Analysis and Prediction of Breast Cancer Using Machine Learning Techniques

Abstract

Breast cancer is one of the second leading causes of cancer death in women. Despite the fact that cancer is preventable and curable in primary stages, the vast number of patients are diagnosed with cancer very late. During their life, among 8% of women are diagnosed with breast cancer after lung cancer. breast cancer is characterized by the mutation of genes, constant pain, changes in the size, colour(redness), skin texture of breasts. Established methods of detecting and diagnosing cancer mainly depend on skilled physicians, with the help of medical imaging, to detect certain symptoms that usually appear in the later stages of cancer. This project presents a novel method to detect breast cancer by employing techniques of Machine Learning.

Classification of breast cancer leads pathologists to find a systematic and objective prognostic, generally the most frequent classification is binary (benign cancer/malign cancer). Today, Machine Learning (ML) techniques are being broadly used in the breast cancer classification problem. They provide high classification accuracy and effective diagnostic capabilities. But the limitations are that either they use faulty dataset or they don’t wrangle the data correctly or select features properly. The objective of this project is to find the smallest subset of features that can guarantee highly accurate classification of breast cancer as either benign or malignant. We will implement different M.L algorithms and evaluate their accuracy using cross validation to find the best model by combining feature pre-processing methods and classifier algorithms like *(Logistic Regression, Decision Tree, Random forest, K-nearest Algorithm, SVM).*