Detecting Pneumonia in Chest X-Rays

Samuel Rahwa

Pneumonia Facts

- Pneumonia is a form of acute respiratory infection that affects the lungs.
- Pneumonia accounts for 14% of all deaths of children under 5 years old, killing 740,180 children in 2019
- For US adults, pneumonia is the most common cause of hospital admissions other than women giving birth.
  - About 1 million adults in the US seek care in a hospital due to pneumonia every year, and 50,000 die from this disease

Normal
Bacterial Pneumonia
Viral Pneumonia

It may shock you to learn that the error rate for radiologists is 4%. On average there are 1 billion radiology exams each year. By this logic, that means there will be 40 million radiologist errors.

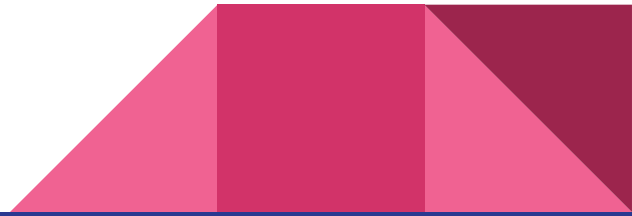
Insurance companies frequently request medical records when evaluating claims. What if we could create a system to double check the radiologist's evaluation. This would increase patient safety/long term care and reduce chances of further complications.

Source of Data

- Guangzhou Women and Children's Medical Center
  - There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal)
  - Selected from retrospective cohorts of pediatric patients of one to five years old
  - All chest X-ray imaging was performed as part of patients' routine clinical care.
- Provided by Kaggle
  - https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

EDA (Exploratory Data Analysis)

- Distribution of Images
  - Train Images
    - A total of 5216 Images
  - Test Images
    - A total of 624 Images
- Both Data Sets are Imbalanced
  - Favoring Pneumonia in both cases
- Normal X-Rays are labeled 0
- Pneumonia X-Rays are labeled 1
  - Regardless of the type of pneumonia

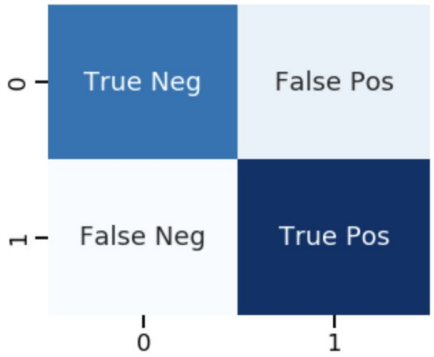


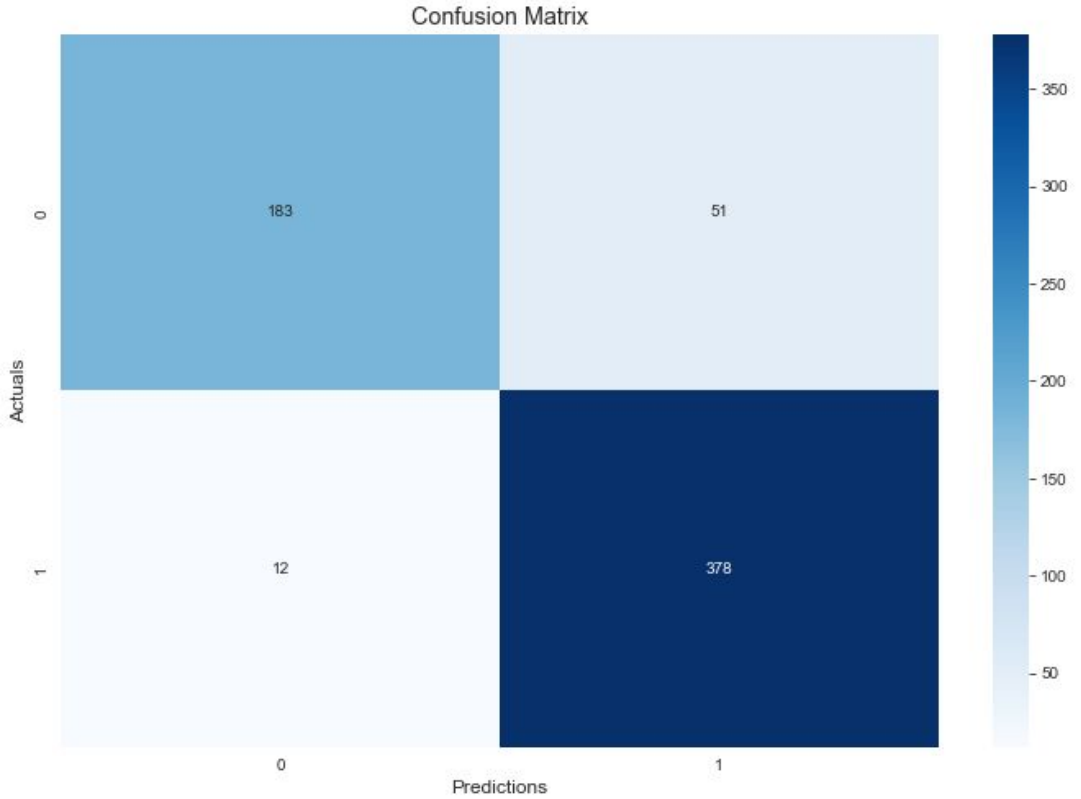
Class Imbalance: Test Data
Normal
Pneumonia
Class
0
50
100
150
200
250
300
350
400
Number of Training Images

Class Imbalance: Training Data
Normal
Pneumonia
Class
0
500
1000
1500
2000
2500
3000
3500
4000
Number of Training Images

Models

- I went through a four part iterative modeling process and here are the Accuracy Scores
  - Baseline Multilayer Perceptron (MLP), 75 %
  - Baseline Convolutional Neural Network (CNN), 78 %
  - Tuned CNN, (In Progress….still)
  - Transfer Learning CNN, 90 %
- The Transfer Learning CNN was the best model
  - Transfer learning generally refers to a process where a model trained on one problem is used in some way on a second related problem.
    - VGG-16 is a dataset of over 14 million images belonging to 1000 classes
  - One or more layers from the trained model are then used in a new model trained on the pneumonia images.
- The Metrics of the Transfer Learning CNN
  - Roughly 1.87 % of the 624 Images are predicted as normal, when they actually have pneumonia
  - About 7.94 % of the 624 Images are predicted as pneumonia, when they actually are normal
  - The model is accurate at detecting normal and pneumonia cases


True Neg
False Pos
False Neg
True Pos
0
1
0
1


Confusion Matrix
183
51
12
378
Actuals
0
1
0
1
Predictions
350
300
250
200
150
100
50

Conclusions

- Insurance companies could use tool to drive down the risks of False Negatives
  - 1.87 % are predicted as normal, when they actually have pneumonia
- By checking the images after receiving a radiologists diagnosis, insurance companies reach out for clarification when they don’t come to the same conclusion.

Next Steps

- We could have doctors and technicians crop out the areas they identify as pneumonia
- Removing the diaphragm in X-Rays, as a standard practice
- Standardizing pixels, resolution and sizes of X-rays images
- Widen the range of ages and geographical distribution of X-Rays Images
- Research more effective tuning methods

Thank you!

Email: samuelaaronrahwa@gmail.com

GitHub: @SamuelRahwa

LinkedIn: https://www.linkedin.com/in/samuelrahwa