TITLE: A selected subset of psychological and functional assessments to efficiently classify Alzheimer's disease from healthy controls using novelty detection

AUTHORS

Shuo Liu[1], Qinyong Ye[2], Damien Coyle[1], Junxiu Liu[1], Yi Cao[3], Li Su[4,$], Xuemei Ding [*1,$]

AFFILIATIONS

[1]School of Computing, Engineering & Intelligent Systems, Ulster University, Derry~Londonderry, BT48 7JL, UK

[2] Department of Neurology, The Union Hospital of Fujian Medical University, Fuzhou 350005, China

[3] Business School, University of Edinburgh, Edinburgh, EH8 9YL, UK

[4]Department of Psychiatry, University of Cambridge, CB2 0SP, UK

*Corresponding e-mail: x.ding@ulster.ac.uk

$ these authors contributed equally

INTRODUCTION

Psychological/functional assessments (PFA) have been widely used in clinical practice for Alzheimer's disease (AD) diagnosis and prognosis due to their lower cost relative to blood test and brain imaging. However, applying a number of sub-assessments is time-consuming, and particularly impracticable for patients with cognitive impairment. This study aims to find a subset of PFA for efficient AD diagnosis.

MATERIALS & METHODS

The data records containing 123 healthy control (HC, Age: 63.8±7.0 (Mean±Std)) and 99 AD (Age: 69.4±8.5) samples were collected from the Union Hospital of Fujian Medical University, China. The dataset (D1) covers 51 AD Cooperative Study - Activities of Daily Living (ADCS-ADL) and 12 Neuropsychological Inventory (NPI) features. Three feature selection algorithms (correlation analysis, information gain, information gain ratio) were carried out in parallel, followed by the aggregation of three feature selection ranks with mild distinction into a single rank using Cross-Entropy Monte Carlo algorithm. The top 20 features (16 ADCS-ADL and 4 NPL) were selected to generate a dimensionality reduced dataset (D2). Then both D1 and D2 were split into a training (including 99 HC and 75 AD samples for 5-fold cross-validation, with only 4-fold HC used for training and the remaining 1-fold HC and 75 AD for validation), and a testing (24 HC and 24 AD for classification) subsets. Finally, four novelty detection algorithms based on K-Nearest Neighbours (KNN), One-class Support Vector Machines (OCSVM), Principal Component Analysis (PCA) and K-means were applied.

RESULTS

The Area Under Receiver Operating Characteristic Curve of testing from KNN, OCSVM, PCA and K-means are 0.799, 0.843, 0.594, 0.796 for D1 and 0.833, 0.924, 0.821, 0.752 for D2.

CONCLUSIONS

The novelty detection using D2 generated an improved or comparable performance of classifying AD from HC, compared to using D1. Future work will investigate the selected 16 ADCS-ADL and 4 NPL sub-assessments to support AD diagnosis accurately and efficiently.