A hybrid machine learning approach for prediction of conversion from mild cognitive impairment to Alzheimer's disease

Magda Bucholc[1], Sofya Titarenko[2], KongFatt Wong-Lin[1], Xuemei Ding[1,3], Paula L. McClean[4], Stephen Todd[5], David P. Finn[6], Liam P. Maguire[1]

[1] Intelligent Systems Research Centre, School of Computing, Engineering & Intelligent Systems, Ulster University, Magee campus, Northern Ireland, United Kingdom
[2] School of Mathematics, University of Leeds, Leeds, United Kingdom
[3] Fujian Provincial Engineering Technology Research Centre for Public Service Big Data Mining and Application, College of Mathematics and Informatics, Fujian Normal University, Fuzhou, Fujian, China
[4] Northern Ireland Centre for Stratified Medicine, Biomedical Sciences Research Institute, Ulster University, Northern Ireland, United Kingdom
[5] Altnagelvin Area Hospital, Western Health and Social Care Trust, Northern Ireland, United Kingdom
[6] Discipline of Pharmacology and Therapeutics, National University of Ireland Galway, Republic of Ireland

Corresponding e-mail: m.bucholc@ulster.ac.uk (Magda Bucholc)

Background

Patients diagnosed with mild cognitive impairment (MCI) face a substantially higher risk of developing dementia, particularly the Alzheimer's disease (AD) type. MCI is often considered a transitional phase between healthy ageing and dementia, with annual rate of conversion between 3-25%. Results of both laboratory experiments and unsuccessful clinical AD trials raise the possibility that interventions implemented earlier in the pathophysiological process of AD, precisely during the stage of MCI, are more likely to modify progression of dementia symptoms. In this study, we developed a prognostic approach for differentiating between MCI individuals at risk of developing AD (MCI-c) and those who are suffering from non-progressive cognitive impairment (MCI-nc) using neuropsychological assessments.

Methods

One of the largest longitudinal datasets for dementia research, obtained from the National Alzheimer's Coordinating Center, was used to develop the prognostic framework. For improved prediction, we combined two machine learning approaches: 1) unsupervised modelling of cognitive trajectories and 2) supervised classification. Trajectory clustering was implemented using the shape-respecting distance and shape-respecting mean as similarity measures. Three classification models, namely, logistic regression with Lasso regularizer (LR), random forest (RF), and support vector machine (SVM) were used for model training. The grid search with 10-fold cross validation was applied to select an optimal set of hyper-parameters. The performance of models was evaluated on an independent testing set using two established measures, i.e., Accuracy and area under the ROC curve (AUC).

Results

The MCI-c vs MCI-nc prediction at Year 4 was determined from clinical scores and cognitive trajectory clusters from three previous years (Year 1-3). Note that we included only patients diagnosed as MCI in Year 1 to 3. The average Accuracy of correctly identifying MCI-c vs MCI-

nc patients was 88.0% for RF, 86.0% for SVM, and 81.3% for LR. The AUC was 74.4% for RF, 73% for SVM, and 79% for LR.

Conclusion

We developed a hybrid computational approach for identifying progressive MCI subjects. Future work will involve systematic testing of our framework on datasets with different number of consecutive patient visits as well as datasets including dropout or intermittent missing values at one or more time measurements.