

Prediction Modeling of Alzheimer's Disease and Its Prodromal Stages from Multimodal Data with Missing Values

Authors : M. Aghili, S. Tabarestani, C. Freytes, M. Shojaie, M. Cabrerizo, A. Barreto, N. Rishe, R. E. Curiel, D. Loewenstein, R. Duara, M. Adjouadi

Abstract : A major challenge in medical studies, especially those that are longitudinal, is the problem of missing measurements which hinders the effective application of many machine learning algorithms. Furthermore, recent Alzheimer's Disease studies have focused on the delineation of Early Mild Cognitive Impairment (EMCI) and Late Mild Cognitive Impairment (LMCI) from cognitively normal controls (CN) which is essential for developing effective and early treatment methods. To address the aforementioned challenges, this paper explores the potential of using the eXtreme Gradient Boosting (XGBoost) algorithm in handling missing values in multiclass classification. We seek a generalized classification scheme where all prodromal stages of the disease are considered simultaneously in the classification and decision-making processes. Given the large number of subjects (1631) included in this study and in the presence of almost 28% missing values, we investigated the performance of XGBoost on the classification of the four classes of AD, NC, EMCI, and LMCI. Using 10-fold cross validation technique, XGBoost is shown to outperform other state-of-the-art classification algorithms by 3% in terms of accuracy and F-score. Our model achieved an accuracy of 80.52%, a precision of 80.62% and recall of 80.51%, supporting the more natural and promising multiclass classification.

Keywords : eXtreme gradient boosting, missing data, Alzheimer disease, early mild cognitive impairment, late mild cognitive impair, multiclass classification, ADNI, support vector machine, random forest

Conference Title : ICADD 2019 : International Conference on Alzheimer

Conference Location : Miami, United States

Conference Dates : March 11-12, 2019