Abstract

Human Immunodeficiency Virus (HIV) is a virus that attacks the immune system, specifically white blood cells called Cluster of Differentiation 4 (CD4) cells. One of the most common opportunistic infections and the mains cause of death in HIV patients is Tuberculosis (TB). According to the World Health Organization (WHO) in 2015, HIV/TB co-infection caused 390,000 deaths worldwide. Detecting HIV/TB co-infection as early as possible allows patients to receive treatments of Anti Tuberculosis Drugs (OAT) and antiretroviral drugs (ARV). Currently, the detection of TB in HIV patients is done by examination of smear culture, namely sputum sampling. One alternative way to detect TB in HIV patients is by using technology microarray. Microarray is a technology to analyze the structure of thousands of gene expressions that found in certain body part simultaneously. Microarray data has large dimensions so a machine learning method is needed to facilitate analysis. The machine learning method used is dimension reduction and classification process, so the microarray data classification model produces good accuracy. The purpose of this study is to analyze the performance results of the TB detection in HIV patients classification model, using the Principal Component Analysis (PCA) method as a dimension reduction and the Support Vector Machine (SVM) method as a classification method. The best model in this study was built using the RBF kernel with seven features that resulted in accuracy and F1 score values of 0.86 and 0.83, respectively.

Keywords: HIV, TB, microarray, PCA, SVM