Performance Analysis on Wisconsin Breast Cancer Dataset Using Ensemble Model

M.Anusha

Department of CSE, G Narayanamma Institute of Technology and Science [For Women], Hyderabad, Telangana, India

Dr.D.V.Lalitha Parameswari

Department of CSE, G Narayanamma Institute of Technology and Science [For Women], Hyderabad, Telangana, India

Abstract

Breast cancer is an invasive cancer that develops from breast tissue. It is one of the most dangerous diseases that are very effective for women in the world. Detecting the cancer through various automatic diagnostic techniques is very necessary. Many machine learning algorithms are available for prediction and diagnosis of breast cancer. Some of the machines learning algorithms are K-Nearest Neighbor (KNN), it suffers from curse of dimensionality, Naïve Bayes, which has disadvantage assumption of independent predictors, Support Vector Machine (SVM), its disadvantage is Extensive memory requirement and Multilayer Perceptron which gives best result for unstructured data but in study structured data is used. Wisconsin Breast Cancer Dataset is used which contains 569 rows and 30 features. Construction of hybrid classifier for Support Vector Machines and decision trees in WEKA gives accuracy of 91%. In this paper Ensemble method is used to compute best method for diagnosing breast cancer disease. Ensemble methods work well to speed up prediction accuracy. The performance parameters like parameters like precision, recall, f1 score were analyzed to identify the best classification method. The primary objective behind using Ensemble learning is to reduce the variables and diagnose the disease effectively. In this approach the features were reduced to sixteen variables. It is ascertained that the classification accuracy has been improved by ensemble method of random forest and gradient boosting algorithm before and after attribute removal. Before the accuracy is 96.07 percent and these selected sixteen variables gave good accuracy of about 97.23% than other set of variables..

ISBN: 978-93-89107-33-3