**PROPOSAL FOR FINAL CAPSTONE**

* **What is the problem you are attempting to solve?**

NIH has huge collection of 112.000 Chest X-ray images from 30,000 unique patients and need a way to accurately classify Chest X-rays using computers. They have labelled the images using NLP into 15 classes (Disease types). We need to develop a method or a model to classify these X-rays into 15 classes with minimum error rate.

* **How is your solution valuable?**

My solution is to use Neural networks to solve the problem of image classification since it has minimal over-fitting and high accuracy. Even more my solution will be using Convolutional Neural Networks (CNN) which is the best known, state-of-the-art technique known till now for image processing and classification.

* **What is your data source and how will you access it?**

The data source is from Kaggle website. They have provided unprocessed images of X-rays which I plan to access by downloading and categorizing with the help of dataframe and then use for neural networks.

* **What techniques from the course do you anticipate using?**

I anticipate using firstly, *image pre-processing* which is a key for learning since we need to present the data in a suitable way for the machine to understand. *Dataframe handling* to handle files to make them *flow into the network* is new technique for me. I plan to use *Keras, layered on Tensor Flow* which I learnt in the Unit 6. Will be using lot of *NumPy* handling of *arrays* to manipulate the network layers.

I will be using a new technique called *Data Augmentation* which will improve model efficiency. Planning to use this Keras ImageDataGenerator. *Fine tuning the hyperparameters* is another big technique, I will be using to improve the model accuracy. I shall try to see if *GridSearchCV* can be applied to find the best parameters since it is used for other supervised models. if I also plan if time permits, to use at least one of the *different flavors or architectures of CNN* (AlexNet, VGG, ResNet, Inception, DenseNet) apart from using Vanilla CNN.

* **What do you anticipate to be the biggest challenge you’ll face?**

The biggest challenge here I anticipate is the tuning of the Convolutional neural network to get maximum accuracy. I plan to overcome this challenge by getting a better understanding of the CNN also by reading papers and coming up with best parameters to tune.

Another challenge would be image-preprocessing i.e. initially getting the image files to be processed and augmented to be fed into the network.