
An Application of Convolutional Neural Networks for COVID-19 CT Scan Classification

Cullen Pu
University of Toronto
cullen.pu@mail.utoronto.ca

Alexander Shih
University of Toronto
alexander.shih@mail.utoronto.ca

Sijia Cai
University of Toronto
s.cai@mail.utoronto.ca

Abstract

The demand for efficient COVID-19 detection methods is evident as the pandemic continues to devastate our communities. In this paper, we compare the performance of several convolutional models, including ResNet [1], DenseNet [2], and InceptionNet [9], with transfer learning to classify between COVID-19 cases and normal cases using a limited dataset of chest CT scans. Our results show that DenseNet-121 works the best for this classification problem, achieving over 94% test accuracy.

1 Introduction

With the advent of the COVID-19 pandemic, the need for fast and reliable diagnostic tools to detect afflicted individuals becomes increasingly urgent not only to maintain the health of the general public, but also to provide early assistance to higher risk individuals. Since COVID-19 is a disease that affects the respiratory system, we propose using chest scans as a possible alternative method to diagnose COVID-19, thereby minimizing healthcare practitioners' direct contact with symptomatic patients and their bodily fluids compared to current methods such as PCR testing.

We propose repurposing, fine-tuning, and comparing the performances of COVID-19 detection on chest CT scans from a variety of existing Convolutional Neural Networks. Specifically, we will use the ResNet-18, ResNet-50 [1], DenseNet-121, DenseNet-161 [2], and Inception [9] models and apply transfer learning. The scans will be sourced from a dataset of 2482 images with 1230 normal chest scans and 1252 COVID-19 chest scans compiled by Eduardo Soares and Plamen Angelov of Lancaster University. It was collected from patients from hospitals in Sao Paulo, Brazil, and made publicly available for research. [7].

2 Related Works

Previous research has been conducted in the area of detecting pneumonia from digital x-ray images. Rahman et. al achieved extremely promising results using four pre-trained CNNs, AlexNet [5], ResNet [1], DenseNet [2], and SqueezeNet [3] in classifying between normal, bacterial pneumonial, and viral pneumonial chest x-ray images, reaching a classification accuracy of 98%, 95%, and 93.3%, respectively [6].

In November of 2020, Wang et. al introduced COVID-Net, one of the first open source CNN architectures specifically designed to detect COVID-19 cases from COVIDx, an aggregation of open source chest x-ray images [10]. COVID-Net was able to achieve 93.3% test accuracy, highlighting the effectiveness of using deep learning to rapidly generate a solution in response to urgent situations.

As mentioned before, the original publishers of the dataset also proposed a deep learning model using an eXplainable Deep Learning approach (xDNN) [11]. They achieved an F1 score of 97.31% [8]. We use their model as on the limited-size dataset as a benchmark to measure the efficacy of our best performing repurposed models.

3 Method

We use CNN models because they are currently the state of the art models for image recognition. AlexNet is considered an outdated model, which is why we chose not to use AlexNet for these tests. ResNet in particular is a powerful deep residual network capable of having many more layers than a traditional CNN, made possible by its use of residual blocks which solve the vanishing gradient problem. There are different ResNet sizes, ranging from the smallest, ResNet-18, being 18 layers deep, up to the 152 layer ResNet-152 [1]. Because ResNet was originally trained to classify between 1000 classes and our COVID-19 classification problem involves binary classification, we decide to use ResNet-18, the smallest variant. In contrast, we also use ResNet-50, a popular middle-ground variant of ResNet, to see if a greater depth achieves better performance in this binary classification problem.

DenseNet was proposed after ResNet and can be thought of as an extension of ResNet. DenseNet uses multi-layer feature concatenation, in which each layer obtains inputs from all previous layers and also connects to all later layers. It requires fewer parameters, provides better accuracy, and allows for more depth than a traditional CNN [2]. Like with ResNet, we train on two different DenseNet sizes, DenseNet-121 and DenseNet-161, and compare which performs better.

InceptionNet is another widely used image recognition model which pushed breakthroughs in image recognition and detection. With a focus on reducing the computational cost in CNNs, it utilizes factorized convolutions and aggressive regularization [9]. It uses less layers and is therefore less prone to overfitting. We will also be using Inception v3 for our COVID-19 classification problem.

Algorithm 1: The following algorithm shows our training loop and how we calculate false positives and negatives.

```

foreach epoch do
  foreach phase in ["train", "validate"] do
    total_loss = total_correct = 0;
    false_positives = false_negatives = total_covid = total_healthy = 0;
    foreach batch in dataloader.phase do
      Run forward pass and get model predictions;
      Calculate Cross Entropy Loss;
      Backpropagate;
      Update total_loss, total_correct;
      Update false_positives, false_negatives, total_healthy, total_covid;
    end
    epoch_loss = total_loss / len(dataloader.phase);
    accuracy = total_correct / len(dataloader.phase);
    false_positive_rate = false_positives / (false_positives + total_healthy);
    false_negative_rate = false_negative / (false_negative + total_covid);
  end
end

```

4 Experiments

We fine tuned ResNet-18 by adding sequential linear layers. For ResNet-18, we replaced the original fully connected layer with three fully connected layers of size 512, 128, and 32 respectively. Since it was the shallowest of the models we tested it is the least prone to overfitting so we added more fully connected layers. When repurposing ResNet-50, we added only two fully connected layers: one of size 2048 and one of size 1024 since it is a larger model than ResNet-18.

We repurposed both DenseNet-121 and DenseNet-161 by replacing the last layer with only a single fully connected layer with two outputs. Loss plots were closely monitored and early stopping was used to prevent overfitting the test dataset. We found that a lower learning rate worked better on DenseNet-161 because it lead to much less fluctuation in the validation loss, usually a telltale sign of overfitting.

Inception v3 was repurposed in the same way as DenseNet, with only a single fully connected layer with two outputs replacing the last layer. We do so because one of the main benefits of Inception v3 is it's efficiency; to preserve this, we add just one fully connected layer to avoid adding many new connections to the model.

All 5 models were trained using the Adam optimizer with weight decay [4].

5 Results

Out of all the models, the one that was able to best classify between COVID-19 and non COVID-19 cases was DenseNet-121, which achieved a test accuracy of 94.38%. Nonetheless, in practicality and especially in the medical field, classification accuracy is not the sole important factor. As a method to quickly determine if a patient may have COVID-19 and require further testing and treatment, this type of model must also achieve a low False Negative Rate (FNR).

Table 1: Model Performance

Model	Val Accuracy	Test Loss	Test Accuracy	False Positive	False Negative
ResNet-18	0.8911	0.2358	0.8835	0.1189	0.0889
ResNet-50	0.8710	0.2132	0.9116	0.1490	0.0077
DenseNet-121	0.8891	0.1944	0.9438	0.0725	0.032
DenseNet-161	0.8810	0.3077	0.8916	0.1194	0.0775
Inception v3	0.8871	0.3145	0.8635	0.1357	0.1049

The model that achieved the lowest FNR is ResNet-50, with a test FNR of 0.77%. Interestingly, ResNet-50 was also the model that achieved the highest False Positive Rate (FPR) of 14.9%. A high False Positive Rate is also undesirable in the medical field because it could lead to a waste of resources and emotional stress on healthy patients. However, due to fluctuations in its loss and accuracy graph, we do not believe that this would be the most effective model. Rather, DenseNet-121 was the most stable in terms of both accuracy and loss and still performed very well on the test data, with a still low false negative rate of 0.032.

Surprisingly, Inception v3 performed the worst overall; it achieved a final test accuracy of 86.35%, FPR of 13.57%, and FNR of 10.49%. This could be because InceptionNet seeks to reduce computational cost in CNNs. However, with a small dataset like ours, computational cost is not very expensive, so this factor was never accounted for.

Another notable result was the effect of overfitting - or lack thereof. The more parameters a model has, the more prone overfitting is to occur. In addition, since our problem is binary classification, a larger model is even more likely to overfit, harming performance. This was the case for DenseNet, as DenseNet-121 outperformed DenseNet-161, and generally for ResNet, where ResNet-18 overall had lower loss and higher accuracy on the validation set. Our results show the models with fewer parameters performed better overall, as expected.

Comparing between models, it is clear that DenseNet was the best performing model as its accuracy increased over the epochs and did not fluctuate as much as the other models. This is likely because DenseNet utilizes connecting layers in a feed-forward method where the feature maps of previous layers are used as inputs for each layer. A benefit is that it drastically reduces the number of parameters for the model, which is likely why it performed so well compared to the other models, as it likely resulted in less overfitting.

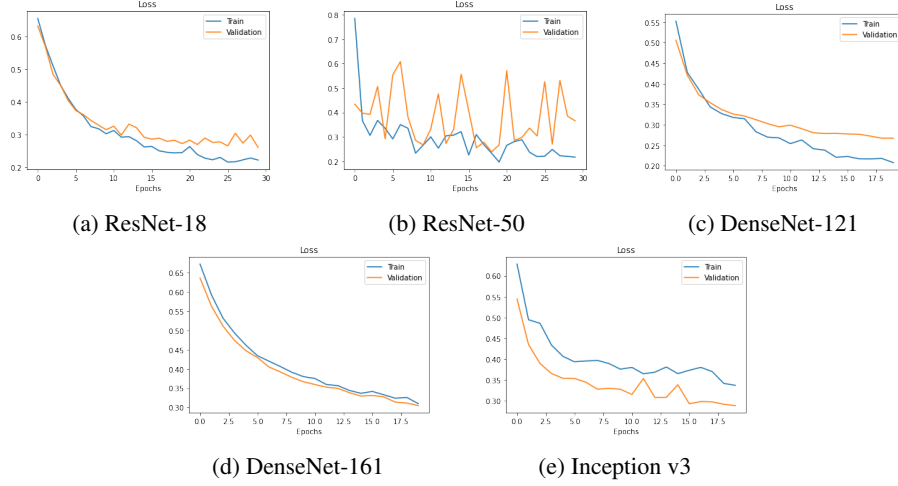


Figure 1: Model Loss

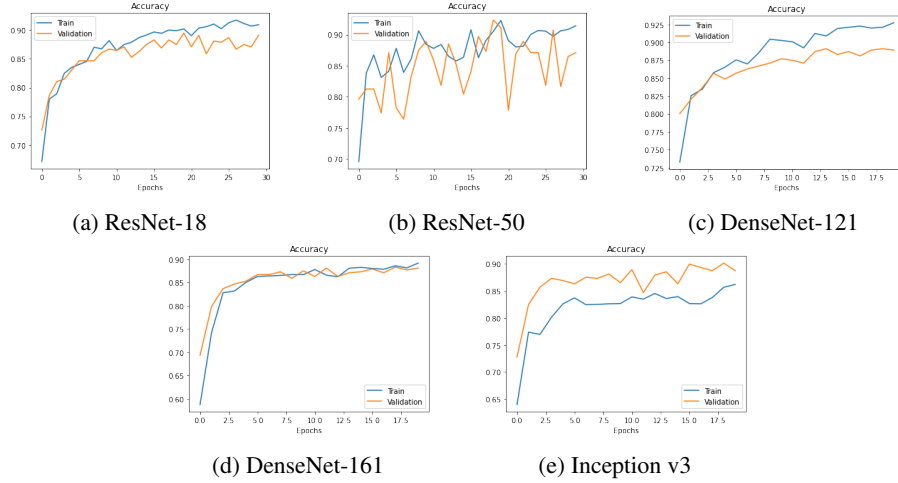


Figure 2: Model Accuracy

6 Summary

The COVID-19 pandemic has caused repercussions around the world that impact our lives. It is evident that we will continue to feel the consequences for years to come. Thus, it is imperative that we continue to develop new methods that can streamline COVID-19 testing in order to mitigate its spread. In this paper, we explored the performance of five pre-trained CNNs, namely ResNet-18, ResNet-50, DenseNet-121, DenseNet-161, and Inception v3, in classifying between COVID-19 and non COVID-19 chest CT scans. From our experiments, we discovered that DenseNet-121 had the highest test accuracy, 94.38%, with reasonable False Positive and False Negative rates. Other models saw tradeoffs in classification accuracy, FPR, and FNR, which would make them less optimized for medical testing. Additionally, we showed that models that have more parameters tend to overfit on smaller datasets like the one used.

In conclusion we demonstrate how several existing architectures can achieve satisfactory on new problems using transfer learning. However, for the purposes of medical testing, more specialized technology should be developed. In future work, researchers can also explore how neural networks specifically designed for examining CT scans and other medical imagery perform in classifying COVID-19 chest scans.

7 Attributions

Cullen Pu wrote the code for the training loop and image visualization. Alex Shih wrote code for loading the dataset and plotting loss. Stella Cai fine tuned many of the models. All three members contributed to the report.

References

- [1] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV].
- [2] Gao Huang et al. *Densely Connected Convolutional Networks*. 2018. arXiv: 1608.06993 [cs.CV].
- [3] Forrest N. Iandola et al. *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size*. 2016. arXiv: 1602.07360 [cs.CV].
- [4] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].
- [5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012. URL: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>.
- [6] Tawsifur Rahman et al. “Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray”. In: *Applied Sciences* 10.9 (May 2020), p. 3233. ISSN: 2076-3417. DOI: 10.3390/app10093233. URL: <http://dx.doi.org/10.3390/app10093233>.
- [7] Eduardo Soares and Plamen Angelov. *SARS-COV-2 Ct-Scan Dataset*. 2020. DOI: 10.34740/KAGGLE/DSV/1199870. URL: <https://www.kaggle.com/dsv/1199870>.
- [8] Eduardo Soares et al. “SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification”. In: *medRxiv* (2020). DOI: 10.1101/2020.04.24.20078584. eprint: <https://www.medrxiv.org/content/early/2020/05/12/2020.04.24.20078584.full.pdf>. URL: <https://www.medrxiv.org/content/early/2020/05/12/2020.04.24.20078584>.
- [9] Christian Szegedy et al. *Rethinking the Inception Architecture for Computer Vision*. 2015. arXiv: 1512.00567 [cs.CV].
- [10] Linda Wang, Zhong Qiu Lin, and Alexander Wong. “COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images”. In: *Scientific Reports* 10.1 (Nov. 2020), p. 19549. ISSN: 2045-2322. DOI: 10.1038/s41598-020-76550-z. URL: <https://doi.org/10.1038/s41598-020-76550-z>.
- [11] Ning Xie et al. *Explainable Deep Learning: A Field Guide for the Uninitiated*. 2020. arXiv: 2004.14545 [cs.LG].