Identifying the Best Image Classification Algorithm for COVID-19 Diagnosis

With a Small, Imbalanced Chest X-Ray Dataset

Joyce Yang
Notre Dame High School, San Jose

Objective

The objective of this project is to train families of deep learning neural networks on small, imbalanced chest X-ray datasets to automate diagnosis of respiratory illnesses. Specifically, the best algorithm will be identified to classify anonymized chest X-ray images to three classes: healthy, COVID-19 and non-COVID pneumonia.

Abstract

This project is regarding which algorithm will perform the best when training a COVID-19 image classifier in the context of a small, imbalanced dataset. Specifically, each algorithm represents the combination of use of domain-relevant dataset for pretraining; method of data sampling; and choice of neural network architecture. Publicly available chest X-ray image datasets are not abundant, and ground truth data of COVID-19 diagnosis is especially hard to come by. To overcome this issue, I pretrained neural networks using the ChestX-ray14 database generated by NIH. Next, to address the imbalance within training data, I implemented two alternative data sampling methods. Third, three families of neural networks that represent state-of-the-art image classification architecture are analyzed: DenseNet, EfficientNet and ResNet. Precision and Recall were the main metrics utilized to evaluate performance. Based on extensive experimentation, the algorithm with pretraining on the ChestX-ray14 database, using fixed-fraction-per-batch sampling method, and trained on the DenseNet family of neural network has been identified to have the highest Recall and Precision for COVID-19 chest X-ray images.

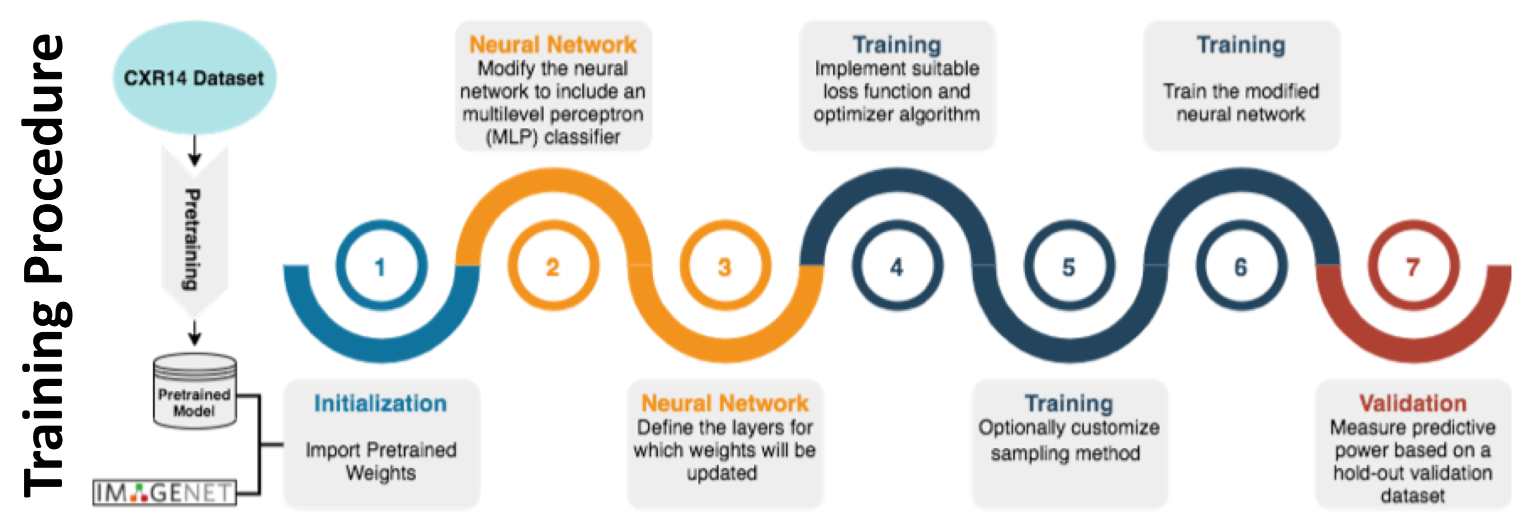
Datasets

- COVID-Net is constructed from five open-source chest X-ray image datasets, containing 13,975 images across 13,870 patient cases with ground truth labels. The ratios of images in three classes – healthy, COVID-19 and non-COVID pneumonia – are **16:1:11**.
- ChestX-ray14 database generated by NIH is used for pretraining. It contains over 100,000 images in 13 classes of respiratory viruses, such as Hernia, Fibrosis and Pneumonia, as well as normal patients.

Materials and Methods

Materials:

- Hardware: AWS p2.xlarge Spot Instance, Nvidia Tesla K80
- Software: torch and torchvision, NumPy, pandas, sklearn
- Development Tools: PyCharm, spotty-cloud



Variables:

For each of the nine neural networks from three families:

1. Selection of pretrained weights (Step 1)
2. Selection of data sampling method (Step 5)

- For Equal-Weight method, during one training epoch, the overall number of images from each class will be equal third of the total number of images.
- For Fixed-Frac method, a fixed number of images from COVID-19 class will be included in each training batch, based on specified fraction, and the rest of images is distributed to the other two classes in proportion to their population.

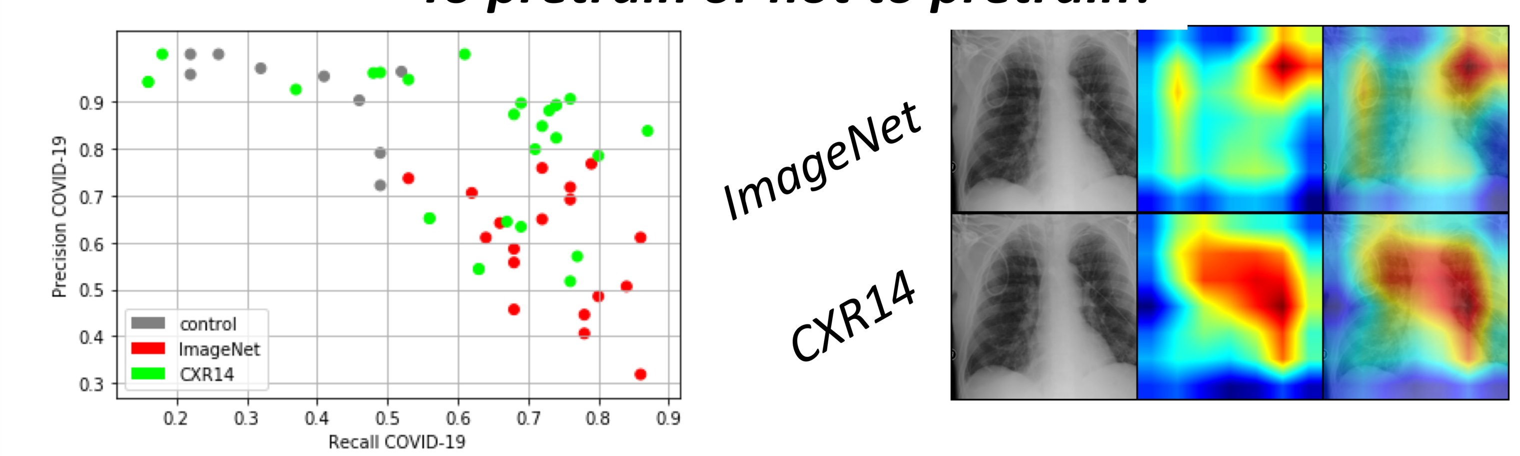
Measurements

For each neural network, 6 sets of performance metrics are collected, including 1 control set, and 5 experiment sets. For both pretraining of neural networks and transfer learning, several classification performance metrics are evaluated:

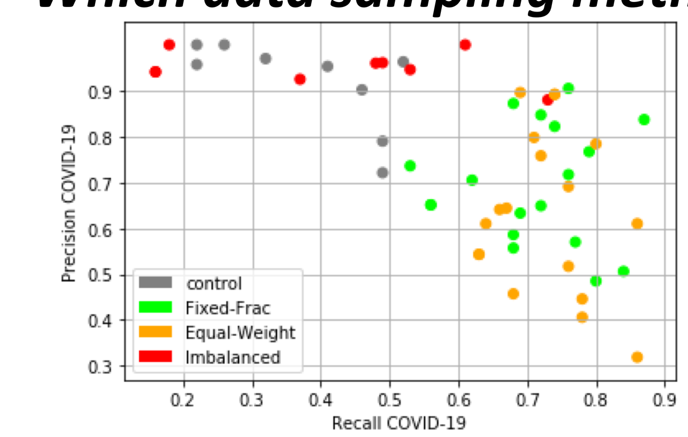
1. Recall – True Positive/(True Positive + False Negative)
2. Precision – True Positive/(True Positive + False Positive)
3. F1 score – (2*Precision*Recall)/(Precision + Recall)
4. Average Precision – area under precision recall curve
5. ROC curve – area under curve of precision and recall values at different thresholds

Results and Interpretation

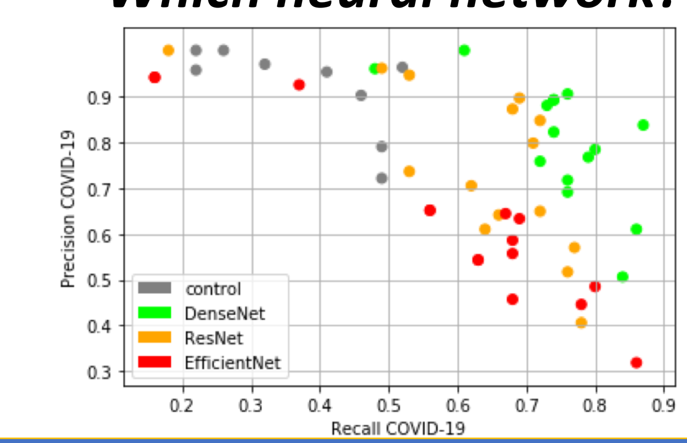
To pretrain or not to pretrain?



Which data sampling method?



Which neural network?



Conclusions

At 87% Recall and 84% Precision for the COVID-19 class, the highest performing algorithm used the CXR14 pretraining, Fixed-Frac sampling method, and DenseNet169 model. In comparison, for the control group, with ImageNet pretrained weights and no customized data sampling method, the Recall metric for COVID-19 is consistently less than 0.6, which is a clear indicator of underperformance on COVID-19 prediction.

- Transfer learning from ImageNet using standard CNN models and corresponding pretrained weights has become a widespread method for deep learning applications to medical imaging. However, pretraining on a domain-specific dataset, in this case ChestX-ray14 database, significantly improves performance during transfer learning.

- The Fixed-Frac sampling is the only method which ensures that within every batch, COVID-19 chest X-ray images are present. For the other sampling methods, because COVID-19 images are not present in many batches, the performance during transfer learning is severely affected.

Relevance to Biotechnology

The utilization of chest X-ray images in medical diagnosis is both a cost-effective and widespread technique for early screening of respiratory illnesses. With more than 82 million COVID-19 cases worldwide, automated chest radiograph interpretation could provide substantial benefit for efficient and accurate diagnosis of COVID-19 patients.

Acknowledgements: Thank you Dr. Koltermann and Ms. Foote for guiding me during the experimentation process, and my parents for their support.