]Department of Computer Science, West Bengal State University, Kolkata, India; {himadrim027, ankita.ankie, kaushik.mrg}@gmail.com

[2]CVPR Unit, Indian Statistical Institute, Kolkata, India; sgcs2005@gmail.com

[3]Department of Computer Science and Engineering, Aliah University, Kolkata, India; sk.oaidullah@gmail.com

[4]Department of Computer Science, The University of South, Dakota, Vermillion, SD, USA; santosh.kc@ieee.org

⋆Corresponding author

Abstract

Among radiological imaging data, chest X-rays are of great use in observing COVID-19 manifestations. For mass screening, using chest X-rays, a computationally efficient AI-driven tool is the must to detect COVID-19 positive cases from non-COVID ones. For this purpose, we proposed a light-weight Convolutional Neural Network (CNN)-tailored shallow architecture that can automatically detect COVID-19 positive cases using chest X-rays, with no false positive. The shallow CNN-tailored architecture was designed with fewer parameters as compared to other deep learning models, which was validated using 130 COVID-19 positive chest X-rays. In this study, in addition to COVID-19 positive cases, another set of non-COVID-19 cases (exactly similar to the size of COVID-19 set) was taken into account, where MERS, SARS, Pneumonia, and healthy chest X-rays were used. In experimental tests, to avoid possible bias, 5-fold cross validation was followed. Using 260 chest X-rays, the proposed model achieved an accuracy of an accuracy of 96.92%, sensitivity of 0.942, where AUC was 0.9869. Further, the reported false positive rate was 0 for 130 COVID-19 positive cases. This stated that proposed tool could possibly be used for mass screening. Note to be confused, it does not include any clinical implications. Using the exact same set of chest X-rays collection, the current results were better than other deep learning models and state-of-the-art works.

Index Terms

COVID-19; Chest X-rays; Deep Learning; Convolutional Neural Network; Mass Screening.

I. Introduction

In December 2019, the novel coronavirus disease (COVID-19) was found in Wuhan Province of China [1], [2]. Unlike the common cold and flu, COVID-19 is much more contagious, and

for humans immune system, it is an absolute unknown. To be more specific, Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) are two well-known coronavirus diseases, which have huge mortality rates of 10% and 37%, respectively [3], [4]. As of now, COVID-19 affected more than 823626 people with more than 40,598 death cases across the world (dated, April 02, 2020) [9]. The COVID-19 outbreak spread rate is exponential and is faster than other respiratory-related diseases. Researchers are now limited to small amount of data to predict possible consequences using AI-driven tools.

In January 2020, Huang C et al. reported some clinical and paraclinical aspects of COVID-19 using 41 patients. Their study stated that abnormalities, such as Ground-Glass Opacity (GGO) can be observed using chest CT scans [5]. CT scans are widely used to identify unusual patterns in confirmed cases of COVID-19 [6], [7], [8]. To be precise, Li Y and Xia L [8] experimented on 51 CT images and in 96.1% cases COVID-19 was successfully detected. Zhou S et al. [11] experimented on 62 COVID-19 and Pneumonia, and their results showed diverse patterns that are visually like lung parenchyma and the interstitial diseases. Also, Zheng Ye et al. [12] stated that typical and atypical CT manifestations help and familiarize radiologists in decision-making.

In a similar fashion, chest X-rays (CXRs) have been widely used to detect COVID-19 positive cases [13], [14], [15], [16]. Soon et al. [13] observed the relationship between CXRs and CT images, where 9 COVID-19 positive cases were used. Besides, others were focused on the use of Neural Network-tailored deep learning (DL) models, such as COVID-Net [14] and resnet50 [15]. COVID-Net was tested only on 8 COVID-19 positive cases, while resnet50 was validated on 50 COVID-19 positive cases. Similarly, Zhang et al. [16] used classical DL model to detect COVID-19 positive cases, where 100 COVID-19 samples were used. As of now, the highest accuracy of 96Motivated by the fact that X-ray imaging systems are more prevalent and cheaper than CT scan systems, in this paper, a shallow Convolutional Neural Network (CNN) is proposed to detect COIVID-19 positive cases from non-COVID-19 ones using CXRs.

II. Materials and methods

A. Data collection

AI-driven tools require enough data so that all possible infestations are trained [18]. However, as of now, we do not have large amount of data for COVID-19 positive cases, unlike other respiratory-related diseases. Radiological imaging data are of great use in observing COVID-19 manifestations, where chest X-rays (CXRs) imaging systems are more prevalent and cheaper than CT scan systems. As an example, Chest X-ray is the first imaging method to diagnose COVID-19 coronavirus infection in Spain (dated, March 20, 2020). A chest X-ray is performed in suspected or confirmed patients through specific circuits.

In this paper, publicly available collection of data that is composed of COVID-19 positive CXRs [19] was used. Altogether, it includes 130 COVID-19 positive and 51 non-COVID-19 cases. The collection includes CXRs related to MERS, SARS, and ARDS. Since AI-driven tool requires balanced dataset, non-COVID-19 cases are required to be added. For this purpose, another collection (publicly available via Kaggle) of CXRs were used [20], where both 48 Pneumonia positive, and 31 healthy CXRs were used. Like COVID-19 CXRs, non-COVID-19 category contains 130 CXRs. Overall, the non-COVID-19 category was composed of multiple

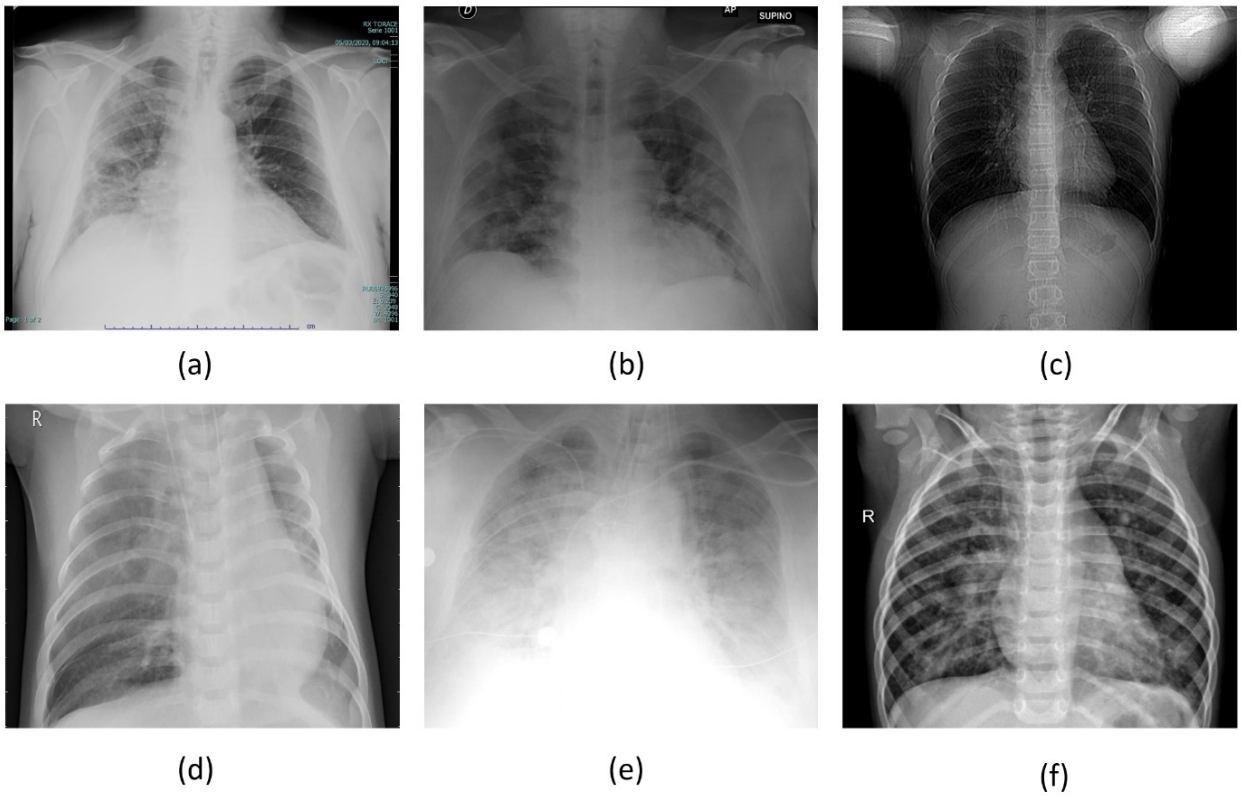

Fig. 1. CXR samples: (a)-(c) COVID-19 positive cases; and (d)-(f) non-COVID-19 cases. CXR image sizes are varied from one collection to another.

abnormalities as well as normal cases. The reason behind this mixed bag in one category is to check whether the proposed model can classify COVID-19 positive case from its counterparts. Further, it was observed that CXRs with Pneumonia were identical to the COVID-19. Few CXR samples are show in Figure 1.

B. Shallow convolutional neural network

Convolutional neural networks (CNNs) [21], [17] are a class of neural networks which work on the principle of deep learning. A basic CNN architecture consists of alternate layers of convolutional and pooling followed by one or more fully connected layers at the final stage.

The convolutional layer is the prime ingredient of this architecture that detects the existence of a set of features from the input. This layer comprises a set of convolutional kernels. The functioning of this layer can be computed as, $f_c^k(m,n) = \sum_d \sum_{r,s} j_d(r,s).i_c^k(v,w)$, where, $j_d(r,s)$ is an instance of the input vector J_d, which is multiplied by $i_c^k(v,w)$ index of the kth kernel of the cthlayer. The output mapping of the kth kernel can me measured as, $F_c^k = [f_c^k(1,1), ..., f_c^k(m,n), ..., f_c^k(M,N)]$.

The pooling layer is arranged between two convolutional layers that reduces the size of the vectors while keeping their relevancy intact. It aggregates the related information in the region of the receptive domain and outputs the feedback within that region using $Y_c^k = 0_p(F_c^k)$, where

Y_c^k determine the pooled feature map of the cth layer for kth kernel and 0_p determines the kind of pooling operation.

The dense layer accepts the input from the previous stages and globally evaluates the output of all the former layers. Hence, makes a non-linear combination of specified features that are used for the classification purpose.

In this paper, a shallow CNN architecture is proposed, which consists of only four layers as compared to deep architectures. The primary motivation behind this was to design light architecture with minimal number of parameters (weights) so that it does not suffer from heavy computational time. As a result, the proposed shallow (or light-weight) CNN architecture is not just computationally efficient but also is able to avoid possible overfitting. More often, deep architectures are prone to overfitting due to their heavy usage of parameters, and of course, longer training period. The proposed shallow CNN architecture is therefore a better fit for mass population screening especially in resource constrained areas.

The network consists of a single convolution layer, followed by a max-pooling layer and a 256-dimensional dense layer. This was finally followed by a 2 dimensional output layer. Initially, the images were scaled down to 50x50 pixels and fed to the network. The convolution layer and the first dense layer had Rectified Linear Unit activation function: $f(x) = max(0, x)$, where x is a input to a neuron. The final dense layer had a softmax activation function: $\sigma(z)_j = \frac{e^{z_j}}{\Sigma_{k=1}^{K} e^{z_k}}$, where z_i is an element of input vector z of size K.

Using the proposed shallow CNN architecture, the generated feature maps for COVID-19 positive and Pneumonia CXRs are shown in Figure 2.

III. Results

To validate the proposed architecture, a 5-fold cross validation was considered for all tests. This provides a thorough statistical analysis of the model. Since the proposed shallow CNN architecture requires several parameters, the first set of experiment tests were based on how well the model can be trained. For this, a few essential parameters, such as image size, number of filters used in convolutional layer and its filter size, pooling window size, and batch size were considered. In what follows, these parameters are discussed.

A. Image size

CXR image size were of different sizes in the dataset. They were, therefore, resized into a fixed dimension. For experimental purpose, the resized dimensions were varied from 50x50 to 150x150 pixels. With the proposed model, better result (accuracy = 96.15%) was obtained from CXRs of size 100x100 as compared to 50x50 (accuracy = 95.00%) and 150x150 (accuracy = 50%).

B. Number of filters in convolution layer

In convolution layer, different the numbers of filters were employed, such as 5, 10, 20, 30. Of all, it is observed best results were obtained from the experimental test, where 10 filters were used.

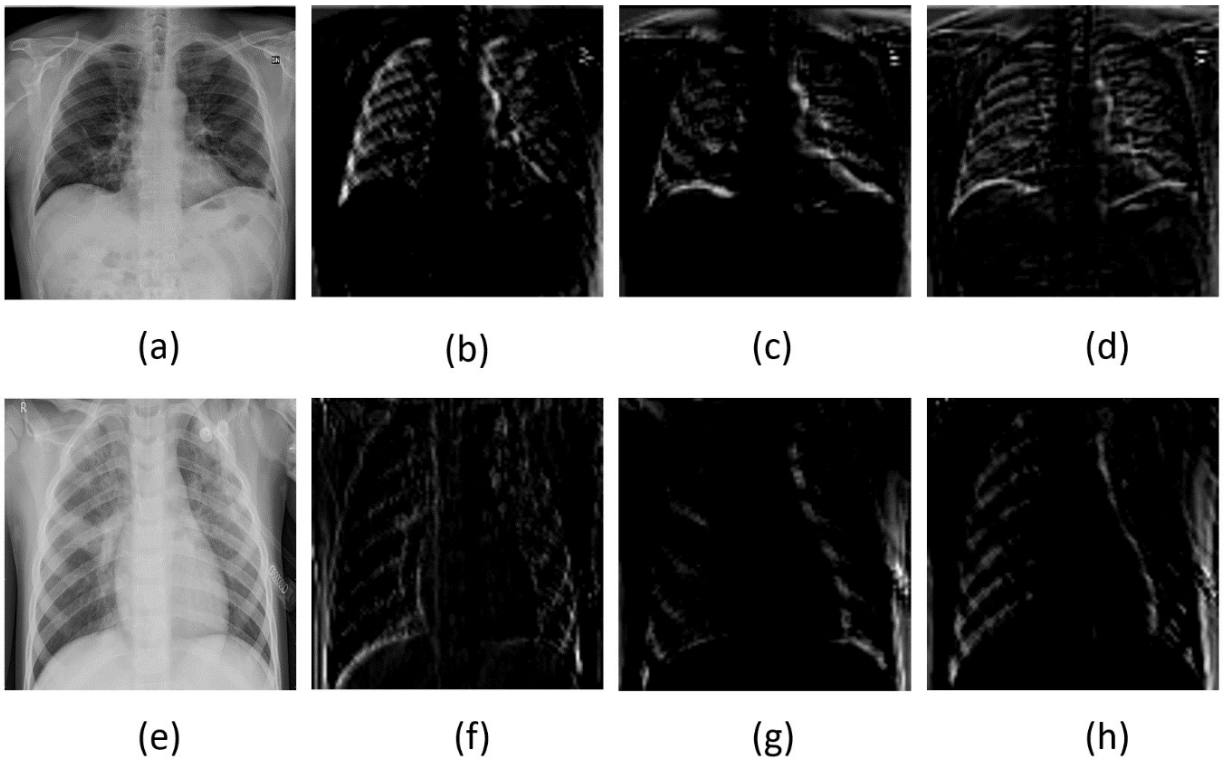

Fig. 2. Feature map visualization: (a) COVID-19 positive case and its corresponding (b)-(D) feature maps; and (e) Pneumonia positive case and its corresponding (f)-(g) feature maps.

C. Filter size

The convolution filter size was also experimented with, from 2 to 5 with an increment of 1: 2x2, 3x3,..., 5x5. Of all, the best results were obtained from the filter of size 4x4 (accuracy = 96.92%).

D. Pooling window size

Like filter size, the pooling windows were varied from 2 to 4, with an increment of 1: 2x2, 3x3, and 4x4. The experimental test results dropped on increasing the pooling window size. Therefore, the pooling window size was fixed at 2 (accuracy = 96.92%).

E. Batch size

During training period, different batch sizes were applied, starting from 25 to 150 instances with the difference of 25. Of all, best results were obtained from both 50 and 75 instances (accuracy = 96.92%). In both batch sizes, 130 COVID-19 positive cases were correctly identified. The closest performance to the best results was obtained from 25 instance batch size and the accuracy was 95.77%. In this case, 2 true negative cases and 9 false positives were identified. The detailed experimental test results for different batch sizes are provided in Table I. The results

TABLE I
CONFUSION MATRICES FOR DIFFERENT BATCH SIZES.

Batch size:25			Batch size:50		
	COVID-19	Non- COVID-19		COVID-19	Non- COVID-19
COVID-19	128	2	COVID-19	130	0
Non-COVID-19	9	121	Non-COVID-19	8	122
Batch size:75			Batch size:100		
	COVID-19	Non- COVID-19		COVID-19	Non- COVID-19
COVID-19	130	0	COVID-19	123	7
Non-COVID-19	8	122	Non-COVID-19	14	116
Batch size:125			Batch size:150		
	COVID-19	Non-COVID-19		COVID-19	Non-COVID-19
COVID-19	129	1	COVID-19	130	0
Non-COVID-19	12	118	Non-COVID-19	11	119

TABLE II
PERFORMANCE METRICS FOR DIFFERENT BATCH SIZES.

Metrics	Batch size					
	25	50	75	100	125	150
Sensitivity	0.9343	0.9420	0.9420	0.8978	0.9149	0.9220
Specificity	0.9837	1.0000	1.0000	0.9431	0.9916	1.0000
Precision	0.9846	1.0000	1.0000	0.9462	0.9923	1.0000
False positive rate	0.0163	0.0000	0.0000	0.0569	0.0084	0.0000
False negative rate	0.0657	0.0580	0.0580	0.1022	0.0851	0.0780
Accuracy (%)	95.77	96.92	96.92	91.92	95.00	95.77
F1 Score	0.9588	0.9701	0.9701	0.9213	0.9520	0.9594
AUC	0.9869	0.9922	0.9921	0.9742	0.9915	0.9908

were further analyzed with respect to several performance metrics, such as sensitivity, specificity, precision, F1 score, and AUC, which are detailed in Table II.

For better understanding, the ROC curves are presented in Figure 3 along with the corresponding AUC values. It is noted that, even though 50 and 75 instance batch sizes reported exact same accuracy, 50 instances batch size achieved a higher AUC of 0.9922 as compared to 75 (AUC = 0.9921).

On the whole, for the proposed shallow CNN-based architecture, parameters were tuned for upcoming test purposes. The best performance scores were achieved when the architecture used 10 filters of size 4x4 in the convolution layer, batch size of 50 for CXR image size of 100X100 pixels, and the window size of 2x2 in the pooling layer. Precisely, the results were provided by considering the following evaluation metrics: sensitivity, specificity, precision, false positive rate, false negative rate, accuracy, F1 score, and AUC (see Table II). The proposed model provided the highest possible accuracy of 96.92% with an AUC of 0.9922. It is important to note that the proposed model received a false positive rate of 0.

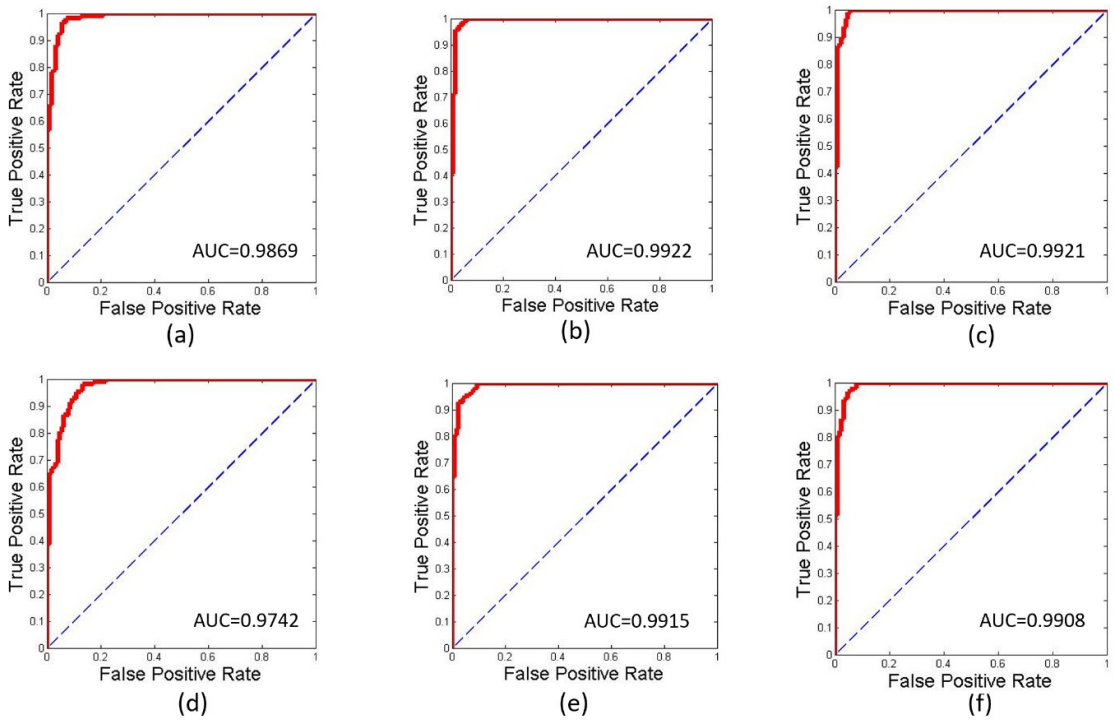

Fig. 3. ROC curves for different batch sizes: (a) 25 batch size; (b) 50 batch size; (c) 75 batch size; (d) 100 batch size; (e) 125 batch size; and (f) 150 batch size. The batch size of 50 was found to be the best of all.

IV. Discussions

For COVID-19 screening, since sensitivity measures the likelihood that the model would not miss to detect COVID-19 positive patients, it plays a crucial role in validating model in early stages of a pandemic. As a consequence, it helps prevent further COIVD-19 spreading. The similar argument lies in computing false positive rate. The proposed model achieved a sensitivity score of 0.9420 (on average), using 5-fold cross validation protocol. Further, precision indicates the probability in detecting COVID-19 positive cases. It is useful as it measures the likelihood that a model would not make a mistake to classify the COVID-19 positive patients as normal and it is important in the later stages of a pandemic, when medical resources are limited to COVID-19 patients. The proposed model reported an average precision of 1.00. This means that false positive rate was 0. As the work was particularly focused on detecting COVID-19 positive cases, the discussion can be limited to sensitivity, false positive rate, and precision. Accuracy was computed to measure the overall performance of the model. AUC is, on the other hand, conveys how stable the system was: degree of measure of separability (between two categories: COVID-19 and non-COVID-19). Similarly, other metrics (as provided in Table II are of importance to test further the robustness of the model.

For a comparison, the exact same set of experimental datasets were applied to other popular Deep Learning (DL) architectures, such as MobileNet [22] and VGG16 [23]. Their performance scores along with the number of generated parameters were presented in Table III. As compared to VGG16, MobileNet performed well, and achieved an overall accuracy of 83.08%. The pro-

TABLE III
PERFORMANCE COMPARISON WITH OTHER DEEP LEARNING MODELS.

Metrics	Architectures		
	MobileNet [22]	VGG16 [23]	Shallow CNN (proposed)
Sensitivity	0.7905	0.0000	0.9420
Specificity	0.8839	0.5000	1.0000
Precision	0.9000	0.0000	1.0000
False positive rate	0.1161	0.5000	0.0000
False negative rate	0.2095	0.0000	0.0580
Accuracy (%)	83.08	50	96.92
F1 Score	0.8417	0.0000	0.9701
AUC	0.9133	0.5000	0.9922
Parameters	7,423,938	16,812,610	5,899,500

TABLE IV
COMPARATIVE STUDY. INDEX: WANG AND WONG [14], SETHY AND BEHERA [15], AND ZHANG ET AL. [16] ARE TAKEN FROM ARXIV.ORG (NON-PEER-REVIEWED). SINCE AUTHORS DID NOT REPORT RESULTS FOR SEVERAL DIFFERENT METRICS, THERE EXISTS SYMBOL — IN THE TABLE.

Metrics	Wang and Wong [14]	Sethy and Behera [15]	Zhang et al. [16]	Shallow CNN (Proposed)
Dataset (# of COVID-19 positive cases)	10	50	100	130
Sensitivity	0.80	0.9729	0.9600	0.9420
Specificity	—	97.4705	0.7065	1.0000
Precision	—	—	—	1.0000
False positive rate	—-	—	—	0.0000
False negative rate	—	—	—	0.0580
Accuracy (%)	92.4	95.38	83.34	96.92
F1 Score	—	0.9552	—	0.9701
AUC	—-	—-	—	0.9922

posed shallow CCN-based architecture outperformed MobileNet by more than 13.84% in terms of accuracy. Not limited to accuracy, the proposed model outperformed others with remarkable difference in terms of other metrics, such as sensitivity, false positive rate, precision, and AUC. Considering computational complexity issue, the proposed model required 5,899,500 number of parameters, which was 1.258 times smaller than MobileNet [22] and 2.850 times smaller than VGG16 [23].

Further, since recently a few researchers worked on exact same set of datasets (but different sizes), they were taken into consideration for a comparison. A comparative study was provided in Table IV. Wang and Wong [14] tested their tool on 10 COVID-19 positive cases. With this small dataset of size 10, their reported accuracy was 92.40% with a sensitivity score of 0.80. This means that all COVID-19 positive cases were not correctly classified. Sethy and Behera [15] tested 50 COVID-19 positive cases and reported an accuracy of 95.38%, where sensitivity was 97.44. Zhang et al. [16] reported an accuracy of 96.00%, where 100 COVID-19 positive cases

were used. For the newly updated dataset collection, which was composed of 130 COVID-19 CXRs, the proposed model achieved higher performance scores in terms of accuracy, precision, sensitivity, false positive rate, F1 score, and AUC (see Table IV). To be precise, test results were higher than state-of-the-art works.

V. Conclusions

On the whole, in this paper, a light-weight CNN-tailored shallow architecture was proposed to detect COVID-19 positive cases using CXRs against non-COVID-19 ones. The experiments were performed on dataset collections of COVID19 positive, Pneumonia positive, SARS positive, MERS positive and healthy CXRs. To validate its robustness, 5-fold cross validation protocol was used, and a comparison study was performed by taking a) popular DL tools, such as MobileNet and VGG16; and b) state-of-the-art works for COVID-19 detection using CXRs, into account. The proposed model outperformed all and is computationally efficient as it requires less number of parameters. As the proposed shallow CNN-tailored architecture has no false positive, it could be used to scree COVID-19 positive cases in chest X-rays.

Authors contributions

a) Conceptualization: Himadri Mukherjee and K.C. Santosh;
b) Methodology, Himadri Mukherjee and Subhankar Ghosh;
c) Validation, Ankita Dhar and K.C. Santosh;
d) Writing–original draft preparation, Himadri Mukherjee and K.C. Santosh;
e) Writing–review and editing, K.C. Santosh, and Sk. Md. Obaidullah; and
d) Supervision, K.C. Santosh and Kaushik Roy.

All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Conflict of interest

The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
COVID-19	Coronavirus Disease 2019
SERS	Severe Acute Respiratory Syndrome
MERS	Middle East Respiratory Syndrome
CXR	Chest X-ray
CNN	Convolutional Neural Network
GGO	Ground-Glass Opacity
DL	Deep learning

REFERENCES

[1] https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

[2] Novel coronavirus China. http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/ Date: Jan 12, 2020 Date accessed: March 25, 2020

[3] Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. https://www.who.int/csr/sars/country/table2004_04_21/en/ Date: Dec 31, 2003 Date accessed: March 25, 2020

[4] Middle East respiratory syndrome coronavirus (MERS-CoV). http://www.who.int/emergencies/mers-cov/en/ Date: November, 2019 Date accessed: March 25, 2020

[5] Huang C et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, in Lancet. 2020 Feb 15;395(10223):497-506. doi: 10.1016/S0140-6736(20)30183-5. Epub 2020 Jan 24.

[6] Fang Y., Zhang H., Xie J., Lin M., Ying L., Pang P., Ji W. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR in Radiology, Feb 19 2020, https://doi.org/10.1148/radiol.2020200432

[7] NG M. et al., Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review in Radiology: Cardiothoracic Imaging, Vol. 2, No. 1, Feb 13 2020, https://doi.org/10.1148/ryct.2020200034

[8] Li Y, Xia L, Coronavirus Disease 2019 (COVID-19): Role of Chest CT in Diagnosis and Management in Am J Roentgenol. 2020 Mar 4:1-7. doi: 10.2214/AJR.20.22954.

[9] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (72)

[10] https://healthcare-in-europe.com/en/news/imaging-the-coronavirus-disease-covid-19.html

[11] Zhou S, Wang Y, Zhu T, Xia L, CT Features of Coronavirus Disease 2019 (COVID-19) Pneumonia in 62 Patients in Wuhan, China, in Am J Roentgenol. 2020 Mar 5:1-8. doi: 10.2214/AJR.20.22975.

[12] Zheng Ye, Yun Zhang, Yi Wang, Zixiang Huang & Bin Song, Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review, in European Radiology, 19 March 2020.

[13] Yoon S. et al. (2020) Chest Radiographic and CT Findings of the 2019 Novel Coronavirus Disease (COVID-19): Analysis of Nine Patients Treated in Korea. In Korean Journal of Radiology. 21(4):494-500

[14] Wang, L. and Wong, A. (2020) COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images. In Arxiv.

[15] Sethy, P.K. and Behera, S.K.. (2020). Detection of Coronavirus Disease (COVID-19) Based on Deep Features. Preprints 2020, 2020030300.

[16] Zhang, J. Xie, Y., Li, Y., Shen, C., Xia, Y. (2020) COVID-19 Screening on Chest X-ray Images Using Deep Learning based Anomaly Detection. In arXiv

[17] Gan, F., Luo, C., Liu, X., Wang, H. and Peng, L. Fast Terahertz Coded-Aperture Imaging Based on Convolutional Neural Network. Appl. Sci. 2020, 10, 2661.

[18] Santosh, K.C.AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data (2020). J Med Syst 44, 93.

[19] https://github.com/ieee8023/covid-chestxray-dataset

[20] https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[21] Krizhevsky A., Sutskever I., and Hinton G.E.. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[22] Chen H. and Su C. An Enhanced Hybrid MobileNet, 2018 9th International Conference on Awareness Science and Technology (iCAST), Fukuoka, 2018, pp. 308-312.

[23] Alippi C., Disabato S., Roveri M. Moving Convolutional Neural Networks to Embedded Systems: The AlexNet and VGG-16 Case, 2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Porto, 2018, pp. 212-223.