

Abstract

Microarray is a modern technique that facilitates simulation analysis of a number of data that describe the expression of genes needed to solve complex biological problems, such as the detection of a particular disease. Microarray data has large dimension characteristics, where the number of response variables is smaller than the predictor variable. Therefore, a scheme is needed in which there is a dimension reduction process and classification process. In this case, the dimension reduction process aims to ease the computational burden and avoid overfitting the classification. The reduction process used in this study is the selection feature of the Genetic Algorithm (GA). Then, the classification process that aims to detect cancer or non-cancer is carried out using the K-Nearest Neighbor (KNN) classification method. The accuracy of the GA-KNN method in data on colon tumor, lung cancer, and leukemia has an average accuracy of 95.01%.

Keywords: K-Nearest Neighbor, Genetic Algorithm, Cancer Detection, Microarray Data