COVID-Net: Covid-19 Detection on Lung CT Scans with Convolutional Neural Networks

Aditya Choudhary **Akhil Sharma** **Shrugal Tayal** **Vaibhav Rajpal**

Abstract

We developed an algorithm to detect COVID-19 from Computerized Tomography (CT) scans .Our algorithm, COVIDNet, is a 16-layer dense convolutional neural network trained on COVID-19 Lung CT Scans, a large publicly available chest CT scan dataset currently available on Kaggle, contains over 349 CT scans that are positive for COVID-19 and 397 CT scans that are negative for COVID-19. The images collected from COVID-19-related papers from medRxiv, bioRxiv, NEJM, JAMA, Lancet, etc., On this dataset, we analyze the performance of COVIDNet. We find the COVIDNet average performance on the F1 metric.

1. Introduction

Globally, as of 4:23pm CET, 2 December 2022, there have been 640,395,651 confirmed cases of COVID-19, including 6,618,579 deaths, reported to WHO. COVID-19 has tremendously impacted patients and medical systems globally. CT scans are a relatively sensitive and fast tool and play a crucial role as a complementary screening method for distinguishing COVID-19 . However, detecting COVID-19 in lung CT scans is a challenging task that relies on the availability of expert radiologists and their interpretation of CT for COVID-19 screening. In this work, we present a model that aids in COVID-19 diagnosis by automatically detecting the impact of COVID-19 on lungs from CT scans..

2. COVIDNet

2.1. Problem Formulation and Model Selection

We have to determine whether a patient has COVID-19 or not from their CT scan image. Since, we have to determine if the patient has COVID-19 or not it becomes a Binary classification problem. As it is a binary classification problem to detect if a person is suffering from COVID-19 or not based upon the images of his lungs, thus we would be using Dense CNN for training our dataset.

2.2. Model Architecture and Training

COVIDNet or VGG16 is a 16-layer Dense Convolutional Network trained on the *COVID-19 Lung CT Scans* dataset. Dense Convolutional Networks improve flow of information and gradients through the network, making the optimization of very deep networks tractable. We replace the final fully connected layer with one that has a single output, after which we apply a sigmoid nonlinearity. The weights of the network are initialized with weights from a model pretrained on ImageNet. The network is trained end-to-end using Adam with standard parameters. We train the model using minibatches of size 32. We use an initial learning rate of 0.001 that is decayed by a factor of 10 each time the validation loss plateaus after an epoch, and pick the model with the lowest validation loss. We ran the model for 50 epochs.

3. Data

3.1. Dataset Acquisition

We have used an openly available database from Kaggle: COVID-19 Lung CT Scans, in which the images are collected from COVID19-related papers from medRxiv, bioRxiv, NEJM, JAMA, Lancet, etc. CT scans containing COVID-19 abnormalities are selected by reading the paper's figure captions. All copyrights of the data belong to the authors and publishers of these papers. The CT scan images of the lungs of some individuals are classified into two categories:

1. Individuals Infected with COVID.

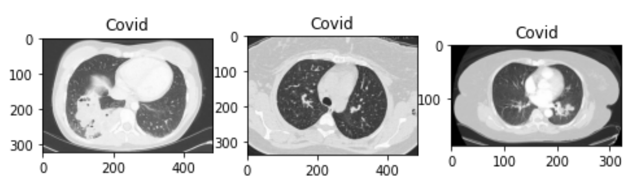
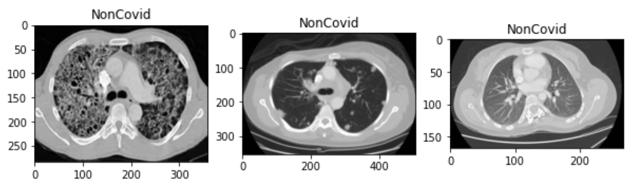


Figure – example of non-covid sample

2. Individuals Not-Infected with COVID.

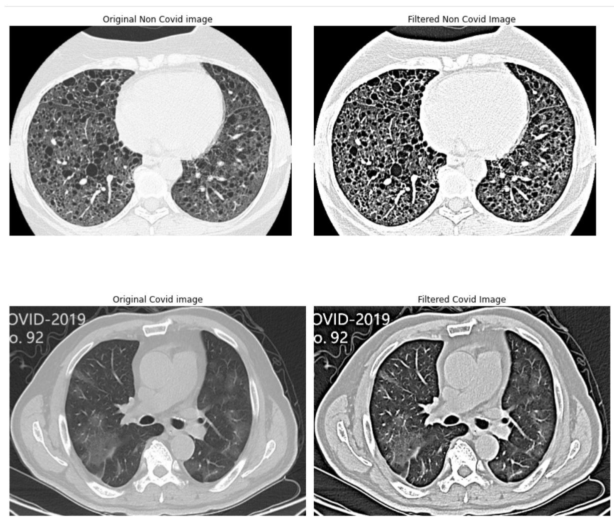
COVID): 318, Testing images (COVID): 69 , Testing images (Non COVID): 79



(b) Figure – example of non-covid sample

3.2. Data Masking and Visualization

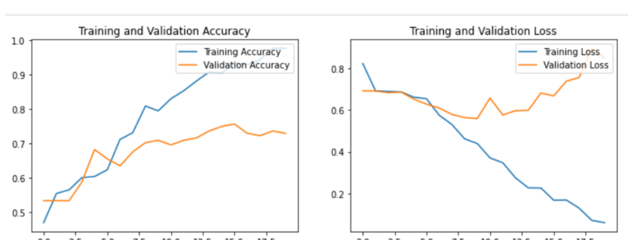
To interpret the network predictions, we produce masks to visualize the areas of the image most indicative of the disease using class activation mappings. To generate the CAMs, we feed an image into the fully trained network and extract the feature maps that are output by the final convolutional layer.



(a) Figure- Used the **Unsharp Mask** filter from skimage we try to create different masks to highlight regions of infection from the Covid Positive CT Scans.

4. Methodology

We did our training on basic CNN model, in which we got a good accuracy of 85% (almost) but F1-score was not quite satisfactory.



(a) Figure- Basic CNN accuracy and loss curves for training vs validation data, what we did is we take; Training images (COVID): 280 Training images (Non COVID): 318, Testing images (COVID): 69 , Testing images (Non COVID): 79

Because of the low F1-score, we trained our model on a special model called VGG16. In this model, we are training our dataset with 16 layers of CNN.

After training on the dataset on VGG16, we got a good accuracy of 97% (almost) and our F1-score also became satisfactory. The F1-score we got is 77% (almost).

5. Results

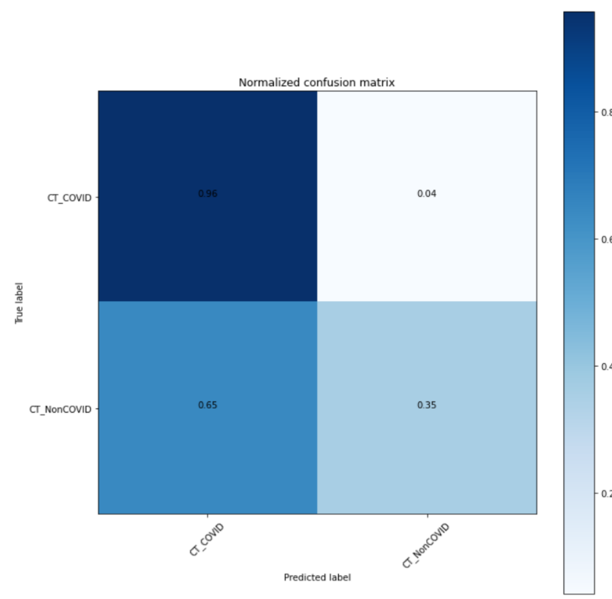
After Applying Basic CNN Model; we got

Accuracy: 0.857463
Precision: 0.492308
Recall: 0.542373
F1 score: 0.516129

As we observe that the f1 Score is quite low, so there can be many false positives which is not good in the medical field. That's why did a big change in our CNN model. And now trained our dataset with 16 layers of CNN.

After applying the VGG16 model; we got

Accuracy: 0.940299
Precision: 0.894737
Recall: 0.684167
F1 score: 0.775410



(a) Generated a normalized confusion matrix for testing. To check accuracy and f1-score of the covid and non covid datasets.

These are the satisfactory results that we got. Our Accuracy has increased drastically and our F1 score is quite reliable for Covid-19 detection.

6. Literary Resources Review

To better understand the project, we selected two research papers. The research papers are as follows:

Paper 1 (COVID-19 detection from lung CT-scan images using a transfer learning approach)

Main Argument: In this research paper, the effect of COVID-19 on human lungs has been discussed. It discusses how technology diagnoses COVID-19 using computed tomography (CT) lung scans [4].

Methodology: Machine learning techniques algorithms(ML) Convolution Neural Networks (CNN) have been applied in the healthcare sector for the analysis and prognosis of COVID-19. The model discussed concluded transfer learning is more beneficial for smaller datasets. The result of the experiments showed that transfer learning gives better results than CNN and other machine learning models.

Paper 2 (COVID-19 Detection on Chest X-Ray and CT Scan: A Review of the Top-100 Most Cited Papers.)

Main Argument: COVID-19 has a high risk to the respiratory system. CXR and CT scans are the most commonly used image testing. During one of the analyses, a single image patch had no match for CNN, but when analyzed with a complete CT piece, the accuracy was less than that of CNN [5].

Methodology: An analysis performed on COVID-19 classification using CNN acted as a feature extraction. This experiment was carried out on a publicly available dataset and gave an accuracy of about 99.4%

7. Conclusion

From our project above, we have learned how Convolution Neural Network (CNN) is used as an image recognition model and how it helped us to classify individuals as affected by COVID-19 or not by their lung CT scan. We observed that when we used the basic CNN model, our accuracy was a bit less but it got improved by the use of the VGG16 model which in turn also gave a better f1 score. The accuracy of the model can be further improved with the help of transfer learning.

F1 score is a popular performance measure for classification and often preferred over, for example, accuracy when data is unbalanced, such as when the quantity of examples belonging to one class significantly outnumbers those found in the other class, so it is suitable in our experiment.

We opted for f1 metric as a main parameter to evaluate the performance of our model because: in case of accuracy the True Positives and True negatives are more important while in case of F1-score the False Negatives and False Positives are crucial i.e. our model should not determine a patient who is covid positive as a covid negative patient.

8. Contributions

1. Aditya Choudhary - VGG16 modeling, analysis and visualization
2. Akhil Sharma - Basic CNN modeling, analysis and visualization.
3. Shrugal Tayal - Literature review, Evaluation of new models to get better accuracy(CNN).
4. Vaibhav Rajpal - Dataset Preprocessing , Description and visualization.

9. Github repository

Code for Basic CNN, VGG16 and Data visualization can be found inside the provided github repository. The training and testing datasets are also available in the below mentioned github repository.

https://github.com/ShrugalTayal/Covid-19-Diagnosis-from-Lung-CT-Scans

References

[1]"CT Scan," *Mayo Clinic*, 06-Jan-2022. [Online]. Available: https://www.mayoclinic.org/tests-procedures/ct-scan/about/pac-20393675 [Accessed: 04-Dec-2022]

[2]W. Zhao, W. Jiang, and X. Qiu, "Deep learning for COVID-19 detection based on CT Images," *Nature News*, 12-Jul-2021. [Online]. Available: https://www.nature.com/articles/s41598-021-93832-2#:~:text=Owing%20to%20its%20high%20sensitivity,prediction%20compared%20to%20RT%2DPCR [Accessed: 05-Dec-2022]

[3]P. M. Sharif, M. Nematizadeh, M. Saghazadeh, A. Saghazadeh, and N. Rezaei, "Computed tomography scan in COVID-19: A systematic review and meta-analysis," *Polish journal of radiology*, 05-Jan-2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8814899/[Accessed: 04-Dec-2022]

[4]A. Halder and B. Datta, "COVID-19 detection from lung CT-scan images using transfer learning approach - IOPscience," *COVID-19 detection from lung CT-scan images using transfer learning approach - IOPscience*, Jul. 19, 2021. [Online]. Available: https://iopscience.iop.org/article/10.1088/2632-2153/abf22c[Accessed: Dec. 04, 2022]

[5]Y. M. Costa, S. A. Silva, L. O. Teixeira, R. M. Pereira, D. Bertolini, A. S. Britto, L. S. Oliveira, and G. D. Cavalcanti, "Covid-19 detection on chest X-ray and CT

scan: A review of the top-100 most cited papers," *Sensors*, vol. 22, no. 19, p. 7303, 2022. [Online] Available: https://www.mdpi.com/1424-8220/22/19/7303/pdf[Accessed: Dec 04,2022]