Abstract

Cancer is one of the main contributors to the cause of death in the world, the World Health Organization (WHO) describes cancer as one of the causes of death with a figure of 9.6 million in 2018. Data mining methods have been found to detect symptoms of cancer through a process of calcification someone's genetic expression data called data microarrays. However, DNA data from microarrays have a large number of dimensions and can have a detrimental effect on the accuracy that results from the microarray classification process. Therefore, a feature selection method is needed before conducting a classification process to eliminate features that do not have important information that can support the classification process. One classification method used in this study is Support Vector Machine (SVM), and the feature selection method used is ReliefF and Correlation-based Feature Selection (CFS). The testing scheme applied in this study aims to make a comparison between the ReliefF and Correlation-based Feature Selection (CFS) methods for the Support Vector Machine (SVM) classification method. Average accuracy obtained based on the ReliefF-SVM and CFS-SVM schemes were 94,87% and 84% respectively.

Keywords: microarray data, classification, feature selection, support vector machine , correlation based feature selection, relieff