Writeup— Deep Learning

Abidemi

Abstract

The goal of this project is to build a classification model to identify chest-ray images with pneumonia. Doctors will be able to use the model to quickly confirm suspected cases of pneumonia from a chest-ray.

Design

Kaggle dataset that has chest x-rays (over 5k unique records) from patients with and without pneumonia. The original dataset was created in March 2018. The JPEG images have various sizes, but one of the largest images has a size of 1857x1317 pixels. The images are grayscale.

Data

The unbalanced dataset has over 5,000 unique images that is divided into into a training set, validation set, and test set. About 74% of the images in the training set belonged to the pneumonia class, but the validation set was split evenly among both classes. The JPEG images were then rescaled to a resolution of 256 x 256. The target feature were binary values (0 for normal class, and 1 for pneumonia class).

Algorithms

1. Logistic Regression model was built using Sklearn. A Simple NN model (hidden layer with 40 nodes, and another hidden layer with 10 nodes) and a CNN model (three convolution layers, multiple pooling layers, with a top section from the simple NN model) were built using TensorFlow and Keras. Recall was used as a metric to select the best model.

2. Dropout layers and number of epochs were tuned. Recall and computational cost were used to determine final model.

3. Heat maps were generated to determine how accurate the best model predicted the classes using test data.

Untuned CNN Model on Test Set:

Normal Recall: 0.46, Pneumonia Recall: 0.99, Accuracy: 0.79

Tuned CNN Model on Test Set:

Normal Recall: 0.65, Pneumonia Recall: 0.96, Accuracy: 0.85

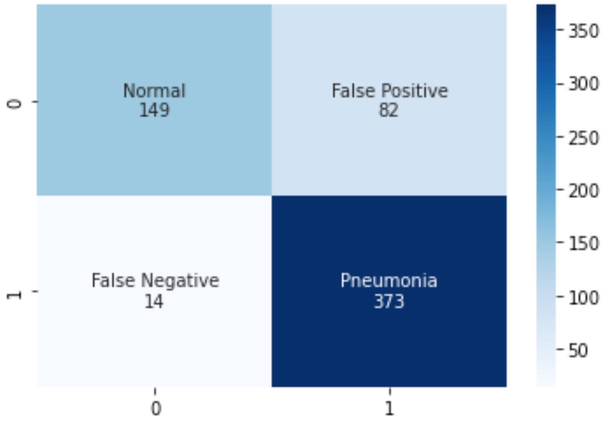

Figure: Heat map of Tuned CNN model on test data

Tools

Python libraries (Numpy, Pandas, Sklearn, TensorFlow,Keras, Mathplotlib, Seaborn).

Communication: Slides and presentation.