INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

COVID-19 Chest X-Ray Case Detection with Ensemble Deep-Learning

R Ashok Kumar[1]*, Shaheda Akthar[2]

1 Lecturer in Computer Science, SCIM Government College(A), Tanuku, West Godavari, AP, India
2 Research Supervisor, Dept. of CSE, Acharya Nagarjuna University, Guntur, AP, India

Check for updates

OPEN ACCESS

Received: 18-11-2022
Accepted: 02-04-2024
Published: 03-05-2024

Citation: Kumar RA, Akthar S (2024) COVID-19 Chest X-Ray Case Detection with Ensemble Deep-Learning. Indian Journal of Science and Technology 17(18): 1898-1905. https://doi.org/10.17485/IJST/v17i18.2220

* **Corresponding author**.

rebbaashok@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: The objective of this research is to enhance accuracy on the COVID-19 case identification using X-ray imagery by addressing the drawbacks of utilising a single deep learning model, such as overfitting, high variance, and generalisation errors, by generating predictions with numerous frameworks as opposed to one model. **Methods**: In this study, secondary data sets from a group of experts from Qatar University in Doha, Qatar, and the University of Dhaka in Bangladesh, together with partners from Pakistan and Malaysia, have produced a dataset of 21,135 CXR pictures for COVID-19 patients, as well as pictures of normal and viral pneumonia. The performance of proposed strategy is, EnDL-COVID-19 is compared with three parameters Accuracy, Sensitivity, PPV Assessment. **Findings**: ENDL-COVID-19 gives good results for COVID-19, instances identification with a performance of 95%, higher than COVID-Net at 93.3%, are according to research observations ENDL-COVID-19 outperformed by a significant margin in a series of experiments using QU&UD test data consisting of 1592 CXR images. It was able to achieve a sensitivity of 96% and a PPV of 94.1% in determining whether or not COVID-19 occurrences were present. **Novelty**: The proposed weighted averaging ensemble technique, which is aware of the various sensitivities of deep-learning frameworks on various category types, is used to combine multiple snapshot frameworks of COVIDNet, which made a breakthrough in an open sourced COVID-19 case identification approach using chest X-ray pictures analyzed by deep neural networks.

Keywords: COVID-19; Ensembling Learning; Deep-Learning; EnDL-COVID-19; X-Ray Pictures

1 Introduction

The new Corona virus infection known as COVID-19 is induced by the SARS corona virus. In order to put a halt to the fast spread of COVID-19, one of the strategy, RT-PCR testing[1], It now also has a number of limitations. This might result in both false-negative and false-positive outcomes. Another way, CXR imagery, as opposed to CT scans, is portable, speedier, and more generally available. Additionally, it may be performed in an isolated setting while still achieving a sufficient COVID-19 disease detection rate[2]. Because of these benefits, a number of recent studies[2,3] are now concentrating on CXR image study for the discovery of COVID-19 cases[2,3]. In particular, a recent study proposes handy CXR imagery as a trustworthy way for identifying COVID-19 cases.

Wang et al.[4] advocate using the COVID-Net convolutional neural network (CNN) model to obtain a diagnosis if you have a chest X-ray and suspect a coronavirus infection. The author relied on the COVIDx CXR imagery, which comprises just 76 pictures of COVID-19 cases, as well as 8066 imagery of normal persons and 5526 imagery of pneumonia patients who did not have COVID-19. The COVID-Net design pattern is PEPX (residual projection extension-projection extension). A fair balance of correctness and computation complexity is attained by attaining a test correctness of 92.4% with only 2260 million multiply–accumulate operations needed to complete case forecast. It should be noted that some research with CT scans instead of CXR imagery to determine COVID-19. This study works with very small data sets, due to the tiny datasets, the performance drastically changes in the case of large datasets.

Ioannis et al.[5] evaluate the effectiveness of modern CNN architectures for categorising medical pictures, such as VGG19, MobileNet v2, Xception, Inception-ResNet-v3. The writer employs transfer learning because to its utility in identifying a range of irregularities in very tiny datasets of medical pictures. Patients with bacterial pneumonia, viral pneumonia, confirmed cases of COVID-19 illness, and 504 healthy controls were among the 1442 X ray datasets used. Based on the data, we may conclude that MobileNet-v2 and VGG19 are the most accurate among the remaining CNNs but significantly accuracy decreases when you run on large data sets. MobileNet-sensitivity v2 accuracy (99.10%) and specificity (97.09%) are also higher than VGG19 accuracy and specificity (98.75%). On large data sets, the accuracy drops to approximately 90% because the deep learning system may be vulnerable to overfitting, excessive variation, and generalization mistakes brought on by a dearth of information.

Individuals like Narin worked hard to solve the issue of a lack of COVID-19 test kits in public hospitals[6] recommends installing an automated detection system as a quick diagnostic backup to avoid the spread of COVID-19 and relieve the pressure on healthcare professionals. The author used 100 chest X-ray pictures and three different CNN origin frameworks (Inception-ResNet-V2, Inception-V3, and ResNetV2) to identify individuals with coronavirus pneumonia (Fifty of normal imagery and 50 of COVID-19 imagery).

The pre-trained ResNet50 model outperforms the other two suggested models by a wide margin, with an accuracy of 92% (InceptionV3 obtains an accurateness of 91 percentage and Inception-ResNet-V2 achieves an accuracy of 86%). This study works with very small data sets because of the tiny datasets; the accuracy and sensitivity drastically change in the case of large datasets because of excessive variation in the deep learning model.

Ozturk et al.[7] describe a novel method for automatically detecting COVID-19 in raw radiography images in order to better precisely identify the virus and alleviate the shortage of experts in rural areas. The real-time object detection system (YOLO) uses a model called DarkNet, which consists of 17 convolutional layers, as its classifier. The authors use a multi-step process, beginning with layer-specific filtering. This approach seeks to provide a perfect diagnosis for both two category (COVID/No-COVID) and three category classifications (COVID/ No-COVID /pneumonia). The accuracy of the three category classifications is 87.02 percent. This model has less accuracy, sensitivity, and PPV assessment in the case of multi classification compared to binary classification.

Sethy and Behra[8] offer an approach for detecting patients with coronavirus infections from X-ray pictures that is base on deep features and SVM. They use SVM-based classifications rather than deep-learning for categorization. The deep features are obtained with CNN model's entirely linked layers. They are then classified using an SVM. SVM is used to categorise corona virus diseased X-rays. Throughout the procedure, COVID-19 X-rays, pneumonia X-rays, and conventional X-rays are employed. With the deep functions of thirteen distinct CNN frameworks, the author assesses the performance of SVM for distinguishing COVID-19. SVM's superior performance may be attributed to its use of ResNet50's extensive characteristics. The combined accuracy of ResNet50 with SVM is 93.66%. According to the outcomes, this model has less accuracy in the case of multi-classification compared to binary classification.

On large datasets, all of the above CNN models produce false-negative and false-positive results. While a few of them claim to have accomplished telltale signs on large data sets, their deep learning systems may be vulnerable to overfitting, excessive variation, and generalization mistakes brought on by noise and a dearth of information.

2 Methodology

In this, Secondary data set from a group of experts from Qatar University in Doha, Qatar, and the University of Dhaka in Bangladesh, together with partners from Pakistan and Malaysia, have produced a dataset of CXR pictures for COVID-19 patients, as well as pictures of normal and viral pneumonia[9,10]. It will have 21,135 CXR images from several individuals, including 1345 Pneumonia cases, 10192 Normal cases, and 3,616 COVID-19 cases and 6012 Lung opacity CXR images. COVID data were gathered from several public datasets, internet sources, and printed publications. Pad chest dataset yielded 2473 CXR images. 183 CXR pictures from a medical school in Germany. Imagery of 559 CXR from SIRM, Github, Kaggle, and Tweeter. 400 CXR photos obtained from another Github repository. 10192 Normal CXR pictures gathered from three distinct sources. From RSNA 8851 images and from Kaggle 1341 images. RSNA CXR dataset contains 6012 lung opacity CXR pictures. Dataset of 1345 CXR Pictures with viral pneumonia.

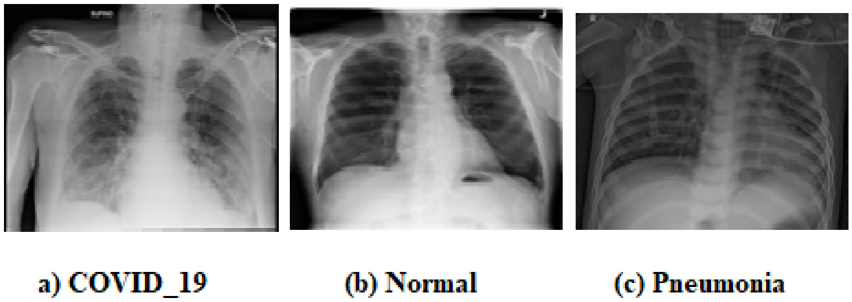

Fig 1. CXR imagery of normal and diseased persons from QU & UD databases, this is further categorized as classes of: (a) COVID-19 Category, (b) Normal Category (c) Pneumonia Category

On the above open-source secondary data set, EnDL-COVID, a quick-look ensembling deep-learning system built on COVID-Net, was presented in this part. Figure 2 depicts the whole training flow for EnDL-COVID, which includes the two stages; one is Quick-Look Model Training and another Assembling Models. The model assembly step takes the snapshots created in the snapshot model training phase and combines them to form a final forecast.

Quick-Look Model Training:

Deep learning ensemble requires pre-trained models. Deep learning ensemble is time and resource-intensive owing to model training. To avoid this, we may train several model snapshots from the same deep learning network in one run. COVID-Net was chosen due to its excellent efficiency in COVID-19 case identification using CXRs. detection and its openly accessible source code.

In order to assemble a comprehensive model, several deep learning models with varying error rates must be used. There is a correlation between the similarities between the created model snapshots and the results in terms of forecasts and mistakes in forecasts. To fix it, we use a fast training rate schedule to adjust model weights and snapshots within the same training run.

We use the cosine annealing learn rate schedule introduced by Loshchilov et al. to aggressively yet systematically modify the learning rate to create distinct model weights throughout training epochs.

$$\alpha(t) = \frac{a_0}{2}\left(cos\left(\frac{\pi mod(t-1[T/M])}{[T/M]}\right)+1\right)$$

Maximum learning rate is $\alpha 0$, current rate at epoch t is α (t), cycles are M, and epochs are T. Following M iterations of training, a brand-new snapshot of the model is generated.

At train a large number of model snapshots, we set COVID-parameters Net's to $\alpha 0$ = 0.002, T = 50, and M = 6. COVIDNet creates 6 different model snapshots. In Figure 4, we see that learning rates aggressively fluctuate throughout model training, leading to training losses. Examining how model weights change as a function of intermediate training losses reveals the impact of a significant change in learning rate.

2.2 Model Ensembling

The models will be assembled into EnDL-COVID-19 in the second step, utilising the snapshots taken in the previous stage as building blocks, as shown in Figure 2. Part II-A of this guide explored the various strategies for assembling a model ensemble. Averaging is a common strategy used in ensemble learning for snapshots. With an input sample, all it does is average out the probability of each class from the multiple models. Let's imagine they have M distinguishing characteristics (i.e., deep learning models). The class probability of the n^{th} class formed by the m^{th} classifier is provided by $P_{m,n}$ for the i^{th} input sample X_i. Then the probability of the class as a whole is

$$\frac{\sum_{m=1}^{NOM} P_{m,n}(X_i)}{NOM}$$

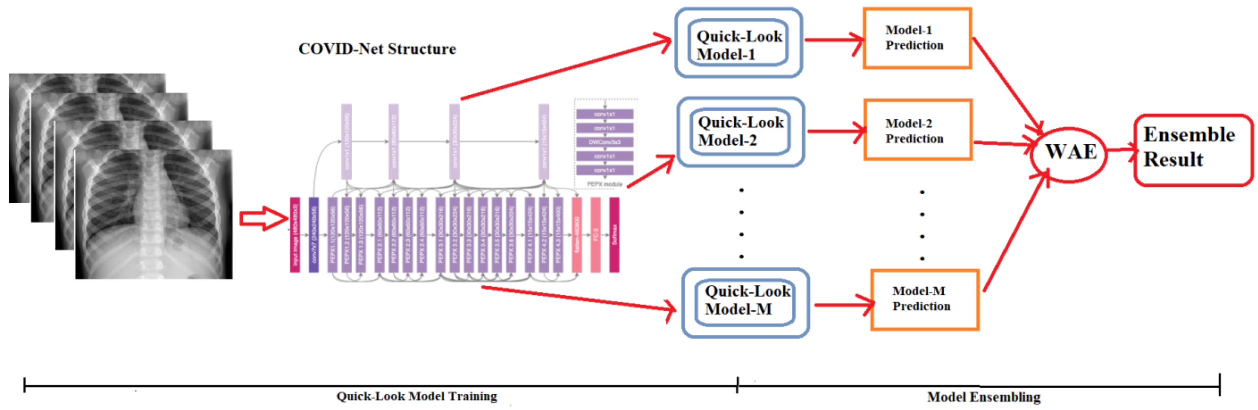

Fig 2. The general procedure for developing the ENDL-COVID-19 ensemble method. It is divided into two stages: Quick-Look Model Training and Assembling Models. As indicated in Algorithm-1, we suggest the WAE technique for model assembly

Where n is any value between [1, N] of the input sample X_i, where N refers number of different classes. Conventional averaging ensembling assumes all models have equal weights. Two essential observations in Conventional averaging ensembling are

- Deep learning models have varied testing accuracies for each class.
- It means we can't assemble models equitably.

Weighted Average Ensembling (WAE) is recommended for snapshot model ensembles based on the preceding evidence.

Algorithm 1

Process of the proposed Weighted Average Ensemble (WAE) strategy.

Input:

1. NOM: the number of classifications available or deep-learning frameworks available.
2. SIZE: the number of pictures in the CXR database CXR.
3. NOC: the number of categories or classifiers available in the input database.
4. $A_{i,j}$: evaluation of the ith model's accurateness for the jth category.
5. Xi: ith CXR image database sample.
6. $WAE_{i,j}$: the normalizing weight of the ith model for WAE over the jth category forecasting, where

$WAE_{i,j} = \frac{A_{i,j}}{\sum_{m=1}^{NOM} A_{m,j}}$.

Output:

7. $PC(X_i)$: EnDL-COVID-19 projected class indexing for the ith source sample
8. for i = 1 to SIZE do
9. for m = 1 to NOM do
10. Get the output (i.e., class probability vector $P_m(X_i) = \{p_{m,n}(X_i)\}$ for all N classes where $1 \leq n \leq N$) of the mth model with respect to the input sample X_i.
11. for n = 1 to N do
12. $P_n(X_i) = \sum_{m=1}^{NOM} P_{m,n}(X_i) .WAE_{i_i}$
13. $PC(X_i)$= arg $\max_{1\leq n\leq N}\{p_k(d_i)\}$

For the ith source sampling, find the class index with the highest category probability.

Algorithm 1 displays a model ensemble. Let $A_{i,j}$ be the test correctness of the ith classifier over CXR test data for the j_{th} class. $WAE_{i,j}$ is the normalized weight of the ith model for WAE over the jth class prediction (Line 6), where

$$WAE_{i,j} = \frac{A_{i,j}}{\sum_{m=1}^{NOM} A_{m,j}}$$

First, we retrieve the class probability $P_{m,n}(X_i)$ for each input sample X_i from the m^{th} model for $\forall m \in NOM$. (Line 8-10).

We may estimate EnDL-COVID's class probability $P_k(X_i)$ by adding all models' weighted class probabilities (lines 11-12) for $\forall n \in N$. Giving the classification index with the highest probability for each input sample gives us the projected class (Line 13).

3 Results and Discussion

On top of Tensor Flow, our own EnDL-COVID-19 framework has been developed. The evaluation of EnDL-COVID-19 was performed using COVID-Net and CXR. Images from the COVIDx collection are rendered using a machine with a GPU.

3.1 Individual Model Performance Measurement

We compare the performance indicators of every model in the following ways.

3.1.1 Evaluation of the Accuracy:

The COVID-Net network architecture was used to train six deep-learning frameworks (COVIDNet Model1, COVIDNet Model2, . . . COVID-Net Model6) using the dataset.

The correctness for all models is shown in Figure 3. It demonstrates that the assembly strategy of our proposed EnDL-COVID model may successfully exceed the other six deep learning methods by 0.3 percentage.

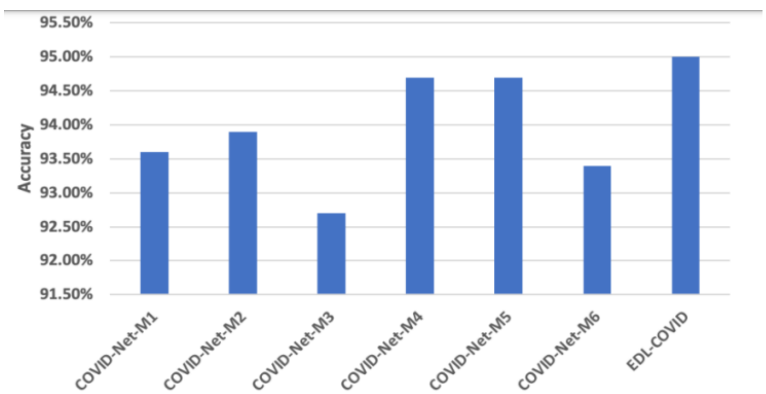

Fig 3. Correctness of Existing Models Comparison with proposed EnDL-COVID model

In practice, we can't assume a more accurate model would perform better for a multiclass situation.

Additionally, we must evaluate important essential category level factors, sensitivity and positive predictive value (PPV).

3.1.2 Sensitivity Analysis:

The delicate nature of a medical examination, a disease's severity may be judged by the percentage of patients who have received an accurate diagnosis. It is crucial to have maximum sensitivity when testing for COVID-19, for example, so that no infected individuals are missed. If we do so, these individuals will be unable to get timely treatment and will instead spread the disease to others. Table 1 presents examining each model's sensitivity for every category. The following observations have been made. It is rare to discover a model that is effective in all three categories. COVID-Net-sensitivity Model5 is greatest in the Normal class, but not in the other two classes. In contrast, EnDL-COVID-Net has the greatest sensitivities for the COVID-19 and Pneumonia classes, but not the Normal class.

Practically, EnDL-COVID should be considered since COVID-19 screening is so sensitive.

Table 1. The sensitivity of various frameworks to every category (for instance, Normal-Category, Pneumonia-Category, and COVID-19-Category) and the best outcomes are emphasized

	Sensitivity		
Model Name	**Normal Class**	**Pneumonia Class**	**COVID-19 Class**
COVIDNet-Model1	93.20%	94.80%	91%
COVIDNet-Model2	93.40%	95.10%	91%

Continued on next page

Table 1 continued

COVIDNet-Model3	94.10%	90.20%	95%
COVIDNet-Model4	94.80%	94.80%	94%
COVIDNet-Model5	95.80%	93.60%	92%
COVIDNet-Model6	95.30%	90.20%	96%
EnDL-COVID-Net	95.00%	94.80%	96%

3.1.3 PPV Assessment:

PPV estimates the percentage of accurate positive test findings. If this score is low, there are likely many false positives, requiring extra testing to replace each positive result with a more trustworthy one. Whether a model's PPV for COVID-19 screening is low, it's impossible to identify if a positive test result is an actual COVID-19 case. Therefore, positive test results need more accurate testing. Table 2 compares each model's PPV for each class. No model works perfectly for every class. COVID-Net-Model6 had the highest PPV for Pneumonia, whereas EnDL-COVID-19 had the highest for Normal and COVID-19.

Table 2. Best PPV findings for each category (for example, Normal Class, Pneumonia Class, and COVID-19 class) are bold

	PPV(%)		
Model Name	**Normal Class**	**Pneumonia Class**	**COVID-19 Class**
COVIDNet-Model1	96.30%	89.90%	94%
COVIDNet-Model2	96.20%	90.50%	92.90%
COVIDNet-Model3	95.20%	92.70%	75.40%
COVIDNet-Model4	96.30%	92.80%	93.10%
COVIDNet-Model5	95.70%	93.60%	92.90%
COVIDNet-Model6	94.80%	94.40%	78.70%
EnDL-COVID-Net	96.40%	93.10%	**94.**10%

In conclusion, while no model surpasses others on all measures for all class types, we feel that EnDL-COVID-19 is the best solution for COVID-19 case identification, because it beats other frameworks in accurateness, sensitivity, and PPV.

3.2 The Outcome of an Assessment EnDL-COVID

Here, we take a look at the EnDL-COVIDNet model and assess it from the aspects listed below.

3.2.1 WAE weights are estimated:

We introduced a model assembly technique called WAE in Section 2 to integrate several deep learning models.

First, deep learning model weights for each WAE class type are required. Figure 4 shows the expected category level weights for all deep-learning frameworks depending on the sensitivity findings for every category type shown in Table 1. Various deep-learning frameworks have various weights for different classes. It is the heaviest for COVID-Net-M6.

The COVID-19 class is severe, whereas the Pneumonia class is mild.

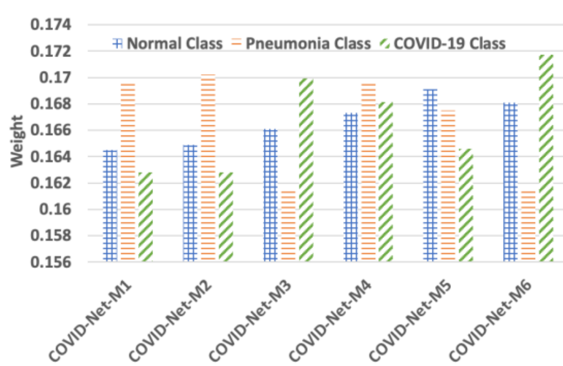

Fig 4. Six deep learning networks' predicted category weights required for WAE's model entanglement

Proposed system results:

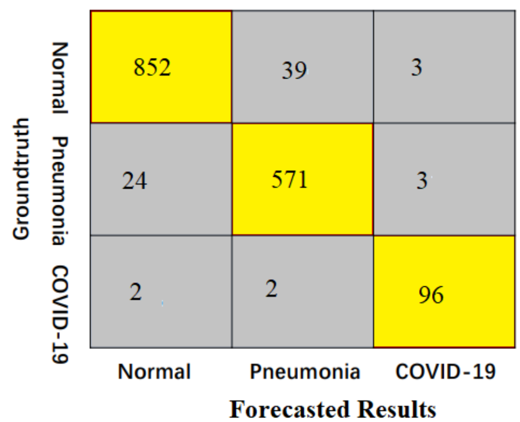

Fig 5. We see the resulting confusion matrix for EnDL-COVID on a database of 100 COVID-19, 598 Pneumonia, and 894 Normal pictures. A correct forecast is shown in yellow, whereas an incorrect forecast is shown in light ash

Figure 5 examines the dataset, which includes CXR pictures of 100 COVID-19, 598 Pneumonia, and 894 Normal patients. For COVID-19 validation, only four out of a possible hundred CXR photos are not filtered out, and six out of 1592 are incorrectly deemed COVID-19, suggesting a modest error ratio relative to the number of CXR pictures for EnDL-COVID-19.

After observation of all the performance metrics of proposed system with comparison to existing study, ENDL-COVID-19 gives good results for COVID-19, instances identification with a performance of 95%, higher than COVID-Net at 93.3%, are according to research observations.

4 Conclusion

In this article, we introduce EnDL-COVID-19, open-source network architecture for recognizing COVID-19 cases in CXR photos. ENDL-COVID-19 is built on the deep learning model that is used by COVID-Ensemble Net. EnDL-COVID-19 is designed to address the limitations of using a single deep learning model, such as over fitting, high variance, and generalisation mistakes, in order to improve performance on the COVID-19 case identification task. This will allow EnDL-COVID-19 to improve overall performance on the COVID-19 case identification task. First, we generate many quick look framework snapshots by training a COVIDNet on top of QU&UD CXR imagery. Next, we merge these frameworks using a WAE ensemble strategy in order to account for the varying sensitivities of several deep-learning frameworks on distinct classes. In a series of experiments using QU&UD test data consisting of 1592 CXR images, ENDL-COVID-19 outperformed every other deep learning model by a significant margin. It was able to achieve a sensitivity of 96% and a PPV of 94.1% in determining whether or not COVID-19 occurrences were present. We have great hopes that our AI-based screening technique will be able to assist radiologists during the continuing COVID-19 epidemic that is occurring all over the globe by helping to speed up patient screening while still keeping a high level of accuracy.

The COVID-Net network architecture and the QU&UD dataset are, on the other hand, highly significant to the inquiry that is now being conducted. Before EnDL-COVID-19 is suitable for general use, there are likely certain improvements that need be made to it. There are a few reasons why we are unable to infer with absolute confidence that the previously trained model snapshots from the COVIDNet design will continue to perform successfully with a maximum degree of accurateness for unviewed COVID-19 CXR photos. In order to make improvements to EnDL-COVID-19, it is necessary to retrain the model snapshots of COVIDNet every time there is a new release of CXR pictures. When trying to improve the functioning of EnDL-COVIDNet, we don't have to depend just on snapshot models; instead, we may use additional deep learning models that are readily accessible to the public.

We seek to create EnDL-COVID-19 for new COVID-19 applications such as risk assessment for COVID-19 case survival study and risk status examination of COVID-19 cases in order to further enhance patient hospitalization and care planning. This will allow us to further improve patient care.

References

1) Alom MZ, Rahman MMS, Nasrin MS, Taha TM, Asari VK. COVID_MTNet: COVID-19 Detection with Multi-Task Deep Learning Approaches. . Available from: https://doi.org/10.48550/arXiv.2004.03747.
2) Rajaraman S, Siegelman J, Alderson PO, Folio LS, Folio LR, Antani SK. Iteratively Pruned Deep Learning Ensembles for COVID-19 Detection in Chest X-Rays. *IEEE Access* . 2020;8:115041 –115050. Available from: https://doi.org/10.1109/ACCESS.2020.3003810.
3) Jacobi A, Chung M, Bernheim A, Eber C. Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. *Clinical Imaging*. 2020;64:35–42. Available from: https://dx.doi.org/10.1016/j.clinimag.2020.04.001.
4) Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Scientific Reports*. 2020;10(1):1–12. Available from: https://dx.doi.org/10.1038/s41598-020-76550-z.
5) Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*. 2020;43(2):635–640. Available from: https://dx.doi.org/10.1007/s13246-020-00865-4.
6) Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*. 2021;24(3):1207–1220. Available from: https://dx.doi.org/10.1007/s10044-021-00984-y.
7) Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*. 2020;121:1–11. Available from: https://dx.doi.org/10.1016/j.compbiomed.2020.103792.
8) Sethy PK, Behera SK. Detection of Coronavirus Disease (COVID-19) Based on Deep Features. *Preprints*. 2020;p. 1–9. Available from: https://www.preprints.org/manuscript/202003.0300/v1.
9) Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, et al. Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access*. 2020;8:132665–132676. Available from: https://dx.doi.org/10.1109/access.2020.3010287.
10) Rahman T, Khandakar A, Qiblawey Y, Tahir A, Kiranyaz S, Kashem SBA, et al. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Computers in Biology and Medicine*. 2021;132:1–16. Available from: https://doi.org/10.1016/j.compbiomed.2021.104319.