# **Capstone Project 02-Proposal**

**CNN For Detecting Pneumonia from X-ray Images**

**1. Problem statement**

**Pneumonia** is an [inflammatory](https://en.wikipedia.org/wiki/Inflammation) condition of the [lung](https://en.wikipedia.org/wiki/Lung) primarily affecting the small air sacs known as [alveoli](https://en.wikipedia.org/wiki/Pulmonary_alveolus). Pneumonia is usually caused by [infection](https://en.wikipedia.org/wiki/Infection) with [viruses](https://en.wikipedia.org/wiki/Virus) or [bacteria](https://en.wikipedia.org/wiki/Bacteria), and less commonly by other [microorganisms](https://en.wikipedia.org/wiki/Microorganism). Pneumonia is a common illness affecting approximately 450 million people a year and occurring in all parts of the world. It is a major cause of death among all age groups resulting in 4 million deaths (7% of the world's total death) yearly. Rates are greatest in children less than five, and adults older than 75 years. It occurs about five times more frequently in the [developing world](https://en.wikipedia.org/wiki/Developing_world) than in the [developed world](https://en.wikipedia.org/wiki/Developed_world). Viral pneumonia accounts for about 200 million cases. In the United States, as of 2009, pneumonia is the 8th leading cause of death. (1)

Its main diagnostic method is chest x-ray examination. Analyzing and classifying chest x-rays can be 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 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**](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 and 2 categories (Pneumonia/Normal).

**4. Approach to solve the problem**

I will develop Convolutional Neural Network from scratch to classify the medical images. The CNNs will be implemented in Python on Google Collab using the Keras interface to TensorFlow. I will be using GCP/AI platform to build and deploy models. I will also use aws sage maker to train and deploy the models.

Workflow of the model:

* Getting Google Collab ready to use
* Build training and validation image data generators in Keras
* Compiling the model
* Run the model and plot training and validation accuracy scores over each epoch
* Optimizing the input parameters and apply regularization techniques as dropout layers, batch normalization, data augmentation
* Evaluate the model on the hold-out dataset
* Use GCP AI platform to train and deploy
* Use aws Sage maker to develop app

**Project Outcomes:**

Project outcomes will include a google Collab notebook containing associated code, reports and presentation in GitHub.

**References** :1. https://en.wikipedia.org/wiki/Pneumonia#Epidemiology