

GENOMIC PROFILES FOR IDENTIFYING BREAST CANCER

Kavindi Oshini Alvitigala

A dissertation submitted in partial fulfilment of the requirement for Bachelor of
Science (Honours) degree in Computer Science

Department of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

University of Westminster, UK

2020

Abstract

Breast cancer is most common cancer in worldwide and it's second leading cause of the cancer death. The gene mutation is the main key factor of the breast cancer. This research was conducted the identified genes expression of the meaningful subtypes of the breast cancer. This research will be grate support for improving the efficiency of the treatment and reducing the toxicity of the breast cancer treatment by identifying the clue to find target therapeutics through the analysis of genetic expression. Gene expression data set was gained form the cBioportal. This analysis of this research was conducted using unsupervised learning and supervised learning techniques. This research identified the 10 gene expression in data preprocessing phase. Then K means clustering, Hierarchical clustering, and PAM clustering are used to identify the meaningful clustering of the selected 10 genes expression. After identified the meaningful clustering, the selected genes set predict the subtypes using several algorithms to get powerful accuracy. Finally, the kernel-SVM algorithm use as the predicting algorithm which have gained the 71% accuracy.

Keyword: gene expression, breast cancer, subtypes, gene clustering , kernel-SVM