

Patient Classification of fMRI Activation Maps

James Ford¹, Hany Farid^{1,2}, Fillia Makedon¹, Laura A. Flashman³, Thomas W. McAllister³, Vasilis Megalooikonomou⁵, and Andrew J. Saykin^{3,4}

¹ Computer Science Department, Dartmouth College, Hanover, NH, USA
jford,farid,makedon@cs.dartmouth.edu

² Center for Cognitive Neuroscience, Dartmouth College

³ Brain Imaging Laboratory, Department of Psychiatry, Dartmouth Medical School

⁴ Department of Radiology, Dartmouth Medical School

⁵ Department of Computer and Information Sciences, Temple University, Philadelphia, PA

Abstract. The analysis of brain activations using functional magnetic resonance imaging (fMRI) is an active area of neuropsychological research. Standard techniques for analysis have traditionally focused on finding the most significant areas of brain activation, and have only recently begun to explore the importance of their spatial characteristics. We compare fMRI contrast images and significance maps to training sets of similar maps using the spatial distribution of activation values. We demonstrate that a Fisher linear discriminant (FLD) classifier for either type of map can differentiate patients from controls accurately for Alzheimer’s disease, schizophrenia, and mild traumatic brain injury (MTBI).

1 Introduction

There is evidence for different fMRI activation patterns in Alzheimer’s disease, schizophrenia, and mild traumatic brain injury (MTBI), suggesting that diagnosis of these conditions may eventually be possible on the basis of brain activations. We present an approach for classifying patients and healthy controls for these three conditions based on fMRI brain activation maps generated using the statistical parametric mapping (SPM) approach [5]. The use of individually generated activation maps has two advantages: first, it allows for better scalability to very large subject pools (*e.g.*, hundreds or thousands of subjects), and

Work by Ford, Makedon, Saykin, and Megalooikonomou was supported by NSF grant IDM-0083423. Farid’s work was supported by an Alfred P. Sloan Fellowship and NSF CAREER Award IIS-99-83806. Work by Ford, Farid, and Makedon was also supported by a departmental NSF Infrastructure Grant (EIA-98-02068). The activation data and brain masks were obtained with the help of Annette Donnelley, B.S., Brenna C. McDonald, Psy.D., Robert M. Roth, Ph.D., and Molly Sparling, B.A., in the Brain Imaging Lab, and work was carried out with the assistance of Sanket Agarwal, Li Shen, and Song Ye in the Computer Science department. Imaging was supported by grants from NARSAD, the Alzheimer’s Association, NINDS, NIA, NIDRR, the Ira DeCamp Foundation, and New Hampshire Hospital.

second, it has the potential to integrate data at the activation map level that would be technically difficult to combine at the raw data level (*e.g.*, due to use of different scanners).

There have been several demonstrations of differences in the brain activations of Alzheimer’s patients compared to controls, including [15]. In schizophrenia, differences in fMRI activation have been found in several brain regions, even in the absence of measurable differences in task performance [14]. Finally, mild traumatic brain injury (MTBI) has been shown to lead to detectable changes in fMRI activation patterns [11].

Earlier studies have demonstrated the feasibility of doing post-analysis on brain activation maps, thereby allowing comparison or combinations of activations between independently-analyzed subjects. For example, in the work of Coulon *et al.* [4], a multi-scale approach is used to describe Positron Emission Tomography (PET) activation maps for a simple motor task using a standardized representation of their structure. Liow *et al.* [10] have demonstrated that a principal component analysis (PCA)/Fisher’s linear discriminant (FLD) approach can be used to classify PET scans of HIV positive patients and healthy controls. That work used 75 scans from 30 HIV+ or AIDS patients and 41 healthy controls, and employed an approach called the scaled subprofile model (SSM) [12] to decompose activations in each voxel into global and subject-specific components: $v_{iq} = g_q(r_i + z_{iq} + \epsilon_{iq})$, where v_{iq} is the value of voxel i in subject q , g_q is a scaling factor for subject q , r_i is the mean value for voxel i after removing g_q , and $z_{iq} + \epsilon_{iq}$ is the subject-specific component (plus noise). PCA is applied only to the second component, and the resulting principal components are termed subject scaling factors (SSFs). Note that PET differs from fMRI in recording a positive activity level for each voxel under a single condition rather than a positive or negative change in activity between two conditions.

There have been some reports in the fMRI literature about the difficulty of reproducing voxel-level significance maps. For example, even the 2% most significant voxels have been found to vary considerably across runs, subjects, and analysis techniques [18]. However, reproducibility of fMRI activations at a regional level has been found to be good across sites, subjects, and techniques [2]. For these reasons it is likely to be more fruitful to compare patterns across activation maps, rather than the location of significance peaks.

In the following sections, we present a method for using a training set of fMRI activation maps to build classifiers for specific brain conditions. The classifiers evaluate activation maps according to their degree of similarity to the maps in the training set and assign them probabilities of being “patient” or “non-patient”. The next section describes the fMRI activation data and methods used.

2 Materials and Methods

Three datasets were used for testing, one consisting of elderly patients diagnosed with early Alzheimer’s disease and associated elderly controls, a second of schizophrenic patients and healthy controls, and a third of patients

who had suffered mild traumatic brain injury (MTBI) and associated controls. The Alzheimer’s disease dataset consists of 9 Alzheimer’s disease patients and 9 elderly controls. The schizophrenia dataset comprises 15 schizophrenic patients and 10 controls. Finally, the MTBI dataset consists of 11 patients and 6 controls. Subjects were tested on the category-exemplar (catx) word pair task (Alzheimer’s and schizophrenia datasets) or the n-back working memory task (MTBI dataset) as described in [15], [16], and [11], respectively. Each subject’s task-related activation (catx or 3-back) was analyzed individually versus a control condition (rest or 2-back, respectively), resulting in individual maps of fMRI signal change at each voxel.

The Statistical Parametric Mapping (SPM) approach analyzes a time series at each normalized, resampled voxel (in this work, $2 \times 2 \times 2$ mm) independently using a regression analysis, thus creating 3D maps of contrast and statistical significance values for pairs of conditions. Contrast values are estimates of the difference in activation between the conditions: a positive contrast value for a voxel is interpreted as an increase in brain activation for the second condition compared to the first, whereas a negative value is often assumed to reflect a decrease [1, 9]. The statistical test applied in this work is the t-test, which determines the probability that the means of two sets of samples with Gaussian distributions are different. Thresholding these values determines which voxels are likely to have had significant changes between conditions, and with what likelihood, assuming a null hypothesis of no changes between conditions.

Activation maps may be thought of as points in a high-dimensional space where each voxel value is a single coordinate. Since an activation map may consist of hundreds of thousands of in-brain voxels, the dimensionality of the space is too high to use most classification methods directly. The work in this chapter follows the approach of Liow *et al.* [10], in using a dimensionality reduction step to represent essential features in a space of much lower dimensionality.

2.1 Dimensionality Reduction

Dimensionality reduction of high dimensional data is useful for three general reasons: it reduces computational requirements for subsequent operations on the data, it eliminates redundancies in the data, and, in cases with small numbers of samples relative to variables, it reduces the number of dimensions needed to model the available data. All three reasons apply here, and motivate the use of principal components analysis (PCA), a standard method for creating uncorrelated variables from best-fitting linear combinations of the variables in the raw data. This approach is equivalent to finding an orthogonal basis such that the projection onto each successive vector (or “principal component”) is of maximal variance (and uncorrelated with each previous vector).

Since all activation maps are normalized to a template brain, they have voxels in register. Let the number of subjects in a training set be n , and let the length m column vectors \mathbf{v}_i , $i = 1, \dots, n$, represent the activation maps. The matrix D is defined as all the \mathbf{v}_i , *i.e.* a matrix with one subject per row and one voxel per column. Let \hat{D} be D with the mean of each column subtracted. The principal

components are then defined as the eigenvectors, \mathbf{e}_i , of the covariance matrix $C = \hat{D}^T \hat{D}$; that is, $C\mathbf{e}_i = \lambda_i \mathbf{e}_i$, where each eigenvalue λ_i is proportional to the variance in the original data represented by the i th principal component \mathbf{e}_i . For computational efficiency, an alternative formulation is preferable in cases like this one where the number of rows (subjects) is far exceeded by the number of columns (voxels). Since the inherent geometric dimensionality of n points of dimension m must be less than n , one can define $C' = \hat{D}\hat{D}^T$, and then relate the eigenvectors $\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n$ of C' to those of C as follows: $\mathbf{v}_i = \frac{\hat{D}^T \mathbf{v}'_i}{\|\mathbf{v}'_i\|}$ for $1 \leq i \leq n$ (the eigenvalues are identical). This allows calculation of the eigenvectors of the $m \times m$ covariance matrix C to be replaced with calculation of the eigenvectors of the smaller $n \times n$ matrix C' .

Typically, many principal components account for only small fractions of the total variance. Dimensionality reduction in addition to the above can be accomplished by sorting components by decreasing eigenvalue and then discarding some of the trailing components (there can be at most $\min(m, n)$ non-zero eigenvalues). The selected principal components form the columns of a matrix P , which is a basis for representing activation data. Projection of the activation matrix D onto this basis is carried out as $\tilde{D} = DP$. Following projection, the reduced representations \tilde{D} can be used to build a classifier.

2.2 Classification

Fisher’s linear discriminant (FLD) is a multi-class linear pattern classification method that is optimal for maximizing the separation of Gaussian populations with a common covariance matrix [7]. FLD finds a linear projection of a training set (consisting of c classes) onto $c - 1$ new dimensions such that a measure of “discriminability” is maximized: the ratio of between-class and within-class variability. Since a simple medical diagnosis classification requires only two classes (patients and controls), only a two-class FLD is described here. The number of variables that can be considered in FLD is limited to at most $N - c$ for N exemplars from c classes; thus, it is necessary to use dimensionality reduction or feature extraction in many cases.

Denote as column vectors \mathbf{x}_i , $i = 1, \dots, n_x$ and \mathbf{y}_j , $j = 1, \dots, n_y$ the exemplars from the two classes making up the training set, which are the transposed rows of the matrix \tilde{D} derived in Section 2.1. The within class means $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ are defined as $\boldsymbol{\mu}_x = \frac{1}{n_x} \sum_{i=1}^{n_x} \mathbf{x}_i$ and $\boldsymbol{\mu}_y = \frac{1}{n_y} \sum_{j=1}^{n_y} \mathbf{y}_j$ for the n_x and n_y members of the two classes in the training set. The global training set mean $\boldsymbol{\mu}$ is defined as: $\boldsymbol{\mu} = \frac{1}{n_x + n_y} \left(\sum_{i=1}^{n_x} \mathbf{x}_i + \sum_{j=1}^{n_y} \mathbf{y}_j \right)$. The between-class scatter matrix is then defined as:

$$S_B = n_x(\boldsymbol{\mu}_x - \boldsymbol{\mu})(\boldsymbol{\mu}_x - \boldsymbol{\mu})^T + n_y(\boldsymbol{\mu}_y - \boldsymbol{\mu})(\boldsymbol{\mu}_y - \boldsymbol{\mu})^T. \quad (1)$$

The within-class scatter matrix is given by:

$$S_W = M_x M_x^T + M_y M_y^T, \quad (2)$$

where the i th column of matrix M_x contains the i th exemplar given by $\mathbf{x}_i - \boldsymbol{\mu}_x$. Similarly, the j th column of matrix M_y contains $\mathbf{y}_j - \boldsymbol{\mu}_y$.

For our two-class discrimination, the FLD basis is the generalized eigenvector for S_B and S_W (*i.e.*, $S_B \mathbf{e} = \lambda S_W \mathbf{e}$) with the maximum eigenvalue. Once an FLD basis has been found, a novel map \mathbf{z} , as represented by the reduced dimensionality row vector $\tilde{\mathbf{z}} = \mathbf{z}P$, can be projected onto the same basis: $u = \tilde{\mathbf{z}}\mathbf{e}$. The resulting scalar value u can then be compared to the projections of the training set.

After creating FLD projections, a practical classifier needs a threshold in order to assign class membership. Various strategies exist for selecting a threshold, including selecting the midpoint between the training set projections or using a Bayesian model to identify an optimal transition point based on assumed distributions for the training set classes (*e.g.*, Gaussian).

In order to avoid the assumption of Gaussianity, or similar assumptions necessary to select any particular classification threshold, we use a threshold-free receiver operating characteristic (ROC) approach that examines the effect of setting the decision boundary at various points between the class means. This overcomes the problem of bias in the threshold by examining classification results over many different threshold values, effectively trading off *sensitivity* (the probability patients are correctly identified) for *specificity* (the probability controls are correctly identified). Sensitivity and specificity are estimated at each threshold as the fraction of these states occurring with the selected bound, and are plotted with successive points connected by line segments [7]. The area under the ROC curve (AUC) can then be used to compare overall discriminative power, although this is best done cautiously since it does not reflect the shape of the ROC curve [10]. The area can be interpreted as the probability that an arbitrary patient will appear more “patient-like” than an arbitrary control [6].

Each trial uses an independent FLD classifier, and may assign different values to its training set. In particular, while a two-class FLD typically remaps one class to a positive mean and the other to a negative mean, the assignment of sign is arbitrary. To account for this, when combining leave-one-out FLD test subject projections the following procedure is used: first, the class means for each trial’s training set are calculated and all projections for that trial are scaled so that the difference between means is exactly 2. Second, all projections are shifted so that the class means fall on -1 and 1. Finally, all projections are sign-flipped if necessary so that the patient class’s mean is positive. Test subject projections can then be combined, since they then correspond to identically aligned training sets in the FLD space.

3 Results

In each case, results with varying numbers of principal components (always selected in order of decreasing eigenvalue) were calculated. Because overtraining can occur when the number of classification features is high, simply using all the principal components is generally not advisable. In a practical classifier, a

heuristic selection of the number of principal components is generally used. Note that the maximum number of components is equal to the number of subjects minus three due to the use of FLD and a leave-one-out cross-validation approach.

Classification was tested using both contrast images and t-maps for each dataset, after removing any voxels missing values in any map. Classification accuracy was estimated with a jackknife using the average of N jackknife trials over the N subjects. In each trial, $N - 1$ of the original N exemplars were used for training and the class assignment of the remaining subject was calculated, resulting in an accuracy of 1 (correct) or 0 (incorrect). The jackknifed estimate of accuracy is then simply the mean of the trial accuracies. Each dataset was processed with PCA to reduce dimensionality as described in Section 2.1, after which FLD was used to build a classifier for each training set as described above, with scaling and if necessary sign-flipping.

In Figure 1, the AUC results for varying numbers of principal components are shown for each dataset. Both maps appear to work fairly well for all the datasets. The t-map might be expected to be superior since it accounts for voxel variability, but it has been argued that basing activation comparisons on signal changes rather than significances of changes better reflects underlying neural activity [3].

4 Conclusions

We have described and demonstrated a straightforward technique for classifying spatial patterns in brain activation maps. Our method consists of selecting appropriate voxels, reducing the dimensionality of voxel values using PCA, and creating a classifier using FLD and a training set of labeled activation maps.

This approach is promising, and can give fairly accurate predictions on the datasets considered (although the exact levels of accuracy must, of course, be interpreted cautiously due to the small dataset sizes [13, 17]). The problem of classifying subjects is very likely complicated by the existence of varying effects of the clinical conditions in the patients, which in Alzheimer’s are known to include diffuse “compensatory activations” [15]. In the case of Alzheimer’s disease, aging-related changes in fMRI response are another complicating factor, since individuals can be affected at a range of ages. Schizophrenia may involve different groupings of symptoms, and thus might conceivably involve different changes in brain activity in different patients. Both Alzheimer’s disease and schizophrenia can vary in severity, resulting in possible individual differences, and effects of MTBI are likely to be idiosyncratic as well. As fMRI data has a low signal to noise ratio, activation patterns may not be completely consistent even across the healthy control subjects. Given these factors, the size of the datasets used here is probably minimal, and better prediction will likely result with larger collections of brain activation maps. This is especially important if this technique is to be applied to detecting early signs of Alzheimer’s disease onset, which is more useful but also quite likely more difficult than identifying clinical Alzheimer’s disease.

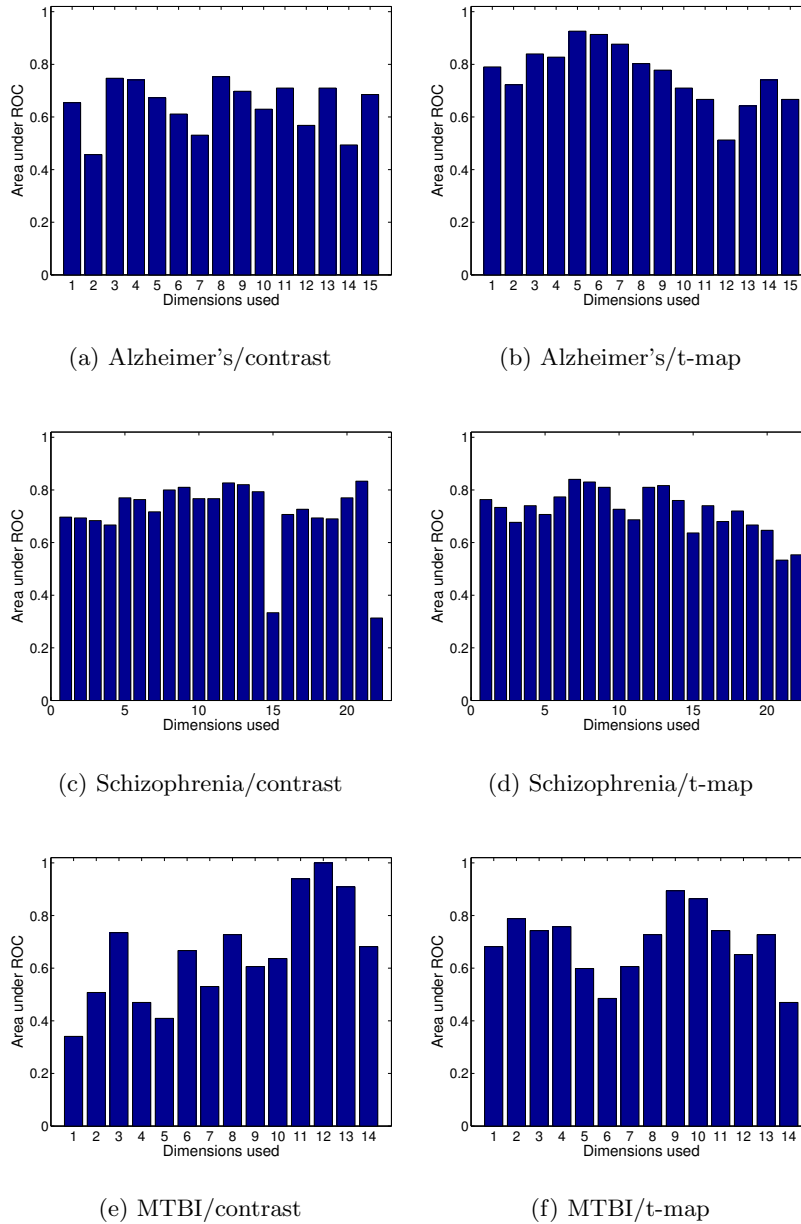


Fig. 1. Discriminative power of FLD classification of activation maps for three datasets, as measured by AUC. The area under the ROC curve (AUC) using leave-one-out FLD classification is shown for varying numbers of principal components of contrast images (left) or significance t-maps (right) for the Alzheimer's/catx, schizophrenia/catx, and MTBI/3back>2back datasets. The AUC can be interpreted as the probability that an arbitrary patient will appear more "patient-like" than an arbitrary control.

References

1. G. Aguirre, E. Zarahn, and M. D’Esposito. The inferential impact of global signal covariates in functional neuroimaging analyses. *NeuroImage*, 8(3):302–306, Oct. 1998.
2. B. Casey, J. Cohen, K. O’Craven, et al. Reproducibility of fMRI results across four institutions using a spatial working memory task. *NeuroImage*, 8(3):249–261, Oct. 1998.
3. M. Cohen and R. DuBois. Stability, repeatability, and the expression of signal magnitude in functional magnetic resonance imaging. *J Mag Res Imag*, 10(1):33–40, July 1999.
4. O. Coulon, J.-F. Mangin, J.-B. Poline, et al. Structural group analysis of functional maps. In A. Kuba et al, ed., *16th Int. Conf. on Information Processing in Medical Imaging, Lec Notes Comp Sci* 1613, 448–53, Visegrad, Hungary, June 1999.
5. K. Friston, A. Holmes, K. Worsley, et al. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210, 1995.
6. J. Hanley and B. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143:29–36, Apr. 1982.
7. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Predictions*. Springer, 2001.
8. G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Trans Information Theory*, IT-14(1):55–63, 1968.
9. M. Hutchinson, W. Schiffer, S. Joseffer, et al. Task-specific deactivation patterns in functional magnetic resonance imaging. *Mag Res Imag*, 17(10):1427–36, Dec. 1999.
10. J. Liow, K. Rehm, S. Strother, et al. Comparison of voxel- and volume-of-interest-based analyses in FDG PET scans of HIV positive and healthy individuals. *J Nucl Med*, 41(4):612–621, Apr. 2000.
11. T. McAllister, A. Saykin, L. Flashman, et al. Brain activation during working memory 1 month after mild traumatic brain injury: A functional MRI study. *Neurology*, 53:1300–1308, 1999.
12. J. Moeller, S. Strother, J. Sidtis, and D. Rottenberg. Scaled subprofile model: A statistical approach to the analysis of functional patterns in positron emission tomographic data. *J Cereb Blood Flow Metab*, 7(5):649–658, Oct. 1987.
13. S. Raudys and A. Jain. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Trans Pattern Analysis Machine Intelligence*, 13(3), Mar. 1991.
14. K. Rubia, T. Russell, E. Bullmore, et al. An fMRI study of reduced left prefrontal activation in schizophrenia during normal inhibitory function. *Schiz Res*, 52(1–2):47–55, Oct. 2001.
15. A. Saykin, L. Flashman, S. Frutiger, et al. Neuroanatomic substrates of semantic memory impairment in Alzheimer’s Disease: Patterns of functional MRI activation. *J Int Neuropsych Soc (JINS)*, 5:377–392, 1999.
16. A. Saykin, L. Flashman, T. McAllister, et al. Semantic, phonological and episodic memory processing in schizophrenia: Functional MRI activation patterns indicate a need for new models of dysfunction. *Schiz Res*, 36(1–3):233, 1999.
17. M. Skurichina and R. Duin. Stabilizing classifiers for very small sample sizes. In *Proc. 13th Int. Conf. on Pattern Recognition*, vol. 2, 891–896, Vienna, Austria, Aug. 1996.
18. C. Tegeler, S. Strother, J. Anderson, and S. Kim. Reproducibility of BOLD-based functional MRI obtained at 4 T. *Hum Brain Map*, 7(4):267–283, 1999.