Problem

Worldwide, breast cancer is the second leading cause of cancer deaths, accounting for 11.6% of them. It was estimated that 6.6 million died of breast cancer in 2018.[1] Although the gold standard for breast imaging and breast cancer detection, approximately 75% of the breast cancer biopsies performed after recommendations from diagnostic mammography are benign.[2] This can lead to unwanted patient anxiety and makes it clear that specificity for diagnostic mammography still needs significant improvements. Artificial intelligence tools have been used to aid with detection and diagnosis, and, in particular, computer-aided diagnosis has gotten much interest and development. Most recently, deep learning techniques have been applied to diagnose lesions/areas of interest directly on mammography images.[3] This was not previously done because of " the scarcity of available data and lack of interpretability." In this project, we will apply the latest deep learning techniques to help improve the specificity of diagnostic mammography.

Who might care?

Clinicians, health care providers, and patients would benefit the most from improved diagnostic mammography. Correctly classifying breast cancer images would prevent unnecessary biopsies, a process that can be emotionally and physically draining.

Data

DDSM Mammography dataset
"The dataset consists of negative images from the DDSM dataset and positive images from the CBIS-DDSM dataset. The data was pre-processed to convert it into 299x299 images...The dataset contains 55,890 training examples, of which 14% are positive and the remaining 86% negative...."

Modeling approach

The goal of this project is to experiment with different approaches in deep learning. Transfer learning with the Keras library, along with Tensorflow backend, will be used to create our deep learning model. Data will be prepared so that it fits the required format for the keras (e.g.,

---

[1] "Global cancer statistics 2018: GLOBOCAN estimates of ... - NCBI." 12 Sep. 2018, https://www.ncbi.nlm.nih.gov/pubmed/30207593.
[2] "Screening for breast cancer: an update for the US Preventive ...." https://www.ncbi.nlm.nih.gov/pubmed/19920273.
[3] "Breast Mass Classification from Mammograms using Deep ...." 2 Dec. 2016, https://arxiv.org/abs/1612.00542.

one-hot encoding the labels, converting our grayscale/single-channel into rgb/3-channel images).

Systematic and random transformation to the images will also be performed. The systematic component (e.g., increase image brightness on all training images) will test whether changing a parameter increases or decreases model accuracy. Performing random transformations on our training images will test whether our model can generalize well/is robust.

Deliverables

- All code created for this project
- A brief scientific paper describing the experiment