Predicting COVID-19 From Chest X-Ray Images Using Deep Transfer Learning

Kavu Maithri Rao

Saarland University, 66123, Germany

Abstract. The COVID-19 pandemic is causing a major outbreak in more than 150 countries around the world, having a severe impact on the health and life of many people globally. One of the crucial ways to fight this pandemic is to detect this disease in its early stages and provide necessary treatment to the patient. Due to overwhelming burden on the healthcare system and lack of availability of RT-PCR kits, there is a need to figure out alternatives by which we can detect this disease in its earlier stages without burdening the current medical system. One of the most viable steps towards achieving this goal is through radiological examination, Chest X-Ray being the most easily available and least expensive option. In this project, we aim to apply deep learning models to detect patients infected with Covid-19, or patients with pneumonia or normal individuals. We first prepare a dataset by collecting images from multiple sources and augment these images to improve the dataset. Multiple state-of-the-art pre-trained CNN models— DenseNet-161, ResNeXt and Inceptionv3, have been adopted in the proposed work. They have been trained individually to make independent predictions. Then the models are combined using ensemble approach to take a mean of their prediction probabilities, to predict a class value. We could achieve a testing accuracy of 72% using the ensemble approach by training 6000 images for 30 epochs. We achieved 100% classification of the classes and confusion was only with pneumonia and normal patient case. While the achieved performance seems encouraging, further analysis is required on a larger set of COVID-19 images, to have a more reliable estimation of accuracy rates.

Keywords: COVID-19 · deep learning · convolutional neural network · Ensemble-CNNs · X-ray scans.

1 Introduction

The outbreak of novel coronavirus has cause catastrophic effects all around the world. The disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is also known as COVID-19. According to the World Health Organization (WHO), more than fifteen million peoples are infected from this virus across 215 countries (Coronavirus Disease (COVID-19) 2020). The early detection of this disease is very important to break the chain of transmission and also

offer early treatment to the patient. Reverse Transcription Polymerase Chain Reaction (RT-PCR) is the definitive test for the recognition of COVID-19 disease. However, RT-PCR test is a time-consuming, laborious, and complicated manual process. In addition, test kits are only available in limited numbers worldwide. On the other hand, the rate of false negatives varies depending on how long the infection has been present. In the false-negative rate was 20% when testing was performed five days after symptoms began, but much higher (up to 100%) earlier in the infection.

Classification with the help of radiographic images, such as chest X-ray (CXR), can be accurate but at the same time much faster and less expensive than the PCR test. Furthermore, chest X-rays are economical than other radiological tests like CT scans and available in almost every clinic. The only perceived challenge in CXR-based detection of COVID-19 patients is that trained doctors may not be available all the time, especially in remote areas. Also, the radiological manifestations related to COVID-19 are new and unfamiliar with many experts not having past experience with COVID-19 positive patient CXRs. So, we have proposed a simple and inexpensive deep learning-based technique to classify COVID-19 +ve and -ve cases using CXR images. Many AI systems based on deep learning have been proposed and their performance has shown promising results in the diagnosis of COVID-19 infection from chest X-ray images.

2 Related Works

So far, due to the lack of availability of public images of COVID-19 patients, detailed studies for automatic detection of COVID-19 from X-ray (or Chest CT) images is not much prominent in public domain. In multiple works, deep learning-based techniques have been developed to identify pneumonia [7, 8], different classes of thoracic diseases [5, 6, 9], etc. from medical images. Some of these works have given promising results with relatively simple architecture [5]. Li et al. [10] implement a fully automatic framework to detect coronavirus affected lungs from chest CT scan images and distinguished it from other lung diseases. However, studies have concluded that CXR images are better than any other means in the detection of COVID-19 because of their promising results along with the availability of CXR machines in most hospitals and their low maintenance cost. Most of the works related to COVID-19 detection from CXR images have utilized individual deep learning models e.g., DenseNet, ResNet, Xception, etc. [11, 2]. Various works done on Ensemble learning with deep neural networks show that ensemble techniques are superior compared to individual model and also prevent overfitting. So, in this project we chose to use transfer learning as well as ensemble techniques in classification of Covid-19, Pneumonia and normal conditions in an individual.

3 Methodology

3.1 Evaluation cases

In our work, we aim to classify the chest X-Rays into three classes, which are:

- Healthy.
- COVID-19.
- Other pneumonia diseases.

To train our models, we collected 404 X-ray scans for each class. In this scenario, we evaluated the performance of three most popular CNN architectures (Densenet-151, Inception-v3, and ResneXt-50) and our proposed Ensemble-CNNs approach. In the training phase, we divided the 404 X-ray scans of each class into training–validation splits (80%–20%). To train the deep learning models, we augmented the original images and improved the training, validation set image count to nearly 6000 images. For the testing part, we collected 207 images that were not included in creation of training and validation splits from entirely different source like the Hospital of Tolga, Algeria.

3.2 Three-class Covid-19 datasets

We considered three classes, which are COVID-19, Pneumonia, and Normal (or Healthy). We used the following resources to create our databases:

1. Chest X-Ray Images (Pneumonia) [13] from Kaggle that contains a lot of images for the classes Pneumonia and Normal. For Pneumonia images, there are two classes, which are Bacterial and Viral. License: CC BY 4.0

2. RSNA Pneumonia Detection Challenge [14] from Kaggle. From this source, we took only Normal and Pneumonia images. In the Pneumonia class there is no distinction between types. License: Open Source

3. CheXpert [15] is a large chest X-ray database from which we took Normal images, and it is the only database that includes Lung Opacity images. License: Public database

4. In addition to the use of above open-source databases, we collected 207 unpublished X-ray samples for the COVID-19 class from the Hospital of Tolga, Algeria. These COVID-19 scans were used as testing data. Most of the testing data classes were taken from new sources that were not used to create the training data.

Since Deep Learning methods require huge amounts of labelled data for training, which is actually not available for the COVID-19 class, we used data augmentation techniques to cope with this issue. By applying the following data augmentation techniques for the training split, we obtained 12 augmented images for each image:

- Color jitter.
- Padding.
- Random Horizontal flip.
- Random Rotation.
- Random Crop making sure that the smaller size remains at least 224px.

Table 1. COVID-19 database sources and statistics for three classes.

Dataset	Train Set(original + Augmented)	Validation Set	Test Set
COVID-19	404 + 1356	100	207
Normal	404 + 1624	100	207
Pneumonia	404 + 1689	100	207
Total	1212 + 4669	300	621

3.3 CNN Architectures

ResNeXt: The structure of ResNeXt architecture is inherited from three CNN architectures i.e., VGG, Inception and ResNet. The input ResNeXt block is split into a few lower-dimensional embeddings (by 1×1 convolutions) with 32 paths each for four channels, then all paths are transformed by the same topology filters of size 3×3. The paths are merged by summation at the end. In our work, ResneXt-50 pre-trained model was trained on ImageNet challenge database. The number of paths inside the ResNeXt block is defined as cardinality. Instead of having high depth and width, we have high cardinality which helps in decreasing validation error.

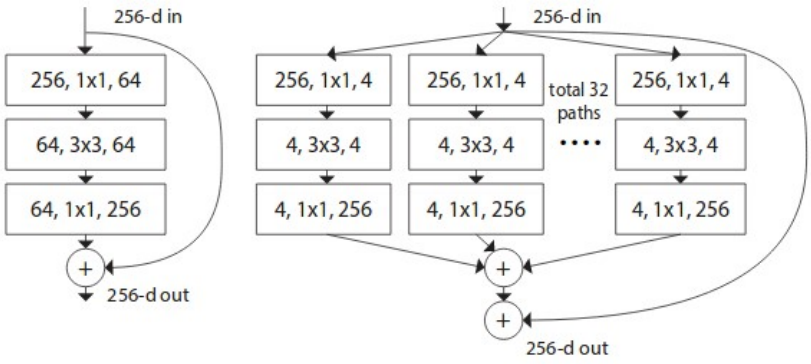

Figure 1. **Left**: A block of ResNet [14]. **Right**: A block of ResNeXt with cardinality = 32, with roughly the same complexity. A layer is shown as (# in channels, filter size, # out channels).

Fig. 1. ResNext Architecture

Inception v3: Very deep networks are prone to overfitting. It also hard to pass gradient updates through the entire network. Naively stacking large convolution operations is computationally expensive. We could have filters with

multiple sizes operate on the same level. The network essentially would get a bit "wider" rather than "deeper". It also offers the following advantages:

- Factorized convolutions;
- Smaller convolutions;
- Asymmetric convolutions;
- Auxiliary classifier;
- Grid size reduction.

DenseNet 161: In this model, each layer obtains additional inputs from previous layers and passes its own feature-maps to all subsequent layers. Hence each layer receives collective knowledge from all previous layers thus making the network thinner and compact, i.e., number of channels can be fewer which in turn achieves higher computational efficiency and memory efficiency.

4 Proposed Approach

In real life, we always tend to prefer the diagnosis or treatment suggested by multiple experts. The combined opinion is much more reliable and helps us faster to reach a particular conclusion. Similarly, in our project, we have used the above philosophy to diagnose Covid patient, pneumonia patient or a normal person from each other. We have used multiple benchmark CNN models to serve our purpose. They have been trained individually to make independent predictions. This newly proposed ensembling method is expected to make the prediction more robust. Our proposed work comprises of three pre-trained CNN models—ResNeXt-50, Inception-v3, and DenseNet-161. Each of these models will make their predictions for an image by assigning prediction probabilities for class A i.e Covid-19 case Or class B i.e Pneumonia case or class C i.e Normal case. For each model, we calculate the average of the prediction probabilities. In more detail, the probabilities of the three models corresponding to all classes give the mean probability for all classes, then the argmax of the mean probabilities will assign the Ensemble-CNNs predicted class.

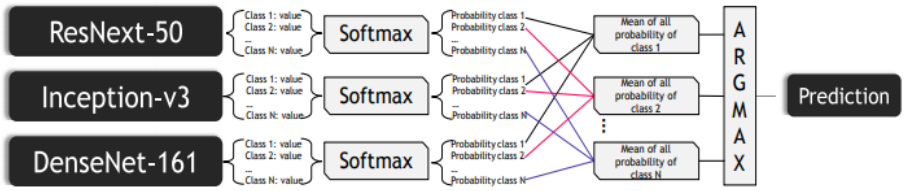

Fig. 2. Ensembling method

5 Experiments and Results

All the networks were trained for 30 epochs with the Adam optimizer and batch size of 30. The initial learning rate was set to 1e-6. The input image size of the Inception-v3 network was 299 × 299 pixels, meanwhile the DenseNet-161 and ReNeXt-50 input sizes were 224 × 224 pixels. Data augmentation was performed by normalizing, resizing, and cropping the input images in order to achieve the correct input size for each network. For the normalization, the following values of mean and standard deviation were used for each channel of the image:

- mean: [0.485, 0.456, 0.406],
- std: [0.229, 0.224, 0.225].

After fully connected layer we added a Dropout layer to avoid overfitting for both DenseNet-161 and ResNeXt-50 with a probability of 0.3. Meanwhile, Inception-v3 already has a default dropout layer with a probability of 0.5

Table 2. Validation Data Results

Model	Accuracy		Loss		F1 Score	
	Train	Val	Train	Val	Train	Val
DenseNet_161	0.99354	0.9726	0.04031	0.09878	0.99354	0.9726
Inception_v3	0.97858	0.93493	0.08578	0.2017	0.97857	0.93493
ResNeXt50	0.93393	0.82333	0.39557	0.60051	0.93415	0.82333

Table 3. Testing Data Results

Model	Accuracy	F1 Score
	Test	Test
DenseNet_161	0.752	0.752
Inception_v3	0.685	0.685
ResNeXt50	0.716	0.716
Ensemble Approach	0.717	0.717 roc_auc_score for covid class: 1.5

From the results of the validation data, the best method is our proposed DenseNet-161 for all of the used evaluation metrics (Training accuracy, Loss, F1 Score). From Table 2, which contains the results of the testing data, DenseNet-161 achieved the best performance for the Accuracy, Loss, and F1-score evaluation metrics, where it is slightly better than our proposed Ensemble-CNNs approach. From these results, we notice that our proposed Ensemble-CNNs approach does not achieve the best result for all the evaluation metrics but still gives a better trade-off between different evaluation metrics' results. In addition, we notice that the performance of the testing data is not good as that of the

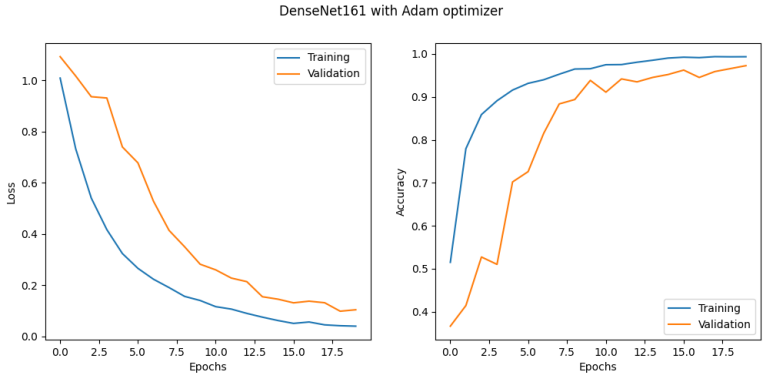

Fig. 3. DenseNet_161 with Adam Optimizer

Fig. 4. Inception_v3 with Adam Optimizer

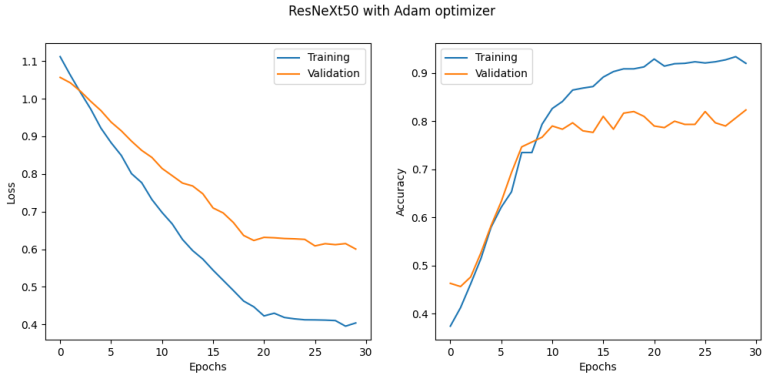

Fig. 5. Resnext50 with Adam Optimizer

validation data. This is because the testing data sources are different from the training and validation ones.

Fig5,Fig6,Fig7 consists of the confusion matrices of the testing data. The main observation is that all models achieved 100% for the classification of COVID-19 samples. The real confusion for all models was in distinguishing between the Normal and Pneumonia classes.

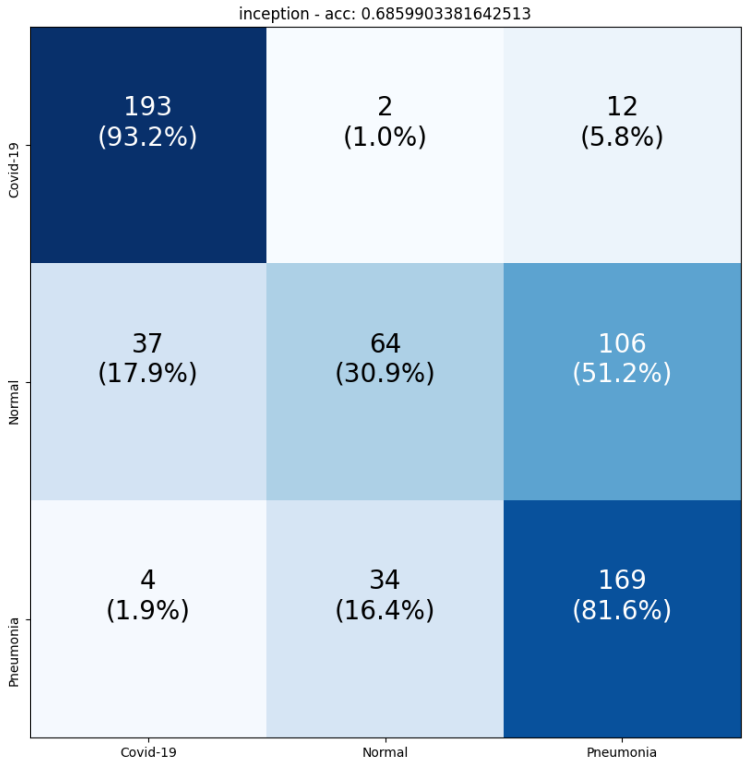

Fig. 6. Confusion matrix of Inception

6 Conclusion and Future Work

Since there is no unified dataset, classes or evaluation criteria or norms in this research area, it is hard to compare different methods and different works in the

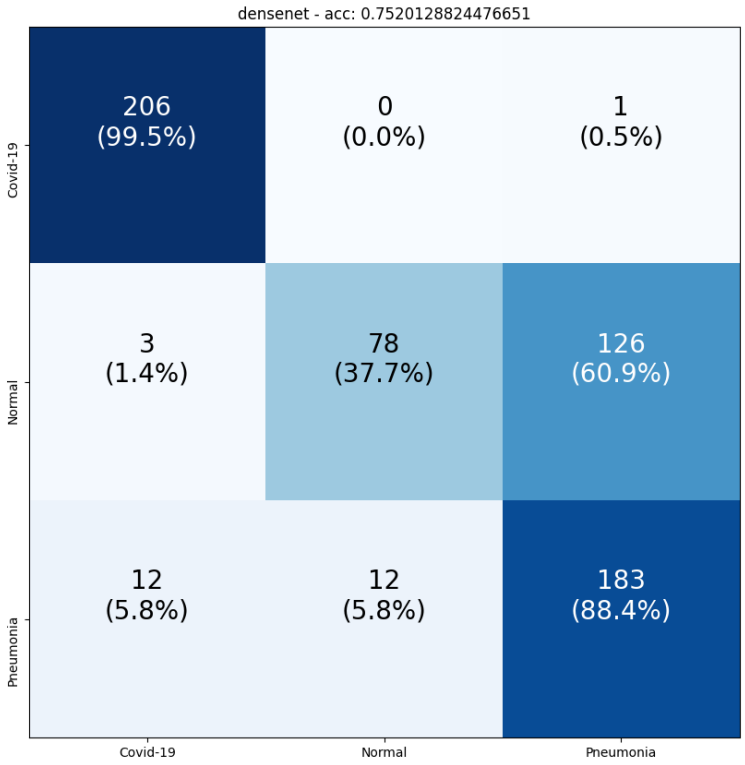

Fig. 7. Confusion matrix of DenseNet

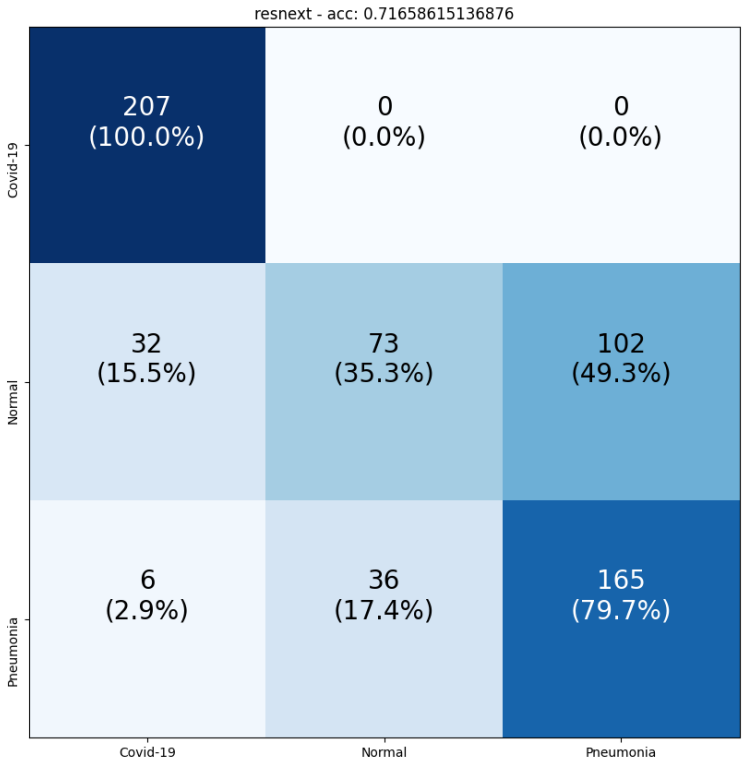

Fig. 8. Confusion matrix of ResNext

same baseline. In this project, we tried some state-of-the-art methods and ensemble approach to detect Covid-19. Since we evaluated three CNN architectures and our proposed Ensemble-CNNs approach on our proposed new databases and scenarios, it is unfair to compare the complexity of our approach with the state-of-the-art methods. However, we were successful to the extent that we were able to see some promising results in a short time. Also, we see a good scope for this area of research if we are able to tune the hyper-parameters, get more data especially with respect to availability of Covid-19 images to create a more balanced dataset. It should be mentioned that the number of X-ray scans used for the training CNN architecture is very limited (404 X-ray scans for each class). One possible solution to improve the performance is to use more X-ray scans for each class. Also, we observe that the required time to test any image is very trivial, which is a major advantage for recognition of COVID-19 infection compared with currently used tests, such as RT-PCR.

References

1. Turkoglu, M.COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble Appl. Intell. (2020).
2. Vantaggiato, E.; Paladini, E.; Bougourzi, F.; Distante, C.; Hadid, A.; Taleb-Ahmed, A. COVID-19 Recognition Using Ensemble-CNNs in Two New Chest X-ray Databases. Sensors 2021, 21, 1742. https:// doi.org/10.3390/s21051742
3. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM (2017) Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classifcation and localization of common thorax diseases. In: Proceedings of the IEEE Conference on com puter vision and pattern recognition, pp 3462–3471
4. Chouhan V, Singh S, Khamparia A, Gupta D, Tiwari P, Moreira C, Damaeviuus R, Albuquerque V (2020) A novel transfer learning based approach for pneumonia detection in chest X-ray images. Appl Sci 10:559
5. Jaiswal A, Tiwari P, Kumar S, Gupta D, Khanna A, Rodrigues JJ (2019) Identifying pneumonia in chest X-rays: a deep learning approach. Measurement 145:511–518
6. Ranjan E, Paul S, Kapoor S, Kar A, Sethuraman R, Sheet D (2018) Jointly learning convolutional representations to compress radiological images and classify thoracic diseases in the compressed domain. Proceedings of the 11th Indian Conference on computer vision, graphics and image processing
7. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, Cao K, Liu D, Wang G, Xu Q, Fang X, Zhang S, Xia J, Xia J (2020) Artifcial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology
8. Jain R, Gupta M, Taneja S, Hemanth DJ (2020) Deep learning based detection and analysis of COVID-19 on chest X-ray images. Appl Intell, pp 1–11. Accessed 29 Mar 2020
9. https://github.com/CUIEMCOVIDProject/COVID-19-Detection-Using-Ensemble-Learning
10. https://github.com/shervinmin/DeepCovid/tree/master/data Oct 2017
11. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
12. https://www.kaggle.com/c/rsna-pneumonia-detection-challenge
13. https://stanfordmlgroup.github.io/competitions/chexpert/