Shape

Description automatically generated

**Breast Cancer Detection**

**Project Proposal**

**by**

Norah Rashed Almarri

**To Instructor:**

Dr. Essam Al Daoud

**October, 2021**

# Project Description

Technology supports many fields to make our life easier such as marketing, education, medicine, and so on. For the medical field, technology becomes more important especially when it is used to make diagnoses that help to discover dangerous diseases early such as cancer. Breast cancer disease is one of the most common diseases in our world. Many women do some diagnoses for their breasts periodically to check if there is cancer or not. Therefore, the hospitals receive a lot of diagnoses to check during the day. The best way to check is to make the diagnoses automatically to get accurate results in a short time. For this reason, it will be helpful if there is an intelligence program that helps to discover breast cancer by implementing the classification technique with high accuracy. The program will insert different information of the breast cells’ images and will produce the diagnoses result. This project will begin to describe the background and discuss the problem statement. In this project, the problem is to implement a classification program for breast cancer. We need a data set that contains information for different breasts cells’ images. Also, our program must classify if the breast cell has cancer or not by implementing a suitable algorithm for classification that provides a result with high accuracy.

# Data Description

The Breast Cancer Wisconsin (Diagnostic) DataSet, acquired from Kaggle, incorporates capabilities computed from a digitized photograph of a first-class needle aspirate (FNA) of a breast mass and describes traits of the cell nuclei exist withinside the photograph.

1. Number of instances: 569
2. Number of attributes: 32 (ID, diagnosis, 30 real-valued enter capabilities)
3. Attributes names:

['id', 'diagnosis', 'radius\_mean', 'texture\_mean', 'perimeter\_mean',

'area\_mean', 'smoothness\_mean', 'compactness\_mean', 'concavity\_mean',

'concave points\_mean', 'symmetry\_mean', 'fractal\_dimension\_mean',

'radius\_se', 'texture\_se', 'perimeter\_se', 'area\_se', 'smoothness\_se',

'compactness\_se', 'concavity\_se', 'concave points\_se', 'symmetry\_se',

'fractal\_dimension\_se', 'radius\_worst', 'texture\_worst',

'perimeter\_worst', 'area\_worst', 'smoothness\_worst',

'compactness\_worst', 'concavity\_worst', 'concave points\_worst',

'symmetry\_worst', 'fractal\_dimension\_worst', 'Unnamed: 32']

In this dataset, I need to exclude two unnecessary columns which are ('id' and 'unnamed') and the target column is "diagnosis".

# Tools

1. Jupyter notebook.
2. Python packages

* Pandas 🡪 to read the dataset.
* Numpy 🡪 to perform mathematical operations.
* Seaborn 🡪 for the graphical view of the images.
* Matplotlib 🡪 to print the images on the panel.
* Sklearn 🡪 to train, test the model.