

MAIS 202 FINAL PROJECT: DELIVERABLE 1

Breast Cancer Detection

Description

1. Dataset used: <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>

Why?

- a. letsfcancer
 - b. Recommended dataset to use (447 scholarly articles cite this dataset)
2. Methodology
 - a. *Data preprocessing*: Analyze features with histograms and heat maps. Since the data set is small (less than 600 rows), I do not need to use PCA. I will follow most of what the following URL recommends: <https://www.kaggle.com/kanncaa1/feature-selection-and-data-visualization>.
 - b. *Machine Learning model*: The objective of this project is to predict whether the cancer is benign or malignant. From my research, it seems that Random Forest, Decision tree and SVM gave good results of around 99% accuracy.
 - c. *Final conceptualization*: Simple web page where you manually enter features of a cell and it predicts whether it is malignant or benign.