[research link](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.56.707&rep=rep1&type=pdf)

Breast Cancer Data Variable Research :

1) ID number: Different data points (patients)

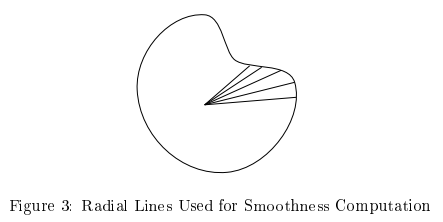
2) Diagnosis (M = malignant, B = benign): Dependent Variable

Ten real-valued features are computed for each cell nucleus:

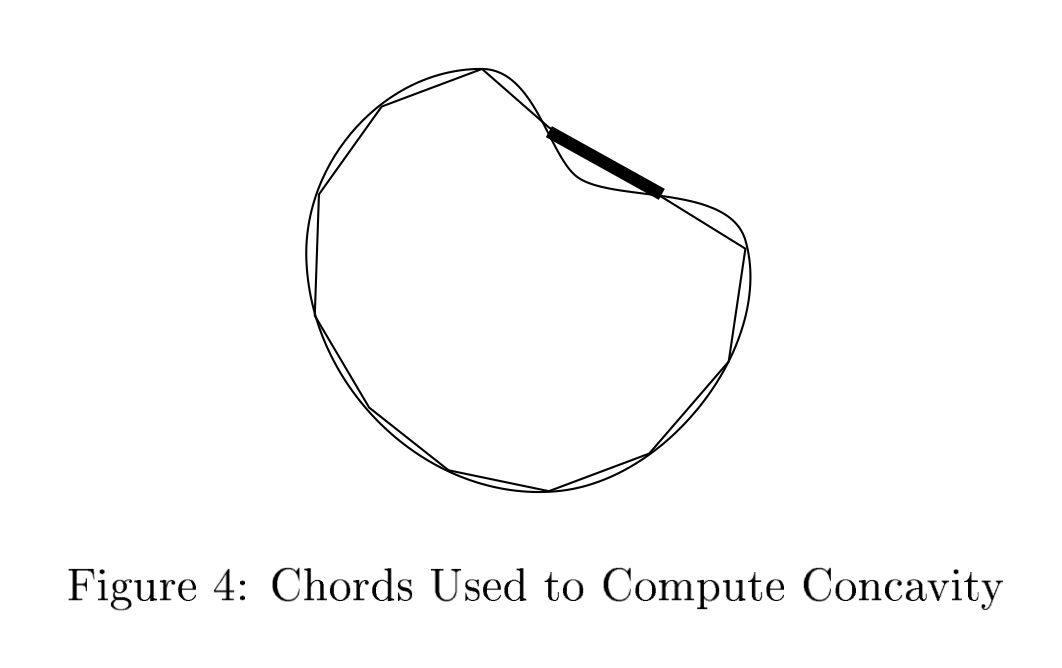
The dataset is research done on the nucleus of the breast cancer cells:

* Nucleus: The center of the cell that contains the cell's DNA. The nucleus of a cancer cell is usually abnormal.

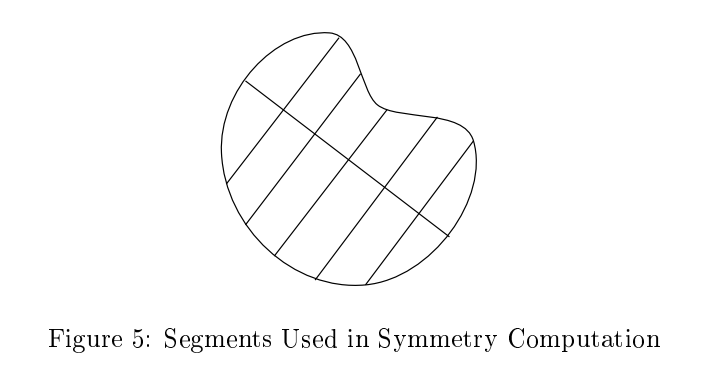
1. radius (mean of distances from the center to points on the perimeter): The radius of an individual nucleus is measured by averaging the length of the radial line segments defined by the **centroid** of the snake and the **individual snake points. (we can think as regular radius measurement)**
2. texture (standard deviation of gray-scale values): measured by finding the variance of the grayscale intensities in the component pixels. Generally, more homogenous texture corresponds to reduced risk while more coarseness denoted increased risk.
3. Perimeter: measured by the total distance between snake points.
4. Area(pixels number on the interior + 1.5 pixels in the perimeter): Nuclear area is measured by counting the number of pixels **on the interior** of the snake and adding one-half of the pixels **in the perimeter**.
5. Smoothness (local variation in radius lengths): measuring the difference between the length of a radial line and the mean length of the lines surrounding it



1. Compactness: Perimeter and area combined to give a measure of compactness of the cell nuclei using perimeter2/area
   1. Dimensionless number increases with irregularity of boundary
      1. Minimized by circular disk
   2. Measurement Bias
      1. Measure of shape increases for elongated cell nuclei, which doesn’t necessarily indicate increased likelihood of malignancy.
      2. Upward bias for small cells because of decreased accuracy imposed by its different shape features
2. Concavity: severity of concave portions of the contour
   1. Draw chords between non-adjacent snake points and measure the extent to which the actual boundary of the nucleus lies on the inside of each chord
   2. Parameter is greatly affected by the length of these chords, as snammer chords better capture small concaitivies.



1. concave points (number of concave portions of the contour)
   1. This feature is similar to Concavity but measures only the number, rather than the magnitude, of contour concavities. The concave points measure how many indentions are on the surface of the nucleus.
2. Symmetry: measured using the length difference between lines perpendicular to the major axis to the cell boundary in both directions



1. fractal dimension ("coastline approximation" - 1) :A [fractal](https://en.wikipedia.org/wiki/Fractal) dimension is an index for characterizing fractal (geometric) patterns or sets by quantifying their complexity as a ratio of the change in detail to the change in scale. A higher value corresponds to a less regular contour and thus to a higher probability of malignancy