NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

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

Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge

Esther E. Bron [a,b,*], Marion Smits [c], Wiesje M. van der Flier [d,e], Hugo Vrenken [f], Frederik Barkhof [f], Philip Scheltens [d], Janne M. Papma [g,c], Rebecca M.E. Steketee [c], Carolina Méndez Orellana [c,g], Rozanna Meijboom [c], Madalena Pinto [h], Joana R. Meireles [h], Carolina Garrett [h,i], António J. Bastos-Leite [j], Ahmed Abdulkadir [k,l,m], Olaf Ronneberger [n,m], Nicola Amoroso [o,p], Roberto Bellotti [o,p], David Cárdenas-Peña [q], Andrés M. Álvarez-Meza [q], Chester V. Dolph [r], Khan M. Iftekharuddin [r], Simon F. Eskildsen [s], Pierrick Coupé [t], Vladimir S. Fonov [u], Katja Franke [v,w], Christian Gaser [v,w], Christian Ledig [x], Ricardo Guerrero [x], Tong Tong [x], Katherine R. Gray [x], Elaheh Moradi [y], Jussi Tohka [y], Alexandre Routier [z,aa], Stanley Durrleman [z,aa], Alessia Sarica [ab], Giuseppe Di Fatta [ac], Francesco Sensi [ad], Andrea Chincarini [ad], Garry M. Smith [ae,ac], Zhivko V. Stoyanov [ae,ac], Lauge Sørensen [af], Mads Nielsen [af], Sabina Tangaro [o], Paolo Inglese [o], Christian Wachinger [ag,ah], Martin Reuter [ag,ah], John C. van Swieten [g], Wiro J. Niessen [a,b,ai], Stefan Klein [a,b], for the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Biomedical Imaging Group Rotterdam, Department of Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
[b] Biomedical Imaging Group Rotterdam, Department of Radiology, Erasmus MC, Rotterdam, The Netherlands
[c] Department of Radiology, Erasmus MC, Rotterdam, The Netherlands
[d] Alzheimer Center, Department of Neurology, VU University Medical Center, Neuroscience Campus Amsterdam, The Netherlands
[e] Department of Epidemiology & Biostatistics, VU University Medical Center, Neuroscience Campus Amsterdam, The Netherlands
[f] Department of Radiology & Nuclear Medicine, VU University Medical Center, Neuroscience Campus Amsterdam, The Netherlands
[g] Department of Neurology, Erasmus MC, Rotterdam, The Netherlands
[h] Department of Neurology, Hospital de São João, Porto, Portugal
[i] Department of Clinical Neurosciences and Mental Health, Faculty of Medicine, University of Porto, Porto, Portugal
[j] Department of Medical Imaging, Faculty of Medicine, University of Porto, Porto, Portugal
[k] Department of Psychiatry & Psychotherapy, University Medical Centre Freiburg, Germany
[l] Department of Neurology, University Medical Centre Freiburg, Germany
[m] Department of Computer Science, University of Freiburg, Germany
[n] BIOSS Centre for Biological Signaling Studies, University of Freiburg, Germany
[o] National Institute of Nuclear Physics, Branch of Bari, Italy
[p] Department of Physics, University of Bari, Italy
[q] Signal Processing and Recognition Group, Universidad Nacional de Colombia, Colombia
[r] Vision Lab, Old Dominion University, Norfolk, VA 23529, USA
[s] Center of Functionally Integrative Neuroscience and MINDLab, Aarhus University, Aarhus, Denmark
[t] Laboratoire Bordelais de Recherche en Informatique, Unit Mixte de Recherche CNRS (UMR 5800), PICTURA Research Group, Bordeaux, France
[u] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada
[v] Structural Brain Mapping Group, Department of Neurology, Jena University Hospital, Germany
[w] Structural Brain Mapping Group, Department of Psychiatry, Jena University Hospital, Germany
[x] Biomedical Image Analysis (BioMedIA) Group, Department of Computing, Imperial College London, UK
[y] Department of Signal Processing, Tampere University of Technology, Finland
[z] Inserm U1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, Inria Paris-Rocquencourt, F-75013 Paris, France
[aa] Centre d'Acquisition et de Traitement des Images (CATI), Paris, France
[ab] Bioinformatics Laboratory, Department of Medical and Surgical Sciences, Magna Graecia University, Catanzaro, Italy
[ac] School of Systems Engineering, University of Reading, Reading RG6 6AY, UK
[ad] National Institute of Nuclear Physics, Branch of Genoa, Italy
[ae] Centre for Integrative Neuroscience and Neurodynamics, University of Reading, RG6 6AH, UK
[af] Department of Computer Science, University of Copenhagen, Denmark
[ag] Computer Science and Artificial Intelligence Lab, MA Institute of Technology, Cambridge, USA
[ah] Massachusetts General Hospital, Harvard Medical School, Cambridge, USA
[ai] Imaging Physics, Applied Sciences, Delft University of Technology, The Netherlands

* Corresponding author at: Biomedical Imaging Group Rotterdam, Departments of Medical Informatics and Radiology, Erasmus MC, Rotterdam, The Netherlands.
E-mail addresses: caddementia@bigr.nl, e.bron@erasmusmc.nl (E.E. Bron).

[1] Some data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.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.

http://dx.doi.org/10.1016/j.neuroimage.2015.01.048

ARTICLE INFO

Article history:
Accepted 24 January 2015
Available online xxxx

Keywords:
Alzheimer's disease
Challenge
Classification
Computer-aided diagnosis
Mild cognitive impairment
Structural MRI

ABSTRACT

Algorithms for computer-aided diagnosis of dementia based on structural MRI have demonstrated high performance in the literature, but are difficult to compare as different data sets and methodology were used for evaluation. In addition, it is unclear how the algorithms would perform on previously unseen data, and thus, how they would perform in clinical practice when there is no real opportunity to adapt the algorithm to the data at hand. To address these comparability, generalizability and clinical applicability issues, we organized a *grand challenge* that aimed to objectively compare algorithms based on a clinically representative multi-center data set. Using clinical practice as the starting point, the goal was to reproduce the clinical diagnosis. Therefore, we evaluated algorithms for multi-class classification of three diagnostic groups: patients with probable Alzheimer's disease, patients with mild cognitive impairment and healthy controls. The diagnosis based on clinical criteria was used as reference standard, as it was the best available reference despite its known limitations. For evaluation, a previously unseen test set was used consisting of 354 T1-weighted MRI scans with the diagnoses blinded. Fifteen research teams participated with a total of 29 algorithms. The algorithms were trained on a small training set ($n = 30$) and optionally on data from other sources (e.g., the Alzheimer's Disease Neuroimaging Initiative, the Australian Imaging Biomarkers and Lifestyle flagship study of aging). The best performing algorithm yielded an accuracy of 63.0% and an area under the receiver-operating-characteristic curve (AUC) of 78.8%. In general, the best performances were achieved using feature extraction based on voxel-based morphometry or a combination of features that included volume, cortical thickness, shape and intensity. The challenge is open for new submissions via the web-based framework: http://caddementia.grand-challenge.org.

Introduction

In 2010, the number of people over 60 years of age living with dementia was estimated at 35.6 million worldwide. This number is expected to almost double every twenty years (Prince et al., 2013). Accordingly, the cost of care for patients with Alzheimer's disease (AD) and other dementias is expected to increase dramatically, making AD one of the costliest chronic diseases to society (Alzheimer's Association, 2014). Early and accurate diagnosis has great potential to reduce the costs related to care and living arrangements as it gives patients access to supportive therapies that can help them maintain their independence for longer and delay institutionalization (Paquerault, 2012; Prince et al., 2011). In addition, early diagnosis supports new research into understanding the disease process and developing new treatments (Paquerault, 2012; Prince et al., 2011).

While early and accurate diagnosis of dementia is challenging, it can be aided by assessment of quantitative biomarkers. The five most commonly investigated biomarkers were recently included in the revised diagnostic criteria for AD (McKhann et al., 2011; Jack et al., 2011) and in the revised diagnostic criteria for mild cognitive impairment (MCI) due to AD (Albert et al., 2011). These five biomarkers can be divided into two categories: 1) measures of brain amyloid, which include cerebrospinal fluid (CSF) measures of Aβ42 and amyloid positron emission tomography (PET) imaging, and 2) measures of neuronal injury and degeneration, which include CSF tau measurement, fluoro deoxyglucose (FDG) PET and structural MRI (Jack et al., 2012). Of these biomarkers, structural MRI is very important as it is widely available and non-invasive. Also, it is a good indicator of progression to AD in an individual subject, because it becomes abnormal in close temporal proximity to the onset of the cognitive impairment (Jack et al., 2010, 2013).

Structural MRI data can be used to train computer-aided diagnosis methods. These methods make use of machine-learning and other multivariate data-analysis techniques that train a model (classifier) to categorize groups (e.g., patients and controls). Computer-aided diagnosis techniques use features derived from neuroimaging or related data, and may therefore benefit from the large amounts of neuroimaging data that have become available over the last years. The techniques may improve diagnosis as they can potentially make use of group differences that are not noted during qualitative visual inspection of brain imaging data, potentially leading towards an earlier and more objective diagnosis than when using clinical criteria (Klöppel et al., 2012). In addition, computer-aided diagnosis algorithms can be used to 1) improve diagnosis in hospitals with limited neurological and neuroradiological expertise, 2) increase the speed of diagnosis, and 3) aid the recruitment of specific, homogeneous patient populations for clinical trials in pharmacological research (Klöppel et al., 2012).

Structural-MRI-based computer-aided diagnosis methods for dementia, mainly for AD and MCI, have previously shown promising results in the literature. A few years ago, Cuingnet et al. (2011) compared the performance of various feature extraction methods (e.g., voxel-based features, cortical thickness, hippocampal shape and volume) for dementia classification using a support vector machine (SVM) based on structural MRI. Using data from 509 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort, three classification experiments were performed: 1) AD versus healthy controls (CN), 2) patients with MCI versus CN, and 3) MCI who had converted to AD within 18 months (MCI converters, MCIc) versus MCI who had not converted to AD within 18 months (MCI non-converters, MCInc). For the AD/CN classification, the best results were obtained with whole-brain methods (voxel-based and cortical thickness) achieving 81% sensitivity and 95% specificity for the best method. The performances of the MCI/CN classifications were much lower than those of AD/CN, and the MCIc/MCInc classifications yielded no performances better than chance. A recent review paper by Falahati et al. (2014) discussed the literature on AD classification and MCI prediction. The research field of computer-aided diagnosis of dementia based on structural MRI is rather extensive, as evidenced by this paper reviewing 50 papers with at least 50 subjects per diagnostic group. The reviewed papers mainly trained a classification model on the AD/CN groups and subsequently tested this model on both AD/CN and MCIc/MCInc classifications. The paper concluded that classification methods are difficult to compare, because the outcome is influenced by many factors, such as feature extraction, feature selection, robustness of the validation approach, image quality, number of training subjects, demographics, and clinical diagnosis criteria. In general, the accuracy obtained for AD/CN classification was 80–90%, and the accuracy for prediction of MCI conversion is somewhat lower. To promote comparison of algorithms, Sabuncu and Konukoglu (2015) published results based on six large publicly available data sets for AD and other diseases (e.g., schizophrenia, autism). A comparison was performed using four feature extraction strategies, including volumetric and cortical thickness features computed with FreeSurfer (Fischl, 2012), and three types of machine learning techniques (SVM, neighborhood approximation forest (Konukoglu et al., 2013), and relevance voxel machine (Sabuncu and Van Leemput, 2012)). Using the ADNI database, the accuracies ranged from 80–87% for AD/CN classification and 58–66% for MCI/CN classification. The authors made all processed data and computational tools available to promote extension of their benchmark results.

Please cite this article as: Bron, E.E., et al., Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.048