Abstract

Breast cancer is the most common cancer worldwide and it is the second leading cause of cancer death. Gene mutation is the main key factor of breast cancer. This research was conducted on identified gene expression of the meaningful subtypes of breast cancer. This research will be a great support for improving the efficiency of the treatment and reducing the toxicity of breast cancer treatment by identifying the clue to find target therapeutics through the analysis of gene expression. Gene expression data set was gained from the "Maharagama Apeksha Cancer Hospital". This analysis of this research was conducted using unsupervised learning and supervised learning techniques. This research identified the 10 gene expression in the 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 has gained 71% accuracy.

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

Master Of Science (MSc) Business Analytics

Genomic Profile for Identifying Breast Cancer

Department of Business

INFORMATICS INSTITUTE OF TECHNOLOGY

Sri Lanka

Validation: April 2017

School of Computing Science and Digital Media

ROBERT GORDON UNIVERSITY

ABERDEEN, SCOTLAND, UNITED KINGDOM