Capstone Project 2 - Proposal

CNN For Detecting Pneumonia from X-ray Images

1. Problem Statement

Pneumonia is a common infectious disease in the world. Globally, 450 million get infected by pneumonia in a year and 4 million people die from the disease. 1 million people each year have to seek care from hospitals and 50 thousand people die from the disease [1] in the United States of America. The numerical difference between the infection rates and death rates shows how crucial the early diagnosis of the disease is. Its main diagnostic method is chest x-ray examination. Analyzing and classifying chest x-rays can be very tedious for radiologists since x-rays are often affected by noise and require domain expertise and experience. Recently, a number of researchers have proposed different artificial intelligence (AI)-based solutions for different medical problems. Although currently, deep learning still cannot replace doctors/clinicians in medical diagnosis, it can provide support for experts in the medical domain in performing time-consuming works, such as examining chest radiographs for the signs of pneumonia.

2. Target Client

The proposed work will help doctors better predict pneumonia in minimal time with high efficiency. The aggregation of this will contribute to the health care system for better patient satisfaction and care.

3. Dataset

Data: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal).

4. Approach to solve the problem

I will develop a Convolutional Neural Network from scratch to classify the medical images. The CNNs will be implemented in Python on Google Colab using the Keras interface to Tensorflow. General workflow for the model will be:

- Getting Google Colab ready to use
- Building training and validation image data generators in Keras

- Compiling the model
- Running the model and plotting training and validation accuracy scores over each epoch
- Optimizing the input parameters such as number of CNN layers, the parameters of callbacks
- Applying regularization techniques such as dropout layers, batch normalization, data augmentation
- Evaluating the model on the hold-out dataset

Project Outcomes

Project outcomes will include a Google Colab notebook containing the associated code, a milestone report and a presentation.

[1] CDC Features. Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 6 Nov. 2017, www.cdc.gov/features/pneumonia/index.html.