A Comparative Study of Model-Based Brain MRI Tissue Classification Methods

Chris A. Cocosco, Alex P. Zijdenbos, Noor Kabani, and Alan C. Evans

McConnell Brain Imaging Centre, Montreal Neurological Institute,
McGill University, Montreal, Canada

email: crisco@bic.mni.mcgill.ca

Abstract. A comparative study of the performance of two different approaches to fully automatic (model-based) brain anatomical MRI tissue classification is presented in this paper. Both simulated and real image data, as well as various subject brain morphologies (young-healthy, elderly, and diseased individuals) were used. The recently proposed "MNI" classification method was found to perform better, and to be more robust, than some of the existing state of the art for this application. Moreover, presented experimental results question the popular assumption that tissue class multi-spectral aMRI intensities have a Normal (Gaussian) distribution.

1 Introduction

Many kinds of computerized analyses can be used to extract information from three-dimensional (3D) MRI data of the human head. The application that concerns this paper is the classification, or labeling, of individual voxels of a 3D anatomical MR image (aMRI) as one of the main tissue classes in the brain: cerebro-spinal fluid (CSF), grey matter, and white matter; a fourth class is defined as "background", denoting everything else: other tissues, as well as air. An accurate and robust tissue classification is the basis for many applications such as: quantitative measurements of tissue volume in normal and diseased populations, morphological analysis (for example, of cortex folding patterns), or visualization. Fully automatic, *robust* tissue classification is required for batch processing the data from large-scale, multi-site clinical trials or research projects, such as [1].

Classification methods operate in a multi-dimensional feature space. Each feature consists of an image intensity at the spatial location (voxel) to be classified. All the features are derived from the same subject (for example, using multi-spectral MRI). An aspect that is ignored by most brain MRI classification schemes is how to fully automatically adapt to a new MRI brain dataset, possibly originating from a new site and MR scanner[1]. Here we present a quantitative comparative study of a method we recently developed [2] against another

[1] The aMRI intensity scale has no absolute meaning; the image values and contrast are dependent on many variable acquisition and post-processing parameters.

method [3] representative for the existing state of the art in fully automatic brain tissue classification.

2 Methods

Expectation-Maximization (EM) is a popular statistical classification scheme for this application. Originally proposed in a brain MRI context by Wells [4], and subsequently improved by others, these methods interleave intensity non-uniformity (INU) field estimation and classification, in an iterative fashion. Recent improvements [3, 5] fully automate this classification scheme by employing a probabilistic brain tissue atlas (a "model"), defined in a "stereotaxic space" (a standard, brain-based coordinate system).

In this comparative study we used the SPM'99 tissue classification (segmentation) software [3]. It employs an EM-style scheme. This software is easily accessible [6] (and at no-cost), easy to use, and popular in the neuroimaging research community. In practice, two of the assumptions made by SPM are questionable:

1. The brain subjected to classification is morphologically similar to the probabilistic model. The iterative approach can compensate for small deviations from the model, which is used for automatic initialization. However, the prior tissue probability at the current spatial location (given by the model) is considered in the classification decision. Consequently, anatomies significantly different than the model can lead to errors.
2. Same as most classification methods proposed in the literature, SPM models the tissue MR image intensity distributions as multi-variate Normal (Gaussian) distributions in feature space.

The tissue classification method recently developed at the Montreal Neurological Institute [2] (to be referred to as the "MNI method") does not rely on the above two assumptions. First, while the MNI method is also model-based, the probabilistic model is only used for initialization, and the method is designed to adapt to significant morphological deviations from the model. Second, this method is non-parametric: it does not assume a particular shape of the tissue class clusters in feature space.

Our study focused on multi-spectral MRI classification experiments. Using more than one MRI contrast of the same subject is attractive because it improves the cluster separation in feature space, especially in the presence of significant imaging artifacts (the noise and the INU are un-correlated between acquisitions).

A comprehensive qualitative visual evaluation of the classification result by a human observer is subjective and difficult, considering the morphological complexity and variability of the human brain. Given a reference classification (a "gold standard"), one can compute a quantitative measure of similarity between the classification result and this standard. We used *Kappa*, which is a chance-corrected similarity measure between two labelings [7]; its maximum value of 1 corresponds to a perfect agreement, and a value of 0 corresponds to agreement due to chance alone. The Kappa was also used by others [1, 3] for this application.

3 Atypical anatomies

Both SPM'99 and the MNI method use probabilistic brain tissue atlases (stereotaxic space tissue probability maps) that were constructed from a young normal (healthy) population. To investigate the robustness against morphological deviations from the probabilistic model, we tested the classification methods on elderly, and on diseased subjects — aging and pathology typically cause significant deviations from a young-normal model.

3.1 Simulated Elderly Brain MRI

These data were produced using a MRI simulator [8] that produces realistic synthetic MRI images based on an anatomical model (a "phantom"). This model was used as the gold standard for computing the *Kappa* figure of merit. 10 different "phantoms" resembling elderly brains (Fig. 1) were created as follows:

1. 10 individual T1 scans were selected from a large database[2]. These individuals (5 males, 5 females) were 60-70 years old, and from the same population.
2. These individual scans were non-linearly spatially registered [9] to a previously available standard phantom [8].
3. The resulting deformation field was inverted and used for deforming the standard phantom, such that it looks similar to the source individual brain.

Then, T1, T2, and PD MRI-s were simulated as $1\,mm$ isotropic voxel acquisitions, with 3% noise and 20% INU (typical artifact amounts). The quantitative classification results on these data are presented in Fig. 2. It should be noted that SPM had an advantage in this experiment: the MNI method did not employ an INU correction procedure; SPM did. In a practical image analysis system, such a correction method (e.g. [10]) is typically employed, thus the INU level may be less than 20%. Also, these simulations [8] model the tissue MR properties as homogeneous throughout the brain. In practice, this is not the case.

3.2 Real Data: Alzheimer's Disease Elderly Patients

Multi-spectral MRI scans (T1-T2-PD) were acquired for 11 elderly patients (aged > 60) diagnosed with Alzheimer's Disease (which causes cortical atrophy), as part of a clinical study. The T1 was acquired at $1\,mm^3$ isotropic resolution. The T2 and PD were acquired as 8 interleaved acquisitions, each with 8 mm sagittal slices ($1\,mm^2$ in slice resolution); the 8 scans were then spatially co-registered, then averaged together in order to obtain a single final 3D image with improved resolution. All three modalities were registered (and re-sampled, at $1\,mm^3$ resolution) to a stereotaxic space using 9-parameter linear registration software. Also, INU correction was performed using N3 [10]. Unlike the MNI method (TPM $\tau = 0.90$), SPM gave poor results on some of the subjects (2 out of 11): Fig. 3.

[2] Data source: Dr. Ryuta Kawashima, Sendai, Japan.

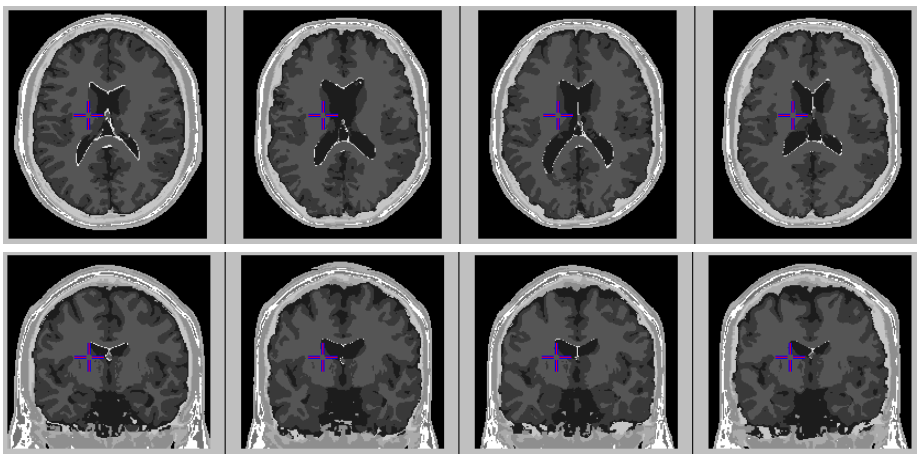

Fig. 1. *(Left to Right)* standard phantom (column 1), and three sample "elderly brain" phantoms (columns 2-4). Each tissue is represented as a different gray shade. Compared to the standard phantom (constructed from a young-normal scan), the "elderly" phantoms exhibit enlarged ventricles and overall brain atrophy (typical of aging brains).

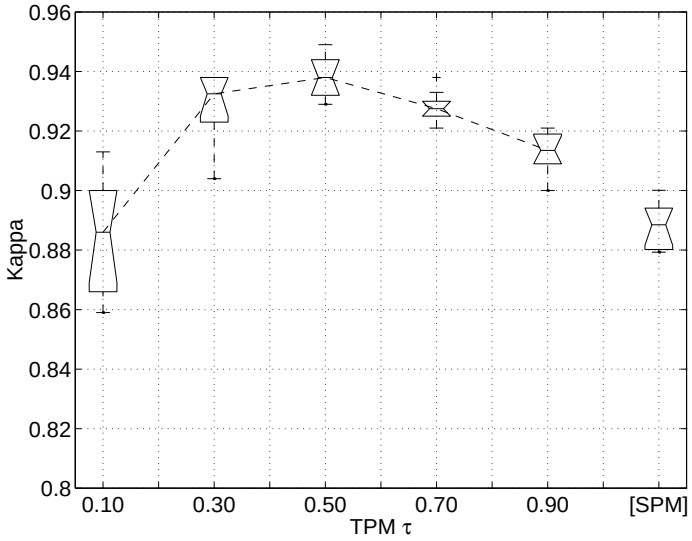

Fig. 2. Simulated (elderly brain) multi-spectral MRI: Kappa figure of merit for tissue classification. 10 repetitions, each with a different "phantom" (anatomical model). *Left:* MNI method (for different values of its parameter "TPM τ" [2]). *Right:* SPM. Note that MNI performs better for all $\tau \geq 0.3$. A Wilcoxon signed rank test for equality of medians (matched samples) between SPM and MNI ($\tau = 0.9$) gives $p = 0.002$.

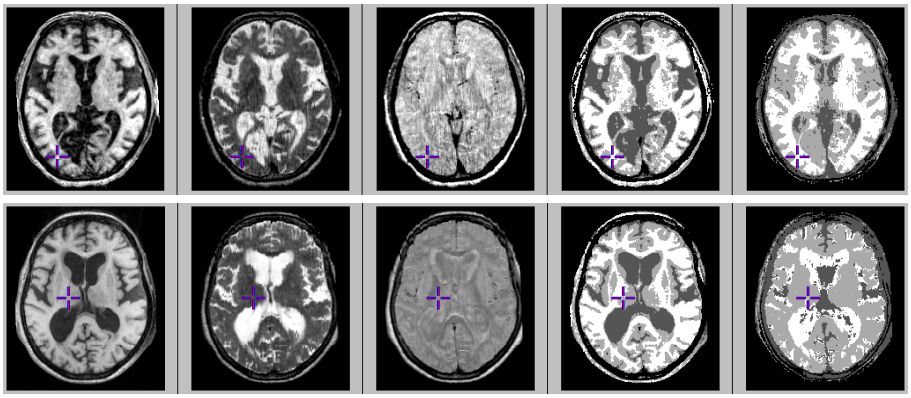

Fig. 3. Two elderly Alzheimer's Disease patients (*top:* #9, *bottom:* #11) poorly classified by SPM. (*left to right:*) T1, T2, PD MRI-s, MNI classification result, SPM classification result (in increasing grey brightness: background, CSF, grey-, white-matter).

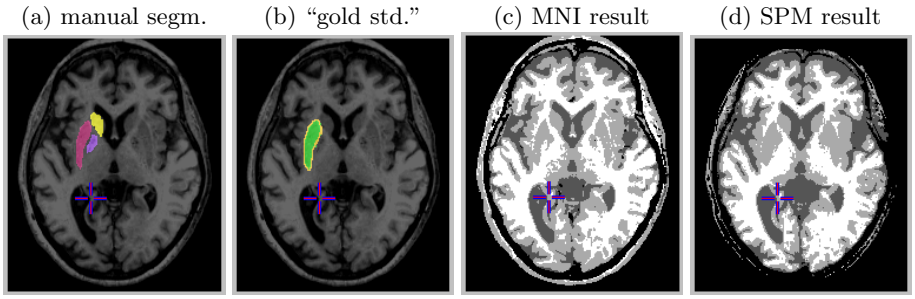

Fig. 4. Elderly Alzheimer's Disease patient #3: *(a)* manual segmentation, and *(b)* the derived "gold standard" superimposed on the T1 MRI; *(c,d)* classified images.

For a quantitative classification evaluation, a trained neuroanatomist (A.D.) manually segmented, using the T1 3D image, the volume of the following left basal ganglia grey-matter structures: Putamen, Globus Pallidus, Caudate, and Nucleus Accumbens [11]. Kappa was computed against the manual segmentation (considered the "gold standard") on a region of interest (ROI) obtained by a 3D "dilate" morphological operation of the manually segmented putamen volume (Fig. 4). Since the putamen is surrounded by white-matter (with the possible exception of "bridges" to the other three segmented basal ganglia grey structures), this ROI allows for a local evaluation of the classification of the grey-white boundary around the putamen (Fig. 5).

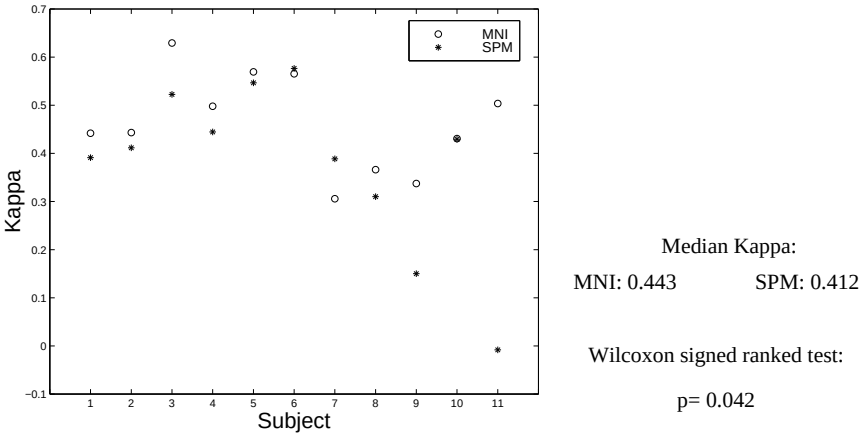

Fig. 5. Multi-spectral MRI, elderly diseased patients: local Kappa figure of merit for tissue classification around the left putamen (Fig. 4). Note the poor SPM results for patients #9 and #11 (see also Fig. 3). MNI gives higher kappa on 9 out of 11 subjects.

4 Feature Space Data Distributions

If the features used by the classifier are MR signal intensities from various MRI modalities (such as T1, T2, PD), then the multi-variate "Normal" (Gaussian) distribution can be a poor model for the data distributions in the (multi-dimensional) feature space. Besides biological causes such as the intrinsic heterogeneity within the tissue classes, the MRI acquisition artifacts also affect the intensity distributions.

A real multi-spectral MRI dataset of a 36 year old normal man was used to study the feature space distributions. The T1-weighted, $1\,mm$ isotropic voxel, 3D scan was completely manually segmented by a human expert neuroanatomist (N.K.). T2 and PD scans were also acquired in two interleaved acquisitions, as $2\,mm$ thick sagittal slices ($1\,mm^2$ in plane); the two paired scans were spatially co-registered and averaged together in order to improve the image resolution. Then, all MR image data was pre-processed as previously described (section 3.2).

In order to experimentally study their effect on feature space distributions, two MRI artifacts were artificially added to this MRI dataset (Fig. 6): an additional multiplicative INU field (estimated from real MR data), and increased partial volume effect corresponding to thicker acquisition slices (simulated by blurring with a 1-dimensional box smoothing kernel).

Using the full-brain (except cerebellum) manual segmentation as "gold standard", a comparative quantitative classification experiment was also performed on this dataset (Fig. 7). It should be noted that, unlike in the other experiments presented here, this subject is from the same population represented by the probabilistic models (young normals). Hence, the better performance of the MNI method over the SPM method is likely due to its non-parametric design.

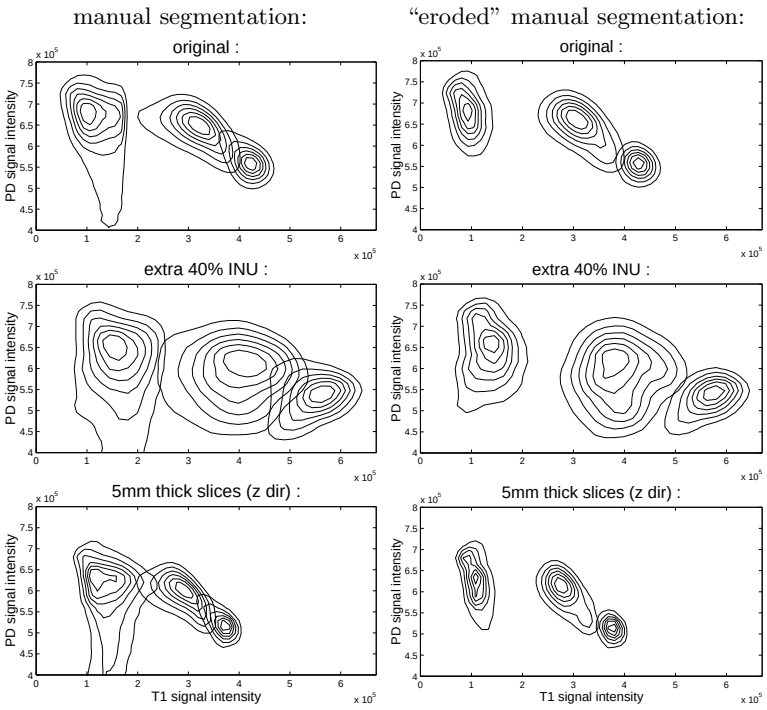

Fig. 6. Feature space probability densities for the three classes (*left to right*): CSF, grey matter, white matter. Densities (represented as iso-contours at equally spaced values between 0 and the cluster's maximum) are estimated from real T1+PD MRI using a full-brain manual segmentation. The INU and the thicker slices noticeably make the clusters further deviate from the Normal shape. A Doornik-Hansen test for multi-variate normality [12] on the grey matter cluster gives $p \approx 0$.

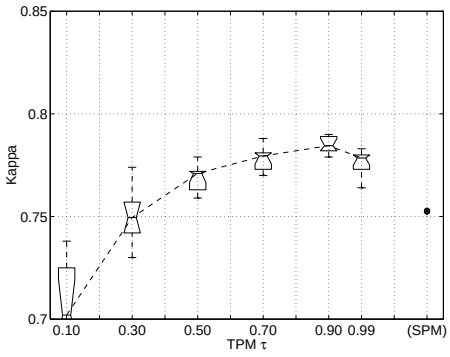

Fig. 7. Real young-normal multi-spectral MRI dataset: Kappa figure of merit for tissue classification. *Left:* MNI method (for different parameters τ [2]), 10 repetitions for each τ. *Right:* SPM. MNI gives a statistically significant improvement over SPM for $\tau \geq 0.5$.

5 Conclusion

We presented a comparative study of a recently developed method for model-based fully automatic brain MRI tissue classification ("MNI" [2]) against another method ("SPM'99" [3]) representative for the existing state of the art. The study was done on both real and simulated multi-spectral data.

The MNI method proved to be more robust against variations in the brain morphology, and in the MRI data. Moreover, in all experiments the MNI method gave a statistically significant improvement over the SPM method in a quantitative figure of merit: a similarity measure computed against a reference classification. In addition, we presented results that question the assumption that the tissue class aMRI intensities have a Normal (Gaussian) distribution. Even if this assumption would be acceptable for some data, it is safer not to make it if the automatic classification method aims to be robust against variability in the imaging data quality. Robustness is especially important for large scale, multi-site data collection research projects or clinical trials.

Acknowledgements: Adina Derrick, Jean-François Mangin, John Sled, Steve Robbins, Peter Neelin, Louis Collins.

References

1. Zijdenbos, A., Forghani, R., Evans, A.: Automatic quantification of MS lesions in 3D MRI brain data sets: Validation of INSECT. In Wells, W.M., Colchester, A.C.F., Delp, S., eds.: Proc. of MICCAI'98. Volume 1496 of LNCS., Springer-Verlag (1998) 439–448
2. Cocosco, C.A., Zijdenbos, A.P., Evans, A.C.: Automatic generation of training data for brain tissue classification from MRI. In Dohi, T., Kikinis, R., eds.: Proceedings of Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2002), Tokyo. Number 2488 in LNCS, Springer-Verlag (2002) 516–523
3. Ashburner, J., Friston, K.J.: Voxel-based morphometry — the methods. NeuroImage **11** (2000) 805–821
4. Wells III, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of MRI data. IEEE Transactions on Medical Imaging **15** (1996) 429–442
5. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE Transactions on Medical Imaging **18** (1999) 897–908
6. : Statistical parametric mapping (spm'99). http://www.fil.ion.ucl.ac.uk/spm/ (2000)
7. Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological measurements **20** (1960) 37–46
8. MNI: Brainweb – Simulated Brain Database. http://www.bic.mni.mcgill.ca/brainweb/ (1997)
9. Collins, D., Evans, A.: Animal: validation and applications of non-linear registration-based segmentation. International Journal of Pattern Recognition and Artificial Intelligence **11** (1997) 1271–1294

10. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Transactions on Medical Imaging **17** (1998) 87–97
11. Buchsbaum, M., Fallon, J., Wei, T., Guich, S., Spiegel-Cohen, J., Hamilton, M., Tang, C.: A method of basal forebrain anatomical standardization for functional image analysis. Psychiatry Research **84** (1998) 113–125
12. Doornik, J.A., Hansen, H.: An omnibus test for univariate and multivariate normality. http://www.nuff.ox.ac.uk/Users/Doornik/ (1994) (working paper).