Project Proposal of Breast Cancer Classification

Asma Ahmed Althagafi

September 2021

1 Introduction:

Recently, The healthcare sector is poised for a radical transformation led by artificial intelligence and machine learning techniques and is powered by an abundance of data sources including electronic medical records, claim form data, genome sequences, medical imaging, and even integrated sensor data. The applications of artificial intelligence and machine learning techniques in the healthcare sector have continued to increase, and research in recent areas such as machine learning continues to highlight the potential of computers to predict results and improve clinical processes in a wide range of settings. In this project, I going to use machine learning techniques to help in the early detection of breast cancer which will increase chances of treatment.

2 Data Description:

I will use the UCI Machine Learning Repository for breast cancer dataset.

2.1 Attribute Information:

1. ID number

2. Diagnosis (M = malignant, B = benign)

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

   - radius (mean of distances from center to points on the perimeter)
   - texture (standard deviation of gray-scale values)
   - perimeter
   - area
   - smoothness (local variation in radius lengths)
   - compactness ($perimeter^2/area - 1.0$)
   - concavity (severity of concave portions of the contour)
   - concave points (number of concave portions of the contour)
   - symmetry
   - fractal dimension ("coastline approximation" - 1)

2.2 Goals:

The goal of this is to observe what features are most useful when trying to predict malignant or benign cancer. This can also help us when we choose the model and in the hyper parameter selection.

3 Tools:

I will use Classification algorithms of supervised learning. There are many different classification models in machine learning:

- Logistic Regression.
- K-Nearest Neighbor.
- Random Forest Classification.