

Abstract

Cancer is one of the leading causes of death worldwide. According to the World Health Organization (WHO) in 2018, about 9.6 million deaths caused by cancer. DNA microarray technology has played an important role in analyzing and diagnosing cancer. The accuracy resulting from the classification of Random Forests is not optimal because microarrays have large dimensional data. Therefore, it is necessary to reduce the dimensions of the Discrete Wavelet Transform (DWT) as a feature to reduce dimensions and increase accuracy in microarray data. Based on simulation, the dimension can be reduced and improve the accuracy of classification up to 8% - 20%. DWT approximation coefficient can improve accuracy better than detailed coefficients for data on colon cancer 100%, lung cancer 100%, ovarian 100%, prostate tumor 80% and central nervous system 83.33%.

Keywords: Cancer, Microarray Data, Dimension Reduction, Discrete Wavelet Transform (DWT), and Random Forest.