# Cancer Diagnosis in Medical Imaging

## Problem statement

Computer vision in medical imaging has many promising applications in cancer diagnosis, and has been shown to even outperform some radiologists [1]. However, an outstanding problem preventing widespread adoption is the 'black-box' nature of these deep neural networks. While these algorithms achieve high accuracy, even a few misdiagnosed images can have catastrophic impacts on patients and hospitals. A burgeoning field of research is in increasing the 'explainability' of these classification models so medical professionals can understand why tumors are classified as benign or malignant. A couple of example ways to do this include:

**Saliency** Reverse-engineering what the models 'look at' when making predictions. This can be accomplished through following gradients back up a neural network to produce saliency maps.

**Generative Models** Unsupervised clustering of features to identify pertinent factors in classification. GANs or auto-encoders have been successful in uncovering relevant details for classification.

**Image-To-Text** (Challenging) Recent models [4] are capable of providing text descriptions of natural scene images. Imagine a model capable of reading out a diagnosis explanation for each image (ie. malignant because of web-like calcification in the top right corner).

**Your own** Find a novel way to increase the explainability of deep learning-based cancer diagnosis. Possibilities can include integrating new datasets or creating a Python package. Discuss with your TF before choosing a new direction. You will need to justify access to data, complexity, and expected hypotheses

## Project goals

1. Reproduce state-of-the-art neural network classifiers for cancer diagnosis, such as U-net [2] or ResNet [3].
2. Familiarize with existing medical imaging databases like DDSM or DREAM.
3. Analyze the potential tradeoffs between explainability and accuracy, and propose methods to counteract this trade off.
4. Create a significant contribution to the existing body of work. The final product can include introducing a useful Python package, Github repo, or presentation. (Optional, but recommended for AC209 students.)

## Data Recources

1. **DDSM Lung Nodule dataset**
   www.ncbi.nlm.nih.gov/pmc/articles/PMC3041807/
2. **DREAM Mammography Dataset**
   www.sagebase.org/in-the-news/digital-mammography-dream-challenge/

## References


1. Solving Cancer: The Use of Artificial Neural Networks in Cancer Diagnosis and Treatment
2. U-Net: Convolutional Networks for Biomedical Image Segmentation
3. Deep Residual Learning for Image Recognition
4. Towards Diverse and Natural Image Descriptions via a Conditional GAN