

ABSTRACT

Cancer is one of the main cause of death in the world with the total of 8,8 million cases of death cause by cancer in 2015 [1]. DNA microarray technology that was developed by Patrick O. Brown, Joseph DeRisi, and David Botstein provides a way to monitor thousands of gen expressions at the same time. But the problem with microarray technology is that microarray data has a high dimensionality. This high dimensionality will effect to the classification performance. In order to solve this issue, this research proposed the use of Minimum Redundancy Maximum Relevance (MRMR) as dimension reduction methods and Support Vector Machine (SVM) as *classifier*. Principal Component Analysis (PCA) method is also used as a comparison to MRMR. MRMR is a dimension reduction method that selects features based on its high relevance values and low redundancy values. While PCA is a method that extracts features based on the principal component obtained from the calculation of eigenvalues and eigenvectors. Data used in this research are colon cancer data, lung cancer data, ovarian cancer data, and leukemia cancer data. Test results on lung cancer data prove that the proposed system is able to provide an F1-score of 1 from the use of MRMR and SVM *classifier* with linear *kernel* function and polynomial *kernel* function, with the number of features used for the classification is 10% of the original features total number. This means the accuracy of classification is 100% and the performance of the system built is very good.

Keywords: microarray data, cancer, classification, SVM, PCA, MRMR