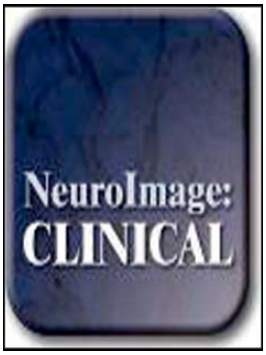

Five-class Differential Diagnostics of Neurodegenerative Diseases using Random Undersampling Boosting

Tong Tong, Christian Ledig, Ricardo Guerrero, Andreas Schuh, Juha Koikkalainen, Antti Tolonen, Hanneke Rhodius, Frederik Barkhof, Betty Tijms, Afina W Lemstra, Hilkka Soininen, Anne M Remes, Gunhild Waldemar, Steen Hasselbalch, Patrizia Mecocci, Marta Baroni, Jyrki Lötjönen, Wiesje van der Flier, Daniel Rueckert

PII: S2213-1582(17)30141-9
DOI: doi:10.1016/j.nicl.2017.06.012
Reference: YNICL 1053

To appear in: *NeuroImage: Clinical*

Received date: 9 March 2017
Revised date: 7 May 2017
Accepted date: 8 June 2017

Please cite this article as: Tong, Tong, Ledig, Christian, Guerrero, Ricardo, Schuh, Andreas, Koikkalainen, Juha, Tolonen, Antti, Rhodius, Hanneke, Barkhof, Frederik, Tijms, Betty, Lemstra, Afina W, Soininen, Hilkka, Remes, Anne M, Waldemar, Gunhild, Hasselbalch, Steen, Mecocci, Patrizia, Baroni, Marta, Lötjönen, Jyrki, Flier, Wiesje van der, Rueckert, Daniel, Five-class Differential Diagnostics of Neurodegenerative Diseases using Random Undersampling Boosting, *NeuroImage: Clinical* (2017), doi:10.1016/j.nicl.2017.06.012

Five-class Differential Diagnostics of Neurodegenerative Diseases using Random Undersampling Boosting

Tong Tong[a,b], Christian Ledig[a], Ricardo Guerrero[a], Andreas Schuh[a], Juha Koikkalainen[c,l], Antti Tolonen[c], Hanneke Rhodius[d], Frederik Barkhof[f,g], Betty Tijms[d], Afina W Lemstra[d], Hilkka Soininen[h,i], Anne M Remes[h,i], Gunhild Waldemar[j], Steen Hasselbalch[j], Patrizia Mecocci[k], Marta Baroni[k], Jyrki Lötjönen[c,l], Wiesje van der Flier[d,e], Daniel Rueckert[a]

[a] *Department of Computing, Imperial College London, London, UK*
[b] *Laboratory for Computational Neuroimaging, Athinoula A. Martinos Center for Biomedical Imaging, MGH/Harvard Medical School, Charlestown, USA*
[c] *VTT Technical Research Centre of Finland, Tampere, Finland*
[d] *Alzheimer Center, Department of Neurology, VU University Medical Centre, Neuroscience Campus Amsterdam, Amsterdam, the Netherlands*
[e] *Department of Epidemiology and Biostatistics, VU University Medical Centre, Neuroscience Campus Amsterdam, Amsterdam, the Netherlands*
[f] *Department of Radiology and Nuclear Medicine, VU University Medical Centre, Neuroscience Campus Amsterdam, Amsterdam, the Netherlands*
[g] *Institutes of Neurology and Healthcare Engineering, UCL, London, UK*
[h] *Institute of Clinical Medicine - Neurology, University of Eastern Finland, Kuopio, Finland*
[i] *Department of Neurology, Kuopio University Hospital, Kuopio, Finland*
[j] *Department of Neurology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark*
[k] *Section of Gerontology and Geriatrics, University of Perugia, Perugia, Italy*
[l] *Combinostics Ltd, Tampere, Finland*

Abstract

Differentiating between different types of neurodegenerative diseases is not only crucial in clinical practice when treatment decisions have to be made, but also has a significant potential for the enrichment of clinical trials. The purpose of this study is to develop a classification framework for distinguishing the four most common neurodegenerative diseases, including Alzheimer's disease, frontotemporal lobe degeneration, Dementia with Lewy bodies and vascular dementia, as well as patients with subjective memory complaints. Different biomarkers including features from images (volume features, region-wise grading features) and non-imaging features (CSF measures) were extracted for each subject. In clinical practice, the prevalence of different dementia types is imbalanced, posing challenges for learning an effective classification model. Therefore, we propose the use of the RUSBoost algorithm in order to train classifiers and to handle the class imbalance training problem. Furthermore, a multi-class feature selection method based on sparsity is integrated into the proposed framework to improve the classification performance. It also provides a way for investigating the importance of different features and regions. Using a dataset of 500 subjects, the proposed framework achieved a high accuracy of 75.2% with a balanced accuracy of 69.3% for the five-class classification using ten-fold cross validation, which is significantly better than the results using support vector machine or random forest, demonstrating the feasibility of the proposed framework to support clinical decision making.

Keywords: Neurodegenerative diseases, Differential diagnosis, MRI, dementia, imbalance learning, multi-class feature selection

1. Introduction

Neurodegeneration is a progressive process that results in the gradual loss of nerve structure and function. The neurodegenerative process occurs with normal aging, but can be accelerated by many neurodegenerative diseases (NDs), including Alzheimer's disease (AD), frontotemporal lobe degeneration (FTLD), dementia