

Breast Cancer Detection Using ML Project

Problem:

Breast cancer is the most commonly occurring cancer among women in Egypt and the second most common cancer overall representing 38% of the total cases diagnosed with cancer. The total number of cases diagnosed with cancer reached 115,000 cases per year according to statistics conducted in 2014 .Our project aims to help both patients and non-patients, and that would be achieved using a variety of platforms and tools such as Mobile application and website.

Firstly, the Mobile Application in which we will provide all necessary information about the disease and how to avoid getting affected by it. The application will ask the user for her history and upon her answers, it would calculate her risk percentage and suggest what to do. It would ask the user to make self-examination and explain by video how to do that and ask for any abnormal things, and if user's answer is normal it would remind her to do self examination again let's say about 2 months later, and if it was upnormal it would forward her to the nearest mammogram lab or doctor.

Secondly, Web applications we uploaded all API in the website like risk assessment tool, Meta analysis and microscopic image . The website also has information about breast cancer and about how to periodically check, and diagnose cancer in his first stage . The user has access to a risk assessment tool and information about breast cancer, but only doctors who have access to Meta analysis and microscopy. All of this APIs work with AI that help in detecting and early diagnosis of cancer.

Introduction:

- Metadata Analysis –Wisconsin Dataset.
- Analysis Approach using Machine learning.
- Objectives:
 - o Comparison between different machine learning algorithms.
 - o Extracting best features from the dataset.
 - o Building a predictive model that used for diagnosing patients.

Dataset Description:

- Wisconsin Breast Cancer Dataset.
- Contains 569 patient entries (357 Benign – 212 Malignant)
- The features were extracted from digitized images of the fine-needle aspirate (FNA) of a breast mass.

- Attributes: Ten cell nucleus features, namely (radius – texture – perimeter – area – smoothness – compactness – concavity – concave points – symmetry fractal dimension).
- Size: 569 rows x 33 columns.

Analysis Stages:

My approach aims to finding the best machine learning model fitting the dataset and then, extracting the most important subset of features that can be used for diagnosis going through some stages as shown in the following flowchart:

