Project Title: Chest X-rays Classification for Disease Detection

Introduction

As a crucial diagnostic imaging tool for identifying abnormalities in the chest, chest radiographs / X-rays are used by health professionals. Chest X-rays are one of the most frequent and cost effective medical imaging examinations available. Chest X-rays can be used to detect many diseases, however, their functionality can be limited by challenges in interpretation, which may lead to wrong diagnosis of diseases. This in turn has an adverse effect on patients.

Objectives

The goal of this project is to develop an algorithm to aid in the classification of chest X-ray images based on the classes of diseases. The accurate classification of these images will help with the early detection of diseases and adequate timely diagnosis, in order to prepare patients for the right treatment, where needed.

Data

The dataset was acquired from the Kaggle official website and comprises 112,120 Chest X-ray images with disease labels from 30,805 unique patients, owned by the National Institute of Health. The dataset is suitable to train a deep network because of its size and intended purpose.There are 15 classes (14 diseases, and one for "No findings"). Images can be classified as "No findings" or one or more disease classes:

- Atelectasis
- Consolidation
- Infiltration
- Pneumothorax
- Edema
- Emphysema
- Fibrosis
- Effusion
- Pneumonia
- Pleural_thickening
- Cardiomegaly
- Nodule Mass
- Hernia

Proposed Analysis & References

To begin, standard MLP with various activiations in the hidden layers and softmax in the output layer will be used. As we progress, we may include backpropagation and various preprocessing techniques. Regularization techniques such as Dropout or BatchNormalization may be adopted to improve model performance

Keras and Sci-kit Learn will be mainly used. Keras contains well developed MLP libraries and it also allows flexibility in designing models for training, and Sci-kit Learn contains the preprocessing libraries we will use (random over sampler, k-fold validation, among others).

Reference materials to obtain sufficient background on applying the chosen network include Neural Network Design II textbook, class lectures, Keras official website and online searches.

The metrics that will be used to judge the performance of the network include **categorical cross-entropy** and **accuracy**.

Proposed timeline for completing project