



Pneumonia Detection Model with Neural Networks

JAN ABDULLAH

Overview

In this project, we build a neural network model that can detect pneumonia in images of human chest X-rays.

Which metrics are we focused on most for a model in this type of domain?

Data

We download labeled chest X-ray images from Kaggle for our problem.

- There are ~5,800 images in total in the dataset.
- The labels are “NORMAL” and “PNEUMONIA”

Data Preprocessing

Steps for data preprocessing are laid out below:

1. Data directory construction.
2. Image reading.
3. Image augmentation.
4. Image reshaping.

Modeling

We iterate through different dense and convolutional neural network structures to find the optimal model.

The optimal model achieved in this project is a convolutional neural network with 1 convolving layer and 2 dense layers. See below:

```
Model: "sequential_13"
Layer (type)                 Output Shape              Param #
-----
conv2d_15 (Conv2D)           (None, 254, 254, 256)    7168
max_pooling2d_14 (MaxPooling (None, 127, 127, 256)    0
flatten_10 (Flatten)         (None, 4129024)          0
dense_51 (Dense)             (None, 25)               103225625
dense_52 (Dense)             (None, 50)               1300
dense_53 (Dense)             (None, 1)                51
-----
Total params: 103,234,144
Trainable params: 103,234,144
Non-trainable params: 0
```

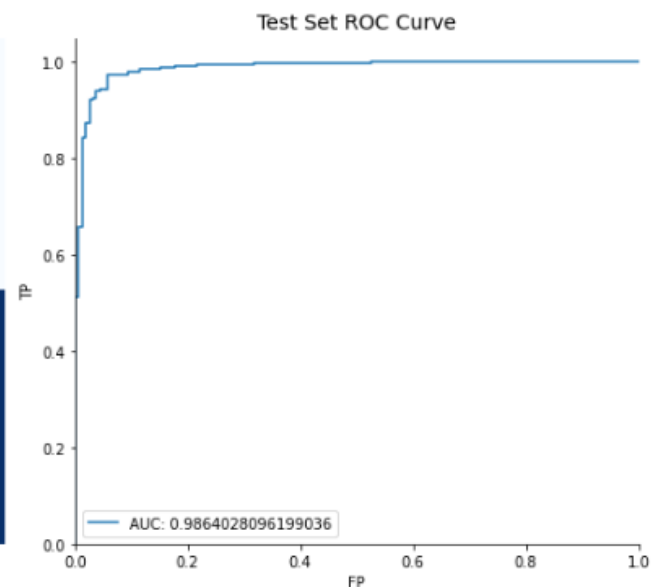
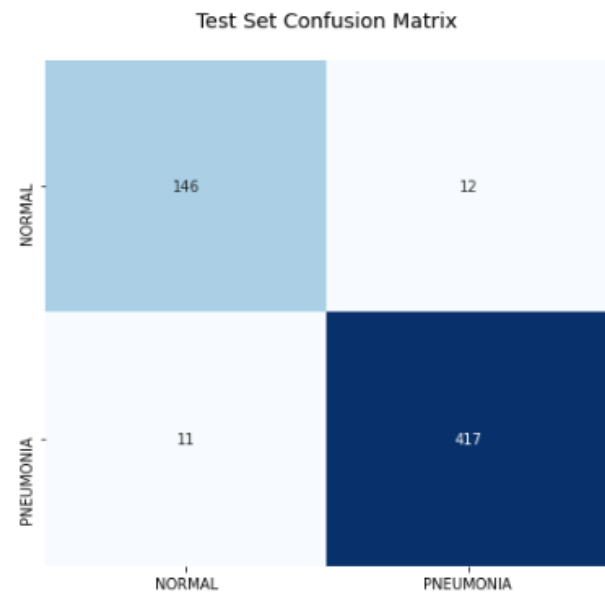
Evaluation

What other metrics should we look at other than the generic accuracy metric?

The false negative rate

```
=====  
=== Evaluating best model on Test set ===  
  
Classification Report  
              precision    recall  f1-score   support  
  
   NORMAL       0.93       0.92       0.93       158  
  PNEUMONIA     0.97       0.97       0.97       428  
  
   accuracy       0.96  
  macro avg       0.95       0.95       0.95       586  
 weighted avg       0.96       0.96       0.96       586
```

```
Loss: 0.11965231597423553  
Accuracy: 0.9607508778572083
```



Future Work

Biggest project bottleneck = processing power.

In the future, can try out (with better resources):

1. More varieties of complex networks and parameters.
2. Expand the project infrastructure by wrapping around it to create more robust and scalable tools.

Thank you!

