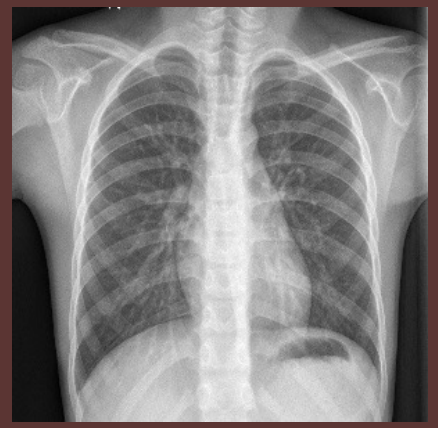
Detecting Pneumonia in X-Ray images using Convolution Neural Networks

Bob Spoonmore

July 2021

Problem Statement

Normal



Pneumonia

How can images of infant chest X-Rays be viewed algorithmically such that Pneumonia can be detected from Normal conditions with a level of confidence above 90%?

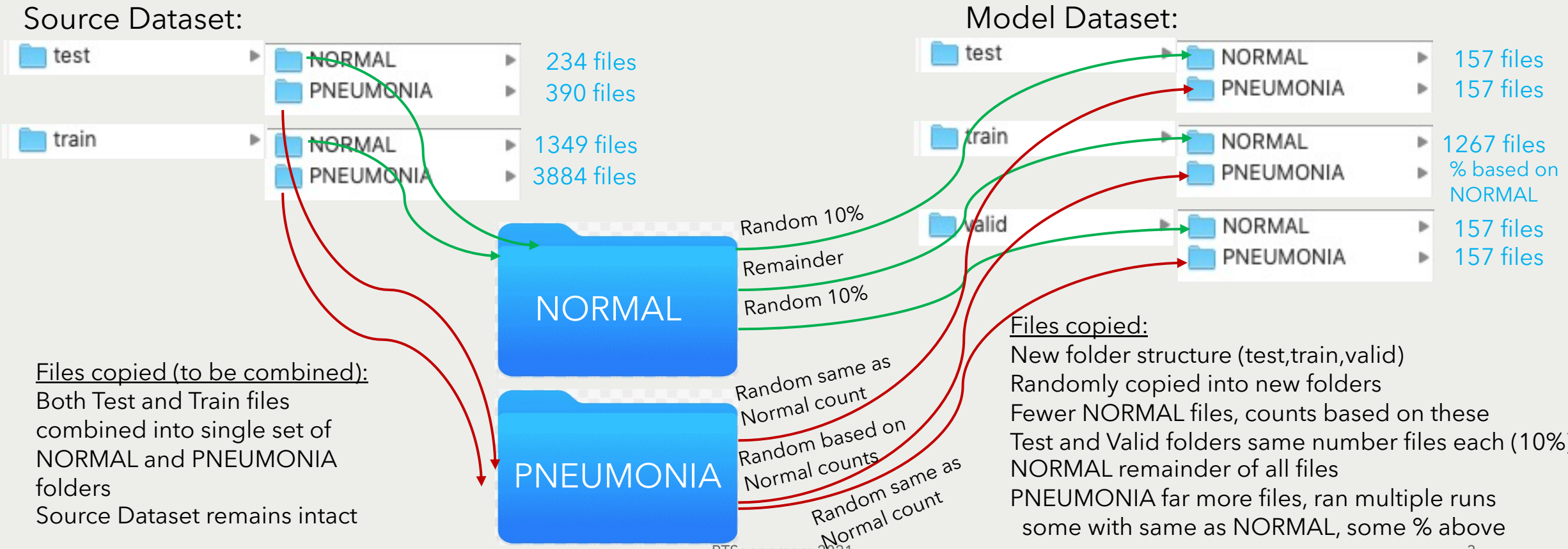
A Deep Learning algorithm of Convolution Neural Networks will be applied to images of pediatric X-Rays separated into Pneumonia and Normal labeled groups to determine if images can predict results based on training a supervised image model

Data

Images from the Kaggle Dataset:
Pediatric Pneumonia Chest X-ray https://www.kaggle.com/andrewmvd/pediatric-pneumonia-chest-xray

Dataset by Daniel Kermany and Michael Goldbaum in 2018. All images labelled as either pneumonia or normal
Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", Mendeley Data, v2 http://dx.doi.org/10.17632/rscbjbr9sj.2

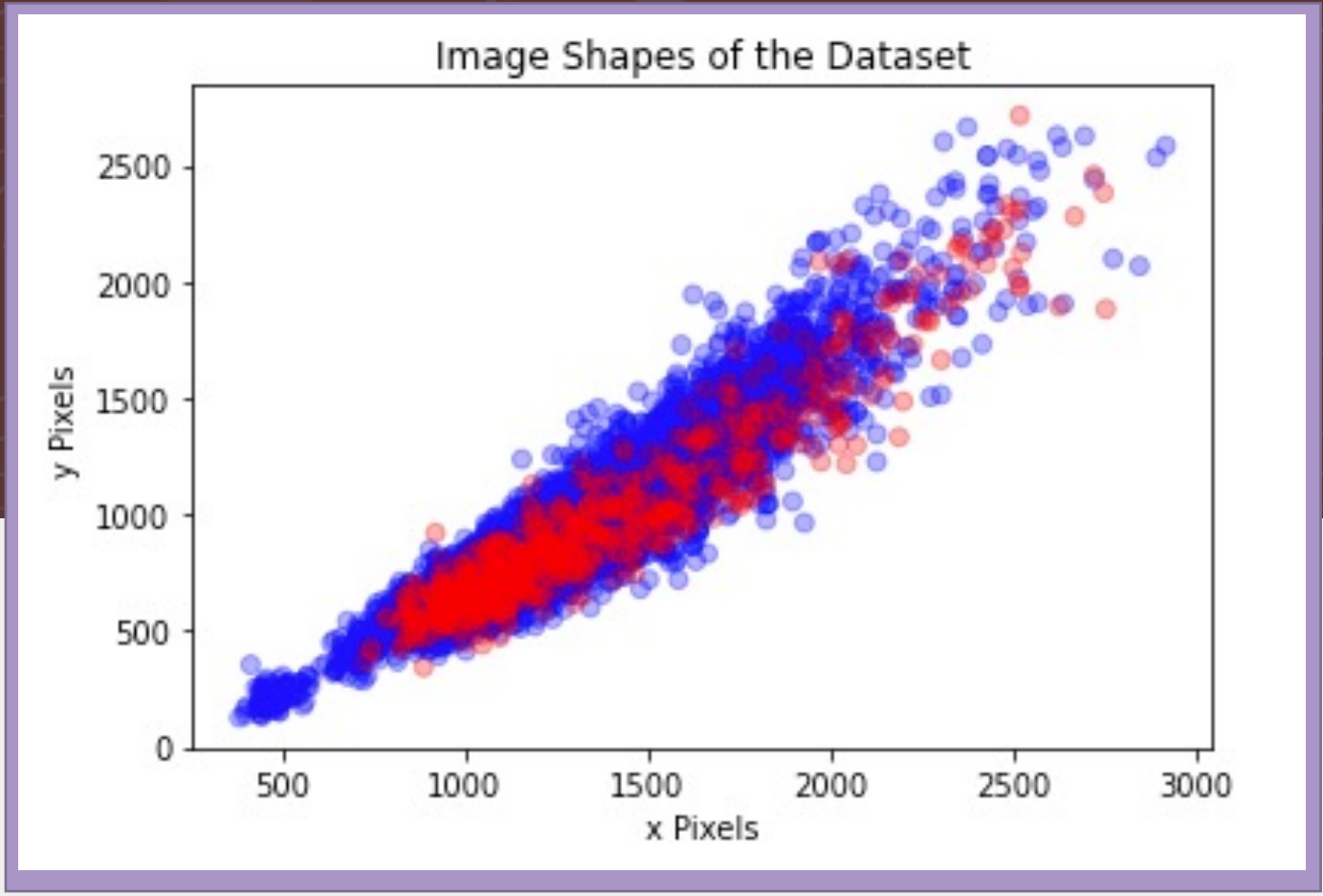
Data set 5856 images in folders: All images jpeg with various resolutions and proportions, no missing data
Final Train, Test, Valid split: 80%, 10%, 10%



Files copied (to be combined):
Both Test and Train files combined into single set of NORMAL and PNEUMONIA folders
Source Dataset remains intact

Files copied:
New folder structure (test,train,valid)
Randomly copied into new folders
Fewer NORMAL files, counts based on these
Test and Valid folders same number files each (10%)
NORMAL remainder of all files
PNEUMONIA far more files, ran multiple runs
some with same as NORMAL, some % above

Exploratory Data Analysis



Read all images and compared sizes
Image Shapes graph:
Red-NORMAL, Blue-PNEUMONIA

Unbalanced number of files:
- Far more pneumonia files – could impact analysis
- Small number of files lower in resolution
- Most image ratios within consistent spread

Set resolution size for analysis to 300x300 pixels
And batch size to 20 (to manage computer load)

Goal: process files at highest resolution/
but tradeoff with RAM size on computer

Initial runs crashed computer until size set to 300

Filtered out images below min size 200 pixels

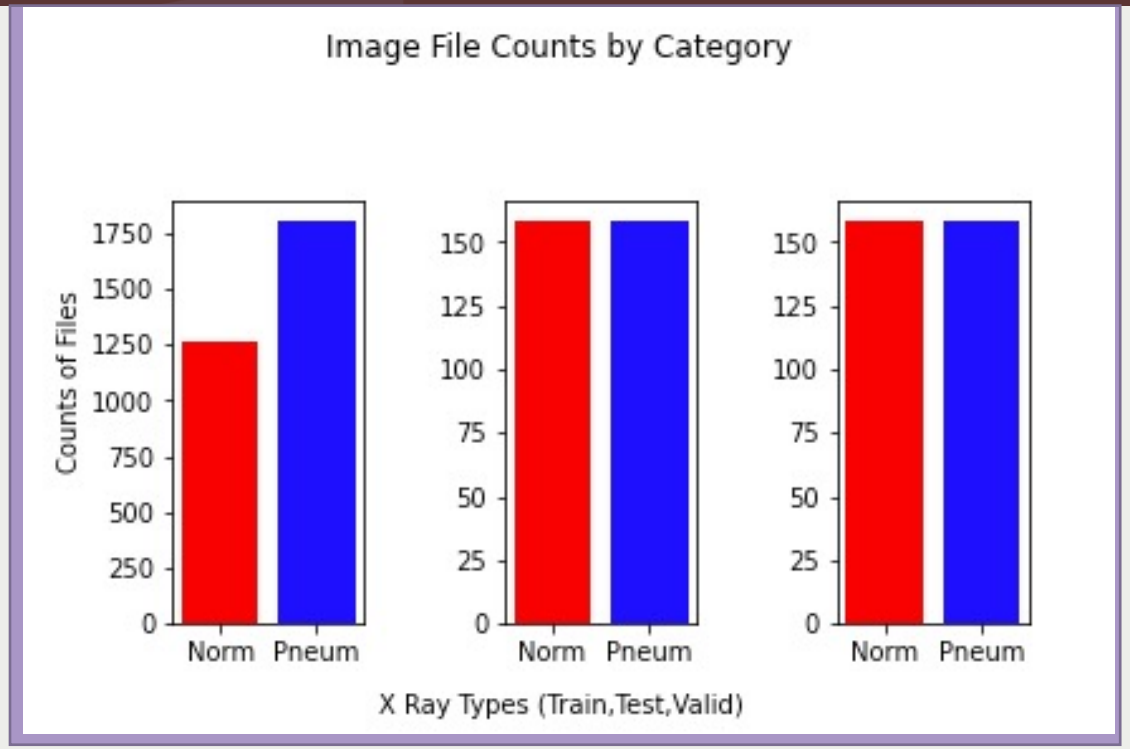
Image Preprocessing

<u>Preprocess to make images similar and comparable</u>
Read each file individually, and applied the following:

- Resize: set all to 300 pixels x 300 pixels. (square)
- Alpha: Contrast, set to 1.0
- Beta: Brightness, set to 1.0

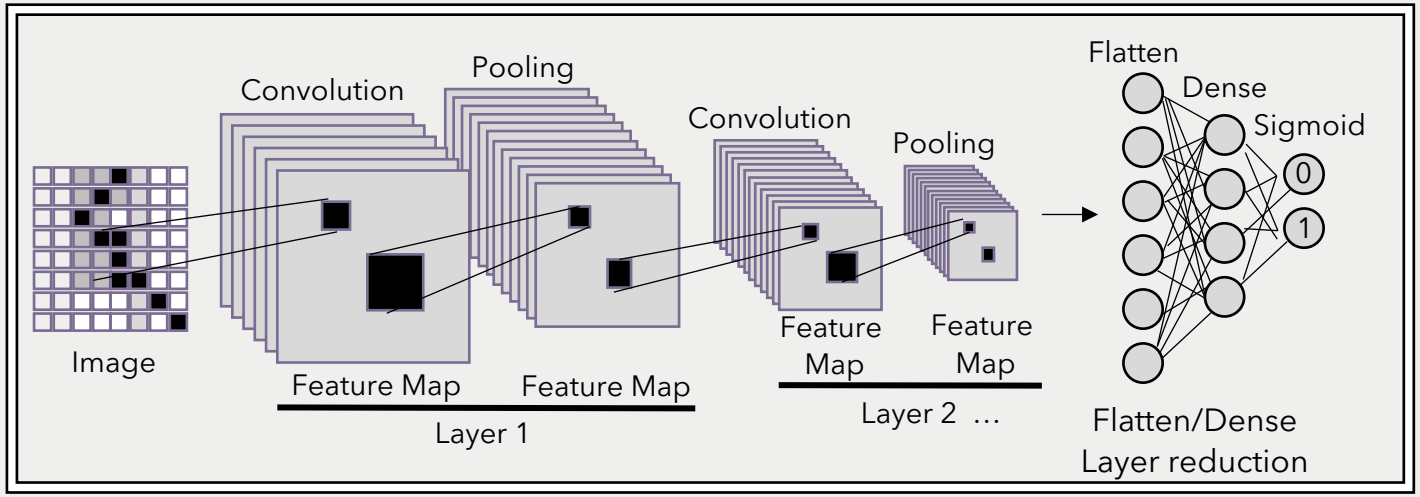
(note: multiple runs tested alpha, beta combos between 0.5 and 1.2 inclusively and best results found for 1.0 each)

- Grayscale: all images changed from RGB to gray
- Normalize: all values normalized to 1 based on #/255



Multiple test sets established: varied between Train folder holding same number of files each to progressive steps of 10% more pneumonia files. Was worried about tradeoff between biasing results and removing files from training.

CNN Model Build




```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 298, 298, 32)      320
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 149, 149, 32)      0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 147, 147, 32)      9248
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 73, 73, 32)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 71, 71, 64)        18496
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 35, 35, 64)        0
_________________________________________________________________
flatten_1 (Flatten)          (None, 78400)             0
_________________________________________________________________
dense_2 (Dense)              (None, 64)                5017664
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 65
=================================================================
Total params: 5,045,793
Trainable params: 5,045,793
Non-trainable params: 0
```

Sequential Model: all outputs lead to inputs next layer
Rectified Linear approach: directed outputs- easy to train
Sigmoid: final reduction results in a 1 or 0 – used for predictions

Three models were applied on each analysis:
- Model 1 = 1 layer
- Model 2 = 3 Layers (shown above)
- Model 3 = 5 Layers

Three models used for analysis on which was better fit: Not known initially

Model fit

Applying the Keras Algorithm

Fit parameters: optimizer = 'adam': stochastic gradient descent algorithm, efficient and low memory demands
loss = 'binary_crossentropy': for binary classification problems (two possible results)

Early Stopping Rules: (Why we made the VALID dataset in the beginning)
Uses the valid dataset to test results as model builds
Epoch (how many times we pass through all images to build model)
Set target Epoch, but model stops when conditions met – prevents overfitting
Parameters:
- target epochs at 25
- based on "accuracy" metric of model
- min_delta = 0.01 (if below this minimum change in accuracy, next epoch prevented)
- patience = 2 (number of epochs min_delta is applied)

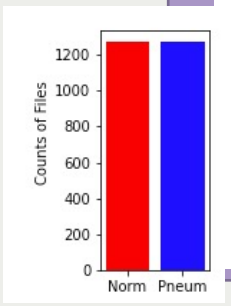
Selecting The Best Model

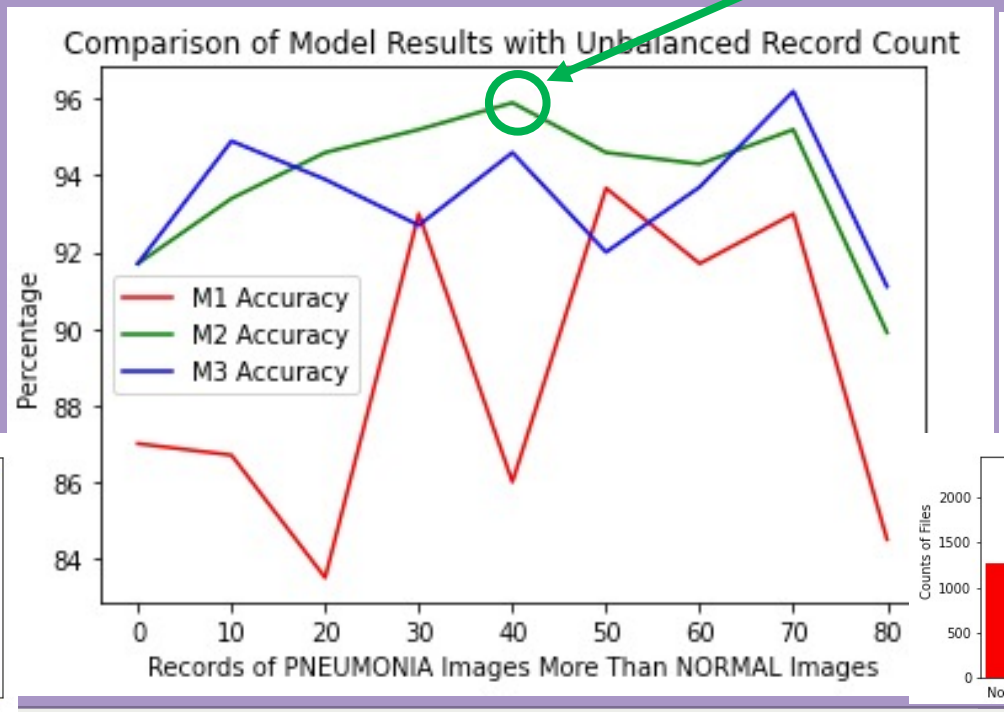
Selected MODEL 2 at 40% file volumes
(Shown in **GREEN** on graphs)

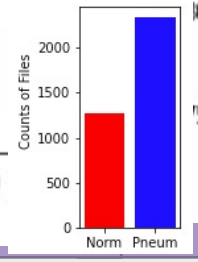
- Best average accuracy
- Best average precision
- Peak accuracy at 40%

Model was run
Successive times
Each time
increasing
Number of pneu
Files by 10%

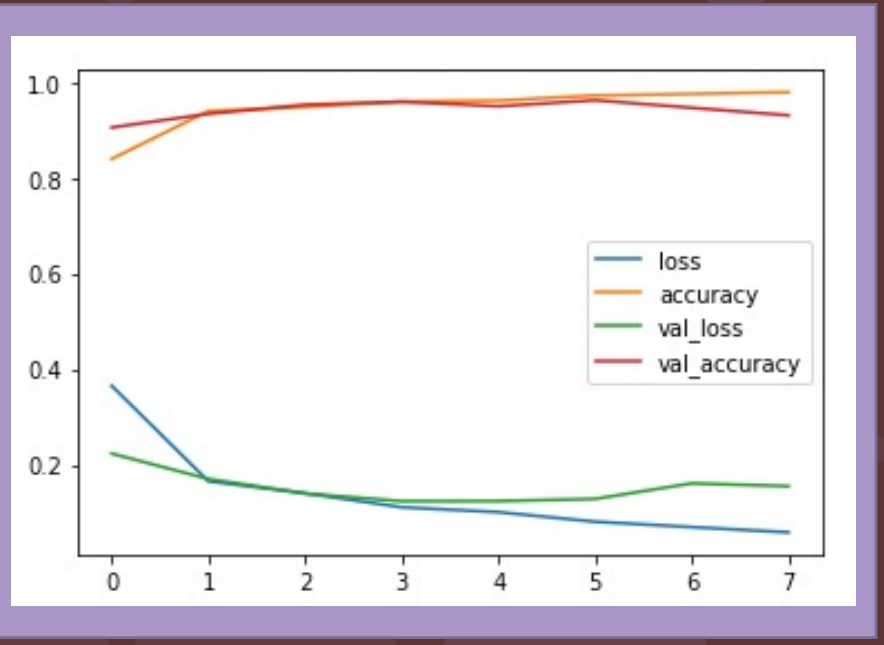
MODEL 2 Best Accuracy with good Precision Balance







Model Performance




```
*****  Classification Report  *****
              precision    recall  f1-score   support

      NORMAL       0.95      0.97      0.96       158
   PNEUMONIA       0.97      0.95      0.96       158

    accuracy                           0.96       316
   macro avg       0.96      0.96      0.96       316
weighted avg       0.96      0.96      0.96       316

*****  Metric Scores  *****
Accuracy : 0.958861
Precision: 0.967742
Recall   : 0.949367
F1 score : 0.958466
```

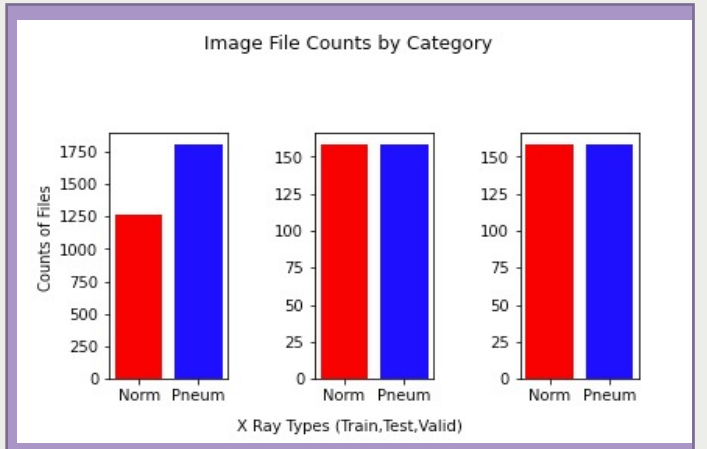
Model 2 chosen as it typically
Performed better for all parameters

Models for all multiple file size inputs
Best result was when pneumonia
Files were 40% more than normal files

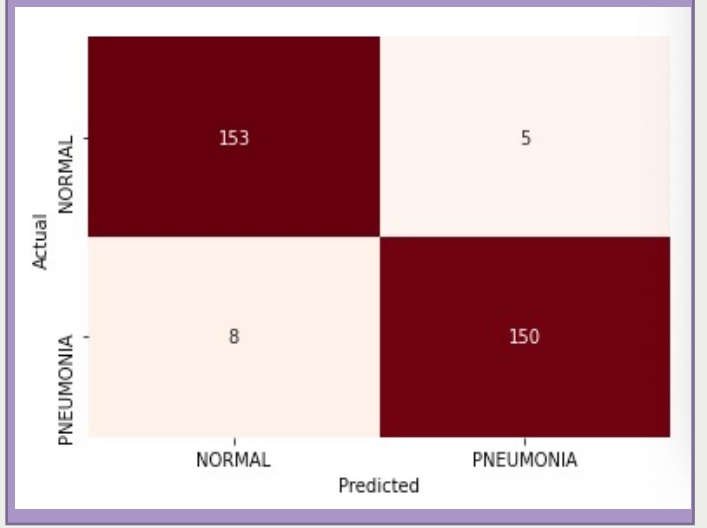
Model stopped at 7 Epochs
Accuracy was 95.8%

Based on these results, we can be
At least 95% confident that a
Predicted value is correct, given
Any file applied and prediction
Given

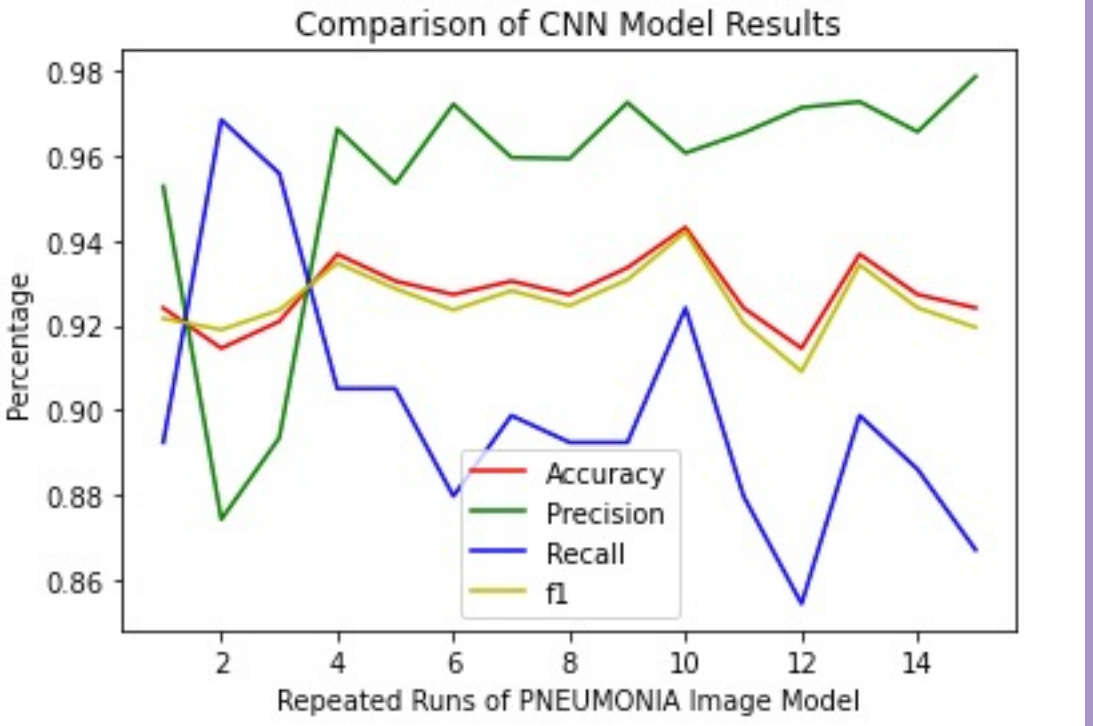
This model accurately predicted
pneumonia 150 times, and incorrectly
Predicted it as normal 5 times.




```
*****  Confusion Matrix  *****
[[153   5]
 [  8 150]]
```



Model Parameter Tuning



Accuracy Range (Max,Min) :	0.94 / 0.91	Range:	0.03	Average:	0.93
Precision Range (Max,Min):	0.98 / 0.87	Range:	0.10	Average:	0.95
Recall Range (Max,Min) :	0.97 / 0.85	Range:	0.11	Average:	0.90
F1 Range (Max,Min) :	0.94 / 0.91				

The chosen model was run completely through model creation to observe the potential variability in successive runs. From the results of 10 runs, it is shown that the accuracy maintained above 91% and the precision and recall had a range of 0.1%.

This variability challenges that model 2 with 40% more data is the best choice, but that model 2 applied with between 20% to 50% data (or at 70%) would product similar results as the predictions were within the error

What was learned from the process

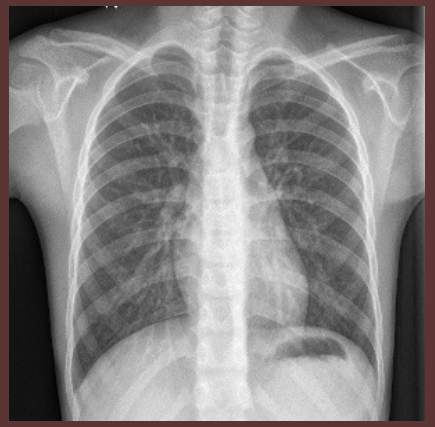
First model runs: no better than 70% - (POOR)
- Adjusted alpha/beta to see 4% accuracy improvement (when 1.0,1.0) X-rays dark and contrasty
- Initially included all files, but large False Negative rate – biased results
- Initially only files as found (RGB) – best model results at this point 78%, changed to grayscale- improvement

Using Grayscale: results above 90% for first time (BETTER)
- Always applied the three models in for each run – Variability
- Did see model 1 sometimes fail completely (No true NORMAL predictions)
- Bias toward PNEUMONIA prediction because more files

Adjusted number of files (lowered file count to make pneumonia/normal closer) (BEST RESULTS)
- Model 2 always outperformed the simpler Model 1, and usually outperformed more complex Model 3
- Continued to see variability in results (up to 3% on accuracy between similar runs)
- Model susceptible to error due to random file placements each run and random file discards of pneumonia
- Variability of final results indicate that range of file inclusions between 20% and 50% more pneu files no effect

Recommendations

Normal



Pneumonia

Model 2 at 40% more pneumonia files than normal files produced results at above 95% accuracy. The model did predict with level of confidence above 90% and met the objectives. This model can be used to accurately predict pneumonia from pediatric X-Ray images.

Based on the variability in successive model runs, it was determined that a potential range of accuracy as low as 91% was possible. Given the variability of future runs, these results still meet the objectives of above 90% accuracy