I was originally going to use in my project the PAMAP2 Physical Activity Monitoring dataset: http://archive.ics.uci.edu/ml/datasets/PAMAP2+Physical+Activity+Monitoring. The dataset consists of measurements from inertial measurement units worn by 9 subjects while performing one of 18 different physical activities. Data from a heart rate monitor worn by each subject is collected.

I was planning to look at classification of activities and intensity of those activities over time. I was planning to use the following tools to complete the project: PCA, Keras, Tensorflow and pytorch. I was not going to use a GPU. Three classes of artificial neural networks that I was planning to use were CNN, RNN and feed-forward neural network.

Well laid out plans are open to change. I ran into an issue where the data would need to be converted from accelerometer data, but I don't have the knowledge to do that. I changed my dataset to use the Pulmonary Chest X-Ray Abnormalities dataset: https://www.kaggle.com/kmader/pulmonary-chest-xray-abnormalities. I used the images in the CXR_png folder from the file ChinaSet_AllFiles.zip which can be downloaded from that page.

I'm using CNN for classification. I am using the following tools to complete the project: Keras, Tensorflow and numpy. I am using a GPU running on an EC2 instance in AWS. My local machine did not have the memory needed to train the dataset. I am using a GPU on an AWS EC2 instance. I had trouble getting my code to run on my EC2 instance. It was supposed to have keras already installed but errored on import keras. I had to install keras and tensorflow. I tried to use AWS Sagemaker and read my image files from AWS S3 but couldn't figure out reading the files from S3.

I sought assistance in my code from the CNN demos we had in class and from the following site: https://github.com/harishanand95/cxr_classification. I used relu and softmax as activations and Adam as an optimizer. The hyperparameters are epochs = 15, batchsize = 2000. Pooling size was 4X4.

In conclusion, I found this project to be extremely challenging. I would start coding much sooner although I did start very early investigating a different dataset.