

Assignment 3

Automated methods to detect pneumonia diseases from medical images.

Pneumonia is inflammatory condition of the lung and a quite common disease. There are two common pneumonias: 1) Bacterial pneumonia 2) Viral Pneumonia.

We want to predict whether an X-ray scan image shows presence of pneumonia or not. This was especially useful during pandemic as COVID-19 is known to cause pneumonia. The dataset is organized into 3 folders (train, test, validation) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal). The data is accessible in Kaggle dataset (<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>). You are going to attend in the competition with other persons in Machine Learning and Deep Learning community. There are several codes in the Kaggle to observe and learn more how to read, make model and train, test and validate the data to predict and detect the disease.



The goal of this project is to train a Convolutional Neural Network (CNN) model to be able to detect the healthy lung from the pneumonia infected one.

There are two main tasks:

- 1) Developing code to train the CNN model.
- 2) Writing a report describing and analyzing the model, regularization, optimization, etc. (1500 - 2000 words).

For task 1, we expected that you consider and decide about the following items:

- a) Using the training data to train the model with CNN which has the minimum loss and maximum accuracy for detecting the images with pneumonia.
- b) In the training choose:
 - Number of convolution-pooling building blocks,
 - The strides, padding and activation function that give you the maximum accuracy,
 - A solution to avoid overfitting problem in your code (regularization).
- c) Use validation dataset if you want to tune the (hyper)parameters, e.g., optimization in your model and use test dataset after the final tuning to obtain the maximum test accuracy.