Abstract

Cancer is a disease that has increased in number every year and is also one of the deadliest diseases in the world. Through the development of machine learning, cancer gene data can be processed using DNA microarrays to detect cancer outbreaks. DNA microarrays will process more genes and classify whether this produces cancer or not in a short time. The problem that requires DNA microarrays is the number of attributes that really need to be done dimension reduction. In this discussion, this study will use the dimensions of the Discrete Wavelet Transform with Classification and Regression Tree (CART) and Random Forest for classification methods. The two classification methods will compare and analyze to find out which classification method produces the best performance when combined with the DWT dimension reduction. The best results obtained for breast cancer data were 76.92% with CART-DWT, colon tumors by 90.1% with RF-DWT, lung cancer by 100% with RF-DWT, prostate tumors by 93.8% with RF-DWT, and ovarian cancer by 100% with RF-DWT. From the results of the data used, four data obtained the best results when using RF-DWT.

Keywords: ***Cancer, Microarray, Dimensions Reduction , Discrete Wavelet Transform (DWT), Classification and Regression Tree(CART), Random Forest***