Training a Multiclass Classifier on Chest X-rays to Distinguish COVID-19-Induced Pneumonia

Sumiran Thakur, Nick Nolan, & Mohammed Abdulwahhab
Spring 2020
University of California, Berkeley
EECS 127/227AT - Optimization Models in Engineering

Abstract— **In this paper, we investigate the potential for four different classifiers to reliably predict instances of COVID-19-induced pneumonia based on patient chest X-rays, and identify that a Convolutional Neural Network trained with Transfer Learning approaches is the best classifier for this purpose. We developed a CNN architecture that can, with very high ($> 90\%$) accuracy, predict a COVID-19-based pneumonia. This paper details the process by which this architecture was developed, and the optimization processes from the course with which we engaged.**

I. BACKGROUND

In the present global pandemic, it is imperative that hospitals are able to conduct diagnostic tests for COVID-19 accurately and quickly in order to facilitate treatment as necessary to save lives. Currently, the most common testing method is genetic and may take up to an hourhowever, access to these testing kits is the greatest bottleneck in the pipeline thus far, preventing people from receiving the treatment they may seriously need. Recent X-ray data suggests that machine learning models may be able to predict the presence of SARS-CoV-2 in the body.

As the title suggests, we would like to investigate how well Convolutional Neural Nets (CNNs) trained on images of chest X-rays can distinguish COVID-induced pneumonia vs. regular viral pneumonia. We hypothesize that there is high predictive power in chest X-ray images of affected patients and would like to determine the extent to which this is true.

To this end, we will train an assortment of different classifiers (Logistic Regression, Support Vector Machine (SVM), RBF-Kernelized SVM, and a CNN) on a suite of aggregated datasets containing healthy chest X-rays, chest X-rays of people with non-COVID-induced pneumonia, and chest X-rays of people with COVID-induced pneumonia. To this end, we can determine which model has the highest test accuracy, and therefore which one is able to differentiate between COVID-induced pneumonia best.

II. MATERIALS & METHODS

A. Datasets

We used a Kaggle dataset to train our model. The dataset contained instances of healthy chest X-rays; chest X-rays from people that have a non-COVID-related viral pneumonia; and those from people with a COVID-related pneumonia. Links to the dataset may be found in the References.

B. Training Models

For all models, 70% of the data is used for training, 15% is used for validation, and 15% for testing. Also of note is the fact that, for any models trained with scikit-learn, we were unable to make use of Google Colab's GPU, which speeds up calculations more than tenfold. This meant that, unless we reduced the data, training for a single model took upwards of 24 hours to run this on Google Colab, the user must be active every 90 minutes. So, the image data was reduced from 1024x1024 down to 224x224.

To train the Logistic Regression model, we utilized scikit-learn's LogisticRegression package; we performed a grid search among C-values ranging from 10^{-4} to 10^{7} with 10-fold cross-validation to find the best model. We used an L2-norm penalty and a Newton-Conjugate Gradient solver.

To train the Linear SVM model, we utilized scikit-learns SVC package; we performed a grid search among C-values ranging from 10^{-4} to 10^{7} with 10-fold cross-validation to find the best model.

To train the Kernelized SVM model, we utilized scikit-learns SVC package; we performed a grid search among C-values ranging from 10^{-4} to 10^{7} with 10-fold cross-validation to find the best model with a Radial Basis Function kernel.

The CNN architecture of our choice was VGG16. To begin with, the architecture was pre-trained on the ImageNet dataset, and the last few layers were fine tuned for the task at hand. The input size of the image to the architecture was 224 by 224 (default VGG16 input) and a 4 by 4 average poling was applied as an input to the fine tuning layers. As with most neural net architectures, a coarse, then fine, grid search was applied to the hyperparameters of the learning rate, batch size, and dropout probability. We found the best values at an initial learning rate of 10^{-3}, a batch size of 8, and a dropout probability of 0.4. As for the learning rate, we also decided to implement a decaying learning rate which decayed at the order of the initial learning rate divided by the number of epochs passed. We decided to use Adaptive Momentum Optimization, or Adam, to optimize parameters; briefly, Adam is an adaptive learning rate method wherein the gradients are squared to scale the learning rate, and the moving average of the gradient instead of the gradient itself, is used. It is a combination of two other well known optimization algorithms known as RMSprop and Stochastic Gradient Descent. As for our linear activation function in the finetuning layers, our choice was ELU, the exponential linear unit. ELU is exactly identical to RELU for non-negative inputs, but for negative inputs does a much better job than RELU as it smooths slowly whereas RELU is much sharper.

III. Results

After training each of the 4 models, their efficiencies were evaluated on the 15% test set of data. Among these, results are seen in Figs. 1-4.

As is evident, the transfer learning VGG16 model performed the best with an overall accuracy of 95.83%, the RBF kernel SVM and linear SVM tied for second with an accuracy of 89.90%, and in last logistic regression with an accuracy of 86.87%. The precision, recall, and f-scores for all the models are also evident in the tables.

Fig. 1. Confusion matrix for logistic regression model

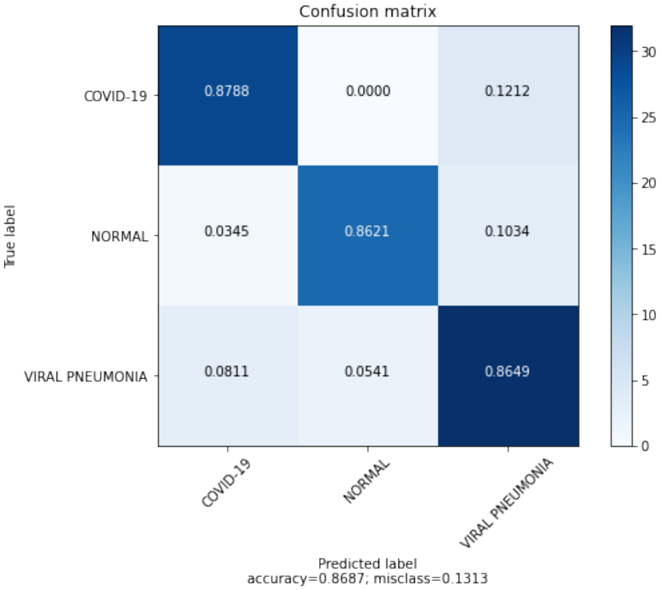


TABLE I
Various statistics for the logistic regression model. Accuracy: 0.87

Classification	Precision	Recall	F1-score
COVID-19	0.88	0.88	0.88
Normal	0.93	0.86	0.89
Viral Pneumonia	0.82	0.86	0.84

TABLE II
Various statistics for the linear SVM model. Accuracy: 0.90

Classification	Precision	Recall	F1-score
COVID-19	0.89	0.94	0.91
Normal	0.87	0.93	0.90
Viral Pneumonia	0.94	0.84	0.89

IV. Discussion

We observe that the transfer learning-based VGG16 model outperformed all other architectures. In studies regarding medical data, even the tiniest percentage poits of accuracy have significant impact. While logistic regression and kernelized SVM did fairly well, they simply don’t

Fig. 2. Confusion matrix for linear SVM model

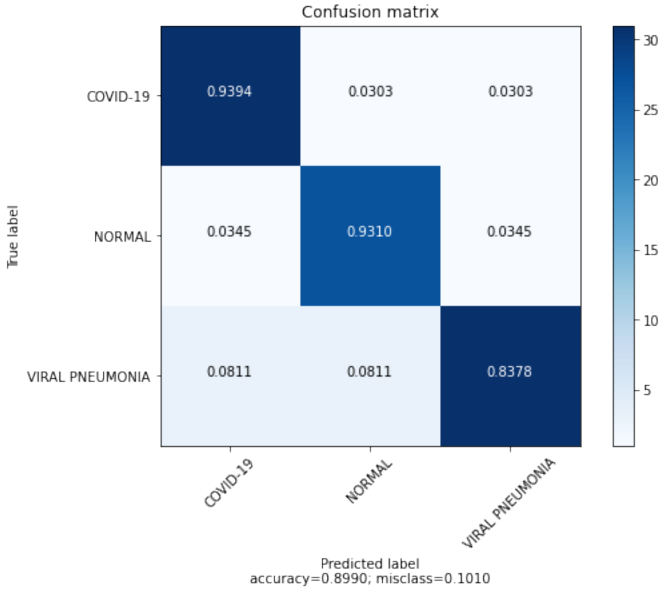
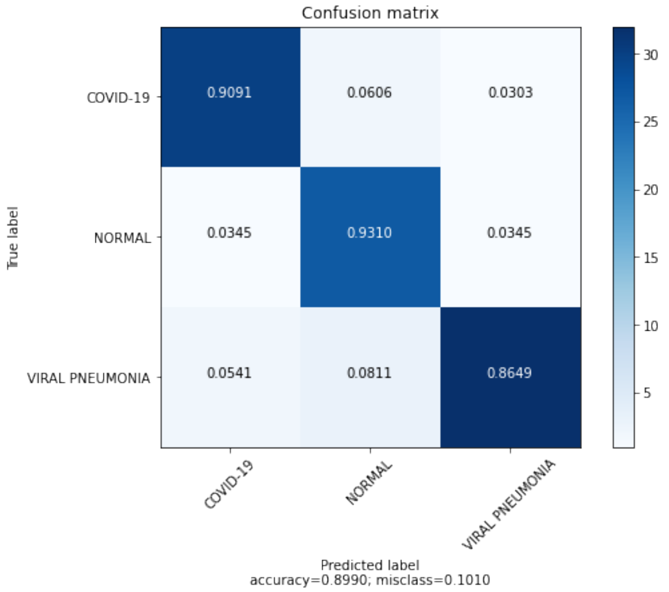


Fig. 3. Confusion matrix for RBF-kernelized SVM model



scale as well with larger datasets, both in terms of accuracy and speed. As mentioned previously scikit-learn is not GPU enabled, and the training time for a GPU enabled keras model will be faster than that of scikit-learn's logistic regression and SVM models for much larger datasets. Since we just used a single Kaggle dataset and scaled it to have an even number of normal, COVID-19, and non-COVID-induced pneumonia chest X-rays, we only end up with a total of 657 samples in the end. A future extension to this could include integrating

TABLE III

VARIOUS STATISTICS FOR THE RBF-KERNELIZED SVM MODEL. ACCURACY: 0.90

Classification	Precision	Recall	F1-score
COVID-19	0.91	0.91	0.91
Normal	0.84	0.93	0.89
Viral Pneumonia	0.94	0.86	0.90

Fig. 4. Confusion matrix for VGG16 model

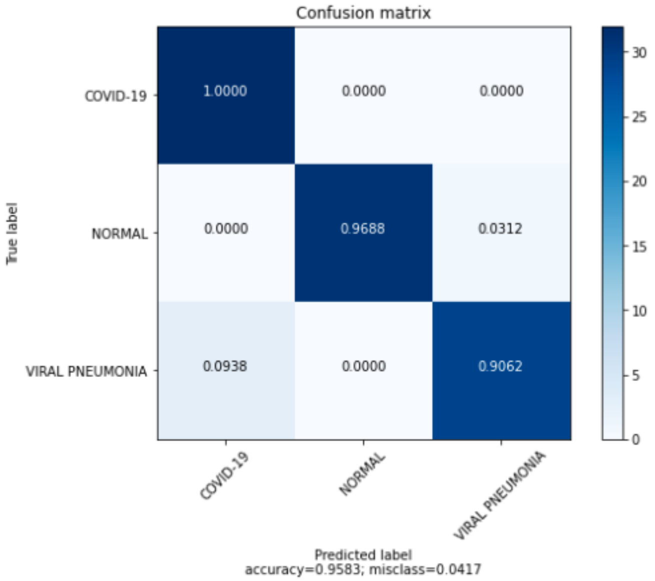


TABLE IV

VARIOUS STATISTICS FOR THE VGG16. ACCURACY: 0.96

Classification	Precision	Recall	F1-score
COVID-19	0.91	1.00	0.96
Normal	1.00	0.97	0.98
Viral Pneumonia	0.97	0.91	0.94

other datasets once data becomes more publicly available, and then checking the robustness of the fine-tuned VGG16 model against a much larger data set.

REFERENCES

[1] Ker, J., Wang, L., Rao, J., Lim, T. (2017). Deep learning applications in medical image analysis. *IEEE Access, 6*, 9375-9389.

[2] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging, 35(5)*, 1285-1298.

[3] Ravishankar, H., Sudhakar, P., Venkataramani, R., Thiruvenkadam, S., Annangi, P., Babu, N., Vaidya, V. (2016). Understanding the mechanisms of deep transfer learning for medical images. *In Deep learning and data labeling for medical applications (pp. 188-196)*. Springer, Cham.

[4] Mooney, P. (2018, March 24). Chest X-Ray Images (Pneumonia). Retrieved May 04, 2020, from https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia