

BREAST CANCER DIAGNOSIS

Supriya. U, Divya. R M, Janani. P & Nivetha. R

*Research Scholar, Department of Computer Science and Engineering, Bannari Amman Institute of Technology,
Sathyamangalam, Erode, TamilNadu, India*

Received: 18 Mar 2020

Accepted: 23 Mar 2020

Published: 31 Mar 2020

ABSTRACT

After lung cancer, Breast Cancer is the most intricate disease diagnosed among 8% of women. Breast Cancer is characterized by the mutation of genes, constant pain, changes in the size, color (redness), skin texture of breasts. The detection and treatment of breast cancer involves screening at an early stage by which the mortality can be reduced by a quarter. Data Mining techniques are used to predict the probability of a person who are affected by Breast Cancer using certain Data Mining Techniques. The techniques include KNN Classification, Support Vector Machine Algorithm, and Confusion Matrix Algorithm. The project is implemented using Python programming and deploys the analytical results as a comparison plot. The data-set has been fetched from the UCI repository, and it contains different attributes or factors such as patient_id, unif_cell-size, unif_cell-shape, marg_adhesion and bare_nuclei. In this project, the results are deployed as a comparison chart, which provides better visualization. Various techniques have been used to understand the behavior patterns and to predict the percentage of occurrence of breast cancer. We infer from our study that KNN is well suited algorithm for diagnosis with a accuracy of 97%.

KEYWORDS: *K-Nearest Neighbors, Support Vector Machine, Confusion Matrix, Feature Extraction, Principle Component Analysis, Algorithm Comparison and Optimization*