Abstract

Parkinson's disease (PD) is a neurodegenerative disease that causes nerve cells in the brain or peripheral nervous system to shrink and lose function. This disease attacks the brain motoric system by causing a condition of tremor that can lead to accidents and even death. There is cure has been found for this disease, resulting in more than 10 million people worldwide living with PD. To date, there is no examination that can confirm the diagnosis of Parkinson's disease early stadium. Microarray datasets present the opportunity to discover new factors to help understand the underlying mechanisms associated with a disease. Based on related studies, machine learning approach using microarray data have a great opportunity to detect diseases, especially PD. This study aims to test the ensemble method (i.e. Random Forest, Adaptive Boosting, and Extreme Gradient Boosting) combined with *Analysis of Variance* (ANOVA) dan *Mutual Information* (MI) as feature selectors. The hyperparameter tuning process plays an important role in improving model performance. The best model performance is the combination of the RF method and the ANOVA feature selection technique which shows the accuracy and F1-Score values of 67% and 70%, respectively.

Keywords: RF, AdaBoost, XGBoost, feature selection, model, PD, machine learning