

# Capstone 3 Proposal – Breast Cancer Prediction

Del

3/1/2021

Most types of breast cancer are easy to diagnose by microscopic analysis of a sample - or biopsy - of the affected area of the breast. The two most commonly used screening methods, physical examination of the breasts by a healthcare provider and mammography, can offer an approximate likelihood that a lump is cancer, and may also detect some other lesions, such as a simple cyst. When these examinations are inconclusive, a healthcare provider can remove a sample of the fluid in the lump for microscopic analysis (a procedure known as fine needle aspiration, or fine needle aspiration and cytology, FNAC) to help establish the diagnosis. A needle aspiration can be performed in a healthcare provider's office or clinic. Together, physical examination of the breasts, mammography, and FNAC can be used to diagnose breast cancer with a good degree of accuracy.

The features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. The question I want to answer is this: can we predict the type of cancer (benign or malignant) using these features with a 95% accuracy or greater?

This dataset was obtained from the University of Wisconsin Hospitals, Madison from [https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(diagnostic\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)).

Number of Attributes: 10 + output attribute

Attribute information:

Input variables:

1 - ID

2 – Clump\_Thickness

3 - Uniformity\_of\_Cell\_Size

4 - Uniformity\_of\_Cell\_Shape

5 - Marginal\_Adhesion

6 - Single\_Epithelial\_Cell\_Size

7 - Bare\_Nuclei

8 - Bland\_Chromatin

9 - Normal\_Nucleoli

10 - Mitoses

Output variable:

13 - Class (Benign or Malignant)

Missing Attribute Values: Bare\_Nuclei is missing 16 values.

**Sources & Citations:**

[https://en.wikipedia.org/wiki/Breast\\_cancer#Diagnosis](https://en.wikipedia.org/wiki/Breast_cancer#Diagnosis)