

Machine Learning-Based Alzheimer's Disease Prediction Using Boruta Feature Selection and Clinical Cognitive Assessment Data

Vaishnavi*

Department of Computer Science & Engineering, Birla Institute of Technology, Mesra, Ranchi, India

Ritesh Jha

Department of Computer Science & Engineering, Birla Institute of Technology, Mesra, Ranchi, India

Siba Mitra

Department of Computer Science & Engineering, Birla Institute of Technology, Mesra, Ranchi, India

Abstract:

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that affects memory, cognitive ability, and daily functioning. Early prediction of AD is important for effective clinical decision-making and treatment planning. This study presents a machine learning-based framework using Boruta feature selection to identify the most significant clinical and cognitive assessment features for AD prediction. Five important features, namely MMSE, Memory complaints, Functional assessment, Behavioral problems and Activities of Daily Living (ADL) were selected by the Boruta algorithm. Mean, median, standard deviation, T-test and the P-value analysis were performed to test for the statistical significance of all the selected features, which were confirmed with the extremely low P-values obtained (<0.05). A number of classifiers based on machine learning algorithms were implemented and compared: Multinomial Naive Bayes, Bernoulli Naive Bayes, Support Vector Machine (SVM), Random Forest, Decision Tree, K-Nearest Neighbor (KNN) and Logistic Regression. The outputs of the experimental work show that the model Boruta + Random Forest is the most suitable with an accuracy of 94.65%, the precision is 93.29%, recall is 91.45%, the F1 score is 92.36% and the AUC score is 94.66%. The ability to classify was further confirmed by the ROC curve analysis, which showed that the Random Forest model outperforms other classifiers. The results showed that the combination of Boruta feature selection with machine learning methods has great potential for boosting prediction accuracy of Alzheimer's disease and eliminating irrelevant features. The suggested framework can aid early diagnosis of AD (in the healthcare system) efficiently and reliably.

Keywords:

AI in healthcare, Alzheimer's disease, Boruta, clinical data, cognitive assessment, healthcare analytics, machine learning.