Breast Cancer Detection and Prediction with Machine and Deep Learning

# **David Kinney Spring 2021 https://github.com/dkinneyBU/breast-cancer**

# Which Domain?

Breast cancer detection and prediction, based on both tabular datasets and histopathology images.

**References**

“Building a Simple Machine Learning Model on Breast Cancer Data” vishabh goel

vishabh goel Sep 29, 2018 <https://towardsdatascience.com/building-a-simple-machine-learning-model-on-breast-cancer-data-eca4b3b99fa3>

“Breast Cancer Images Classification” Salah Sammari n.d. <https://www.kaggle.com/midouazerty/breast-cancer-images-classification>

“Breast Histopathology Images” Paul Mooney n.d. <https://www.kaggle.com/paultimothymooney/breast-histopathology-images>

“Breast Cancer Wisconsin Data Set” <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)>

# Which Data?

At this stage I have found two datasets to explore:

* Breast Cancer Wisconsin (Diagnostic) Data Set <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>
* Breast Histopathology Images https://www.kaggle.com/paultimothymooney/breast-histopathology-images

# Research Questions? Benefits? Why analyze these data?

How are you proposing to analyze this dataset? This is about your approach. Here, you’ll be proposing your research questions as well as justifications for why you’d offer these data in this way.

# What Method?

I plan to explore this project by leveraging both Machine Learning for prediction and Deep Learning for detection.

* Machine Learning – for this effort I will leverage the Breast Cancer Wisconsin dataset, initially applying exploratory data analysis (EDA) to get a feel for the data schema, and then various Machine Learning Classification algorithms to train a model to predict breast cancer detection.
* Deep Learning – while also a Classification application, this approach will be based on image analysis.

# Potential Issues?

At this stage I do not anticipate any major issues, given the wide availability of both data and research on this subject.

# Concluding Remarks

I originally planned to do my final project on predicting mortality rates based on how much someone abused their body during their lifetime—smoking, alcohol, obesity, etc. But much to my chagrin, finding relevant datasets became a lost cause. I had taken a day off from work and spent a fair amount of time in my (newly renovated) back yard, just thinking. Cancer in general is devastating, but for some reason, breast cancer has always struck a deeper emotional chord with me. While we humans hurtle forth and evolve towards gender equality, even so, on a visceral level I feel that a woman recovering from a mastectomy must be particularly devastating, when that part of a woman’s body speaks to so many things, perhaps the greatest being motherhood. Again, not a woman, so all I can do is imagine how I would feel in that situation.

And if I now switch over to the Data Scientist, there is a veritable plethora of data as well as activity on Kaggle and Towards Data Science regarding this subject. So, unlike my original idea, there is ample data available, as well as research to review.