**Diagnostic Dataset**

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

a) radius (mean of distances from center to points on the perimeter)

b) texture (standard deviation of gray-scale values)

c) perimeter (of nucleus)

d) area (of nucleus)

e) smoothness (local variation in radius lengths)

f) compactness (perimeter^2 / area - 1.0)

g) concavity (severity of concave portions of the contour)

h) concave points (number of concave portions of the contour)

i) symmetry

j) fractal dimension ("coastline approximation" - 1) the fractal dimension, is an objective and reproducible measure of the complexity of the tissue architecture of the biopsy specimen. The higher the number, the more abnormal the tissue is.

(not sure if the units for some are pixel values or not)

The mean, standard error, and "worst" or largest (mean of the three

largest values) of these features were computed for each image,

resulting in 30 features. For instance, field 3 is Mean Radius, field

13 is Radius SE, field 23 is Worst Radius.

All feature values are recoded with four significant digits.

9. Class distribution: 357 benign, 212 malignant

**Original Dataset**

1. Clump Thickness: Benign cells tend to be group in a monolayer, while cancerous cells are often grouped in a multilayer.
2. Uniformity of Cell Size (🡪 3)
3. Uniformity of Cell Shape

Cancer cells tend to vary drastically in size and shape, thus a lower uniformity correlates with a higher possibility of cancer cells.

1. Marginal Adhesion: Cancer cells tend not to stick to one another as well as normal cells, so less adhesion correlates to a higher malignancy.
2. Single epithelial Cell Size: The size is related to uniformity, but enlarged epithelial cells may be malignant.
3. Bare Nuclei: It is an index of nuclei not surrounded by a cell, which is present in malignant tumors.
4. Bland Chromatin: Uniformity of “texture” appears in benign cells, while malignant tumors are typically coarse-textured. A lower number corresponds to more unity.
5. Normal Nucleoli: The rate of occurrence of normal nucleoli; abnormal nucleoli indicate possible mutated DNA, thus possible genetic expression for cancer reproduction. Thus, the smaller the rate of occurrence, the larger the chance of malignancy becomes.
6. Mitoses: Cancer cells tend to replicate faster which contributes to a tumor and leads to increased potential in harmful consequences. Thus, a set of cells with a higher rate of mitoses has an increased chance of being malignant.