

# Breast Cancer Datasets

- **Breast cancer** is the most common cancer among women in the world. It accounts for 25% of all cancer cases and affected over 2.1 Million people in 2015 alone. It starts when cells in the breast begin to grow out of control. These cells usually form tumors that can be seen via X-ray or felt as lumps in the breast area.
- **The key challenge** against its detection is how to classify tumors into malignant (cancerous) or benign(non-cancerous).
- **Attribute Information:**
  - 1) ID number
  - 2) Diagnosis (M = malignant (1), B = benign (0) )
- **Ten real-valued features are computed for each cell nucleus:**
  - a) radius (mean of distances from center to points on the perimeter)
  - b) texture (standard deviation of gray-scale values)
  - c) perimeter
  - d) area
  - e) smoothness (local variation in radius lengths)
  - f) compactness ( $\text{perimeter}^2 / \text{area} - 1.0$ )
  - 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)