**The Challenge:**

To build algorithms that computationally determine whether a patient’s tumor is metastatic or benign.

This challenge consists of two parts, from two separate datasets:

1. The first is a data table of features extracted from digitized images of patient cancer cell nuclei
2. The second is a set of raw images of patient cancer cells

The aim of the challenge in both cases is to be able to algorithmically predict the malignant/benign classification.

For both parts of the challenge individually:

* Report on success of the algorithm(s)
* Report any interesting observations
* Report the code used and build an annotated copy for review (all code must be written in Python or R)

Finally, compare and contrast the effectiveness of the two datasets in diagnosing patients.

**Dataset 1**

The data is a CSV file of tumor cell features from 569 patients.

Columns 1-2 are Patient ID and Malignant/Benign classification

Columns 3-12, 13-22, and 23-32 contain the means, standard errors, and largest values (respectively) of the following 10 features:

* Radius (mean of centre to perimeter)
* Texture (SD of gray-scale values)
* Perimeter
* Area
* Smoothness (variation in radius)
* Compactness (perimeter squared, divided by area - 1)
* Concavity (severity of concavity in contour)
* Concave points (number of concavities in contour)
* Symmetry
* Fractal dimension (coastline approximation – 1)

(e.g. column 3 is mean radius, column 13 is standard error of radius, column 23 is largest radius observed, etc)

**Dataset 2**

The data is a zip folder of GIF images of tumor cells from 20 patients.

The following diagnoses were recorded for the following patient IDs:

Benign:

92\_751

92\_5311

92\_5292

92\_5291

92\_5277

92\_5236

92\_4964

92\_4934

92\_4632

92\_4342

Malignant:

92\_7241

92\_6954

92\_6682

92\_6424

92\_6125

92\_5622

91\_9555

91\_6838

91\_6799

91\_5691