

Abstract

Cancer is a world-famous deadly disease. According to the World Health Organization (WHO), cancer is the second leading cause of death in the world and is responsible for an estimated 9.6 million deaths in 2018. One well-known technique for cancer detection is the DNA microarray technique. DNA microarray technique provides an opportunity for researchers to analyze thousands of gene expression profiles at the same time to determine whether a person has cancer or not. However, one of the problems that exist in DNA microarray data is the large number of features that require feature selection. In overcoming these problems, this study will use the feature selection method Support Vector Machine-Recursive Feature Elimination (SVM-RFE) and Chi-Square and use the Naïve Bayes classification method. The accuracy results from using feature selection with those that are not will be compared. The results of the accuracy between the use of the two feature selection methods will also be compared, to find which feature selection method is better when combined with the Naïve Bayes classification method. To get an overall picture of the performance comparison, this study also considers precision, recall, and F1-score. The best accuracy results obtained were 100% lung cancer data with SVM-RFE and Chi-Square, 99.6% ovarian cancer with SVM-RFE, 93.7% breast cancer with SVM-RFE, and 90% colon cancer with SVM- RFE.

Keywords: Cancer, Microarray, Feature Selection, Support Vector Machine-Recursive Feature Elimination (SVM-RFE), Chi-Square, Naïve Bayes.