Angelique Alexander

Problem Statement:

How can breast cancer be detected and classified with at least 95% accuracy as malignant vs benign based on the metrics of the tumors found in the breast?

Context:

Breast cancer is the most common cancer amongst women in the world. It accounts for 25% of all cancer cases, and affected over 2.1 Million people in 2015 alone. It starts when cells in the breast begin to grow out of control. These cells usually form tumors that can be seen via X-ray or felt as lumps in the breast area. The key challenge against its detection is how to classify tumors into malignant (cancerous) or benign(non cancerous). As someone who has a close friend recently diagnosed with Breast Cancer, I want to make detection and classification more accurate for patients to feel a higher confidence in their diagnosis.

Criteria for success:

Create an accurate model using machine learning that can predict if a tumor is benign or malignant based on different metrics about the tumor.

Scope of solution:

The scope of the solution encompasses various aspects of data collection, using comparative analysis based on a variety of metrics for tumor classification, showing the differences between these metrics for benign vs malignant tumors visually and using a subset of the data to create a model that can use these metrics to accurately predict if the tumor present is benign or malignant.

Constraints:

There is a possible lack of data but the preliminary explorations suggest there is an adequate feature amount.

Stakeholders to provide key insights:

The stakeholders that need to be involved in this project are patients, and health care providers.

Key Data Source:

The key pieces of data I'm using are from the kaggle website
https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data/data