ABSTRACT

Cancer is the second largest cause of death worldwide. In 2018, the global cancer burden has risen to 18.1 million new cases and 9.6 million mortalities were recorded. Globally, around 1 in 6 deaths is caused by cancer. Therefore, it is important to detect this disease early. In medical field, there are many methods that can be used to detect cancer. One of the methods is microarray data technology. This method allows us to read thousands of gene expressions on the same time. However, this method has a major problem, namely high dimension data. This problem can affect classification performance and need a high computational time. On previous research, the system using Principal Component Analysis (PCA) as the feature extraction and logistic regression as the classifier gave the average accuracy of 72.58% for all of the cancer datasets. This low accuracy has to be risen due to the fact that cancer is a serious disease. Furthermore, this research objective is to improve the accuracy of cancer detection system. The dimension reduction method that will be used on this research is Minimum Redundancy Maximum Relevance(mRMR). Meanwhile, the classifier that will be applied is Modified Logistic Regression. The result shows an improvement of accuracy, with the average accuracy of 93.33%.

Keywords: microarray data; cancer; classification; mRMR; logistic regression