**Predicting Breast Cancer**

**A.**

A random forest classifier was used to predict the malignant tumors. The data received from the scanned tumors were used for the perdition. After careful consideration the prediction model was very fairly successful. We used the area under the curve method for measuring the classifiers performance. The score from this model was 99% which is a good score but the accuracy was 86%.

**B**

1. The model is fairly good. The score for the model is detailed below:

Accuracy is 0.86

RandomForest AUC = 0.99

precision recall f1-score support

0 0.83 0.99 0.90 458

1 0.98 0.61 0.76 241

avg / total 0.88 0.86 0.85 699

2. 99 % of tumors predicted to be non-malignant were actually non-malignant. 83% of tumors predicted as non-malignant tumors were accurately predicted.

3. From the data in the table above, of all tumors that were malignant, only 61% were accurately predicted although 98 percent of the tumors prediction as malignant were actually malignant.

**C**

According to the model used the most important variables are the following (starting with the one with the highest importance)

1. Uniformity of cell size

2. Uniformity of cell shape

3. bare nuclei

4. dump thickness

A full table of the importance of features is listed below:

