CALIFORNIA SCIENCE & ENGINEERING FAIR
2019 PROJECT SUMMARY

Name(s)

Andrew Nazareth

Project Number

S0822

Project Title

Predicting the Presence of Pneumonia in Chest X-rays Using Deep Learning with Convolutional Neural Networks

Abstract

Objectives

Pneumonia is an infection that causes lung inflammation. In the US, 1 million people with pneumonia are hospitalized annually resulting in 50,000 deaths. A 2017 Stanford ChexNet study suggested that radiologists have a 95% accuracy in detecting pneumonia from chest X-rays. My goal is to train a Convolutional Neural Network (CNN) to meet or exceed this threshold.

Methods

- 8000 chest pre-classified (NORMAL, PNEUMONIA) X-rays from Kaggle.
- The resnet set (resnet34, resnet50) of CNN's from fastai pretrained on regular (non-medical) images,
- Linux hardware with a Nvidia GPU from Paperspace
- Software utilities: FastAI a framework for fast training CNN s, Python, Jupyter Notebook.

1. Pre-process the X--rays, randomly separating them into training (80%) and validation (20%) sets.
2. Select and train the resnet34 CNN to recognize X-rays that have pneumonia:
   - Measure the prediction accuracy of the pre-trained network,
   - Train the outer layers; re-measure the accuracy and loss rates
3. Improve the accuracy of the pre-trained model.
   - Identify a good learning rate.
   - Unfreeze the hidden layers and retrain the network.
4. Use input and test time data augmentation to improve the prediction accuracy
5. Repeat steps 2-4 to see if a deeper CNN s (resnet50) can provide better accuracy.
6. Validate results on random chest X-rays and correlate results with practicing radiologists.

Results

By training the outer layers only, I achieved a prediction accuracy of 95.6%. Using a graph of learning rate versus validation loss, I selected a learning rate of 0.05. With this learning rate, the prediction accuracy decreased marginally to 95.3%. With the addition of data augmentation and training the network for 3 epochs, the prediction accuracy increased to 97.1%. Furthermore, unfreezing the hidden layers and adding a differential learning rate yielded an accuracy of 98.1%.

Conclusions

CNN's can be used to predict the presence of pneumonia in a chest X-ray with > 98% accuracy. After tuning, false negatives were under 2% and false positives were 1%.

Summary Statement

I developed a Convolutional Neural Network that accurately predicts the pressence of pneumonia in chest X-rays.

Help Received

Erik Perkins is my project advisor at school. Dhar Rawal mentored me and provided me with the machine learning knowledge to help me to successfully undertake this project.