Breast Cancer Prediction

In [38]:
```python
# Import necessary packages
import os , glob
from imutils import paths
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
import seaborn as sns
import random
import subprocess, sys

import skimage
from skimage.io import imread , imread_collection

from sklearn.metrics import accuracy_score , classification_report
from sklearn.preprocessing import LabelEncoder,StandardScaler

import keras
from keras.applications import mobilenet
from keras.layers import Dense ,Dropout
from keras.models import Sequential
```

In [2]:
```python
# Get some high level stats from the images
images_path ='../Data/Images/*/'
patient_list = list(glob.glob(images_path))
print(f'Number of the patient :{len(patient_list)}')

images_count = 0
for i in range(0 , len(patient_list)):
    images_count+=len(glob.glob(patient_list[i]+'/*/*'))
print(f'Number of the images :{images_count}')

benign_count = 0
for i in range(0 , len(patient_list)):
    benign_count+=len(glob.glob(patient_list[i]+'/0/*'))
print(f'Number of the benign (non-cancerous) cells :{benign_count}')

malignant_count = 0
for i in range(0 , len(patient_list)):
    malignant_count+=len(glob.glob(patient_list[i]+'/1/*'))
print(f'Number of the malignant (cancerous) cells :{malignant_count}')
```

```
Number of the patient :279
Number of the images :277524
Number of the benign (non-cancerous) cells :198738
Number of the malignant (cancerous) cells :78786
```

In [20]:
```python
# Build the Dataframe with counts
BC_list = [["0", benign_count], ["1", malignant_count]]
BC_df = pd.DataFrame(BC_list, columns=['Diagnosis', 'Count'])
BC_df.head()
```