2006

Detecting Cognitive States from fMRI Images by Machine Learning and Multivariante Classification

Yong Fan
University of Pennsylvania, yong.fan@uphs.upenn.edu

Dinggang Shen
University of Pennsylvania, Dinggang.Shen@uphs.upenn.edu

Christos Davatzikos
University of Pennsylvania, christos.davatzikos@uphs.penn.edu

Suggested Citation:
Yong, F., D. Shen and C. Davatzikos. (2006). "Detecting Cognitive States from fMRI Images by Machine Learning and Multivariate Classification." *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop.* 17-22 June 2006.

Detecting Cognitive States from fMRI Images by Machine Learning and Multivariante Classification

Abstract
The major obstacle in building classifiers that robustly detect a particular cognitive state across different subjects using fMRI images has been the high inter-subject functional variability in brain activation patterns. To overcome this obstacle, firstly, the brain regions that are relevant to the problem under study are determined from the training data; then, statistical information of each brain region is extracted to form regional features, which are robust to inter-subject functional variations within the brain region; finally, the regional feature statistical variations across different samples are further alleviated by a PCA technique. To improve the generalization ability and efficiency of the classification, from the extracted regional features, a hybrid feature selection method is utilized to select the most discriminative features, which are used to train a SVM classifier for decoding brain states from fMRI images. The performance of this method is validated in a deception fMRI study. The proposed method yielded better results compared to other commonly used fMRI image classification methods.

Disciplines
Biomedical Engineering and Bioengineering | Engineering

Comments
Suggested Citation:
Yong, F., D. Shen and C. Davatzikos. (2006). "Detecting Cognitive States from fMRI Images by Machine Learning and Multivariate Classification." *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop.* 17-22 June 2006.

Detecting Cognitive States from fMRI Images by Machine Learning and Multivariate Classification

Yong Fan, Dinggang Shen, and Christos Davatzikos
Section for Biomedical Image Analysis (SBIA), Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA
{yong.fan, dinggang.shen, christos.davatzikos}@uphs.upenn.edu

Abstract

The major obstacle in building classifiers that robustly detect a particular cognitive state across different subjects using fMRI images has been the high inter-subject functional variability in brain activation patterns. To overcome this obstacle, firstly, the brain regions that are relevant to the problem under study are determined from the training data; then, statistical information of each brain region is extracted to form regional features, which are robust to inter-subject functional variations within the brain region; finally, the regional feature statistical variations across different samples are further alleviated by a PCA technique. To improve the generalization ability and efficiency of the classification, from the extracted regional features, a hybrid feature selection method is utilized to select the most discriminative features, which are used to train a SVM classifier for decoding brain states from fMRI images. The performance of this method is validated in a deception fMRI study. The proposed method yielded better results compared to other commonly used fMRI image classification methods.

1. Introduction

Functional magnetic resonance imaging (fMRI) has been playing an important role in neuroscience research, for exploring the regional brain activity associated with perception, cognition, or emotion. Most fMRI image analysis methods are mass-univariate analysis techniques, such as general linear model (GLM), which have been successfully applied to identifying the brain regions involved in specific tasks [1, 2]. More recently, some researchers have begun to explore an inverse problem, i.e., using multivariate classification methods to decode the cognitive states from fMRI images [3-7]. Also, fMRI images have been used in [8-10] for disease diagnosis.

One of the major problems in using fMRI images to decode the cognitive states is the high inter-subject variability [2], in the spatial locations and functional activation degrees. Although the inter-subject variability can be partially alleviated by stereotactic normalization and spatial smoothing, the remaining inter-subject variability after spatial normalization and smoothing can be still critical [11]. This is true in our deception fMRI study. Group analysis was implemented within a random-effects model for two groups of different subjects, each of them consisting of 11 subjects who were randomly selected. As illustrated in Figure 1, the difference of activation patterns in these two groups can be observed both in spatial locations and degrees of activation, which indicates that there exists significant inter-subject variability. Such inter-subject variability has been a major problem for successful classification of fMRI images, since it degrades the generalization of a classifier in predicting the cognitive states of unseen subjects.

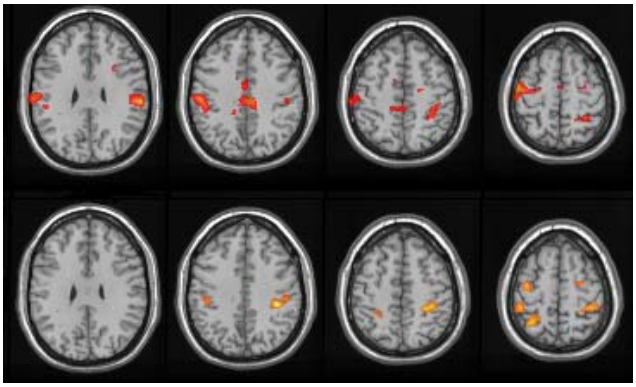

Figure 1: Group analysis results of lying and truth-telling tasks for two groups of different subjects, displaying different, albeit similar activation patterns. (Top: group1, bottom: group2)

Moreover, the sheer dimensionality of fMRI data is another major problem in fMRI image classification. Usually, the dimensionality of fMRI images, including both temporal and spatial dimensions, is much high, often in tens of millions. On the other hand, the number of training samples used in a typical fMRI-based study is very limited, i.e., a few dozens or at most hundreds. In this case, the performance and generalization ability of a multivariate

Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)
0-7695-2646-2/06 $20.00 © 2006 IEEE

pattern classifier might be severely affected, due to this small sample size problem.

To solve these problems in fMRI image classification of multiple-subjects, we propose a comprehensive framework, with major steps summarized in Figure 2. *First*, fMRI image of each subject are spatially normalized to a standard space, and further smoothed by a Gaussian filter to partially alleviate the inter-subject variation of functional activations. *Second*, by assuming that only part of brain regions are involved in specific tasks, a robust feature extraction method is proposed to extract from those active brain regions the regional features, invariant to the inter-subject variations within each active brain region. This regional feature extraction step consists of three components. The component ① is proposed to partition the brain in the standard space into a number of different regions, with each region having similar functional activations across different training samples. However, for each brain region involved in specific tasks, it can not be guaranteed that brain activations of different subjects exactly overlay at same positions in this brain region, as indicated in Figure 1. Therefore, the component ② is proposed to characterize each brain region by a set of regional features that capture the statistical information, i.e., probability distribution of fMRI image intensity values within this region. Such regional features are invariant to the inter-subject spatial variability within a brain region. In order to capture the inter-subject statistical variations of those regional features and also to compactly represent them, the component ③ is proposed to estimate the subspace of distribution of those regional features from the training samples by a PCA technique. *Third*, after extraction of regional features from each generated brain region, a hybrid feature selection approach is utilized to choose a set of discriminative features for classification. *Finally*, based on those selected features, a support vector machine (SVM) based classifier is trained, and it is used to perform the fMRI image classification on new testing samples. The proposed fMRI image classification method has been tested on a deception fMRI study with 22 subjects, and achieved superior performance compared to other methods using similar framework but different feature extraction techniques.

The remainder of this paper is organized as follows. In Section 2, related work is reviewed. Subsequently, the proposed method is detailed in Section 3, which includes feature extraction, feature selection, and nonlinear classification. Experimental results and comparisons with other methods are presented in Section 4. This paper concludes in Section 5.

2. Related Work

The fMRI image classification problem is typically solved by standard pattern recognition methods. Thus, the related work is mainly on feature extraction, feature dimensionality reduction, and feature-based classification of fMRI images, as summarized next.

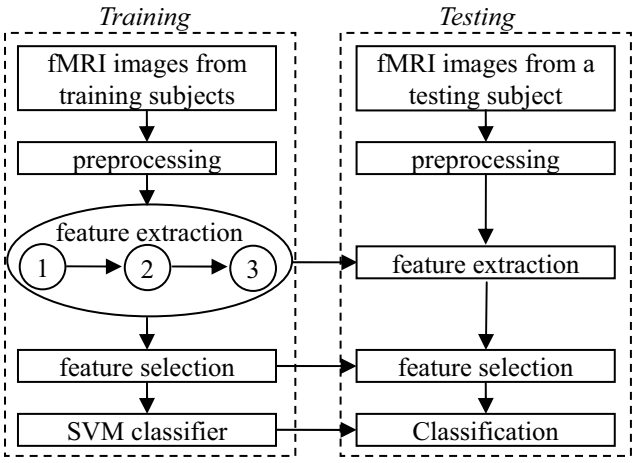

Figure 2: A framework for fMRI image classification. In particular, there are three components in the feature extraction step, ① brain template space partition, ② regional feature extraction, and ③ statistical and compact representation of regional features. The finally constructed classifier from the training stage is used to classify a new testing sample as descried on the right.

The intensities of fMRI images reflect the functional activities in different brain regions, thereby they can be directly utilized as features for classification [4-6]. Since there always exists inter-subject and intra-subject variability in fMRI images, spatial normalization and smoothing is generally used as a preprocessing step before performing fMRI image classification. However, it is impossible to completely remove inter-subject and intra-subject variability by using spatial normalization and smoothing, so classification directly based on fMRI images might lead to suboptimal results [11].

To further reduce intra-subject variability and improve the signal to noise ratio (SNR) of fMRI images, the mean image or the mean statistical map of fMRI images over a certain time interval is typically adopted as raw features to classify fMRI images in many applications [3, 5-7]. Compared to the method that directly uses all temporal fMRI images for classification (in order to completely use the temporal information) [10], there are several advantages of using this mean or statistical map for fMRI image classification, for example, the impact of hemodynamic delay can be alleviated and also the size of data can be significantly reduced [3].

To relieve the curse of dimensionality in fMRI image classification, dimensionality reduction is typically performed before classification, which is significant for successful classification of fMRI images. The principal component analysis (PCA) and the region of interest (ROI)

based feature reduction techniques are two most widely used dimensionality reduction methods [4-6]. They are detailed in the next two paragraphs, respectively.

PCA is a standard feature reduction method that transforms high dimensional data onto a linear eigen-space learned from the training dataset [12]. Due to its simplicity, PCA has been applied to a variety of problems, i.e., feature representation, face recognition, statistical model building, and feature dimensionality reduction [4, 12-14]. However, the performance of PCA is generally limited due to insufficient learning when the dimensionality of original data is much higher than the number of available training samples [13]. Furthermore, to make PCA valid, the spatial correspondence across samples should be assumed. However, in practice, the spatial correspondence across fMRI images of different subjects is generally very poor.

Brain regions that are relevant to the problem under study must first be selected from a background of brain activity. This is because the features from brain regions that are less relevant to the specific task under study only add confounding information to the training of a classifier and thus degrade the performance of the finally constructed classifier. An intuitive way for selecting the informative voxels for classification is to simply pick the most activated voxels within specific ROIs, which are typically determined by experts with the help of structural brain images [5, 6]. Alternatively, all voxels within each ROI can be averaged to be a supervoxel and used for classification. Importantly, the latter has several advantages, i.e., the feature dimensionality is greatly reduced and also the generated features might be better corresponding across different subjects. However, *a prior* knowledge about the ROIs that will be activated by the specific tasks is generally required. This limits the application of this technique to a broad range of studies where no good prior knowledge about the ROIs is available. Importantly, the structural ROIs might be inconsistent with the actual functional activation regions of the brain, thus affecting the overall performance of a classification method, by adding the confounding features from the assumed ROIs and missing the important features from the actual activated ROIs.

With the most informative features extracted by the techniques as reviewed above, a classifier can be built by a multivariate pattern classification method. So far, a number of classification methods have been used for fMRI image classification, i.e., Support Vector Machine (SVM), k Nearest Neighbor (kNN), Gaussian Naïve Bayes (GNB), and Linear Discriminative Analysis (LDA). Generally, the SVM-based classifier has superior performance [4-7]. Besides these methods, an Adaboost algorithm has been recently proposed to separate the drug-addicted subjects from the controls, by simultaneously selecting features and training a classifier [9].

3. Methods

Our classification method consists of three major steps, i.e., feature extraction, feature selection, and nonlinear multivariate classification, as detailed next.

3.1. Feature Extraction

To alleviate the inter-subject variability in the fMRI images, all fMRI images of different subjects are spatially normalized into a standard space and further smoothed by a Gaussian filter, as indicated in Figure 2. Although this preprocessing step is suboptimal to achieve spatial correspondence in voxel level across subjects [11], it partially alleviates the inter-subject variability. This preprocessing step has been the starting point of our method.

In order to extract robust regional features for fMRI image classification, we propose a three-step method. *First*, the brain template space is adaptively partitioned into a number of brain regions, based on the smoothed training samples normalized in the template space. It is assumed that the voxels with similar functional activities in the spatial neighborhood are grouped into a brain region, and also this brain region has similar functional activities across different training samples. This method has been successfully utilized in structural brain image classification [15]. *Second*, statistical information, i.e., the probability distribution of fMRI image intensities, is computed within each generated brain region for each sample. Importantly, this statistical information, which forms regional features, is independent of the spatial locations of functional activation peaks within each brain region, thereby inter-subject spatial variation within a brain region can be solved. *Finally*, in order to capture the statistical variations of those regional features across different samples and also to compactly represent them, those regional features are further represented as coefficients in a low-dimensional subspace, learned from the training samples by using PCA.

3.1.1 Adaptive partition of brain regions

The entire template brain is adaptively partitioned into a number of separate brain regions by performing a watershed segmentation algorithm on a score map of classification power, defined for each voxel in the template image, and estimated from all spatially normalized and smoothed training samples. The watershed segmentation algorithm is a traditional image segmentation method, which has been widely used in medical image analysis for partitioning images into different regions according to local intensity [16, 17].

The classification power of each voxel feature in the template image is related to the relevance of this voxel to the specific task under study. It is also related to the

generalization ability of using this voxel for classification, which is important to infer the cognitive states from the unseen subjects.

The relevance of each voxel to the specific task can be measured by a correlation measurement between the feature in this voxel and the corresponding class label in the training samples, i.e., lying or telling truth in our simulated deception study. Although a lot of non-linear correlation measurements, such as mutual information, are available, we choose a linear correlation method to measure the relevance of each voxel to the specific task, since it is easier to compute even for continuous features and it is robust to over-fitting [18]. Here, we used the Pearson correlation coefficient definition, which is closely related to the t-test [18]. Intuitively, the larger the absolute value of Pearson correlation coefficient is, the more relevant to classification this feature is. Given a location u in the template space, the Pearson correlation coefficient between a feature $f(u)$ in this location and the corresponding class label y can be defined as

$$\rho(u) = \frac{\sum_j \left(f_j(u) - \overline{f(u)}\right)\left(y_j - \overline{y}\right)}{\sqrt{\sum_j \left(f_j(u) - \overline{f(u)}\right)^2 \sum_j \left(y_j - \overline{y}\right)^2}}, \qquad (1)$$

where j denotes the j-th sample in the training dataset. Thus, $f_j(u)$ is the image intensity value in the location u of the j-th sample, and $\overline{f(u)}$ is the mean of $f_j(u)$ over all samples. Similarly, y_j is a corresponding class label of the j-th sample, and $\overline{y}$ is the mean of y_j over all samples.

In addition to the relevance, the generalization ability of the feature in each voxel is equally important for classification, particularly in the applications where the dimensionality of data is much higher than the size of training samples, such as our deception study. Thus, a leave-one-out cross-validation strategy is adopted to take the generalization ability into account when measuring the overall correlation of a feature to class label by the Pearson correlation coefficient. That is, given n training samples, the worst Pearson correlation coefficient resulting from n leave-one-out correlation measurements is selected as the overall correlation coefficient of this feature to class label. This overall correlation coefficient is defined as the classification power of a feature $f(u)$, as mathematically given below:

$$P(u) = \arg\min_{\{\rho_j(u) | 1 \le j \le n\}} \left|\rho_j(u)\right|, \qquad (2)$$

where $\rho_j(u)$ is a Pearson correlation coefficient between the feature $f(u)$ and the class label y, obtained from the j-th leave-one-out case where the j-th sample is excluded.

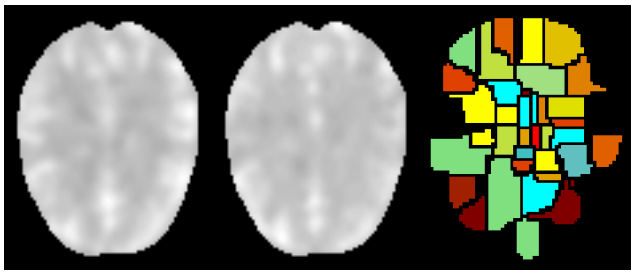

Figure 3: From left to right are cross-sectional views of "lie" image, "truth" image, and generated brain regions.

The definition of $\rho_j(u)$ is similar to the definition of $\rho(u)$ in equation (1), except that the j-th sample is excluded from correlation computation. This definition of classification power for each voxel feature is particularly important when examining a very large number of features; otherwise, outlier features can be found just by a chance.

By calculating a gradient map of the score map of classification power, $P(u)$, and using it in conjunction with a watershed segmentation algorithm, we can adaptively partition a brain into R different regions, i.e., $r^i, i = 1, \ldots R$. In order to obtain the relatively large brain regions, Gaussian smoothing is applied to the score map before computing its gradient map. Typical fMRI images, with different cognitive states, and the generated brain regions are shown in Figure 3.

3.1.2 Regional features

It is important to extract the regional features that are relatively invariant to inter-subject variations of spatial locations of functional activation in each generated brain region. Note that, if we simply average the fMRI image intensities within each region, the obtained average value might be not sufficiently discriminative, as later shown in our experiments. Also, picking the most active voxels within each region [5, 6] obviously does not consider the inter-subject spatial variations within each region. In this paper, we propose to extract statistical information from each region, such as computing the probability distribution of fMRI image intensities within each region, which is invariant to the spatial locations of functional activations within each region and also retains the details of functional information.

To estimate the probability distribution of fMRI image intensities within a brain region, a histogram of fMRI image intensities is calculated for each training sample. Note that the histogram-based regional feature representation method is computationally efficient and robust, thereby it has been successfully and extensively studied in the computer vision area for object recognition and image retrieval [19, 20]. Although a variety of methods, i.e., parametric estimation and non-parametric kernel based

estimation [12], can be used for estimation of probability distribution of fMRI image intensities, they are limited in our study since the parametric estimation methods typically require good prior knowledge about the probability distribution to be estimated, and the non-parametric estimation methods generally require a large number of samples to obtain an accurate estimation.

In our study, the number of bins used for calculating the histogram is 32. Thus, each region is represented by a 32-dimensional regional feature vector $h(r)$. In this way, the *j*-th training sample can be represented by a set of R vectors, $\{h(r_j^i), i = 1, \ldots, R\}$, where R is the number of our generated brain regions.

3.1.3 Statistical representation of regional features

The regional features, represented by a feature vector, for each region, are robust to the inter-subject spatial variations within each brain region. In order to robustly and efficiently compare the similarity or the difference of the feature vectors in the same region from different samples, it is important to capture the statistical variations of each feature vector across different samples, and also to compactly represent them.

PCA is used to estimate the subspace of feature vectors in each region from all training samples, and then each feature vector is further represented by the coefficients in the subspace constructed. In particular, PCA is applied to each region r^i, to estimate the eigen-space from the covariance matrix of the feature vector, $h(r^i)$, from n training samples, $\{ h(r_j^i), j = 1, \ldots, n \}$. We found the original feature vectors can be represented in a lower-dimensional subspace estimated by PCA. More importantly, since only a few coefficients, corresponding to the eigenvectors with largest eigen-values, are most relevant to classification, we finally use 5 coefficients to statistically represent each region, i.e., $f(r_j^i), \ i = 1, \ldots, R$, for the *j*-th sample. Thus, this representation is more compact and it considers the statistical variation of regional features across different samples.

3.2. Feature Selection

By extracting a compact set of regional features from each automatically generated brain region, we can efficiently represent the features in each training sample. However, some features are less effective, irrelevant and redundant for classification, compared to others. Therefore, it is important to select a small set of most effective features, in order to improve the generalization ability and the performance of the finally constructed classifier.

Thus, a hybrid feature selection algorithm [15] is used in this study for achieving good performance in a reasonable time cost. In this hybrid feature selection algorithm, we first use a correlation-based feature ranking method to select a set of the most relevant features, from which we further select a subset by a SVM-based subset feature selection algorithm [21].

This hybrid feature selection algorithm simply integrates the advantages of the ranking based feature selection method and the subset feature selection method. The ranking based feature selection method is computationally efficient and thus preferable for high dimensional problems, since it simply selects the top ranked features according to their ranking scores defined by classification powers. However, the selected features might contain a lot of redundant features, since the ranking score is computed independently for each feature and inter-feature correlation is ignored. On the contrary, the feature subset selection method focuses on selecting a subset of features that jointly have better discriminative power. However, its requirement on high computational cost usually limits their applications to problems with high dimensional features. Thus, by using this hybrid feature extraction method, we can have advantages of both methods, and finally select a subset of most effective regional features within a reasonable time.

3.3. SVM-based Classification

A nonlinear support vector machine (SVM) [22] is employed in our study for fMRI image classification, based on the regional features selected above. There are a number of reasons to use the SVM for fMRI image classification. First, SVM can construct a maximal margin linear classifier in a high (often infinite) dimensional feature space, by mapping the original features via a kernel function. Second, SVM is not only empirically demonstrated to be one of the most powerful pattern classification algorithms, but also has provided many theoretical bounds on the generalization to estimate its capacity, for example, the radius/margin bound, which could be used in feature selection. Finally, SVM has an inherent sample selection mechanism, i.e., only support vectors affect the decision function, which may help us find the subtle differences between two groups. In this study, the Gaussian radial basis function kernel is used.

4. Experimental Results

4.1. Classification Performance

We applied the proposed approach to one of the long-standing challenges in the applied psychophysiology, namely lie-detection, based on fMRI images of twenty-two right-handed male undergraduate students (Mean age = 19.36, Standard deviation (SD) = 0.5). Written informed consent was obtained from all participants after complete

description of the deception study, in which participants will perform lying and truth-telling tasks. The experimental procedure and image acquisition has been detailed in [23].

Functional data were processed with SPM2 (Wellcome Department of Cognitive Neurology, London, UK) following a standard procedure, including slice-time correction, motion-correction, spatial normalization and spatial Gaussian smoothing. Subject-level statistical analyses were performed using a regression model which consisted of 50 columns, 24 columns for "lie", 24 columns for "truth", 1 column for the repeat distracter and 1 column for the distracters, which produced 50 beta images with dimension of 79x95x68. The generated 48 beta images, corresponding "lie" condition and "truth" condition, were used for the nonlinear pattern classification. In order to reduce the effects of various types of noise, we formed averages of all beta images of truthful and non-truthful conditions for all subjects, thereby ending up with 44 images in total, 2 for each subject.

We investigate the generalization performance of our classifier by training it on the 42 images of the 21 subjects, then testing it on the 2 images of the left out subject. This procedure is repeated for 22 times, each time leaving the 2 images of a different subject out. During the training stage, all steps of adaptive regional feature extraction, feature selection, and classifier training are completely based on the training data. Then, the classification result on the testing subject using the trained SVM classifier was compared with the ground-truth class label, to evaluate the classification performance. Finally, these experiments were repeated for different numbers of features used for classification, in order to test the stability of classification results with respect to the number of features used.

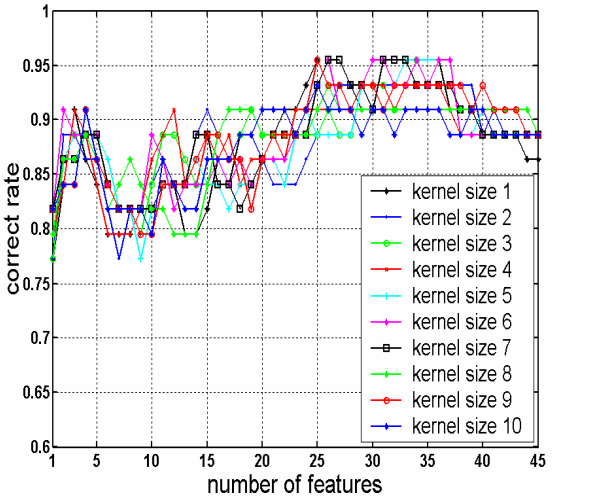

Figure 4: The performance of our trained classifier with respect to different number of features used and different size of kernels used in SVM. These results are cross-validated.

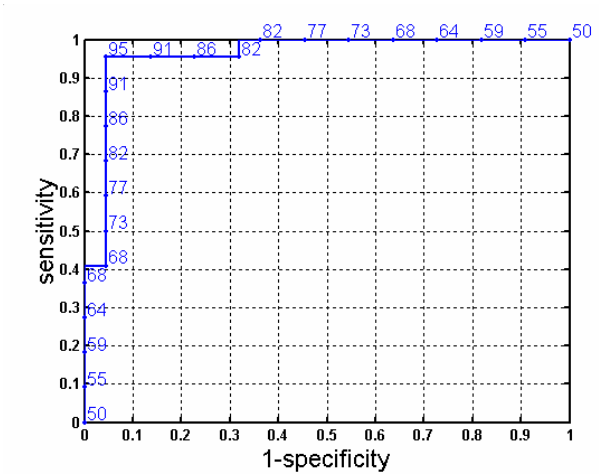

Figure 5: ROC curve. For each point on the ROC curve, the overall accuracy is shown in blue numbers (0.95 = 95%).

The best cross-validated correct classification rate was 95.5%, which can be achieved by only using 25 features, as shown in Figure 4. These results also indicate that the stable performance was achieved with respect to the size of Gaussian kernel used in SVM. In addition, the average correct classification rate is 92.4% when using 25 to 35 features and various kernel sizes in SVM. To further show the performance of our method, the receiver operating characteristics (ROC) curve of the classifiers that yield the best classification results is constructed [24]. As shown in Figure 5, the area under the ROC curve is 0.96, indicating the good performance of our classifier.

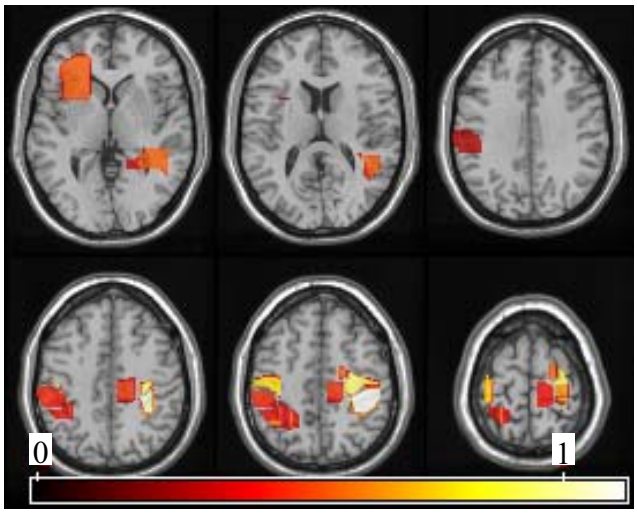

Figure 6: Regions of most representative of the group differences, found via decision function gradient (high value indicates more significant).

Table 1. Comparison on different feature extraction methods in lie detection (%).

Methods	PCA	ROI(active 20)	ROI(mean)	Anatomical ROI	Our method
Classification rates	88.6	90.9	90.9	90.9	93.2

In order to interpret the classification results, we utilize a *discriminative direction* method, as used in [15, 25], to estimate the difference of functional patterns (features used in classification) between lying and telling truth from the constructed SVM classifier. Afterwards, the feature differences are mapped to regions from which the features are extracted. Since a leave-one-out cross validation is performed in our experiments to test the generalization ability of our classifier, the overall group difference is constructed by averaging all group difference obtained from all leave-one-out cases. In Figure 6, group differences are overlaid on the template image, highlighting the most significant and frequently detected group differences in our leave-one-out experiments.

4.2. Comparison with other Methods

The performance of our method is compared with the performances of three types of classification methods, which use different feature extraction methods, but a very similar classification framework as depicted in Figure 2.

The first method uses the global feature extraction method, i.e., PCA, to extract features from all voxels in the entire brain. The traditional eigenvalue-based feature ranking method is used for feature selection.

The second type of methods are very similar to ours, except extracting from each of our automatically generated regions (1) the 20 most active voxels, which we call the method ROI(20) in Table 1; (2) the mean fMRI image intensity value, which we called the method ROI(mean) in Table 1.

The third method is also very similar to ours, except using the structural ROIs, such as 116 ROIs [26], to replace our automatically generated ROIs. This comparison is to show the importance of generating ROIs according to the data under study, in order to achieve a better classification performance.

For the second and third methods, we use only a feature ranking based method, not a hybrid feature selection method, for feature selection. For fair comparison, the feature selection in our approach is also completed by using the feature ranking based method. Therefore, the result of our method (93.2%) shown in Table 1 is a little bit worse than the result of our method (95.5%) provided in Section 4.1, where a hybrid feature selection method was used. This indicates the importance of using the hybrid feature selection method for fMRI image classification.

For all methods listed in Table 1, a nonlinear SVM with Gaussian radial basis function kernel was trained and tested, with a full leave one subject out cross-validation procedure. Different feature numbers and different SVM kernel sizes were tested for determining the best parameters for each classification method, and the best classification result for each classification method is reported in Table 1. The PCA-based method obviously obtained the worst classification result, indicating that the PCA method is unable to fully capture informative features from a limited number of training samples with images at this dimensionality [13]. The other three methods, all based on ROIs, have the same performance. But, their performance is worse than ours, although only a simple feature ranking method was used in our classifier. This indicates that (1) it is important to obtain data-adaptive ROIs for fMRI image classification, since structural ROIs do not necessarily coincide with functional ROIs; (2) it is important to extract our suggested regional features from each automatically generated brain region, not simply an average feature or most active voxel features in the region.

5. Conclusion

We have presented a comprehensive framework for decoding the cognitive states from the fMRI images of multiple subjects, by using a nonlinear multivariate classification method.

High inter-subject variability of functional activity and sheer dimensionality of fMRI data are the two major problems in fMRI image classification. Thereby, we proposed three particular steps to extract the most effective regional features for classification, based on SVM. *First*, the brain regions were adaptively partitioned, according to the classification power defined for each voxel in the brain. Our experimental results have shown that the adaptively generated brain regions are better in capturing regional features for classification, compared to the structural ROIs. *Second*, statistical information from each generated brain region forms regional features, which are invariant to the inter-subject variations within the brain region. This regional feature extraction method has proved to be better for fMRI image classification in our experiments, compared with others, such as extracting the mean or the

most active voxels from each region. *Finally*, a hybrid feature selection method, integrating the advantages of both ranking-based and subset-based feature selection methods, was used for selection of the most effective regional features. Our experiments also show that this hybrid feature selection method leads to a higher classification rate, compared to the use of a simple ranking-based feature selection method.

Our experimental results have shown that the proposed method can produce a high classification rate for a simulated deception study. It is worth noting that our method is applicable to other applications, such as inferring the cognitive states from the spatiotemporal patterns of brain activity.

References

[1] K. J. Friston, A. P. Holmes, K. Worsley, J. B. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *Human Brain Mapping*, vol. 2, pp. 189-210, 1995.

[2] R. S. J. Frackowiak, K. J. Friston, C. Frith, R. Dolan, C. J. Price, S. Zeki, J. Ashburner, and W. D. Penny., *Human Brain Function*, 2nd ed: Academic Press, 2003.

[3] D. D. Cox and R. L. Savoya, "Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex," *NeuroImage*, vol. 19, pp. 261–270, 2003.

[4] J. Mourão-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data," *NeuroImage*, vol. 28, pp. 980-995, 2005.

[5] T. M. Mitchell, R. Hutchinson, R. S. Niculescu, F. Pereira, and X. Wang, "Learning to Decode Cognitive States from Brain Images," *Machine Learning*, vol. 57, pp. 145-175, 2004.

[6] X. Wang, R. Hutchinson, and T. M. Mitchell, "Training fmri classifers to detect cognitive states across multiple human subjects," presented at NIPS03, 2003.

[7] C. Davatzikos, K. Ruparel, Y. Fan, D. Shen, M. Acharyya, J. Loughead, R. C. Gur, and D. Langleben, "Classifying spatial patterns of brain activity with machine learning methods: application to lie detection," *NeuroImage*, vol. 28, pp. 663-668, 2005.

[8] J. Ford, H. Frarid, F. Makedon, L. A. Flashman, T. W. McAllister, V. Megalooikonomou, and A. J. Saykin, "Patient classifiation of fmri activation maps," presented at MICCAI03, 2003.

[9] L. Zhang, D. Samaras, D. Tomasi, N. Volkow, and R. Goldstein, "Machine learning for clinical diagnosis from functional magnetic resonance imaging," presented at CVPR 2005, 2005.

[10] L. Zhang, D. Samaras, D. Tomasi, N. Alia-Klein, L. Cottone, A. Leskovjan, N. Volkow, and R. Goldstein, "Exploriting temporal information in functional magnetic resonance imaging brain data," presented at MICCAI 2005, 2005.

[11] B. Thirion, P. Pinel, and J.-B. Poline, "Finding landmarks in the functional brain:detection and use for group characterization," presented at MICCAI 2005, 2005.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*: John Wiley and Sons, Inc., 2001.

[13] C. Davatzikos, X. Tao, and D. G. Shen, "Hierarchical active shape models using the wavelet transform," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 414-423, 2003.

[14] T. F. Cootes and C. J. Taylor, "Combining point distribution models with shape models based on finite element analysis," *Image and Vision Computing*, vol. 13, pp. 403-409, 1995.

[15] Y. Fan, D. Shen, and C. Davatzikos, "Classification of Structural Images via High-Dimensional Image Warping, Robust Feature Extraction, and SVM," presented at MICCAI, Palm Springs, California, USA, 2005.

[16] V. Grau, A. U. J. Mewes, M. Alcañiz, R. Kikinis, and S. K. Warfield, "Improved Watershed Transform for Medical Image Segmentation Using Prior Information," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 447-458, 2004.

[17] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 583-589, 1991.

[18] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.

[19] B. Schiele and J. L. Crowley, "Recognition without Correspondence using Multidimensional Receptive Field Histograms," *International Journal of Computer Vision*, vol. 36, pp. 31-50, 2000.

[20] D. G. Lowe, "Distinctive image features from scale invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, 2004.

[21] A. Rakotomamonjy, "Variable Selection using SVM-based criteria," *Journal of Machine Learning Research*, vol. 3, pp. 1357-1370, 2003.

[22] V. N. Vapnik, *The Nature of Statistical Learning Theory (Statistics for Engineering and Information Science)*, 2nd edition ed: Springer-Verlag, 1999.

[23] D. Langleben, J. Loughead, W. Bilker, K. Ruparel, A. Childress, S. Busch, and R. C. Gur, "Telling truth from lie in individual subjects with fast event-related fMRI," *Human Brain Mapping*, vol. 26, pp. 262-272, 2005.

[24] T. Fawcett, "ROC graphs: notes and practical considerations for data mining researchers," HP Laboratories Palo Alto HPL-2003-4, 01-07-2003 2003.

[25] P. Golland, W. E. L. Grimson, M. E. Shenton, and R. Kikinis, "Deformation Analysis for Shape Based Classification," *Lecture Notes in Computer Science*, vol. 2082, pp. 517-530, 2001.

[26] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, and N. Delcroix, "Automated anatomical labelling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single subject brain," *Neuroimage*, vol. 15, pp. 273-289, 2002.