MET 521 Final Project Submission
Advanced Topic: Chest Xray classification through Deep Learning

Author: Tzu-Yu Chiang (Diana)

Introduction:
In this project, I set out to compare two different approaches to chest X-ray classifications. While this task has been achieved in previous attempts using purely machine learning techniques and features developed from well-known deep learning algorithms, I seek to implement scientific observations as stratification parameters to increase the explainability of predictions provided by machine learning in medical applications.

In this project, I have created two jupyter notebooks that takes the user through the steps I utilized to run two algorithms:

1. Importing dataset
2. Data processing for using Keras image-generator to augment and process data
3. Data visualization
4. Transfer learning and building a model
5. Run each model for 100 epochs
6. Visual display of prediction results
7. VGG16 – data post-processing

As the data are chest X-ray images with labels, visualization of the images and research into possible scientifically backed disease presentation in X-rays are vital in increasing the explainability of machine learning predictions.

Downloading and processing data:
To process the data utilizing prior clinical knowledge, each chest X-ray image must be separated in two along X-axis. As research [1]states that bacterial pneumonia typically presents in one instead of both lobes as opposed to viral pneumonia, this means to process each lobe separately to determine whether it is healthy or diseased. Each image file was separated and renamed with original filename with extension of _L or _R prior to file type extension. The original data file was then moved to a new folder to reduce unnecessary code changes as original method utilized Keras “Flow from Directory” method with folder names as labels.

Data augmentation utilizing Keras image generator and separating of training and validation (80-20 split) set was utilized to prevent overfitting or memorization of data by model. A test set was loaded separately to ensure that no test data was utilized in training. This ensures the integrity of the model.

---

[1] https://www.pfizer.com/news/articles/viral_vs_bacterial_pneumonia_understanding_the_difference

Clinical indication and Data exploration:
Covid Pneumonia typically presents with “ground glass opacity” in chest Xrays which refers to a grey hazy area on radiography of increased density. Differentiating between normal vs COVID pneumonia CXR has been relatively easy to achieve for deep learning models. However, as research shown that bacterial pneumonia typically only shows in one lobe with increased density vs both lobes, this suggests, classifying each lobe on its own then categorizing with ++ vs +- can assist in differentiating between bacterial and viral pneumonia. The same can be observed in parts of the current data as shown in figure 1.

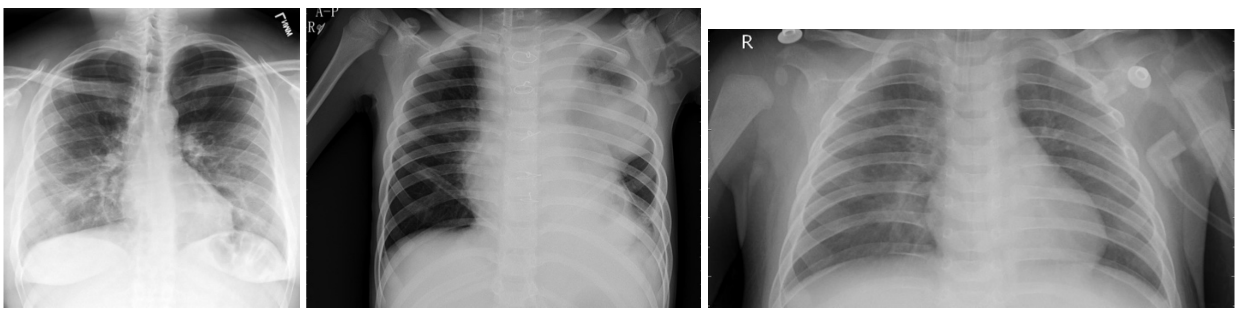

Figure 1: (Left) Covid Pneumonia from radiopaedia.org (Center)Bacterial Pneumonia from data set. (Right)Viral pneumonia from data set

ML Algorithm Selection:
VGG16 was the transfer learning base model selected as this model utilizes convolutional network of 3x3 filters to find features that are utilized to categorize images. This model was tested in 2014 on ImageNet database of 14 million images belonging to 1000 classes. The network is 16 layers deep with many 3x3 filters and max pooling layers. The small size of filters allows for better feature identification; thus, better differentiating between different images. The output layer was modified to be a Dense layer of 2 as the desired output is to identify each lobe as “normal” or “abnormal”. Prediction made with this ML model was then sorted through to find each L and R lobe of the same file and if they have been identified as same, the CXR image is classified as viral pneumonia; if each lobe is identified as different, the image is classified as bacterial pneumonia.

MobileNetV2 was utilized as the transfer learning base model for the traditional approach. This model was boosted as a robust but light weight model. The main feature of MobileNet is that it utilizes depth -wise convolution which drastically reduce complexity cost and model size. This means that the model can be updated more frequently as symptoms and disease presentation changes with mutations.

Results:
VGG16 with scientific backing model shows 44% accuracy in identifying bacterial pneumonia and 88.8% in viral pneumonia. The combined accuracy was 60% for distinguishing between viral and bacterial pneumonia without utilizing more advanced techniques such as decreasing step size closer to local minimum in gradient descent, utilizing GPU for more training rounds,

etc. The accuracy for classification between normal and abnormal was over 90% without advance techniques. Average accuracy over all 4 classes, averaged out to be just above 75%.

MobileNet was used with traditional approach to distinguish between all 4 classes with accuracy just over 50 percent. A closer look at the results tells the story that the model simply chooses between class 1 normal and class 3 viral pneumonia with only 1 test case where it chose class 0 covid. The model has completely ignored label of bacterial pneumonia.

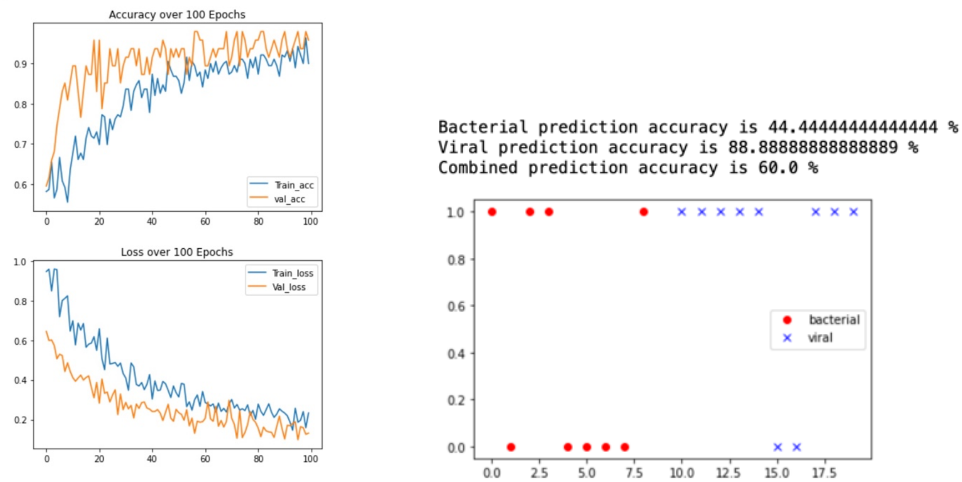

Figure 2: VGG16 (Left) Accuracy per epoch for training and validation set of normal vs abnormal (normal vs covid) data. (Right): Result of using trained model on test data to distinguish between viral and bacterial pneumonia. Red “o” are bacterial, blue “x” are viral with 0 being incorrectly classified and 1 being correctly classified

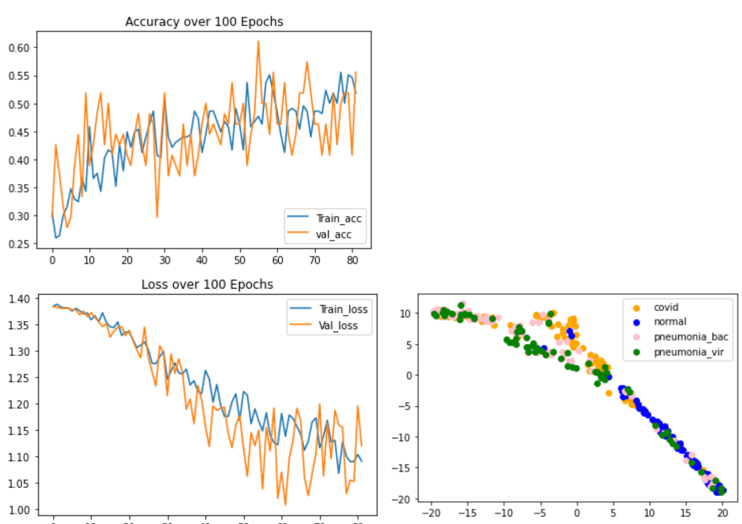

Figure 3: MobileNetV2 (Left) Accuracy per epoch for training and validation set of all 4 (Right): TSNE showed poor separation between all 4 classes

Conclusion:
While ML assisted chest X-ray may not be perfect, the utilization of knowledge gained through research showed significantly more intuitive results and may provide more acceptance of ML assisted diagnostics tools for alleviation of tremendous workload on hospital radiologist by prescreening as a rule-in mechanism and quickly identify normal CXR.

The utilization of evidence backed research in classification also follows physicians' trainings of utilizing latest and rigorously demonstrated clinical evidence and research in practice. Making the adaptation of research-based ML classification techniques a better alternative to traditional purely ML algorithm-based prediction.

Future work and Direction:
I intend to continue this work and dive deeper into explainabilities of the very deep learning models and algorithms. While many Blackbox algorithms produce amazing accuracy and results, the mechanism in which these models make predictions remain unknown and such unknown poses significant risk to treatment and patient care based on these tools. But I firmly believe that when done right, such tools can help increase clinical throughput and reduce healthcare costs.