**BREAST CANCER Dataset**

This is a dataset for classifying whether patient has got breast cancer (or not) from diagnostic data.

As it is part of your assignment to describe the dataset, only the following high-level information is given.

The data consist of features that have been computed from a digitised image of a fine needle aspirate (FNA) of a breast mass, and the diagnosis (M = Malignant, B= Benign). The features describe characteristics of the cell nuclei present in the image.

While you are not expected to understand the attributes in terms of their medical meanings they are described as follows:

1. ID

2-31. FNA features

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

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

c) perimeter

d) area

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)

32. Diagnosis

The mean, standard error and "worst" or largest (mean of the three largest values) of the FNA features were computed for each image, resulting in 30 features (2-31).

Note that this dataset contains missing data, which you would need to handle as part of the assignment.

It is a modified version of this dataset, available online:

<https://archive-beta.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+diagnostic>