Pediatric Pneumonia Image Classification with Deep Learning

By: Andres Ramirez
Flatiron School - Full time Data Science Program

Business Problem

Goal

Build a model that can classify whether a given patient has pneumonia, given a chest x-ray image.

Accomplished using Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images

Context

Pediatric pneumonia is responsible for the deaths of more than 800,000 young children worldwide each year, according to the United Nations Children's Fund (UNICEF). [1] These deaths occur almost exclusively in children with underlying conditions, such as chronic lung disease of prematurity, congenital heart disease, and immunosuppression. Although most fatalities occur in developing countries, pneumonia (see the image below) remains a significant cause of morbidity in industrialized nations.

Challenges deep-dive

Challenge 1

- Alter the dataset in a way that will allow it to be usable and trainable
- Images can be large which leads to expensive computation!

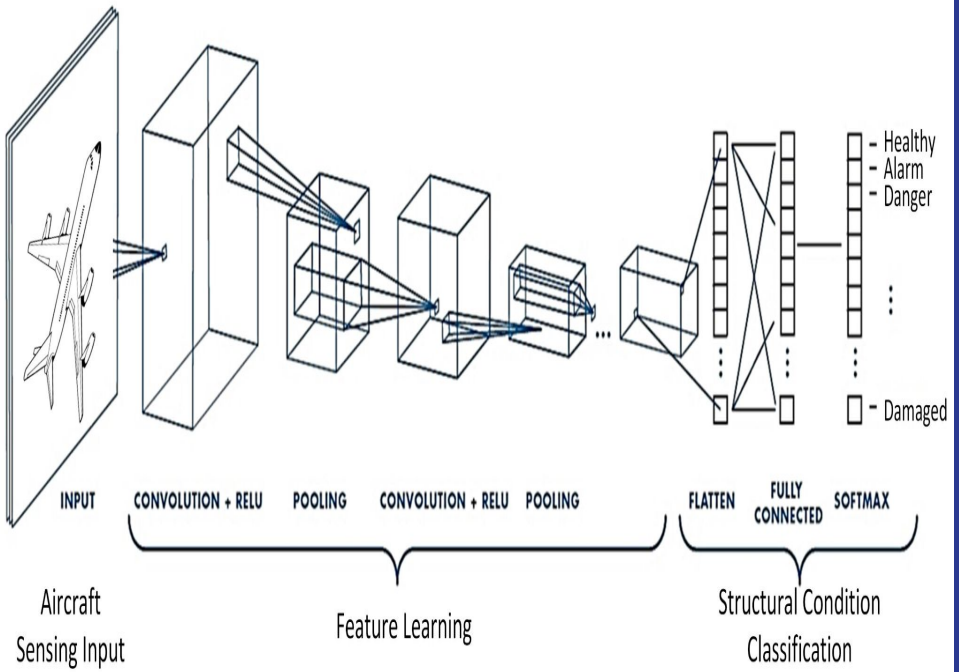
Challenge 2

- Choose the modalities that we wish to use in order to analyze our data.
- While there are many processes we can use it is important to fit them with our end goal

Challenge 3

- Obtain usable results with appropriate accuracy without overfitting our model

Solution



Utilizing Convolutional Neural Networks

- Assess Multiple models
- Check important parameters
  - Loss
  - Accuracy
  - Validation Loss
  - Validation Accuracy

Normal

Pneumonia

Our models tried to differentiate between the images to create patterns that help identify pneumonia.

Strategies were used to help solve class imbalance, which tried to address the challenge of having too many pneumonia x-rays compared to normal x-rays in our training set.

Results

Our model has 81.7% accuracy against our test set. This means that there were some instances when the patient had pneumonia but was diagnosed as healthy. It was helpful to grayscale and limit the epochs by using early stopping to prevent overfitting of the model.

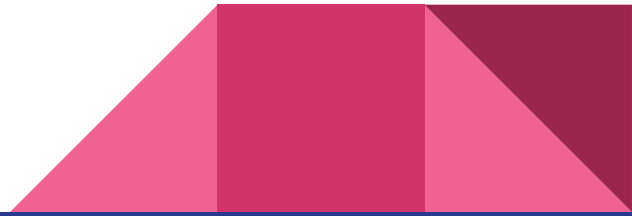
Future Work

It is important to consider that there may be a potential for human error as well, which would be seen as a misdiagnosis. Work with cross-validating with a team of radiologist and/or pulmonologist would be helpful in making sure that our images where labeled correctly.

Future Work

Future work would benefit from utilizing more data from populations such as adults and the elderly as well as from different parts of the world. This would allow us to generalize the data and increase the accuracy of diagnosis.

This model may be helpful as a primary screening test for patients. Since it has usable accuracy, this could flag a physician to examine the models determination and make the final confirmatory call on the diagnosis.



Thank You!

Questions and comments are appreciated

Andres Ramirez
Email: aj.ramirez23@gmail.com
Github: https://github.com/6cats1dog