Below are three early ideas for potential capstone projects.

Breast Cancer Diagnosis

Problem:

This project would aim to develop an algorithm that can predict whether a diagnostic image of a cancer biopsy shows a malignant or benign tumor. Secondarily, it could also explore whether recurrence could be predicted based on image characteristics.

Data:

[Breast Cancer Wisconsin (Diagnostic) Data Set](http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29)

For each of 569 instances, the data set contains 30 attributes derived from a diagnostic image, an ID number and a diagnosis (malignant or benign).

A second data set is provided with 169 instances that specify whether the cancer recurred and after how long.

Arrhythmia Classification

Problem:

This project would aim to predict whether an electrocardiograph (ECG) recording is normal or arrhythmic, and possibly classify the recording based on the type of arrhythmia present. This type of program could be implemented into an ECG to give automatic feedback to the patient or health care practitioner.

Data:

[Arrhythmia](http://archive.ics.uci.edu/ml/datasets/Arrhythmia)

The data set includes values for 279 attributes derived from ECG signal. There is information for 452 different recordings, each one classified by type as either normal or arrhythmic and by type of arrhythmia.

Parkinson’s Progression Monitoring

Problem:

This project would aim to predict a measure of the severity of Parkinson’s from characteristics derived from a voice recording. This could aid in diagnosis and disease progression monitoring, especially for remote patients.

Data:

[Parkinsons Telemonitoring Data Set](https://archive.ics.uci.edu/ml/datasets/parkinsons+telemonitoring)

The dataset contains data from 5875 voice recordings. For each, UPDRS (a measure of Parkinson’s severity), age, sex, and several voice recording characteristics are included.