DeepHealth

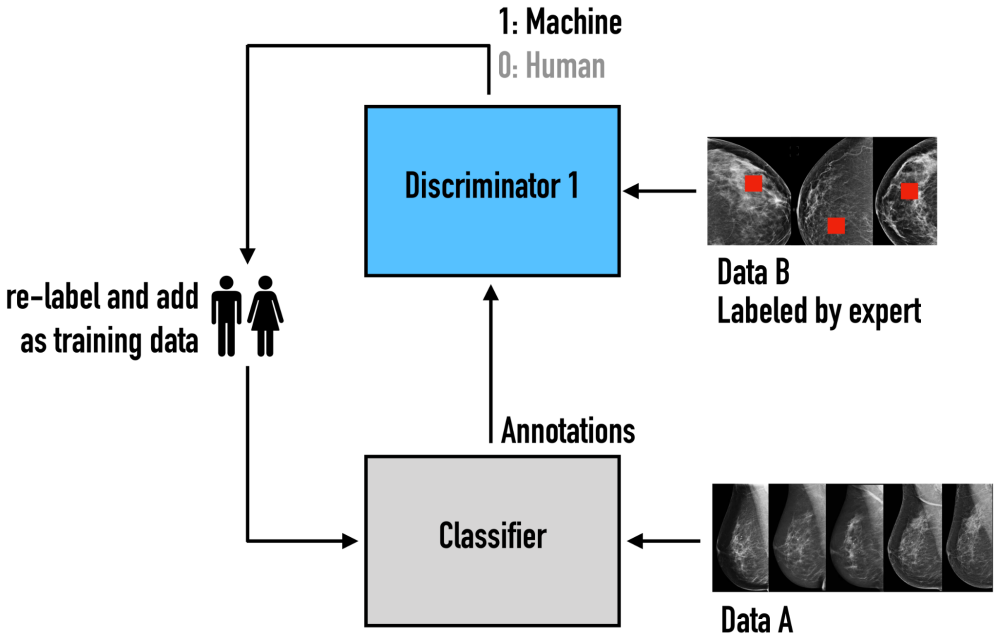
# Data Curation Software for the Oregon/Massachusetts Mammography Database

Supported by a recent research grant, DeepHealth works with researchers in the UMass Boston Computer Science department to create the world's largest annotated mammography database. With such a database, we can likely improve existing breast cancer detection algorithms. More information about the grant is available here.

This is a multi-faceted research project with many moving components and parts. There is active development in the field of **anomaly detection** where you can take part in developing a multi step algorithm to filter unwanted images from a massive dataset. This component focuses on using modern machine learning libraries and you will gain skills in python programming, using hyperparameter optimization frameworks, implementing experiment tracking, and you will get a hands-on understanding of the software development life cycle.

Another very crucial component of this project is the active development of a **classifier tuning framework** we call GP2. Those working on this will learn to use Python frameworks for machine learning such as TensorFlow and Keras. You will get to help build out and use an existing API which will give you training and experience that will be beneficial for real world projects/work, as you will first need to understand the code base that is already in place before you can start using/adding functionality. This component once finished will be able to tune any existing machine learning classifier, but for this particular project we are focused on fine tuning an existing breast cancer detection classifier called DeepSight. DeepSight has been FDA approved and already works well, but we will make it even BETTER! Also, you will get hands-on experience in reverse engineering an already developed and

packaged piece of software in order to understand how it works so that we can then re-tune it with our GP2 framework.
This two-network setup aims to fine-tune existing segmentation algorithms using a pair of artificial neural networks work together for quality control. Here is the a brief summary of this setup:




Yet another facet of this project is exploring and using other existing datasets such as those from Kaggle, where we plan on adding this data to our experiments to help build robustness and diversify our training and testing data. For this individuals will learn valuable skills in data exploration.

This project is a great opportunity for those involved to get hands-on experience in software development, feature engineering, research best practices, working with existing code bases, machine learning, reverse engineering, and most of all working on a team of diverse and motivated individuals who love to code and build things.

Resources
PyOD python library used for the outlier detection
kaggle Dataset