doi:10.1093/brain/awn239 *Brain* (2008), **131**, 2969–2974

Accuracy of dementia diagnosis—a direct comparison between radiologists and a computerized method

Stefan Klöppel,[1,2] Cynthia M. Stonnington,[2,3] Josephine Barnes,[4] Frederick Chen,[5] Carlton Chu,[2] Catriona D. Good,[6] Irina Mader,[7] L. Anne Mitchell,[4,8,9] Ameet C. Patel,[5] Catherine C. Roberts,[5] Nick C. Fox,[4] Clifford R. Jack Jr,[10] John Ashburner[2] and Richard S. J. Frackowiak[2,11,12]

[1]Department of Psychiatry and Psychotherapy and Freiburg Brain Imaging, University Clinic Freiburg, Freiburg, Germany, [2]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, UK, [3]Department of Psychiatry and Psychology, Mayo Clinic, Scottsdale, AZ, USA, [4]Dementia Research Centre, University College London, Institute of Neurology, London, UK, [5]Department of Radiology, Mayo Clinic, Scottsdale, AZ, USA, [6]Department of Neuroradiology, Hurstwood Park Neurosciences Centre, Brighton & Sussex Universities Hospital NHS Trust, Haywards Heath, West Sussex, UK, [7]Department of Neuroradiology and Department of Neurology, Freiburg Brain Imaging, Neurocenter, University Clinic Freiburg, Freiburg, Germany, [8]Department of Radiology, Austin Health, Heidelberg, [9]Department of Radiology, University of Melbourne, Melbourne, Australia, [10]Department of Radiology, Mayo Clinic, Rochester, USA, [11]Département d'études cognitives, Ecole Normale Supérieure, Paris, France and [12]Laboratory of Neuroimaging, IRCCS Santa Lucia, Roma, Italy

Correspondence to: Stefan Klöppel, MD, Department of Psychiatry and Psychotherapy, University Clinic Freiburg, Freiburg, Germany
E-mail: stefan.kloeppel@uniklinik-freiburg.de

There has been recent interest in the application of machine learning techniques to neuroimaging-based diagnosis. These methods promise fully automated, standard PC-based clinical decisions, unbiased by variable radiological expertise. We recently used support vector machines (SVMs) to separate sporadic Alzheimer's disease from normal ageing and from fronto-temporal lobar degeneration (FTLD). In this study, we compare the results to those obtained by radiologists. A binary diagnostic classification was made by six radiologists with different levels of experience on the same scans and information that had been previously analysed with SVM. SVMs correctly classified 95% (sensitivity/specificity: 95/95) of sporadic Alzheimer's disease and controls into their respective groups. Radiologists correctly classified 65–95% (median 89%; sensitivity/specificity: 88/90) of scans. SVM correctly classified another set of sporadic Alzheimer's disease in 93% (sensitivity/specificity: 100/86) of cases, whereas radiologists ranged between 80% and 90% (median 83%; sensitivity/specificity: 80/85). SVMs were better at separating patients with sporadic Alzheimer's disease from those with FTLD (SVM 89%; sensitivity/specificity: 83/95; compared to radiological range from 63% to 83%; median 71%; sensitivity/specificity: 64/76). Radiologists were always accurate when they reported a high degree of diagnostic confidence. The results show that well-trained neuroradiologists classify typical Alzheimer's disease-associated scans comparable to SVMs. However, SVMs require no expert knowledge and trained SVMs can readily be exchanged between centres for use in diagnostic classification. These results are encouraging and indicate a role for computerized diagnostic methods in clinical practice.

Keywords: MRI; diagnosis; dementia; support vector machine

Abbreviations: FTLD = fronto-temporal lobar degeneration; GM = grey matter; MCI = mild cognitive impairment; MMSE = mini mental state examination; NIA-RIA = National Institute on Aging and the Reagan Institute of the Alzheimer's Association; sAD = sporadic Alzheimer's Disease; SVM = support vector machine

Received April 15, 2008. Revised August 3, 2008. Accepted September 3, 2008. Advance Access publication October 3, 2008

Introduction

There has been recent interest in the application of machine learning techniques to neuroimaging diagnosis. These methods promise fully automated, standard PC-based clinical decisions, unaffected by individual neuroradiological expertise. Such methods have increasingly been applied

to a number of diagnostic problems in neuroimaging ranging from gender-based classification (Lao *et al.*, 2004) to a variety of diseases including Alzheimer's disease and mild cognitive impairment (MCI) (Kawasaki *et al.*, 2007; Teipel *et al.*, 2007*a*, *b*; Davatzikos *et al.*, 2008b; Fan *et al.*, 2008).

We recently used support vector machines (SVMs) to solve the problem of separating mild to severe sporadic Alzheimer's disease from normal ageing and from frontotemporal lobar degeneration (FTLD) with structural MRI (Kloppel *et al.*, 2008b). SVM-based classification can be seen as a two-step procedure. In a first step, SVMs learn the differences between two diagnostic groups (i.e. Alzheimer's disease and healthy controls or Alzheimer's disease and FTLD). This knowledge is then tested on a new brain scan, not used in the training procedure. This test scan will only be assigned to the correct group if group separation is based on disease-related changes. Neuropathological examination served as a gold standard against which classification accuracy was compared. Given that neurodegeneration primarily affects grey matter (GM), we first extracted GM segments from T1-weighted brain scans and normalized them into standard anatomical space. Such segments contain several thousand voxels. After pre-processing, each voxel reflects the magnitude of the local GM volume. SVMs are multivariate in nature and so information from all available voxels is combined to reflect differences between groups. During this training process, those subjects that are most difficult to separate are used to define the 'boundary' between the diagnostic groups.

Using a whole brain, GM-based SVM, we achieved up to 95% correct classification (range = 87.5–95%) depending on the patient group. These results have been subsequently confirmed by others (Davatzikos *et al.*, 2008c) and indicate a performance equal to or better than that achieved by radiologists (Wahlund *et al.*, 2005). A retrospective comparison of SVM results with radiological expertise reported in the literature would be imprecise. The severity of degeneration, technical image quality, the availability of additional clinical information and the frequency of each diagnostic group in any sample could all bias such a comparison. We therefore undertook a direct prospective, blinded comparison of diagnostic accuracy between radiologists and the automated SVM-based method using the same scans and associated clinical information.

Material and Methods

Three imaging datasets were rated by six radiologists with different levels of experience in the diagnosis of dementia. Three radiologists were based at the Mayo Clinic (Scottsdale, USA), one was a visiting radiologist from Melbourne University (Australia) to the Dementia Research Centre (UCL Institute of Neurology, London, UK), one worked at the Hurstwood Park Neurosciences Centre (UK) and one at the Department of Neuroradiology (Freiburg, Germany). All had at least 6 years experience of clinical radiology. Four were neuroradiologists who routinely saw brain scans, which comprised >40% of their workload; two were general radiologists who estimated that brain scans comprised <5% of their daily workload. They thus reflected the range of expertise encountered in typical clinical settings.

The first image set, from a community and referral-based sample in Rochester, Minnesota, USA, comprised 20 sporadic Alzheimer's disease and 20 matched controls with a mean age of 80 years (range: 51–102); the great majority were older than 75 years. All sporadic Alzheimer's disease diagnoses were neuropathologically confirmed according to criteria formulated by a working group of the National Institute on Aging and the Reagan Institute of the Alzheimer's Association (NIA-RIA, 1997). Scans were excluded from analysis if they showed gross structural abnormalities other than atrophy. Diagnostic assignation was based on the combined results of medical history, clinical examination, psychometry and neuropathology. Criteria for diagnosis of normal cognition were independently functioning community membership with (i) no active neurological or psychiatric disorder; (ii) no psychoactive medication; (iii) a normal neurological examination; and (iv) no ongoing medical problem and (v) no associated treatment that might interfere with cognitive function (Jack *et al.*, 2004).

The second set comprised 18 pathologically proven sporadic Alzheimer's disease scans and 19 scans from patients with pathologically proven FTLD, matched for age, scanner and mini mental state examination (MMSE) score. The Alzheimer's disease-patients all fulfilled NINCDS-ADRDA criteria for definite Alzheimer's disease in that the clinical diagnoses were confirmed histopathologically after cerebral biopsy or at autopsy (McKhann *et al.*, 1984) according to Consortium to Establish a Registry for Alzheimer's Disease (CERAD) (Mirra *et al.*, 1991) and NIA-RIA criteria (NIA-RIA, 1997). There was no clear family history in any subject. All FTLD patients were diagnosed according to consensus criteria (Neary *et al.*, 1998) into one of the three FTLD subtypes during life: nine patients had behavioural-variant FTLD, eight had semantic dementia and two had progressive non-fluent aphasia. In total, there were eight patients with tau-positive pathology and 11 patients with ubiquitin-positive, tau-negative pathology, diagnosed according to consensus pathological criteria (McKhann *et al.*, 2001); behavioural-variant FTLD (five tau-positive, four ubiquitin-positive), semantic dementia (two tau-positive, six ubiquitin-positive) and progressive non-fluent aphasia (one tau-positive, one ubiquitin-positive). FTLD patients tended to be younger than Alzheimer's disease patients in this group, but not significantly so ($P = 0.1$).

The third set comprised patients referred to a specialist centre who underwent appropriate specialist diagnostic workup and for whom there was a pathological diagnosis. There were 14 sporadic Alzheimer's disease cases in this set (taken from the second set), most of whom were younger than 75 years at onset, and matched controls (mean age 64, range 51–85). The 14 sporadic Alzheimer's disease cases were selected to match them to the controls in terms of age and scanning equipment (see below). Controls were considered cognitively normal if there was no evidence of abnormality on clinical examination at follow-up or if histological material confirmed the absence of Alzheimer-related change. Demographic details are presented in Table 1.

Scans from the first set were obtained with 11 different General Electric Signa 1.5T scanners (T1-weighted image parameters: TR = 23–27 ms, TE = 6–10 ms, flip angle 25° or 45°, voxel size 0.86 mm × 0.86 mm × 1.6 mm or 0.94 mm × 0.94 mm × 1.6 mm, matrix dimensions 256 × 192). The major hardware elements

Table 1 Demographic information

Group (n)	Sporadic Alzheimer's disease set1/controls		Sporadic Alzheimer's disease set2/FTLD		Sporadic Alzheimer's disease set3/controls	
	Alzheimer's disease (20)	Controls (20)	Alzheimer's disease (18)	FTLD (19)	Alzheimer's disease (14)	Controls (14)
Sex (F/M)	11/9	10/10	6/12	8/11	5/9	5/9
Age (mean, range) at MRI-scan	81.0 (51–102)	79.5 (55–91)	66.0 (53–85)	61.7 (46–73)	65.0 (53–85)	63.0 (51–81)
MMSE-score (mean, range)	16.7 (7–29)	29.0 (27–30)	16.2[a] (5–29)	18.0 (0–26)	16.1[a] (10–20)	29.2 (28–30)
Years from MRI-scan to death (mean, range)	1.7 (0.2–3.4)	NA	3.5 (0.3–7.2)	5.8 (1.3–11.0)	3.6 (0.3–7.2)	NA

[a]MMSE scores obtained around the time of scanning only available from 12 subjects.

(body resonance module, gradient coil and birdcage–head transmit–receive volume coil) were unchanged throughout time and across all scanners. For sets two and three, data were acquired from three different 1.5T scanners from different manufactures. Image parameters were TR = 35 or 15, TE = 5 or 5.4 or 7, flip angle 35° or 15°. Scanners and scanning parameters were balanced across groups and within groups as well as between Alzheimer's disease patients and FTLD patients. This criterion was achieved by excluding four Alzheimer's disease subjects from the third set compared to the control group. Because the mix of scanners used was different for normal elderly controls and FTLD subjects, the same four Alzheimer's disease subjects were included for comparison between Alzheimer's disease and FTLD subjects (second set) to maintain an equal balance of scanners between groups. See Kloppel *et al.* (2008b) for further details.

To allow a fair comparison, radiologists were provided information about the age range of patients and controls and informed that the two diagnostic categories for differentiation were age matched and equal in number. In other words, radiologists were not told the age for each scan but only for the group as a whole (but see our additional analysis below). The radiologists made categorical decisions on each scan to mimic SVM-based diagnoses. They were also asked to rate their level of diagnostic confidence (low, intermediate or high). Radiologists rated the datasets in the order they are listed above. We disclosed the diagnosis of a third of patients and controls to the radiologists for the third dataset so as to mimic the training of an SVM, which uses exemplars for that purpose that are themselves similar to those categorized. Diagnoses were disclosed just before radiologists started with this third dataset but not before completing categorization of the other two sets. Disclosed cases were randomly chosen ensuring that equal numbers of cases and controls were selected and these subjects were not included in scoring. No time limit was set. All radiologists viewed all scans and were asked to diagnose scans in order and to avoid a comparison of their results with the opinions of other radiologists. We report accuracy, sensitivity and specificity (considering a detected Alzheimer's disease patient a true positive) with median and range for each set separately. The SVM results are taken from our previously reported study for comparison (Kloppel *et al.*, 2008b).

Results

One radiologist separated the first set of sporadic Alzheimer's disease cases from cognitively normal subjects as accurately as the SVM; otherwise, the SVM performed better than radiologists (see Table 2 and Fig. 1). Radiological diagnostic accuracy increased substantially (reaching 100%) when high diagnostic confidence was expressed. To evaluate the effect of experience, defined as the percentage of brain scans in their daily workload, we correlated this figure with diagnostic accuracy. A significant correlation indicated that classification accuracy improved with the level of experience (Fig. 2) for sporadic Alzheimer's disease cases in set 1 (Spearman's $r = 0.90$; $P = 0.007$, one-tailed) and when sporadic Alzheimer's disease had to be separated from FTLD ($r = 0.77$; $P = 0.036$). No such correlation was found for the third dataset. The time required to classify all three datasets ranged from 70 min to 510 min (median = 198).

Since we sought to compare diagnostic ability directly between radiologists and the computer-based method, we did not provide information usually available in clinical practice, such as age. To exclude any significant effect of such an omission, five radiologists repeated their classification on the second dataset (FTLD versus Alzheimer's disease) after learning the age of each subject and the diagnosis of a third of them. Wilcoxon signed rank test shows no significant improvement in accuracy [with age information and training: median accuracy 72.40% (range 62.7–82.8); without age information or training: 64.9% (range 56.8–83.8)].

Discussion

The results we present indicate that computer-based diagnosis is equal to or better than that achieved by radiologists. This result has a number of implications that suggest a general adoption of computer-assisted methods for MRI scan-based dementia diagnosis, which should be seriously considered. The most important of these are: (i) improving diagnosis in places where trained neuroradiologists or cognitive neurologists are scarce; (ii) increasing speed of diagnosis without compromising accuracy by eschewing lengthy specialist investigations and (iii) recruitment of clinically homogeneous patient

Table 2 Diagnostic performance of radiologists reported with median and range (in square brackets)

	Sporadic Alzheimer's disease set 1/controls			Sporadic Alzheimer's disease set 2/FTLD			Sporadic Alzheimer's disease set 3/controls		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Radiologists (95% CI) [range]	88.8 (75.4–96.8) [65.0–95.0]	87.5 (72.4–95.3) [70.0–95.0]	90.0 (75.4–96.7) [60.0–95.0]	68.6 (50.1–81.5) [56.8–83.8]	64.1 (47.4–79.3) [57.9–90.0]	71.0 (55.6–85.6) [55.6–83.8]	82.5 (62.4–93.2) [80.0–90.0]	80.0 (58.5–91.0) [80.0–90.0]	85.0 (66.4–95.3) [80.0–100.0]
SVM (95% CI)	95.0 (81.8–99.1)	95.0 (73.1–99.7)	95.0 (73.1–99.7)	89.2 (73.6–96.5)	83.3 (57.7–95.6)	94.7 (71.9–99.7)	92.9 (75.1–98.8)	100.0 (73.2–100)	85.7 (56.2–97.5)

95% confidence intervals (CIs, in parentheses) are calculated according to the efficient-score method (Newcombe, 1998; http://faculty.vassar.edu/lowry/clin1.html). For radiologists, CIs are reported for the median of all radiologists. CIs were based on 20 subjects in the third dataset for radiologists as one-third of diagnosis was disclosed.

Fig. 1 Illustration of performance. Positions are jittered to indicate overlap. Grey error bars display 95% confidence intervals for SVM accuracy. The fourth column illustrates the performance when sets 1 and 3 are combined. Note the shrinking CIs. sAD = sporadic Alzheimer's Disease.

populations for pharmacological trials. A significant correlation between accuracy and experience is very understandable and speaks to a need for specialization, especially in dedicated cognitive neurology clinics. However, primary care and local referral play an important role in the diagnosis of a disease as common as Alzheimer's disease. In this context, computerized methods may be especially helpful for screening purposes. Screening of the worried and those with MCI are other possible applications, but direct validating studies will be necessary to extend the conclusions of our results to these diagnoses.

All radiologists achieved a relatively high accuracy on the third dataset (see Table 2 and Fig. 1). Figure 2 suggests that disclosing the diagnosis of a third of cases in set 3 may have helped the less-experienced radiologists as they score in a range similar to that of experienced radiologists. Interestingly, this effect was absent when these radiologists repeated the separation of Alzheimer's disease and FTLD after a subset was disclosed. It is possible that pathological changes separating Alzheimer's disease from healthy aging can be picked up more easily from a few examples than the differences between two types of dementia.

Given the diagnostic accuracy achievable, machine learning-based categorization methods, such as the SVM technique, we have evaluated and now compared to radiological expertise, substantially extend the role of computers in clinical decision making (Ashburner *et al.*, 2003). Clearly, experienced radiologists working under ideal conditions are very accurate if confident of a diagnosis. However, the computerized method does not

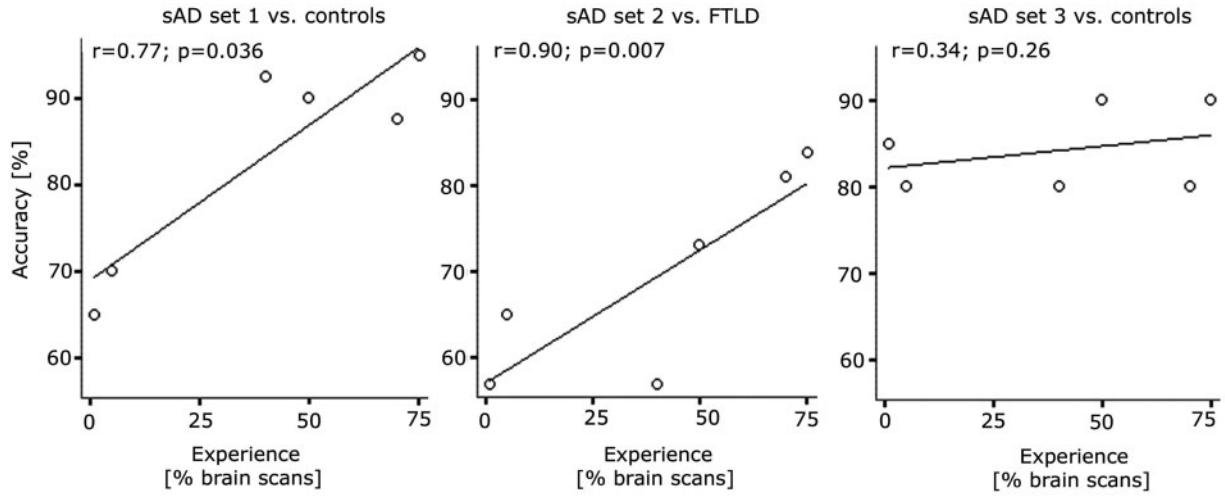

Fig. 2 Illustration of the correlation between experience (given as the percentage of brain scans out of all scans in daily routine practice) and accuracy. sAD = sporadic Alzheimer's Disease.

depend on experience, though crucially it does depend on an appropriately and accurately validated image data for training. Our previous study (Kloppel *et al.*, 2008b) was the first to use pathological confirmation of diagnosis as a criterion for inclusion of scans into the training set. This choice of gold standard limited the number of mildly affected subjects in the training set and hence any conclusion can be drawn from this study. The separation of moderately affected sporadic Alzheimer's disease subjects from controls and other types of dementia is clinically useful but extension of inferences to the mildly affected is less certain. While it can be argued that the 100% sensitivity of SVMs on the third dataset indicates that even the few mild sporadic Alzheimer's disease cases were correctly identified, the inclusion of such cases made group separation more difficult in the eldest and resulted in the two controls of greatest age being misclassified as Alzheimer's disease sufferers. From the first set, a single Alzheimer's disease case was wrongly identified as a control by SVMs. Interestingly, an MMSE score of 29 indicated that this subject was very mildly affected. The application of automatic classification methods to a large set of mildly affected Alzheimer's disease cases will be an important next step in the development of the method.

Our proof-of-principle study (Kloppel *et al.*, 2008b) shows that an SVM trained on validated scans from one imaging centre can be used successfully to classify structural images obtained elsewhere. The results of this study underline the importance of MRI scans in the diagnosis of dementia (Scheltens *et al.*, 2002; Ashburner *et al.*, 2003; Frisoni *et al.*, 2003; Teipel *et al.*, 2008). They indicate that a validated MRI-based SVM is more accurate than the average clinical diagnosis of probable Alzheimer's disease even when based on well-established diagnostic criteria (Knopman *et al.*, 2001).

What is needed now are large numbers of well, ideally pathologically, defined scans from a range of different diseases. Those scans could then be used to train SVMs either for differential diagnosis or for comparison against healthy controls. The trained SVMs could then be exchanged between different imaging centres to aid clinical diagnosis. Such well-defined scans are also needed to identify the optimal number of scans required to construct an SVM that performs optimally in classification. Our preliminary analysis suggests that performance declines when fewer than 20 subjects are included from each diagnostic group (Kloppel *et al.*, 2008*a*) but the number is likely to depend on the extent and variability of disease-specific changes in brain structure. Large-scale imaging initiatives, such as the Alzheimer's Disease Neuroimaging Initiative (Mueller *et al.*, 2005) which provide unrestricted access to image data (Butcher, 2007) are an important resource in this regard. While they are not currently associated with pathological features, there are increasing numbers of MCI patients who convert to Alzheimer's disease. The detection of MCI is another important potential application of classification-based methods (Fan *et al.*, 2008).

In the clinical setting, a new scan could be compared against a number of trained SVMs to further explore their diagnostic usefulness and the sensitivity of the classification method. Although good clinical screening will limit the number of diagnostic choices, they are often larger than in this study where the choice was limited to two types of dementia and to healthy aging.

There is already a variety of fully automatic scan-based diagnostic tools described (Davatzikos *et al.*, 2008a; Teipel *et al.*, 2007*b*; Vemuri *et al.*, 2008) and it is likely that even better methods may become available in the future. New methods are needed to provide a probabilistic rather

than categorical classification framework that adds a level of confidence to a diagnostic decision.

Acknowledgements

The authors would like to thank Rachael Scahill and Jonathan Rohrer for contributing demographic and neuropathological data.

Funding

The Dementia Research Centre is an Alzheimer's Research Trust Co-ordinating Centre. Some of this work was undertaken at UCLH/UCL who received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. Wellcome Trust (grant 075696 2/04/2 to R.S.J.F. and J.A.); Mayo Clinic (grant to C.M.S.); the National Institute on Aging (grants P50 AG16574, U01 AG06786, and AG11378 to Mayo Clinic Rochester, MN); the Robert H. and Clarice Smith and Abigail Van Buren Alzheimer's Disease Research Program of the Mayo Foundation (to Mayo Rochester, MN); UK Medical Research Council (grant G9626876 to N.C.F.); Alzheimer's Research Trust (ART) research fellowship (to J.B.) and the German Research Foundation (grant WE 1352/14-1 to I.M.). Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust.

References

Ashburner J, Csernansky JG, Davatzikos C, Fox NC, Frisoni GB, Thompson PM. Computer-assisted imaging to assess brain structure in healthy and diseased brains. Lancet Neurol 2003; 2: 79–88.

Butcher J. Alzheimer's researchers open the doors to data sharing. Lancet Neurol 2007; 6: 480–1.

Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of MRI. Neurobiol Aging 2008a; 29: 514–23.

Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol Aging 2008b; 29: 514–23.

Davatzikos C, Resnick SM, Wu X, Parmpi P, Clark CM. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. Neuroimage 2008c; 41: 1220–7.

Fan Y, Batmanghelich N, Clark CM, Davatzikos C. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage 2008; 39: 1731–43.

Frisoni GB, Scheltens P, Galluzzi S, Nobili FM, Fox NC, Robert PH, et al. Neuroimaging tools to rate regional atrophy, subcortical cerebrovascular disease, and regional cerebral blood flow and metabolism: consensus paper of the EADC. J Neurol Neurosurg Psychiatry 2003; 74: 1371–81.

Jack CR Jr, Shiung MM, Gunter JL, O'Brien PC, Weigand SD, Knopman DS, et al. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. Neurology 2004; 62: 591–600.

Kawasaki Y, Suzuki M, Kherif F, Takahashi T, Zhou SY, Nakamura K, et al. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. Neuroimage 2007; 34: 235–42.

Kloppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. A plea for confidence intervals and consideration of generalizability in diagnostic studies. Brain 2008a (in press).

Kloppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008b; 131: 681–9.

Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the quality standards subcommittee of the American academy of Neurology. Neurology 2001; 56: 1143–53.

Lao Z, Shen D, Xue Z, Karacali B, Resnick SM, Davatzikos C. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. Neuroimage 2004; 21: 46–57.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. Neurology 1984; 34: 939–44.

McKhann GM, Albert MS, Grossman M, Miller B, Dickson D, Trojanowski JQ. Clinical and pathological diagnosis of frontotemporal dementia: report of the work group on frontotemporal Dementia and Pick's disease. Arch Neurol 2001; 58: 1803–9.

Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 1991; 41: 479–86.

Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack C, Jagust W, et al. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin N Am 2005; 15: 869–77, xi–xii.

Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology 1998; 51: 1546–54.

Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998; 17: 857–72.

NIA-RIA. Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. The National Institute on Aging, and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease. Neurobiol Aging 1997; 18: S1–2.

Scheltens P, Fox N, Barkhof F, De Carli C. Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion. Lancet Neurol 2002; 1: 13–21.

Teipel SJ, Born C, Ewers M, Bokde AL, Reiser MF, Moller HJ, et al. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. Neuroimage 2007a; 38: 13–24.

Teipel SJ, Stahl R, Dietrich O, Schoenberg SO, Perneczky R, Bokde AL, et al. Multivariate network analysis of fiber tract integrity in Alzheimer's disease. Neuroimage 2007b; 34: 985–95.

Teipel SJ, Meindl T, Grinberg L, Heinsen H, Hampel H. Novel MRI techniques in the assessment of dementia. Eur J Nucl Med Mol Imaging 2008; 35 (Suppl 1): S58–69.

Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, et al. Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. Neuroimage 2008; 39: 1186–97.

Wahlund LO, Almkvist O, Blennow K, Engedahl K, Johansson A, Waldemar G, et al. Evidence-based evaluation of magnetic resonance imaging as a diagnostic tool in dementia workup. Top Magn Reson Imaging 2005; 16: 427–37.