NeuroImage: Clinical 11 (2016) 435–449

Contents lists available at ScienceDirect

NeuroImage: Clinical

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

# Differential diagnosis of neurodegenerative diseases using structural MRI data

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

[a] VTT Technical Research Centre of Finland, Tampere, Finland
[b] Alzheimer Center, Department of Neurology, VU University Medical Centre, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands
[c] Department of Radiology and Nuclear Medicine, VU University Medical Centre, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands
[d] Department of Computing, Imperial College London, London, UK
[e] Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland
[f] Department of Epidemiology and Biostatistics, VU University Medical Centre, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands
[g] Department of Neurology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[h] Section of Gerontology and Geriatrics, University of Perugia, Perugia, Italy
[i] Combinostics Ltd., Tampere, Finland

ARTICLE INFO

*Article history:*
Received 2 November 2015
Received in revised form 2 February 2016
Accepted 29 February 2016
Available online 5 March 2016

*Keywords:*
MRI
Neurodegenerative diseases
Classification
Volumetry
TBM
VBM
Alzheimer's disease
Frontotemporal lobar degeneration
Vascular dementia
Dementia with Lewy bodies

ABSTRACT

Different neurodegenerative diseases can cause memory disorders and other cognitive impairments. The early detection and the stratification of patients according to the underlying disease are essential for an efficient approach to this healthcare challenge. This emphasizes the importance of differential diagnostics. Most studies compare patients and controls, or Alzheimer's disease with one other type of dementia. Such a bilateral comparison does not resemble clinical practice, where a clinician is faced with a number of different possible types of dementia.

Here we studied which features in structural magnetic resonance imaging (MRI) scans could best distinguish four types of dementia, Alzheimer's disease, frontotemporal dementia, vascular dementia, and dementia with Lewy bodies, and control subjects. We extracted an extensive set of features quantifying volumetric and morphometric characteristics from T1 images, and vascular characteristics from FLAIR images. Classification was performed using a multi-class classifier based on Disease State Index methodology. The classifier provided continuous probability indices for each disease to support clinical decision making.

A dataset of 504 individuals was used for evaluation. The cross-validated classification accuracy was 70.6% and balanced accuracy was 69.1% for the five disease groups using only automatically determined MRI features. Vascular dementia patients could be detected with high sensitivity (96%) using features from FLAIR images. Controls (sensitivity 82%) and Alzheimer's disease patients (sensitivity 74%) could be accurately classified using T1-based features, whereas the most difficult group was the dementia with Lewy bodies (sensitivity 32%). These results were notable better than the classification accuracies obtained with visual MRI ratings (accuracy 44.6%, balanced accuracy 51.6%). Different quantification methods provided complementary information, and consequently, the best results were obtained by utilizing several quantification methods.

The results prove that automatic quantification methods and computerized decision support methods are feasible for clinical practice and provide comprehensive information that may help clinicians in the diagnosis making.



## 1. Introduction

Dementia is a general term to describe a syndrome involving loss of cognitive abilities. Most often dementia is caused by a progressive neurodegenerative disease. Dementia is a major health issue in our society both from the economic and human point of view.

Alzheimer's disease (AD) is the most common type of dementia that may account for 60–75% of dementia cases. Vascular dementia (VaD) and dementia with Lewy bodies (DLB) also occur frequently in elderly patients, while frontotemporal dementia (FTD) is relatively more common in dementia patients with early onset. Characteristic structural pathologies in these diseases include atrophy of the medial temporal lobe

* Corresponding author at: Combinostics Ltd., Hatanpään valtatie 24, 33100, Tampere, Finland.
*E-mail address:* juha.koikkalainen@combinostics.com (J. Koikkalainen).

http://dx.doi.org/10.1016/j.nicl.2016.02.019

in AD and atrophy of the frontal and temporal lobes in FTD. In DLB the brain structure is typically less affected. Absence of medial temporal lobe atrophy and findings of infarcts or white matter changes are typical to VaD. The atrophy patterns can be detected with T1-weighted images. Cortical and lacunar infarcts and white matter changes that are typical to VaD are identified on T1-weighted images and T2-weighted, dual-echo Turbo Spin Echo (TSE) or Fluid-Attenuated Inversion Recovery (FLAIR) images.

Early and accurate differential diagnostics of neurodegenerative diseases is essential for two reasons. First, it has been shown that early diagnosis combined with current treatments can delay hospitalization (Feldman et al., 2009), and the importance of the early diagnosis will dramatically increase as soon as disease-modifying drugs become available (Siemers et al., 2015). Second, developing new treatments requires early and accurate identification of correct target populations. It has been hypothesized that too heterogeneous study populations may explain the failure of some previous pharmaceutical trials (Falahati et al., 2014).

The studies on structural MRI that have characterized distinct neurodegenerative diseases are mostly based on visual ratings (Barber et al., 1999; Burton et al., 2009; Meyer et al., 2007; Varma et al., 2002), volumetry (Meyer et al., 2007; Frisoni et al., 1999; Barber et al., 2000; Munoz-Ruiz et al., 2012; Ishii et al., 2007), and local morphometry analyses (Munoz-Ruiz et al., 2012; Laakso et al., 2000; Burton et al., 2002; Barber et al., 2002; Ballmaier et al., 2004; Whitwell et al., 2007; Rabinovici et al., 2008; Klöppel et al., 2008). Typical findings on the differences between different dementia types include: 1) the hippocampal volume and medial temporal lobe are relatively preserved in FTD as compared to AD (Duara et al., 1999; Frisoni et al., 1999), 2) FTD-specific atrophy of the frontal and temporal lobes (Duara et al., 1999; Varma et al., 2002; Klöppel et al., 2008), 3) relatively preserved brain anatomy in DLB as compared to AD and FTD (Meyer et al., 2007; Barber et al., 1999, 2000; Burton et al., 2002, 2009; Kantarci et al., 2012; Ishii et al., 2007; Whitwell et al., 2007), and 4) extensive white matter changes with lacunar and cortical infarcts in VaD (Meyer et al., 2007).

There are extensive literature comparing the dementia types with controls, but far less studies have been done on comparing the different dementia types with each other. In clinical practice, the actual question is to determine to which type of dementia a patient with cognitive complaints should be diagnosed. The guidelines for the early detection of neurodegenerative diseases (Román et al., 1993; Neary et al., 1998; McKeith et al., 2005; Dubois et al., 2007; Waldemar et al., 2007; McKhann et al., 2011) are relatively general and do not provide specific and uniform information for accurate differential diagnostics of neurodegenerative diseases. Therefore, the current diagnostic processes involve a certain degree of subjective assessment and require significant expertise from clinicians. Automatic image quantification methods and computerized decision support methods are able to objectively extract lots of information, more than the human eye can see, and evaluate how the patient data relates to typical data from different dementias. Such data are likely to be useful in clinical diagnosis making, especially supporting the decisions of unexperienced clinicians.

The objective of this paper is to perform an extensive study on differential diagnostics of dementias utilizing only structural MRI data. We evaluate several state of the art automatic quantification methods in order to find out which of the methods or what combination gives optimal classification accuracy. We utilize a dataset of 504 patients divided into five different groups: controls (CN), AD, FTD, DLB, and VaD. Both T1 and FLAIR data are used in the analysis.

## 2. Material and methods

### 2.1. Patient groups

We study a total of 504 patients from the Amsterdam Dementia Cohort who had visited the Alzheimer center of the VU University Medical Center between 2004 and 2014 (van der Flier et al., 2014). The patients were included if MRI and mini mental state examination (MMSE) (Folstein et al., 1975) were present. At baseline, all patients received a standardized and multi-disciplinary work-up, including medical history, physical, neurological and neuropsychological examination, MRI, laboratory test and lumbar puncture to collect cerebrospinal fluid. Diagnoses were made in a multidisciplinary consensus meeting.

In this study, patients with subjective cognitive decline (SCD) were regarded as the control subjects. Patients were diagnosed as having SCD when cognitive complaints could not be confirmed by cognitive testing and criteria for MCI, dementia or other neurological or psychiatric disorder known to cause cognitive complaints were not met. Patients were diagnosed with probable AD using the criteria of the National Institute for Neurological and Communicative Diseases Alzheimer's Disease and Related Disorders Association; all patients also met the core clinical criteria of the National Institute on Aging-Alzheimer's Association guidelines for AD (McKhann et al., 1984; McKhann et al., 2011). FTD was diagnosed using the Neary criteria; patients also met the core criteria from Rasckovsky (Neary et al., 1998; Rascovsky et al., 2011). VaD was diagnosed using the National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherché et l'Enseignement en Neurosciences criteria (Román et al., 1993), and DLB using the McKeith criteria (McKeith et al., 1996; McKeith et al., 2005). The study was approved by the local Medical Ethical Committee. All patients have signed written informed consent for their clinical data to be used for research purposes.

The normal cognition of all the SCD patients was confirmed at 9 months follow-up. Follow-up took place by annual routine visits to the memory clinic in which patient history, cognitive tests and a general physical and neurologic examination were repeated. Follow-up data was available in all SCD subjects, with a mean of $2.5 \pm 1.4$ years.

### 2.2. Imaging

Subjects were scanned routinely on either 1.0 T, 1.5 T or 3.0 T MRI devices. All scans include a 3-dimensional T1-weighted gradient echo sequence and a fast FLAIR sequence. The voxel size of the T1-images varies between $0.9 \times 0.9 \times 0.9$ mm$^3$ and $1.1 \times 1.1 \times 1.5$ mm$^3$. For FLAIR images there is much more variation in the slice thickness, as the voxel size varies between $0.4 \times 0.4 \times 1.0$ mm$^3$ and $1.2 \times 1.2 \times 5.0$ mm$^3$. 86 patients were imaged using 1.0 T device, whereas 1.5 T and 3.0 T devices were used for the remaining 97 and 321 patients, respectively. Detailed information on the imaging parameters for each disease group is available in Appendix A.

Imaging data were assessed visually for atrophy and vascular changes. Visual rating of medial temporal lobe atrophy was performed on coronal T1-weighted images according to the 5-point (0–4) rating scale for medial temporal lobe atrophy (MTA) from the average score of the left and right sides (Scheltens et al., 1995). Global cortical atrophy (GCA) was assessed visually on axial FLAIR images (possible range of scores 0–3) (Pasquier et al., 1996). The degree of severity of white matter hyperintensities was rated on axial FLAIR images using Fazekas' scale (possible range of scores 0–3) (Fazekas et al., 1987). The number of lacunes (# of lacunes) was defined as T1-hypointense and T2-hyperintense CSF-like lesions surrounded by white matter or subcortical gray matter. Next to an overall count of lacunes, the presence of ≥1 lacunes in the basal ganglia (BG lacunes) was determined. Finally, the presence of infarcts ≥1 (Infarcts) was visually evaluated.

In this study, the visual scores serve as reference values: the multi-class classification is performed using visual scores (Section 3.1) and the results obtained with automatic image quantification methods (Section 3.2) are compared against these results.