In this project, I will be working on the [MIMIC III dataset](https://mimic.physionet.org/gettingstarted/access/), which is the largest public repository of ICU patient records. The data is a time-series data.

There are various analyses that can be done on this data. I choose to do the following-

1. Forecasting decompensation and length of stay:

I will build a neural network to forecast the status of a patient based on available records of their stay. This will help in predicting the length of stay of the patient in the ICU, which in turn can be used to manage the ICU admissions based on the beds that would be available. Even though this is in a healthcare domain, these models can be used in forecasting datasets in finance space.

1. Phenotyping Analysis:

A typical classification problem has one label associated with the data. But, in this case, I will be associating a number of diseases with the input patient data. This is similar to a correlation analysis among diseases, but the difference here is that this is done on the output, and not on the input. So the complexity of the output disease space and dependencies between diseases need to be understood. Also, since this is real world data, there may exist variability across the population.

Temporal convolution networks (e.g.: dilated convolution) have been used in EEG data, and will be explored for use here as well.

1. Understanding ML’s decision making process:

Since Machine Learning models are black box, there is a critical need to understand why a decision is made in order to trust these neural networks. Explainable AI is even more important in a field like healthcare, where the decision of a model needs to be justified. I will do a sensitivity analysis to understand the model. For example, I will change input parameters to notice differences in the output prediction and try to correlate them.