CDT Journal Club - First Written Review

Automatic classification of MR scans in Alzheimer's disease

Stefan Klppel , Cynthia M. Stonnington , Carlton Chu , Bogdan Draganski , Rachael I. Scahill , Jonathan D. Rohrer , Nick C. Fox , Clifford R. Jack , John Ashburner , Richard S. J. Frackowiak

Review Author:
Răzvan Valentin Marinescu
razvan.marinescu.14@ucl.ac.uk

Paper chosen by:
Prof. Sebastien Ourselin

EPSRC Centre for Doctoral Training in Medical Imaging
University College London

February 18, 2015

Paper background

Dementia and associated diseases such as Alzheimer's disease (AD) affect 36 million people globally [1] and account for 486,000 deaths [2]. Definitive diagnosis of AD can only be made with histopathological confirmation of amyloid plaques and neurofibrillary tangles, normally at autopsy. Early accurate detection of AD is important because the treatment is most effective if undertaken as early as possible. Medical imaging such as MRI has been used as a tool for AD diagnosis by measuring the atrophy rates of cortical volumes [3, 4] but these methods have not yet been introduced in clinical practice.

Introduction

This paper uses a machine-learning classifier called *Support Vector Machine* (SVM) for the purpose of AD diagnosis. It is the first study to use SVMs for diagnosing AD using pathologically confirmed cases. The authors futher use SVMs to differentially diagnose AD and frontotemporal lobal degeneration (FTLD). The study uses 4 different cohorts of patients: (1) the first set is a community-based cohort where AD was confirmed with neuropathology (2) the second set consisted of neuropathologically-confirmed AD subjects from a different center than the first group (3) the third set consisted of probable mild-AD patients limited to 80 years of age or younger (4) finally, the fourth set was made of subjects with neuropathologically proven FTLD having comparable Mini Mental State Examination (MMSE) scores with the first two groups. The scans were collected over a period up to 10 years with a total of around 13 different scanners in order to test the ability of SVMs to diagnose patinents using training data from different centers.

Several image processing steps were taken in order to remove artefacts and ultimatelly register the scans. Images were segmented into grey matter, white matter and cerebro-spinal fluid (CSF) using SPM5 [5]. The GM segments were further normalised to the population templates using a diffeomorphic registration algorithm [6], which was a sophisticated technique at the time the paper was published. In order to ensure the overall ammount of each tissue class remained constant a separate *'modulation'* step was undertaken as described by Ashburner and Friston [7].

Classification of patients was performed using Support-Vector Machine (SVM) which is a method of supervised, binary classification where voxels from MR images are treated as points in a high-dimensional space. Once projected in this higher-dimensional space, it tries to find the optimal separating hyperplane (OSH) that separates the two classes as efficient as possible. The OSH is defined by the voxels that are the closest to the separating boundary, called *support vectors*. [8, 9] The OSH, wich is produced during training, can then be used to classify a new patient by finding on which side of the OSH it's corresponding data point lies.

Method

Key results

A linear SVM is used to diagnose AD vs controls (cohorts I,II and III) and AD (cohort II) vs FTLD (cohort IV). In all cases, classification accuracy is very high, reaching 96% in some cases. Classification accuracy is also high for probable mild-AD (89%), in line with the diagnosis rates of the best clinical centers around the world. Authors have also tried using different datasets for training and testing, and again the observed accuracy is very high (87.5% and 96.4%), given the fact that different scanners were used for producing the images. A nice feature of SVMs is that it allows the localisation of the voxels that were relevant for classification (i.e. the support vectors). Authors show that for AD vs controls, most of these vectors were clustered around

the parahyppocampal gyrus and parietal cortex. Authors also tried to use non-linear kernel functions but this failed to improve performance. The results they obtained were equally good or better than other methods used in the field based on MR images. [10, 11, 4, 12, 13]

Contributions

The work of Klöppel et al shows that automated, data-driven techniques are equally good or even better than clinical experts at diagnosing AD patients, as NINCDS-ADRDA or DSM-IIIR criteria have an average sensitivity and specificity of 81% and 70% respectively, which are lower than the values reported in the paper. [14]

Limitations

One methodological limitation is that the SVM is a binary classifier, where only two classes can be distinguished at the same time. Although there exist ways to implement a multiclass-SVM by reducing it to several binary problems [15] or using directed acyclic graphs (DAGs) [16], other multiclass classifiers might provide a more natural solution such as logistic regression [17] or a parzen-window classifier with a linear kernel. [18] Moreover, the method of Klöppel assumes there is no head motion during the scan and does not correct it. Another limitation is that the diffeomorphic registration technique by Ashburner, 2007 [6] which is based on sums of squares is very sensitive to affine initialization. [19]

Impact

The work of Klöppel et al is the first to perform automatic classification of MRI scans in pathologically-confirmed AD and to differentiate different forms of dementia. It clearly made a huge impact when it was published in 2008 and got 569 citations so far. Although it is a methodological paper, it managed to get published in a clinical journal like Brain probably due to the unique datasets it analysed which contained pathologically-confirmed AD patients. Since then, the field has evolved and now people look at diagnosing various subtypes of AD or other types of dementia together. $\text{nned}_t o_c ite_h ere$

Later in 2008 the author did a more detailed study of the diagnostic accuracy between the SVM method and radiologists and found SVMs to have similar results to expert radiologists, the advantage of SVMs being that they don't require expert-knowledge and can easily be exchanged between centers. [**?**]

Bibliography

[1] World Health Organization et al. Dementia: Fact sheet no. 362. *Retrieved June*, 30:2012, 2012.

[2] Rafael Lozano, Mohsen Naghavi, Kyle Foreman, Stephen Lim, Kenji Shibuya, Victor Aboyans, Jerry Abraham, Timothy Adair, Rakesh Aggarwal, Stephanie Y Ahn, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *The Lancet*, 380(9859):2095–2128, 2013.

[3] Nick C Fox and Jonathan M Schott. Imaging cerebral atrophy: normal ageing to alzheimer's disease. *The Lancet*, 363(9406):392–394, 2004.

[4] Josephine Barnes, Rachael I Scahill, Richard G Boyes, Chris Frost, Emma B Lewis, Charlotte L Rossor, Martin N Rossor, and Nick C Fox. Differentiating ad from aging using semiautomated measurement of hippocampal atrophy rates. *Neuroimage*, 23(2):574–581, 2004.

[5] SPM5; Wellcome Trust Centre for Neuroimaging, London, UK. `http://www.fil.ion.ucl.ac.uk/spm`.

[6] John Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007.

[7] John Ashburner and Karl J Friston. Voxel-based morphometrythe methods. *Neuroimage*, 11(6):805–821, 2000.

[8] Vladimir Naumovich Vapnik and Vlamimir Vapnik. *Statistical learning theory*, volume 1. Wiley New York, 1998.

[9] Christopher M Bishop et al. *Pattern recognition and machine learning*, volume 4. springer New York, 2006.

[10] KM Gosche, JA Mortimer, CD Smith, WR Markesbery, and DA Snowdon. Hippocampal volume as an index of alzheimer neuropathology findings from the nun study. *Neurology*, 58(10):1476–1482, 2002.

[11] CR Jack, DW Dickson, JE Parisi, YC Xu, RH Cha, PC Obrien, SD Edland, GE Smith, BF Boeve, EG Tangalos, et al. Antemortem mri findings correlate with hippocampal neuropathology in typical aging and dementia. *Neurology*, 58(5):750–757, 2002.

[12] John G Csernansky, Julia Hamstra, Lei Wang, Daniel McKeel, Joseph L Price, Mokhtar Gado, and John C Morris. Correlations between antemortem hippocampal volume and postmortem neuropathology in ad subjects. *Alzheimer Disease & Associated Disorders*, 18(4):190–195, 2004.

[13] Lars-Olof Wahlund, Ove Almkvist, Kaj Blennow, Knut Engedahl, Aki Johansson, Gunhild Waldemar, and Henrike Wolf. Evidence-based evaluation of magnetic resonance imaging as a diagnostic tool in dementia workup. *Topics in Magnetic Resonance Imaging*, 16(6):427–437, 2005.

[14] DS Knopman, St T DeKosky, JL Cummings, H Chui, J Corey-Bloom, N Relkin, GW Small, B Miller, and JC Stevens. Practice parameter: Diagnosis of dementia (an evidence-based review) report of the quality standards subcommittee of the american academy of neurology. *Neurology*, 56(9):1143–1153, 2001.

[15] Kai-Bo Duan and S Sathiya Keerthi. Which is the best multiclass svm method? an empirical study. In *Multiple Classifier Systems*, pages 278–285. Springer, 2005.

[16] John C Platt, Nello Cristianini, and John Shawe-Taylor. Large margin dags for multiclass classification. In *nips*, volume 12, pages 547–553, 1999.

[17] Rahul S Desikan, Howard J Cabral, Christopher P Hess, William P Dillon, Christine M Glastonbury, Michael W Weiner, Nicholas J Schmansky, Douglas N Greve, David H Salat, Randy L Buckner, et al. Automated mri measures identify individuals with mild cognitive impairment and alzheimer's disease. *Brain*, 132(8):2048–2057, 2009.

[18] John Shawe-Taylor and Nello Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[19] Brian B Avants, Nicholas J Tustison, Gang Song, Philip A Cook, Arno Klein, and James C Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044, 2011.