BBM469 Data Intensive Applications Lab
Data Science Capstone Project Proposal

Group Number : 27
Group Member(s) : Okan Alan - Batuhan Mete
Project Title : Pneumonia Detection from X-Ray Images

1. The Problem

Pneumonia is an infection that inflames the air sacs in one or both lungs. The air sacs may fill with fluid or pus (purulent material), causing cough with phlegm or pus, fever, chills, and difficulty breathing. A variety of organisms, including bacteria, viruses and fungi, can cause pneumonia.

Our goal is detecting the pneumonia by the chest x-ray images to more precise and fast detection.

2. Data

Our dataset "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification" contains healthy, bacterial pneumonia and viral pneumonia x-ray images. The OCT Images are split into a training set and a testing set of independent patients. OCT Images are labeled as (disease)-(randomized patient ID)-(image number by this patient) and split into 4 directories: CNV, DME, DRUSEN, and NORMAL.

3. Methodology

Generally our goal is classifying the images, we will use classification methods such as CNN, KNN, Restnet and VGG. We will going to try and find the most suitable method for our problem.

We will start testing with KNN. If the KNN results won't satisfy us, we will continue with VGG.

4. References

https://www.who.int/news-room/fact-sheets/detail/pneumonia
https://data.mendeley.com/datasets/rscbjbr9sj/2
https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia