Architectural Components Overview

1.1 Data Source

- Technology Choice: we have used a dataset provided by kaggle.
- Justification : The data set provides us with ample numbers of images of both pneumonia patients and healthy lung x-ray, clearly divided into three groups- train, test and validation.

1.2 Data Quality Assessment

- Technology : Not performed
- Justification : data has been retrieved from a well formed Kaggle dataset, and does not require quality assessment.

1.3 Data Preparation

- Technology : numpy library has been used to transform and reshape the data
- Justification : to make the data into a form that can be fed into our model, we have to convert it into an array and reshape the dimensions of the images. Numpy library provides functions that allow us to do it efficiently.

1.4 Model/Algorithms Used

- Technology : Convolution Neural Network, and SVM are used.
- Justification : In this project, we are using x-ray of patients to determine if or not they suffer from pneumonia. For this we have to analyse the x-ray of patients to make a prediction, so it is an image classification problem. Convolution Neural Network is the de facto standard for image classification modeling in today's date. Hence, for deep learning model we have used CNN's for building our model. Another machine learning model that we have utilised is SVM, since it is a binary classification of whether or not a patient is healthy or not. We have utilised both LinearSVC and SVC with rbf kernel for modeling.

1.5 Framework

- Technology : Keras and Scikit Learn
- Justification : Keras is an open source deep learning API that can be easily scaled to a large cluster of GPUs. Scikit learn is another open source machine learning tool that is effective and easy to use.

1.6 Model Evaluation

- Technology : Accuracy and F1 Score is used

- Justification : Accuracy in CNN gives us how accurate our model is, based upon how many of predicted values are equal to the real values when applied to the test data set. F1 score gives us the weighted average of precision and recall which is a good indicator of how good our model is.