NeuroImage: Clinical 6 (2014) 115–125

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness ☆

A.V. Lebedev[a,*], E. Westman[b], G.J.P. Van Westen[c], M.G. Kramberger[d], A. Lundervold[e,f], D. Aarsland[a,b], H. Soininen[g], I. Kłoszewska[h], P. Mecocci[i], M. Tsolaki[j], B. Vellas[k], S. Lovestone[l,m], A. Simmons[l,m], for the Alzheimer's Disease Neuroimaging Initiative and the AddNeuroMed consortium

[a]Centre for Age-Related Medicine, Stavanger University Hospital, Stavanger, Norway
[b]Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, Alzheimer's Disease Research Centre, Karolinska Institute, Stockholm, Sweden
[c]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom
[d]Department of Neurology, University Medical Center Ljubljana, Slovenia
[e]Neuroinformatics and Image Analysis Laboratory, Department of Biomedicine, University of Bergen, Bergen, Norway
[f]Department of Radiology, Haukeland University Hospital, Bergen, Norway
[g]Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland
[h]Department of Old Age Psychiatry and Psychotic Disorders, Medical University of Lodz, Lódz, Poland
[i]Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy
[j]Aristotle University of Thessaloniki, Thessaloniki, Greece
[k]GERONTOPOLE, UMR INSERM 1027, CHU, University of Toulouse, France
[l]King's College London, Institute of Psychiatry, London, UK
[m]NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia, London, UK

ARTICLE INFO

Article history:
Received 2 February 2014
Received in revised form 6 June 2014
Accepted 26 August 2014
Available online 29 August 2014

Keywords:
Alzheimer's disease
Mild cognitive impairment
Structural MRI
Random Forest
Computer-aided diagnosis
Multi-center study
ADNI
AddNeuroMed

ABSTRACT

Computer-aided diagnosis of Alzheimer's disease (AD) is a rapidly developing field of neuroimaging with strong potential to be used in practice. In this context, assessment of models' robustness to noise and imaging protocol differences together with post-processing and tuning strategies are key tasks to be addressed in order to move towards successful clinical applications. In this study, we investigated the efficacy of Random Forest classifiers trained using different structural MRI measures, with and without neuroanatomical constraints in the detection and prediction of AD in terms of accuracy and between-cohort robustness.

From The ADNI database, 185 AD, and 225 healthy controls (HC) were randomly split into training and testing datasets. 165 subjects with mild cognitive impairment (MCI) were distributed according to the month of conversion to dementia (4-year follow-up). Structural 1.5-T MRI-scans were processed using Freesurfer segmentation and cortical reconstruction. Using the resulting output, AD/HC classifiers were trained. Training included model tuning and performance assessment using out-of-bag estimation. Subsequently the classifiers were validated on the AD/HC test set and for the ability to predict MCI-to-AD conversion. Models' between-cohort robustness was additionally assessed using the AddNeuroMed dataset acquired with harmonized clinical and imaging protocols.

In the ADNI set, the best AD/HC sensitivity/specificity (88.6%/92.0% — test set) was achieved by combining cortical thickness and volumetric measures. The Random Forest model resulted in significantly higher accuracy compared to the reference classifier (linear Support Vector Machine). The models trained using parcelled and high-dimensional (HD) input demonstrated equivalent performance, but the former was more effective in terms of computation/memory and time costs. The sensitivity/specificity for detecting MCI-to-AD conversion (but not AD/HC classification performance) was further improved from 79.5%/75%–83.3%/81.3% by a combination of morphometric measurements with ApoE-genotype and demographics (age, sex, education). When applied to the independent AddNeuroMed cohort, the best ADNI models produced equivalent performance without substantial accuracy drop, suggesting good robustness sufficient for future clinical implementation.

☆ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

* Corresponding author at: Centre for Age-Related Medicine, Stavanger University Hospital, PO Box 8100, 4068 Stavanger, Norway.
E-mail address: alexander.vl.lebedev@gmail.com (A.V. Lebedev).

http://dx.doi.org/10.1016/j.nicl.2014.08.023

1. Introduction

The application of pattern recognition approaches to neuroimaging offers the potential for diagnostically relevant analysis techniques, in particular for magnetic resonance imaging (MRI), which has already been demonstrated to provide relevant support in the diagnosis of Alzheimer's disease (AD) (O'Brien, 2007). A large number of studies addressing the use of pattern recognition methods in image-based detection of AD have been published in recent years (Gray et al., 2013; Liu et al., 2012; Cuingnet et al., 2011; Klöppel et al., 2008).

The advantage of these methods over visual assessment by a medical expert is that they are fully automated and therefore unbiased towards human mistakes and can be incorporated into computerized medical decision-support systems, a growing field with especially fast research progress in radiology (Stivaros et al., 2010; Belle et al., 2013).

However, such methods do have limitations. Our previous work demonstrated that pattern recognition methods are sensitive to MR-protocol differences (Westman et al., 2011; Lebedev et al., 2013) and that a harmonization step is therefore required. Another relevant issue pertains to the comparison of high-dimensional imaging data input versus measurements extracted by neuroanatomical parcellation atlases, with the areas separated according to functional and histological maps of the human cortex (for simplicity, we will use the term "parcelled data"). Parcelled input has some obvious advantages in terms of lower computation, memory cost and processing time. However, it is possible that it could be biased by these landmarks. Normalized high-dimensional measurements without parcellation, in contrast, are unbiased, but at the same time are more difficult to handle using multivariate and machine learning approaches due to computation and memory costs. Moreover, situations where the number of measurements is much larger than the number of observations ($p \gg n$) are often associated with the so-called "curse of dimensionality" (Bellman, 1961). This refers to a number of events that happen when dealing with high-dimensional input (due to increasing sparsity of the data), significantly hampering modeling efficacy. Such cases often require a preparatory step of dimensionality reduction.

Random Forest (RF) is an ensemble machine learning algorithm, which is best defined as a "combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest" (Breiman, 2001).

In many applications this algorithm produces one of the best accuracies to date and has important advantages over other techniques in terms of ability to handle highly non-linear biological data, robustness to noise, tuning simplicity (compared to other ensemble learning algorithms) and opportunity for efficient parallel processing (De Bruyn et al., 2013; Caruana and Niculescu-Mizil, 2006; Menze et al., 2009). These factors also make RF an ideal candidate for handling high-dimensional problems, where the number of features is often redundant. Although RF can itself be considered as an effective feature selection algorithm, several approaches for feature set reduction within and outside the context of RF have been proposed to further improve its performance (Tuv et al., 2009). In the current study, we use recursive feature elimination (Kuhn, 2012a) to optimize the models.

Our previous work revealed that parcelled cortical thickness together with subcortical volumetric measurements (used as an input to a multivariate model) resulted in the best performance, compared to other modalities (Westman et al., 2013). Here, we aimed not only to assess the accuracies of the classifiers trained with different morphometric modalities, but also to analyze the impact of dimensionality, parcellation strategy on models' accuracy, computation/memory/time costs of model training and feature selection. Finally, previous studies have successfully employed pattern recognition techniques to classify MRI images from different cohorts only within the combined sets (Westman et al., 2011; Lebedev et al., 2014). The present study was planned as one of the first to assess classifiers' between-cohort robustness in two independent large-scale datasets.

We hypothesized that with the use of more disease-specific parcellation atlases (in this case, when the measurements are extracted from the predefined regions, known to be affected by Alzheimer's disease), it would be possible to achieve AD-detection accuracy equivalent to that of the models trained with high-dimensional input without parcellation with shorter computational time. In addition, we hypothesized that it is possible to achieve good between-cohort generalization of the models if the MRI protocols are harmonized.

2. Methods

2.1. Subjects

The study was based on two cohorts. The first set of clinical and MRI data was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI-1) database (*http://adni.loni.ucla.edu*). In short, ADNI-1 includes more than 800 subjects with up to 5 years of annual follow-up with comprehensive clinical, neuropsychological, imaging and laboratory evaluations, performed at the 57 specialized ADNI sites in North America. For details, see Aisen et al. (2010) and ADNI-Core (2011). The present cross-sectional study is focused on baseline imaging data and longitudinal information regarding conversion to dementia.

In total, 3D T1 baseline brain scans from 809 subjects passed our image quality control criteria. From this group we selected 575 subjects – 185 AD, 225 healthy controls (HC) and 165 patients with mild cognitive impairment (MCI) and long term follow up information – who met the inclusion criteria (see below).

In order to test the impact of different cohorts, we additionally included 321 subjects (AD 107, 114 MCI and 100 HCs) from the AddNeuroMed study with harmonized clinical and imaging protocols (*http://www.innomed-addneuromed.com/*). The standardized study harmonization workflow (described in previous publications) particularly included careful MR protocol alignment evaluated by phantom scanning and careful quality control (Simmons et al., 2011).

2.2. Inclusion criteria and clinical assessment procedures

All AD patients met the NINCDS/ADRDA criteria for probable AD, had mild level of dementia, defined as the Mini-Mental State Examination (MMSE) score between 20 and 26, and had the Clinical Dementia Rating (CDR) score of 1.0.

Inclusion criteria for MCI were: 1) MMSE score between 24 and 30, 2) memory complaints and objective memory impairment measured by the Logical Memory II subscale of the Wechsler Memory Scale (education adjusted), 3) CDR of 0.5, 4) absence of significant levels of impairment in other cognitive domains, 5) preserved activities of daily living, and 6) absence of dementia. MCI converters had to meet the criteria for Alzheimer's disease during at least two sequential evaluations (e.g., at 24 and 36 month follow-ups). Those MCI subjects who did not have the required follow-up information or had their diagnoses changed back from AD to MCI (or to HC) were excluded ($n = 232$ out of 397). To consider MCI subjects as being non-converters we required that their clinical status remained stable for at least 3 years of follow-up.

Controls (general inclusion/exclusion criteria): 1) MMSE scores between 28 and 30, 2) CDR of 0, and 3) they did not meet the criteria for clinical depression at baseline, MCI or dementia within 3 years of follow-up. One HC subject (ID # 0223) was excluded from the sample due to conversion to AD at follow-up. One AD subject (ID # 0805) was