

Automatic Classification of SPECT Images of Alzheimer's Disease Patients and Control Subjects

Jonathan Stoeckel¹, Nicholas Ayache¹, Grégoire Malandain¹,
Pierre M. Koulibaly², Klaus P. Ebmeier³, and Jacques Darcourt²

¹ EPIDAURE Project, INRIA,
2004 route des Lucioles, Sophia Antipolis 06902, France
Jonathan.Stoeckel@siemens.com, {na, greg}@sophia.inria.fr
<http://www-sop.inria.fr/epidaure>

² Service de Médecine Nucléaire - Centre Antoine Lacassagne
33 avenue de Valombrose, 06189 NICE cedex 2, France

³ Royal Edinburgh Hospital, Morningside Park
Edinburgh EH10 5HF, United Kingdom

Abstract. In this article we study the use of SPECT perfusion imaging for the diagnosis of Alzheimer's disease. We present a classifier based approach that does not need any explicit knowledge about the pathology. We directly use the voxel intensities as features. This approach is compared with three classical approaches: regions of interests, statistical parametric mapping and visual analysis which is the most commonly used method. We tested our method both on simulated and on real data. The realistic simulations give us total control about the ground truth. On real data, our method was more sensitive than the human experts, while having an acceptable specificity. We conclude that an automatic method can be a useful help for clinicians.

1 Introduction

Alzheimer's disease (AD) is the most frequent type of dementia for elderly patients. Due to aging populations its occurrence will still increase. Even though no definitive cure has been found for this disease, reliable diagnosis is useful for excluding other dementias, choosing the right treatment and for the development of new treatments.

AD is diagnosed using the criteria from the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [1]. In practice the main tool for evaluating patients are neuro-psychologic tests, that test abilities like memory and language. The Mini Mental State Examination (MMSE) is the most widely used of these tests [2].

¹ JS is now with the CAD Solutions Group, Siemens Medical Solutions USA Inc., Malvern, PA

Brain images can provide some indication of AD. Magnetic resonance imaging (MRI) is used to study possible anatomical changes of the brain [3]. Images showing the local perfusion of the brain can be used for the diagnosis of AD because the perfusion pattern is affected by the disease. In this article we will look into the use of cerebral perfusion imaging acquired by single photon emitting computer tomography (SPECT) using Technetium-99m Hexamethyl Propylene Amine Oxime (HMPAO) as the tracer. Even though the perfusion pattern and its evolution are not the same for all patients some hypo-perfusion patterns seem to be typical for the disease. In the literature, three main regions are described as showing hypo-perfusion[4] signals, 1. the temporo-parietal region, 2. the posterior cingulate gyri and precunei, and 3. the medial temporal lobe. The first region is known as the predominant pattern for AD, however this region was not found for early AD [5]. The second region is probably more specific and more frequent in early AD [6]. Previous pathological studies have suggested that the third region is the first affected by the disease [7], however in practice it is only observed in more advanced stages of the disease [6].

There is not one single perfusion pattern that differentiates AD patients from healthy subjects. Thus it might be useful to have tools that could assist physicians in this difficult task. In this article we will present a method that does not need any explicit knowledge about the perfusion pattern of AD patients.

Some approaches for a computer aided diagnosis (CAD) system for the analysis of SPECT images for AD can be found in literature. The first family is based on the analysis of regions of interest. The mean values for these regions are analyzed using some discriminant functions (see e.g. [8][9]).

The second approach is statistical parametric mapping (SPM) and its numerous variants. Statistical parametric mapping is widely used in the neuro-sciences. Its framework was first developed for the analysis of SPECT and PET studies, but is now mainly used for the analysis of functional MRI data. It was not developed specifically to study a single image, but for comparing groups of images. One can use it for diagnostics by comparing the image under study to a group of normal images.

Statistical parametric mapping consists of doing a voxel-wise statistical test, in our case a t-test, comparing the values of the image under study to the mean values of the group of normal images. Subsequently the significant voxels are inferred by using the random field theory (see e.g. [10] for a full description of SPM). A largely used freely available implementation called SPM99 [11] has been developed and will be used in this article to compare our approach.

In this article we will propose another approach using as less a-priori information about the pathology as possible, by obtaining it implicitly from image databases. The other important aspect of our approach is the use of a global approach. This means that we will not provide the clinician with information about where the hypo-perfused areas are situated (e.g. as given by SPM) but we will only give a global answer to the question whether or not the image under study belongs to an AD patient. This has the advantage over a more local approach that all the information in the image can be used at once in contrast

to the mono-variate methods like SPM. A multi-variate approach generally increases sensitivity at the price of losing regional specificity (e.g. depicting local hypo-perfusion regions).

In [12] some very preliminary results were presented. In this article we present results on a large set of real data. This data comes from several centers to show the robustness of our method. However only limited knowledge about the perfusion pattern is available for real data. We only know if the perfusion pattern was that of an Alzheimer's subject or not. The locations of the hypo-perfused regions are not known, neither the amount of the hypo-perfusions. Therefore we chose to simulate images as to have complete local control, and thus knowledge of the ground truth.

The following section first discusses the pre-processing of the data, followed by the analysis of the data using classifiers. In section 3 we present the data we used for our experiments. We used both real data and simulated data. The results on the data are presented in section 4 and discussed in section 5.

2 Methods

Spatial Normalization. In the classifier based approach we need the assumption that the same position in the volume coordinate system within different volumes corresponds to the same anatomical position.

This means that we need to spatially register the volumes. Because of the limited anatomical information available in the volumes we chose to estimate affine transformations between the volumes. We used the correlation ratio as the similarity measure [13] that we minimized using Powell optimization [14]. To obtain a more robust result we used the following procedure. First of all, we registered all volumes to a single volume, this was done to obtain an average volume. This average volume was first put on the mid-sagittal plane by registering it with a flipped version (see [15]). Subsequently it was made to be symmetrical by taking the mean of itself with a flipped version around the mid-sagittal plane. Finally all volumes were matched to this symmetric average volume.

Intensity Normalization. HMPAO brain uptake is proportional to regional cerebral blood flow. However absolute measurement is not possible and only the relative distribution of the tracer can be studied. The cerebellum is usually not affected in AD and can therefore be chosen as reference area. We normalized the intensities by dividing by the summed intensity in the cerebellum. This was done automatically by using a template on the registered volumes.

Classification. Because the hypo-perfusion pattern for early AD is not very well defined we chose to develop a method where we did not use any explicit knowledge about the typical perfusion patterns. We used implicit knowledge about the perfusion patterns by using a database of images of AD patients and normal subjects.

To separate the images we used a classifier using the voxel intensities as features and this database to train the classifier. Using the voxel intensities as features makes it possible not to introduce any particular knowledge about the exact location of the hypo-perfusion area(s). Thus by using a database of images and the voxel intensities we circumvent the problem of the exact definition of the typical perfusion pattern for early AD.

In general the number of images available in the training databases is significantly smaller (< 100) than the number of voxels (> 1000). Thus the number of features (voxels) is much larger than the number of samples (training images). The number of samples is considered to be small if it is about the same or smaller than the number of dimensions. In this case we speak of almost empty spaces, the small sample size problem or the so called curse of dimensionality. In classical pattern recognition it was believed that no good generalization could be obtained for this cases. Generalization is the capacity of a classifier to rightly classify a sample never seen before.

Even though, traditionally it was thought that due to the curse of dimensionality it is necessary to fill the feature space with many more objects than its dimensionality in order to obtain a classifier that generalizes well. Recently it has become clear however, that there are several ways to construct classifiers in almost empty spaces (see the following review [16]).

We chose to use the pseudo-Fisher linear discriminant classifier (PFLDC) as it has already been applied to the small sample size problem for some non-medical applications [17][18][16]. We also did some experiments with the nearest mean classifier (NMC)[19].

All experiments carried out in this article (except noted otherwise) were carried out using the leave-one-out approach to have an independent test and training set, and thus this provides an unbiased estimate of the classification performance.

3 Materials

Simulations. We chose to use a so called photon simulator to generate simulated images. We used the freely available SimSET simulation package developed by the University of Washington Imaging Research Laboratory. More information about this package as well as further references can be found on its website¹.

The freely available Zubal phantom [20] was used to define the activities and the attenuation coefficients for each segmented region. Instead of using this phantom directly to generate volumes, we chose to deform it, to simulate the differences in morphology that exist between different subjects. To obtain realistic deformations we used the free-form deformation fields as generated by the Pasha algorithm (see [21] when registering the phantom with 23 MRI volumes of healthy subjects).

To simulate hypo-perfusion areas we lowered the activity by 20 % percent in the following areas: 1) internal temporal, 2) cingulum and precuneus, 3)

¹ http://depts.washington.edu/~simset/html/simset_main.html

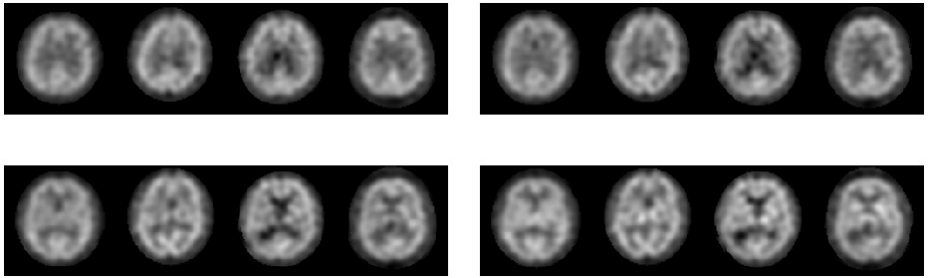


Fig. 1. Examples of four volumes from the Cologne center after intensity and spatial normalization. In each column the first two small images show two normal subjects, the last two images show slices of AD subjects. The sets of slices are ordered from left to right and from top to bottom. Strong hypo-perfusion can be seen for the first AD patient, whereas the hypo-perfusion is more subtle for the second patient.

temporal-parietal region. These regions were manually segmented in the Zubal atlas because the segmentation of the atlas was not fine enough. The three regions had respectively a volume of 72319 mm^3 , 32010 mm^3 and 18642 mm^3 . For each simulated AD image, only one region was hypo-perfused.

We simulated 13 control volumes and 10 volumes per hypo-perfusion region and value (2 hypo-perfusion values times 3 regions times 10 images, gives 60 AD volumes). The volumes were reconstructed using filtered-back projection, without attenuation correction. The reconstructed volumes had a voxel size of 2.5 mm by 2.5 mm by 3.6 mm.

Real Data. The real images we used for our experiments were taken from a concurrent study investigating the use of SPECT as a diagnostic tool for the early onset of AD. A detailed description of this data can be found in [22]. Subjects of four different centers, Edinburgh (Scotland), Nice (France), Genoa (Italy), and Cologne (Germany) were included for this study. In total 158 subjects participated, including 99 patients with AD, 28 patients suffering from depression (not used in this article), and 31 healthy volunteers. An example of this data is seen in figure 1.

Applying the registration procedure as described above resulted in images of 128 by 128 by 89 voxels, with a voxelsize of 1.71 mm by 1.71 mm by 1.88 mm for all four centers.

Experts. All real images were rated in four categories (very probable, probably, probably not and very unlikely to have AD) by sixteen European expert nuclear medicine physicians .

To be able to compare the data from the experts with that of the automatic methods, we considered the first two ratings as positive and the other two as negative. The different observers turn out to have quite different sensitivities and specificities therefore we will also just show the results of the Nice experts.

4 Experiments

Simulated Data. We tested the nearest mean classifier (NMC) as well as the Fisher linear discriminant classifier (PFLDC) on the simulated data. The images were subsampled by a factor two in each dimension as to obtain a number of voxels that does not need excessive amounts of computing time and computer memory. This resulted in 9150 features.

The FLDC was able to classify the images with the following results per region 1: 90.0%, 84.6% 2: 70.0%, 76.9% 3: 80.0%, 84.6% sensitivity and specificity. The NMC classifier had slightly worse results. However SPM was not able to detect the hypo-perfusion areas at all, different significance levels were tried, at each level SPM showed approximately the same number of regions containing hypo-perfusions (true positives) as not containing hypo perfusions (false positives).

Real Data. Figure 2 shows the results on all the real data using the leave-one-out error estimate. The right side of the figure shows results when training on just data from one site and using the classifier for the remaining data. The FLDC ROI approach consisted of using a regions of interest approach (21 regions) similar to that found in literature [8][9] defined by the Zubal phantom and carrying out classification using the FLDC by using the mean intensities of those areas as features. In the SPM cluster approach we used SPM at a significance level of 0.1 at the cluster level. We considered each image where some significant clusters were found to be a positive result. The FLDC column shows the results for the Fisher linear classifier on the images after a subsampling of a factor two in each dimension resulting in 26950 features. All experts show the results of all 16 experts, whereas Nice experts show only the four experts from Nice.

On all the data the FLDC is statistically significantly more sensitive than the experts from all centers ($p < 3.10^{-10}$), however the experts are significantly more specific than the FLDC ($p < 0.05$). When training the FLDC on the data from Edinburgh and testing on the remaining data the FLDC is nearly significantly more specific ($p < 0.07$) but there is no significant difference in sensitivity.

The difference in the results seen between the left and the right of figure 2 are due to the much smaller training set in the right case and also probably due to differences in the acquisition protocols between the different sites. However note that the classification approach still provides useful results even when using a small training set not acquired in exactly the same way. This shows that for clinical use larger databases are needed, however that classification of images with different acquisition parameters is possible.

5 Conclusion

Based on the experiments described in this article we conclude that automatic approaches to the classification of images perform at least as well as human observers. In general our automatic method is more sensitive while still being

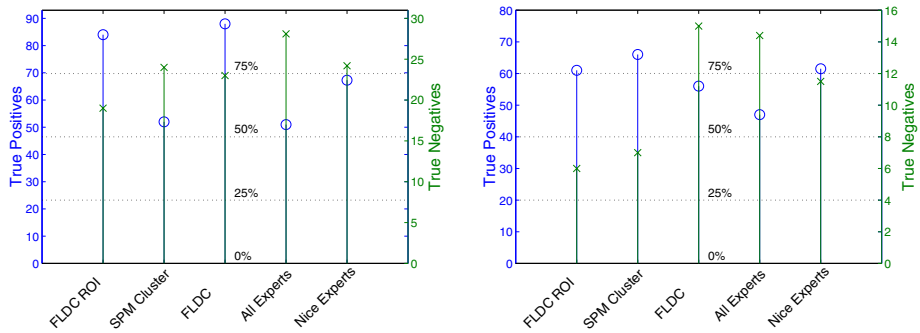


Fig. 2. Left: Graphical representation of the main results when using **all** the real data (AD: 93, normal: 31 cases), Right: results of testing on **all data except Edinburgh** (80 AD, 16 controls) when training on the data from the Edinburgh center only. The number of true positives is given by circles. The number of true negatives is given by crosses. Because the axes are scaled to the number of AD (93) and normal (31) images, sensitivity (circles) and specificity (crosses) are indicated by the dotted horizontal lines.

sufficiently specific. One would need more data, especially of control subjects to be able to state that automatic methods outperform human observers.

We have shown that classification of images using the voxel values as features equals or outperforms the other automatic methods. Thus classification without using any specific knowledge related to the pathology is shown to be possible.

The automatic methods used in this article, with the exception of SPM give only global information on the image. The best of them do outperform SPM. However only providing global information might not be sufficient for clinicians. Therefore in practice, results as those obtained by SPM might be used as additional information to the global classification.

For future work one might want to try the presented approach for differential diagnosis (e.g. AD versus Pick's disease) which might be an even more important clinical issue. ROC analysis of the classifier as well as of the experts will be useful to better compare performances, especially when observing the probable differences in position on the ROC curves (operating points) in the current study between the experts (see e.g. sensitivity differences between all the experts and a given group of experts only).

References

1. G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E.M. Stadlan. Mental and clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7):939–944, July 1984.
2. M.F. Folstein, S.E. Folstein, and P.R. McHugh. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198, November 1975.

3. K.M. Gosche, J.A. Mortimer, C.D. Smith, W.R. Markesbery, and D.A. Snowdon. Hippocampal volume as an index of Alzheimer neuropathology: findings from the Nun Study. *Neurology*, 58(10):1476–1482, May 2002.
4. I. Goethals, C. van de Wiele, D. Slosman, and R. Dierckx. Brain SPET perfusion in early Alzheimer's disease: where to look? *European Journal of Nuclear Medicine*, 29(8):975–978, August 2002.
5. W.A. Van Gool, G.J. Walstra, S. Teunisse, F.M. Van der Zant, H.C. Weinstein, and E.A. Van Royen. Diagnosing Alzheimer's disease in elderly, mildly demented patients: the impact of routine single photon emission computed tomography. *Journal of Neurology*, 242(6):401–405, June 1995.
6. D. Kogure, H. Matsuda, T. Ohnishi, T. Asada, M. Uno, T. Kunihiro, S. Nakano, and M. Takasaki. Longitudinal evaluation of early Alzheimer's disease using brain perfusion SPECT. *Journal of Nuclear Medicine*, 41(7):1155–1162, July 2000.
7. H. Braak and E. Braak. Diagnostic criteria for neuropathologic assessment of Alzheimer's disease. *Neurobiology and Aging*, 18(4):S85–S88, July 1997.
8. M.R. Dawson, A. Dobbs, H.R. Hooper, A.J. McEwan, J. Triscott, and J. Cooney. Artificial neural networks that use single-photon emission tomography to identify patients with probable Alzheimer's disease. *European Journal of Nuclear Medicine*, 21(12):1303–1311, December 1994.
9. D. Hamilton, D. O'Mahony, J. Coffey, J. Murphy, N. O'Hare, P. Freyne, B. Walsh, and D. Coakley. Classification of mild Alzheimer's disease by artificial neural network analysis of SPET data. *Nuclear Medicine Communications*, 18(9):805–810, September 1997.
10. R.S.J. Frackowiak, K.J. Friston, C.D. Frith, and R. Dolan. *Human Brain Function*. Academic Press, 1997.
11. J. Ashburner, K. Friston, A. Holmes, and J.-B. Poline. Statistical Parametric Mapping, SPM'99. The Wellcome Department of Cognitive Neurology. Institute of Neurology, University College London, 1999. Freely available at <http://www.fil.ion.ucl.ac.uk/spm>.
12. J. Stoeckel, G. Malandain, O. Migneco, P.M. Kouloubaly, P. Robert, N. Ayache, and J. Darcourt. Classification of SPECT images of normal subjects versus images of Alzheimer's disease patients. In *Proc. MICCAI'01*, volume 2208 of *LNCS*, pages 666–674, October 2001.
13. A. Roche, G. Malandain, X. Pennec, and N. Ayache. The Correlation Ratio as a New Similarity Metric for Multimodal Image Registration. In W. M. Wells, A. C. F. Colchester, and S. Delp, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, volume 1496 of *Lecture Notes in Computer Science*, pages 1115–1124, Boston, USA, October 1998.
14. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1997.
15. S. Prima, S. Ourselin, and N. Ayache. Computation of the mid-sagittal plane in 3D brain images. *IEEE Transaction on Medical Imaging*, 21(2):122–138, February 2002.
16. R.P.W. Duin. Classifiers in Almost Empty Spaces. In *Proceedings of the 15th International Conference on Pattern Recognition (ICPR'00)*, volume 2, pages 1–7, 2000.
17. J. Shürmann. *Polynomklassifikatoren für Zeichenerkennung*. R. Oldenburg Verlag, 1977.

18. S.J. Raudys and R.P.W. Duin. Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix. *Pattern Recognition Letters*, 19:385–392, 1998.
19. R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. John Wiley & Sons, 2nd edition, 2001.
20. I.G. Zubal, C.R. Harrell, E.O. Smith, Z. Rattner, G. Gindi, and P.B. Hoffer. Computerized 3-Dimensional Segmented Human Anatomy. *Medical Physics*, 21(2):299–302, February 1994.
21. P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache. Iconic Feature Based Nonrigid Registration: The PASHA Algorithm. *Computer Vision and Image Understanding*, 89(1-2), February 2003.
22. D. Soonawala, T. Amin, K.P. Ebmeier, J.D. Steele, N.J. Dougall, J. Best, O. Migneco, F. Nobili, and K. Scheidhauer. Statistical parametric mapping of (99m)Tc-HMPAO-SPECT images for the diagnosis of Alzheimer's disease: normalizing to cerebellar tracer uptake. *Neuroimage*, 17(3):1193–1202, November 2002.