Pediatric Pneumonia Classification

By Aaron Lee

The Problem

PNEUMONIA CASES DIAGNOSED AS MIMICS:
Upper respiratory infections
Influenza
Cold
Pleurisy
Sinus infection
Seasonal viruses

SERIOUS CONDITIONS DIAGNOSED AS PNEUMONIA
Acute respiratory failure
COPD, Bronchitis
Lung Cancer
Myocarditis / pericarditis
Pulmonary Embolism, Heart attack
Bronchitis
Legionnaire's, Tuberculosis
Sepsis, Septicimia

PNEUMONIA (ALL AGES)
3.3% Mortality Rate
No formal consensus guideline
1.1 million cases per year
74% Sensitivity (Feldman Et al..)

Increased Mortality
Delay in Treatment
Overuse of antibiotics
Increased Cost
Multiple Visits
Increased Malpractice

https://www.sciencedaily.com/releases/2010/10/101022123749.html.
http://www.rimed.org/rimedicaljournal/2014/08/2014-08-20-cont-maughan.pdf
https://pneumonia.biomedcentral.com/articles/10.1186/s41479-016-0002-1
https://pneumonia.biomedcentral.com/articles/10.15172/pneu.2014.5/464 (feldman)

Our Project

Create a user friendly Machine Learning model that, when provided a frontal chest pediatric x-ray image, can accurately predict whether a child has pneumonia to assist with clinical diagnosis.

Project Goals

1. Build and train a machine learning model which maximizes prediction accuracy.
2. Make the model accessible and easy to use.

Our Data

Our model is trained using a dataset from Kaggle[1] is taken adapted from the original images at Mendeley Data[2].

Dataset features

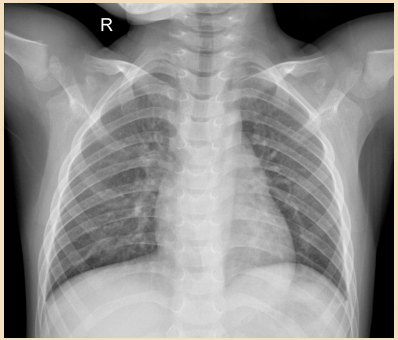
1. 5800 pediatric frontal chest x-rays (black and white medical quality images)
2. Labeled pneumonia 75%, normal images 25%
3. 1200x1200 images or larger with varying aspect ratio
4. Children from infants to adolescents
5. Contains both bacterial and viral pneumonia

1. Kaggle dataset at https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
2. Mendely Dataset at https://data.mendeley.com/datasets/

Image Classification

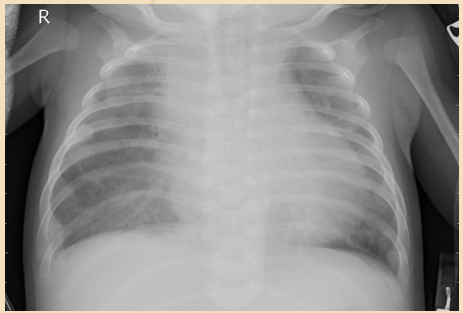
Infant

Adolescent

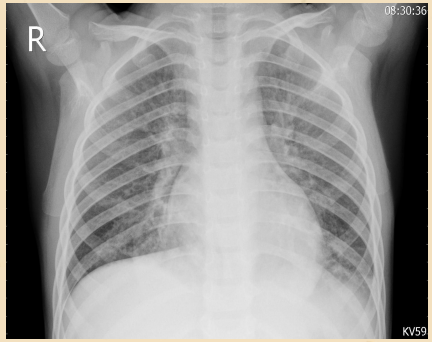


Normal

Bacteria



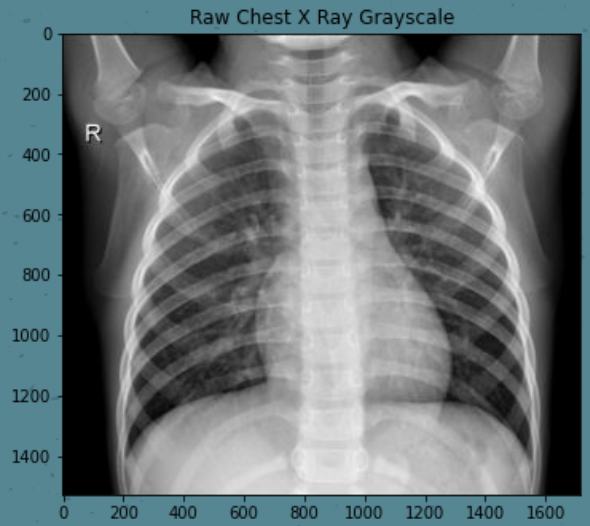
Virus



Pneumonia[3]

3. British Medical Journal “Differentiation of bacterial and viral pneumonia in children”: https://thorax.bmj.com/content/57/5/438

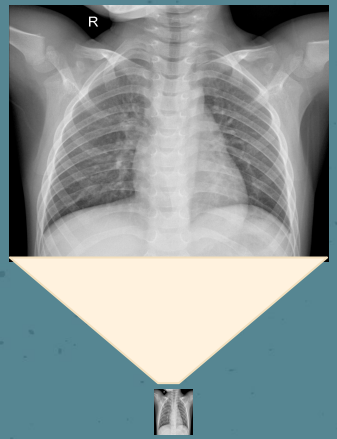
Preprocessing our data



To prepare our images for training:

- Images are sorted by classification
- Images are randomized
- Split dataset: 80% images used for training, 20% for validation and testing
- Images are resized (loses >90% of original)

Original Image 1600x1400

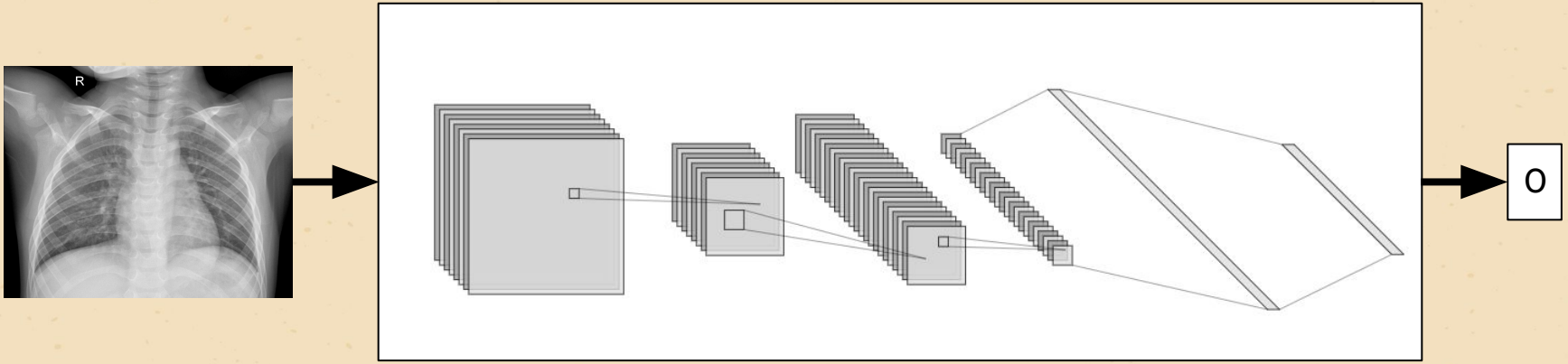


Reduced Image 150x150 (Loss of 98% of image)

Convolutional Neural Network (CNN)

A CNN was used to train this model

- Layers extract features of the image
- Model improves from training image feedback
- Trains over many iterations to improve the prediction accuracy

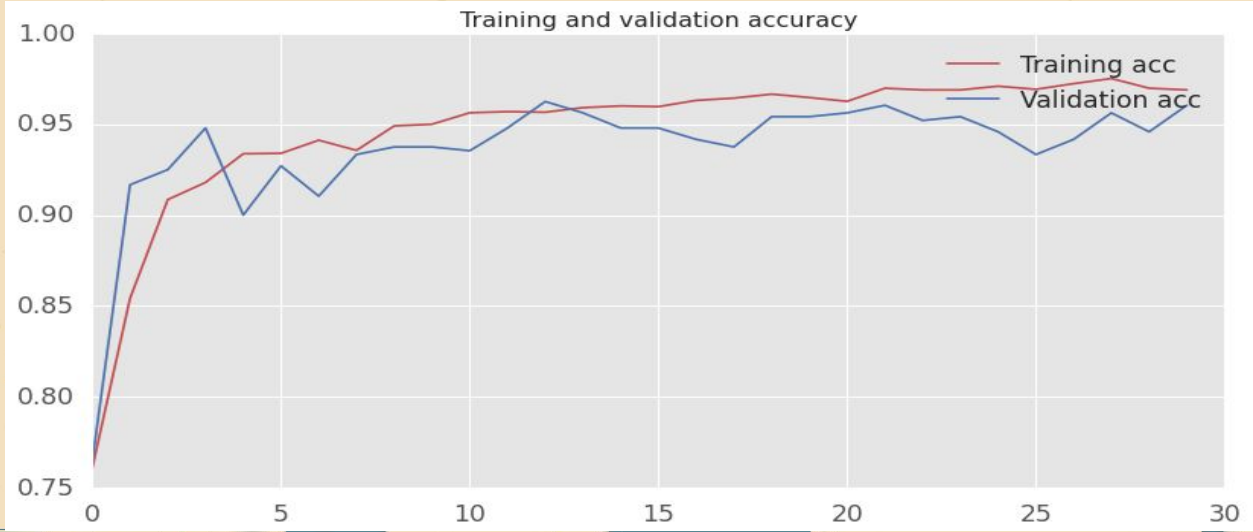


https://alexlenail.me/NN-SVG/LeNet.html

Convolutional Neural Network (CNN)

A CNN was used to train this model

- Layers extract features of the image
- Model improves from training image feedback
- Trains over many iterations to improve the prediction accuracy

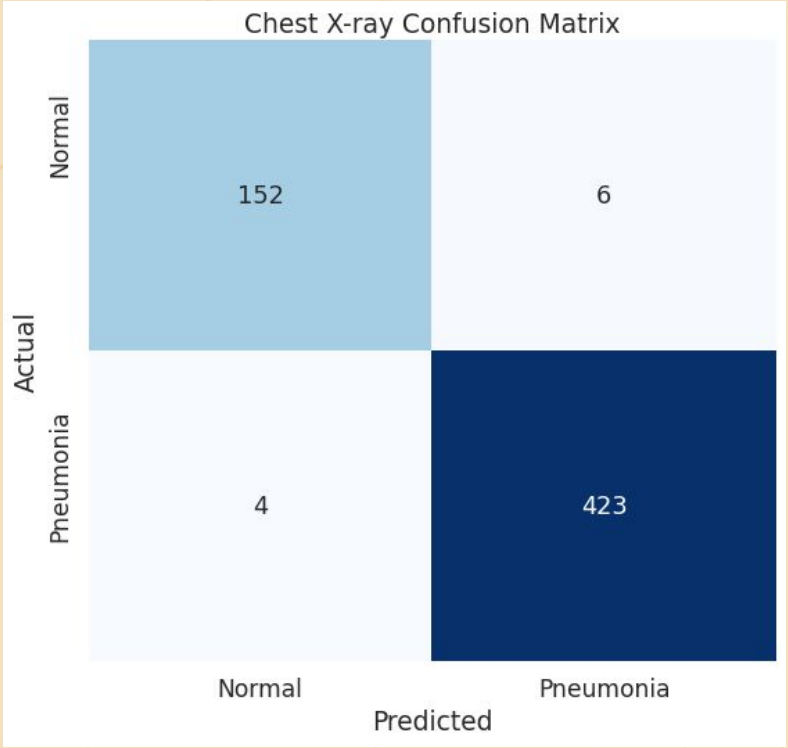


Our Model Results

Test accuracy **98.3%**

Precision* **99.1%**

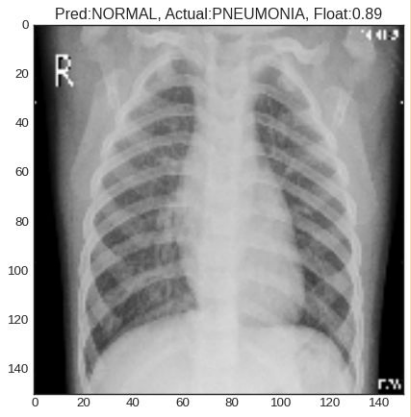
* Missed only 0.9% of positive cases
N test images = 585 (missed 10)

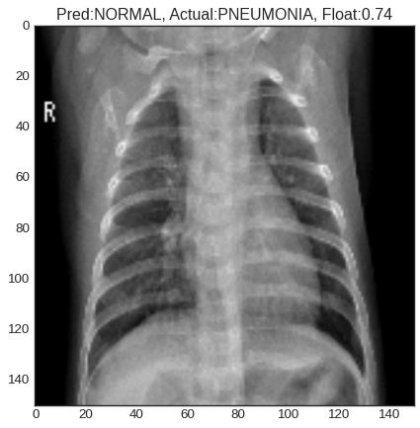


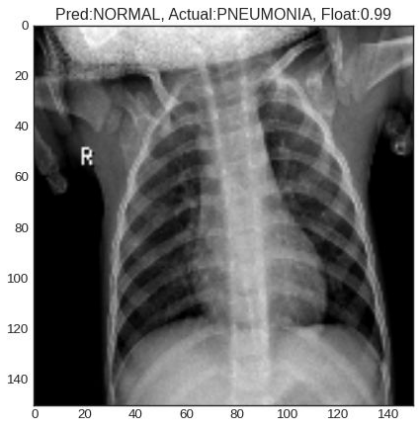
Missed images

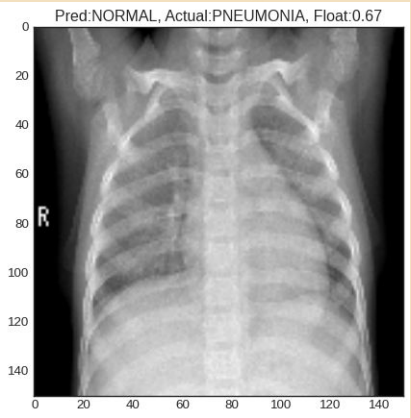
Characteristics of images miscategorized by the model.

- Mostly “Landscape” oriented (skinny lungs after resize)
- Significant black edges compared to most images
- Infants over represented
- Numerical uncertainty by model
- Blurry or low contrast images
- Most can be corrected by image preprocessing (crop)



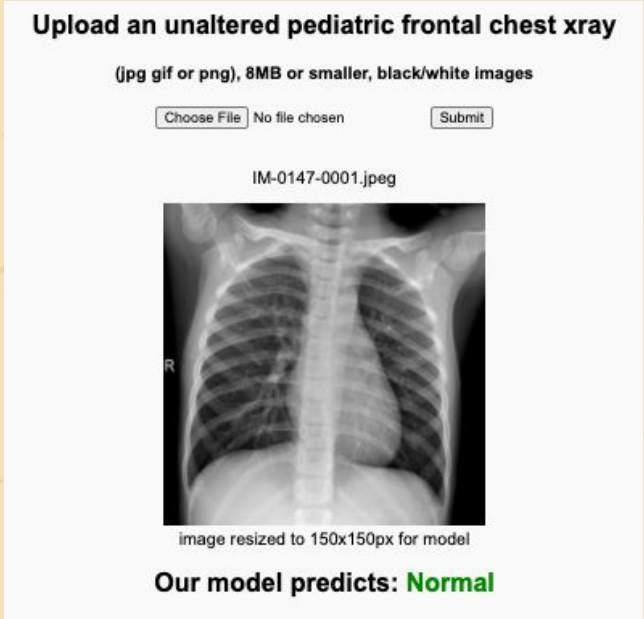


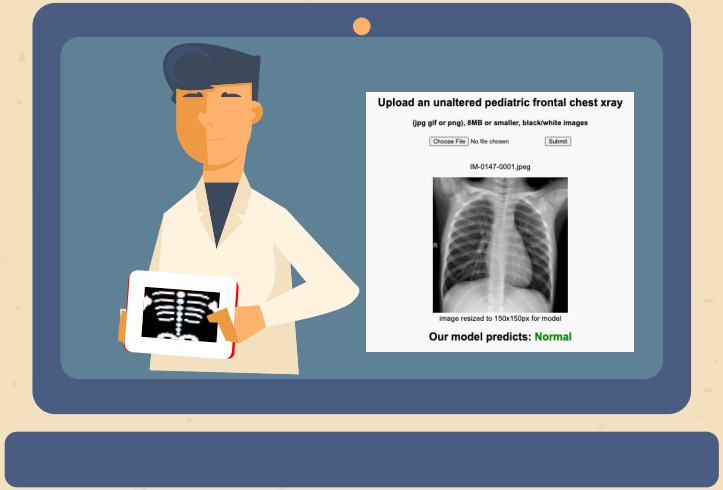




Web App of Model

The model is live
https://xray-pred.herokuapp.com/





Conclusion and Next Steps

Model can be used to assist in identification of pediatric pneumonia

Ways to improve the model

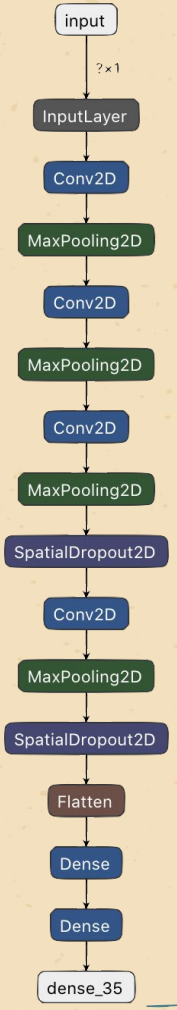
- Increased accuracy and precision
  - More training images
  - More layers to neural network
  - Tuning our model
  - Higher image quality used in model (currently 150x150)
  - More computer resources and time required
  - More advanced techniques (early model)
- Make different models for age groups (infants are different)
- Classify viral and bacterial pneumonia (perhaps COVID-19)

Questions?

Appendix (models)

Final Model
- CNN with Dropouts
- 3x3 kernel
- 3 Layers (Conv + MaxPooling + Dropout(20%)), relu
- Flatten > Dense 512 > Dense (sigmoid) (binary output)

Other Models considered
- Baseline model
  - Sequential 1D
  - 3 Dense Layers, relu
  - 92% acc
- CNN model (no dropout
  - 3x3 kernel
  - 3 Layers (Conv + MaxPooling)
  - Flatten > Dense 512> Dense 1 (sigmoid)



Appendix (resources)

Computing resources:

- 4 hours on Google Colab servers to run all models created.
- Final model 1.5 hours by itself on Google with GPU. (No GPU ~5hr)
- Larger image size would require Colab Pro services for extra RAM

Appendix (web app)

Hosted on Heroku (limits of free account)
Made with Python (Flask)
Uses Gunicorn WSGI
Uploaded files are held in session cookies, but never saved to file or database