Identification of preclinical dementia according to ATN classification for stratified trial recruitment A machine learning approach

Ivan Koychev, Evgeniy Marinov, Simon Young, Sophia Lazarova, Denitsa Grigorova, Dean Palejev

Introduction: The Amyloid/Tau/Neurodegeneration (ATN) framework was proposed to identify the preclinical biological state of Alzheimer's disease (AD). We investigated whether ATN phenotype can be predicted using routinely collected research cohort data.

Methods: 927 EPAD LCS cohort participants free of dementia or Mild Cognitive Impairment were separated into 5 ATN categories. We used machine learning (ML) methods to identify a set of significant features separating each neurodegeneration-related group from controls (A-T-(N)-). Random Forest and linear-kernel SVM with stratified 5-fold cross validations were used to optimize model whose performance was then tested in the ADNI database.

Results: Our optimal results outperformed ATN cross-validated logistic regression models by between 2.2% and 8.3%. The optimal feature sets were not consistent across the 4 models with the AD pathologic change vs controls set differing the most from the rest. Because of that we have identified a subset of 10 features that yield results very close or identical to the optimal.

Discussion: Our study demonstrates the gains offered by ML in generating ATN risk prediction over logistic regression models among pre-dementia individuals.