

Data Dictionary

1. Id – unique id for each record
2. Diagnosis – whether Benign(B) or Malignant(M)

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

3. Radius - mean of distances from center to points on the perimeter
4. Texture - standard deviation of gray-scale values
5. Perimeter – distance around the nuclear border
6. Area – no. of pixels in the interior of the nuclear border & adding one half of the pixels on the perimeter
7. Smoothness - mean of the local variation in radius lengths
8. Compactness - $\text{perimeter}^2 / \text{area} - 1.0$
9. Concavity - severity of concave portions of the contour
10. Concave.points - number of concave portions of the contour
11. Symmetry - relative difference in length between line segments perpendicular to and on either side of the major axis.
12. Fractal_dimension - “coastline approximation” – nuclear border irregularity

NOTE:

Se – standard error

Worst – worst cell/largest cell

Mean – simple mean