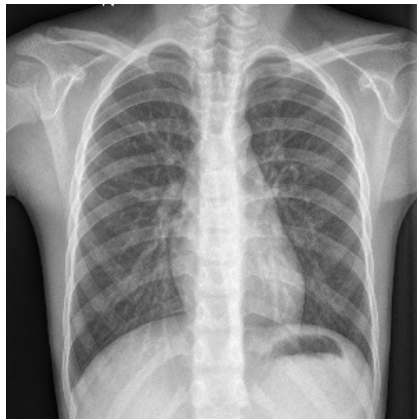# Detecting Pneumonia in X-Ray images using Convolution Neural Networks
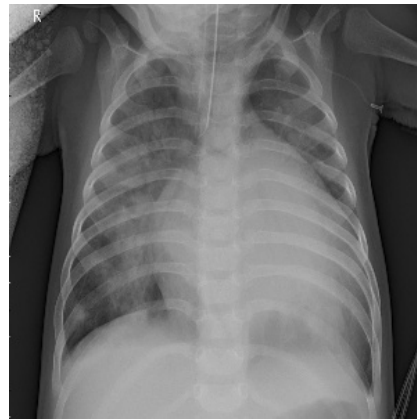
By Bob Spoonmore

Problem Statement

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%?

Normal   Pneumonia

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

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

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

Labeled data set of 5856 images in folders labeled pneumonia or normal: All images jpeg with various resolutions and proportions, no missing data: Final Train, Test, Valid split: 80%, 10%, 10%

Model Dataset:

|			
test	NORMAL	▸	157 files
	PNEUMONIA	▸	157 files
train	NORMAL	▸	1267 files
	PNEUMONIA	▸	1799 files
valid	NORMAL	▸	157 files
	PNEUMONIA	▸	157 files

Unbalanced number of files:
- **Pics: 1583 Normal, 4274 pneumonia**
- **Small number of files lower in resolution**
- **Most image ratios within consistent spread**
- **Copied files into modeling folders**
    - **Test: 10% of data, equal # norm, pneu**
    - **Train: 80% of data, rest of norm, % above pneu**
    - **Valid: 10% of data, equal # norm, pneu**
- **To minimize bias, set number of images for each category to be similar. Original data set very overloaded for pneumonia images**

Image File Counts by Category
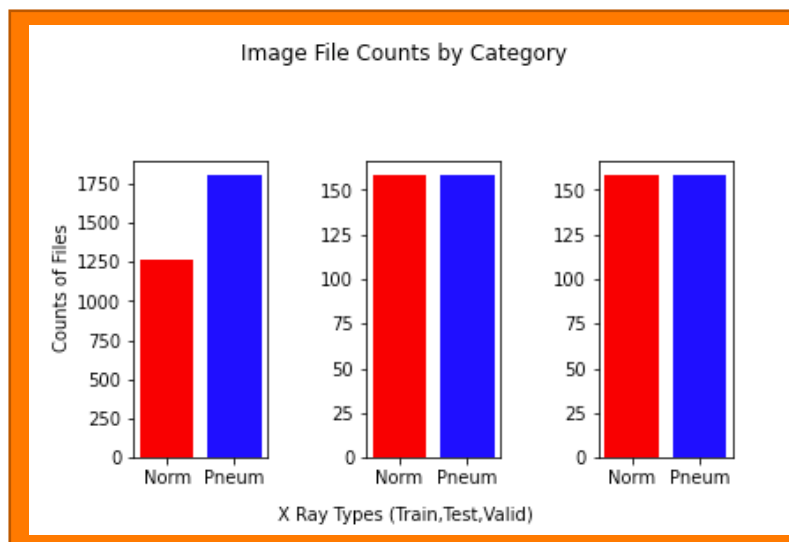
Exploratory Data Analysis

Analysis of the image files indicates a similarity in proportions, with a small number of files low in resolution. All files below
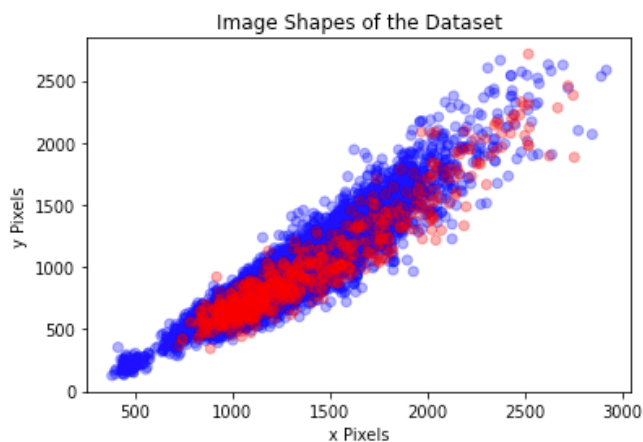
Read all images and compared sizes
Image Shapes graph:
        Red-NORMAL, Blue-PNEUMONIA

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, 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

All images were color with values for RGB. For improved effectiveness of the model, the images will be converted to one dimension as grayscale.

Data Preparation for Modeling

Preprocess to make images similar and comparable

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
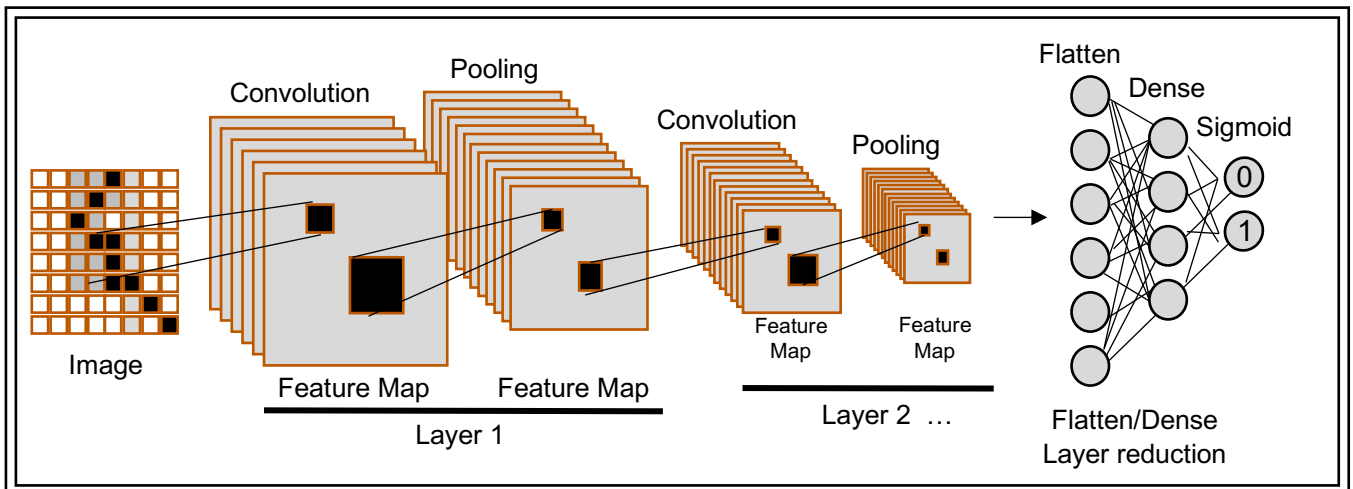

```
Model: "sequential_1"

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
```

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)

Set default number of epochs to 25, but used valid dataset for early stopping

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 min in accuracy, next epoch prevented)

            - patience = 2 (number of epochs min_delta is applied)

Model Selection

Three different Models applied:

M1: 1 layer (simple)

M2: 2 layer (typical approach in complexity)

M3: 5 layer (added complexity to see if feature identification improvement)

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
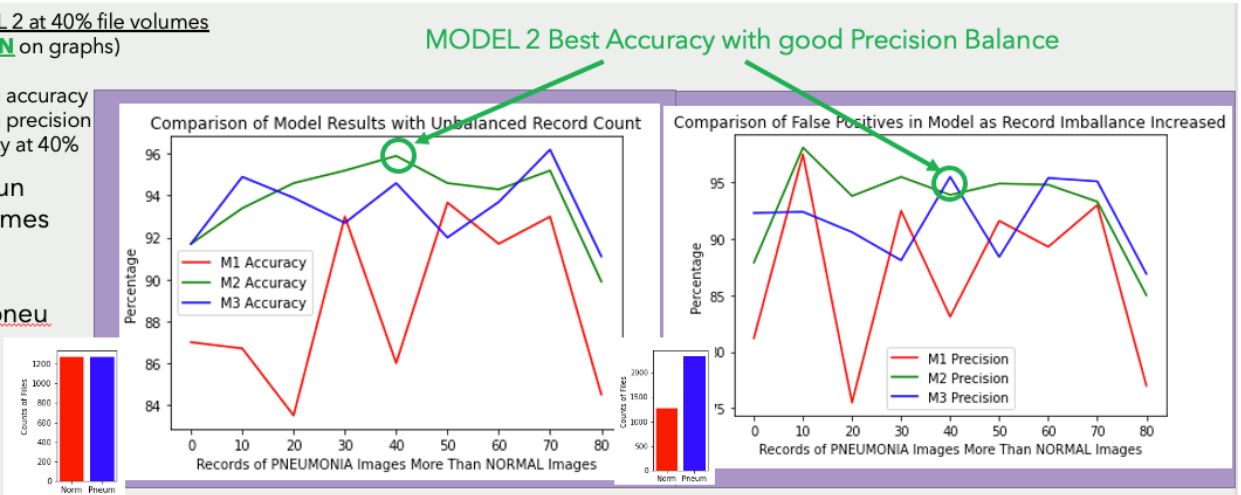
First model runs: Accuracy no better than 70% - (POOR)

   - 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

Predictions and Validation

Model 2 chosen:

Performed better for all parameters

Best for dataset with 40% more pneumonia files than normal files

<u>Model stopped at 7 Epochs</u>

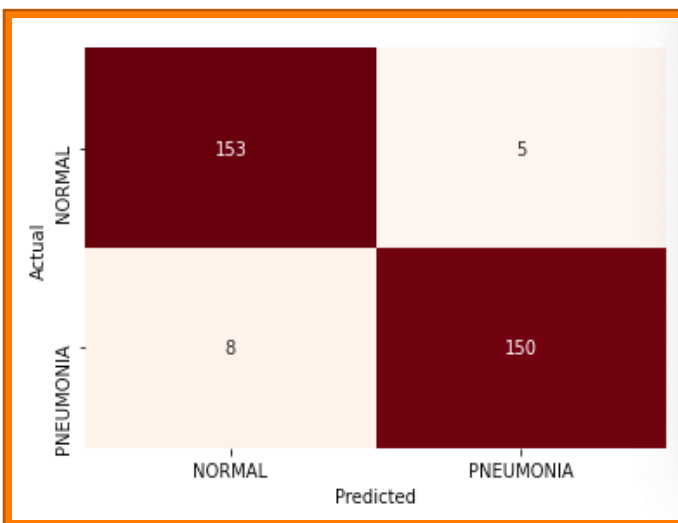Accuracy : 95.8%
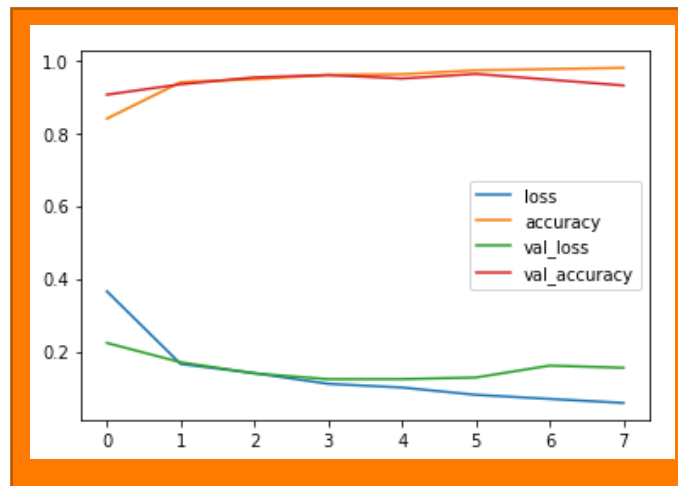
Precision : 96.8%

Recall   : 94.9%

F1 score : 95.8%

This model accurately predicted pneumonia 150 times, incorrectly

Predicted it as normal only 8 times.

This model accurately predicted normal 153 times, incorrectly predicted it as pneumonia only 5 times.

Hyperparameter Selection

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. Optimal values chosen over 15 runs, did not apply Gridsearch due to limitations on CPU of machine.
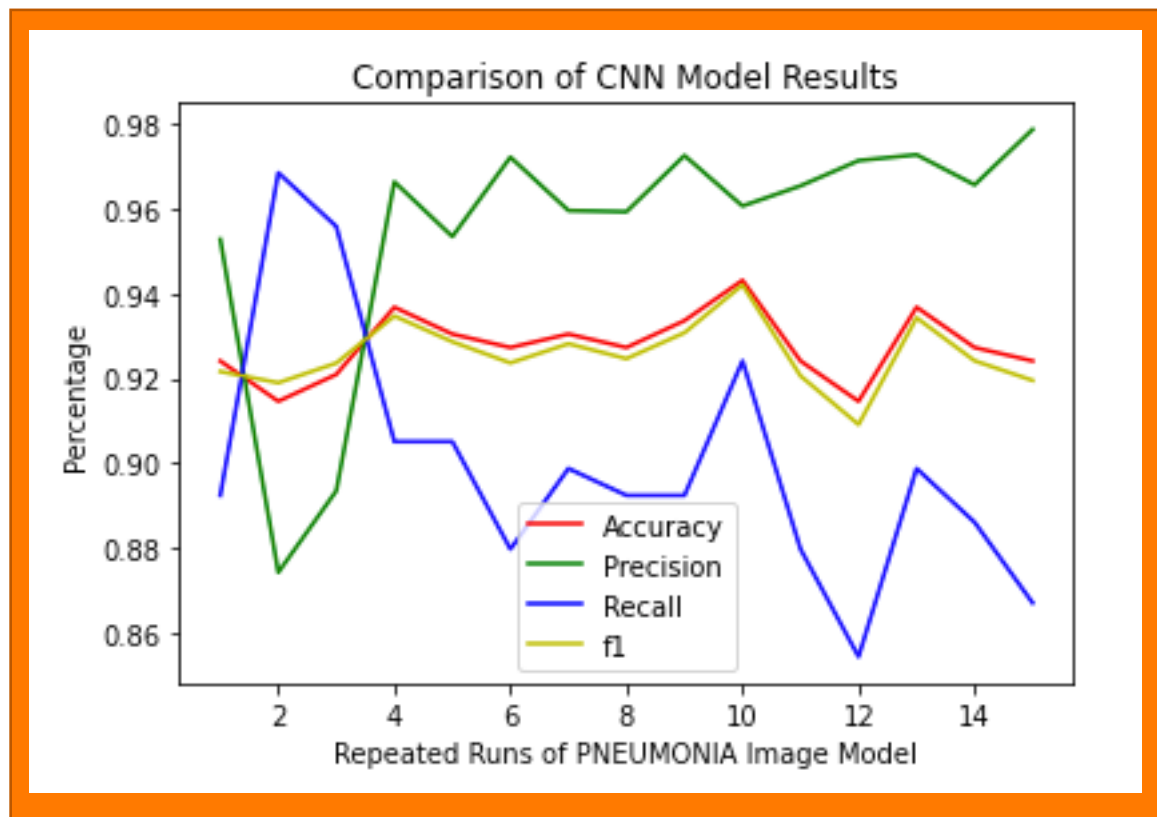


```
Accuracy Range (Max,Min)  :  0.94 / 0.91    Range:  0.03  Average:  0.93
Precision Range (Max,Min):   0.98 / 0.87    Range:  0.10  Average:  0.95
Recall Range (Max,Min)    :  0.97 / 0.85    Range:  0.11  Average:  0.90
F1 Range (Max,Min)        :  0.94 / 0.91
```

Model Results

The image evaluation model produced accurate results with 94.7% accuracy

The model met the objectives and did predict with a level of confidence above 90%. The model was run 10 consecutive times on the same dataset, it was shown that a potential range of accuracy of 93.6% to 95.8% was possible. The average accuracy over these 10 runs was 94.7% with a standard deviation of 0.8%. This model had a balanced level of precision and recall to avoid bias towards normal or pneumonia.
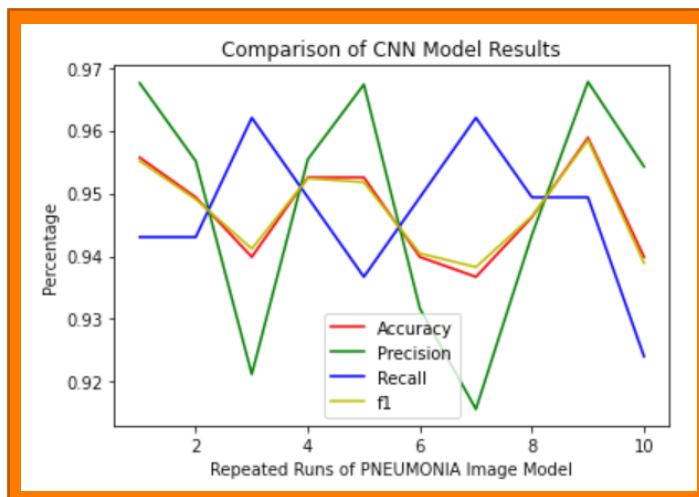

```
*****  Classification Report  *****
               precision    recall  f1-score   support

      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
```

Recommendations

The image evaluation model produced accurate results with 94.7% accuracy and can be used as a method for detection for similar X Ray images.

Further improvements can be accomplished through optimizing the model with Gridsearch running on a more powerful system. Future optimization improvements could improve the accuracy above 95%.