**EENG 860-Assignment3-G5**

**Due date: March 3, 2023**

Consider the attached Breast Cancer Wisconsin dataset from the UCI machine learning repository:

https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29. The study of classification of 568 breast Cancers into two classes of Malignant(M) and Benine(B) based on the following 30 attributes. Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

**Attribute Information:**

1) ID number (Do not use for classification)

2) Diagnosis (M = malignant, B = benign) (Class label)

3-32)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 (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)

The mean, standard error, and "worst" or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features (features 3-32 in the attached file). For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius

Use 5-fold cross validation to evaluate the classification performance of a decision tree and a random forest classifier. Describe which classifier gives you the best performance. Provide the confusion matrix, sensitivity, specificity, total accuracy, F1-score, Roc curve, and area under curve.