Project: Deep learning models, image processing, and NLP in medical applications  
Prerequisites: Ability to develop ML models using python/R  and open source deep learning/image/NLP libraries.

Description: This project will focus on end-to-end medical applications of deep learning in cancer research. We will be considering 2 types of applications:

* Exploring Cancer Genomics/Imaging Data available in cBioPortal
* Using deep learning for translating and segmenting tissue images with *immunohistochemical (IHC) staining* based on MSK’s DeepLIIF project (<https://github.com/nadeemlab/DeepLIIF>)

**Project Approach**

In cancer research, machine learning can be used for:

* Identifying the types of cancer cells (clustering)
* Classifying the tumors as benign or malignant
* Predicting the survival based on the data about the cancer as well as doctors’ notes about the decease progression
* Identifying the key factors related to cancer identification and survival prediction

MSK team will provide Columbia students with references to a publicly available data sets that include genomic, radiology, and other types of data.

In this project, Columbia students will be working to answer advanced research questions in applying data science to cancer data:

- How genomic data can be used in predicting tumor types

- Given the built models, what are key factors that influence the decision or how can the model’s decision be explained?

At minimum, Columbia students need to present their analysis of the available data and demonstrate how machine learning model can be built with the available data to predict the survival or type of tumor.

Ideally, as a result of this project students would come up with an end-to-end ML solution and how it can support the medical professional in finding insights in multidimensional data. DeepLIIF project will be used as an example of using deep learning in cancer research.

This project will give students an opportunity to develop advanced ML and deep learning skills along with exposure to medical applications.

\*\*\*