



ELSEVIER

## Multivariate examination of brain abnormality using both structural and functional MRI

Yong Fan,<sup>a,\*</sup> Hengyi Rao,<sup>b</sup> Hallam Hurt,<sup>c</sup> Joan Giannetta,<sup>c</sup> Marc Korczykowski,<sup>b</sup> David Shera,<sup>c</sup> Brian B. Avants,<sup>a</sup> James C. Gee,<sup>a</sup> Jiongjiong Wang,<sup>b</sup> and Dinggang Shen<sup>a,\*</sup>

<sup>a</sup>Department of Radiology, University of Pennsylvania, PA 19104, USA

<sup>b</sup>Department of Neurology, Center for Functional Neuroimaging, University of Pennsylvania, PA 19104, USA

<sup>c</sup>Department of Pediatrics, Division of Neonatology, The Children's Hospital of Philadelphia, PA 19104, USA

Received 25 October 2006; revised 2 April 2007; accepted 10 April 2007

Available online 19 April 2007

A multivariate classification approach has been presented to examine the brain abnormalities, i.e., due to prenatal cocaine exposure, using both structural and functional brain images. First, a regional statistical feature extraction scheme was adopted to capture discriminative features from voxel-wise morphometric and functional representations of brain images, in order to reduce the dimensionality of the features used for classification, as well as to achieve the robustness to registration error and inter-subject variations. Then, this feature extraction method was used in conjunction with a hybrid feature selection method and a nonlinear support vector machine for the classification of brain abnormalities. This brain classification approach has been applied to detecting the brain abnormality associated with prenatal cocaine exposure in adolescents. A promising classification performance was achieved on a data set of 49 subjects (24 normal and 25 prenatally cocaine-exposed teenagers), with a leave-one-out cross-validation. Experimental results demonstrated the efficacy of our method, as well as the importance of incorporating both structural and functional images for brain classification. Moreover, spatial patterns of group difference derived from the constructed classifier were mostly consistent with the results of the conventional statistical analysis method. Therefore, the proposed approach provided not only a multivariate classification method for detecting brain abnormalities, but also an alternative way for group analysis of multimodality images. © 2007 Elsevier Inc. All rights reserved.

### Introduction

Recently, machine learning and pattern recognition techniques have been playing an increasingly important role in analyzing structural and functional brain images (Dehmeshki et al., 2002; Golland et al., 2002; Timoner et al., 2002; Cox and Savoya, 2003; Ford et al., 2003; Yushkevich et al., 2003; Lao et al., 2004; Liu

et al., 2004; Mitchell et al., 2004; Thomaz et al., 2004; Davatzikos et al., 2005; Fan et al., 2005, 2006a,b; LaConte et al., 2005; Mourão-Miranda et al., 2005; Zhang et al., 2005). Multivariate classification methods employed in these studies typically consist of three components, i.e., feature extraction, feature dimensionality reduction, and feature-based classification. Feature extraction is a crucial one among these three components. Once effective features have been extracted, feature dimensionality reduction and feature-based classification can be implemented in a straightforward means by using suitable methods developed in machine learning area. For example, for feature dimensionality reduction, a principal component analysis (PCA) method or a feature selection technique can be employed (Cox and Savoya, 2003; Mitchell et al., 2004; Fan et al., 2005; Mourão-Miranda et al., 2005); whereas for feature-based classification, a support vector machine (SVM)-based classifier or a linear discrimination analysis method can be applied (Cox and Savoya, 2003; Davatzikos et al., 2005; Mourão-Miranda et al., 2005).

Voxel-wise features have been widely used as extracted features for neuroimage-based classification (Cox and Savoya, 2003; Ford et al., 2003; Mourão-Miranda et al., 2005). These voxel-wise features are typically computed from the normalized individual brain images in a template space, or the deformation fields that bring those individual brain images into the template. However, such voxel-wise features are sensitive to registration errors and inter-subject variations of structure and function, particularly for functional image analyses as indicated in (Thirion et al., 2005). On the other hand, regional effects, i.e., considering both the size of region and the total amount of statistical power in the region, are of interest in most voxel-wise analysis studies, although the positions and shapes of the regions are generally not known in advance (Friston, 1997; Poline et al., 1997; Hayasaka and Nichols, 2004). This implies that we should extract features from the regions with voxels of similar correlation to diseases, i.e., obtaining regional features that are robust to noise, registration error, and inter-subject variation (Fan et al., 2005, 2006a,b).

\* Corresponding author.

E-mail addresses: yong.fan@uphs.upenn.edu (Y. Fan), dinggang.shen@uphs.upenn.edu (D. Shen).

Available online on ScienceDirect (www.sciencedirect.com).

Features used in brain abnormality detection are typically extracted from either structural or functional neuroimages, respectively (Golland et al., 2001; Dehmeshki et al., 2002; Timoner et al., 2002; Ford et al., 2003; Lao et al., 2004; Liu et al., 2004; Fan et al., 2005). Although in the above studies promising results have been reported by classification methods that focused on single modality images, i.e., either structural or functional neuroimages, it is potentially advantageous and physiologically significant/important to use both structural and functional images for detecting brain abnormalities, because generally pathological changes should have characteristic markers in both brain anatomy and function. Accordingly, in this study, we provided the methodological framework to extract regional features simultaneously from both structural and functional images, which was then applied in conjunction with a hybrid feature selection method and a SVM classifier. This method was subsequently applied and validated in a preclinical cohort of adolescents with prenatal drug exposure and a socioeconomically matched control group (Avants et al., in press; Rao et al., in press). High-resolution 3D structural MRI and arterial spin labeling perfusion MRI were used as the two imaging modalities in our study. Preliminary results of this work were presented in Fan et al. (2006a,b).

## Methods

The classification method consists of four major components: preprocessing step, feature extraction, feature selection, and classification, as shown in Fig. 1. In the preprocessing step, morphological and functional representations of each subject are computed by spatially normalizing its brain images to a standard space, and further smoothed by a Gaussian filter to partially alleviate the inter-subject morphological and functional variation. By assuming that only part of brain regions are affected by the

problem under study, a robust feature extraction method is proposed to extract, from the morphological and functional representations, the regional features, invariant to the inter-subject variations within each small brain region. This regional feature extraction step consists of three components. The first component is proposed to partition the brain in the standard space into a number of different regions, with each region having voxels of similar characteristics with respect to the classification variable across different training samples. However, for each brain region affected by a specific problem, it cannot be guaranteed that the brain morphological or functional changes of different subjects exactly overlay at same positions in this brain region. Therefore, the second component is proposed to characterize each brain region by a set of regional features that capture the statistical information, i.e., probability distribution of values of morphological and functional representations 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 third component is proposed to estimate the subspace of distribution of those regional features from the training samples by a PCA technique. 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 SVM-based classifier is trained, and it is used to perform the classification on new testing samples. All of these components are detailed in the following.

### *Morphometric and functional representations of brain images*

For direct comparison of brains across different subjects, all subjects are spatially normalized to one of the subjects chosen as a template, by registering their structural images with the template's structural image using an automatic algorithm termed HAMMER (Shen and Davatzikos, 2002). Functional images are directly warped to the template space by the deformation fields estimated for their corresponding structural images after being co-registered with corresponding structural images by maximizing mutual information (Collignon et al., 1995). Thus, the morphometric and functional representation of each subject in the template space can be computed.

The morphometric representation of each subject can be derived from the deformation field which warps this subject's structural image into the template space, including the Jacobin determinant of deformation field and the modulated local concentration of brain tissues (Frackowiak et al., 2003). In this paper, we adopted tissue density maps as the morphometric representation (Davatzikos et al., 2001). The brain tissue density maps are obtained by first segmenting each structural MR image into three tissues, i.e., white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) (Pham and Prince, 1999), and then warping these tissue maps into the template space in a mass-preserving way (Davatzikos et al., 2001), which is achieved by increasing the respective density when a region is compressed, and vice versa. As a result, three tissue density maps are generated in the template space, each reflecting local volumetric measurements corresponding to WM, GM, and CSF, respectively, and giving a quantitative representation of spatial tissue distribution. In this study, only WM and GM tissue density maps ( $m_1$ ,  $m_2$ ) are used for brain classification.

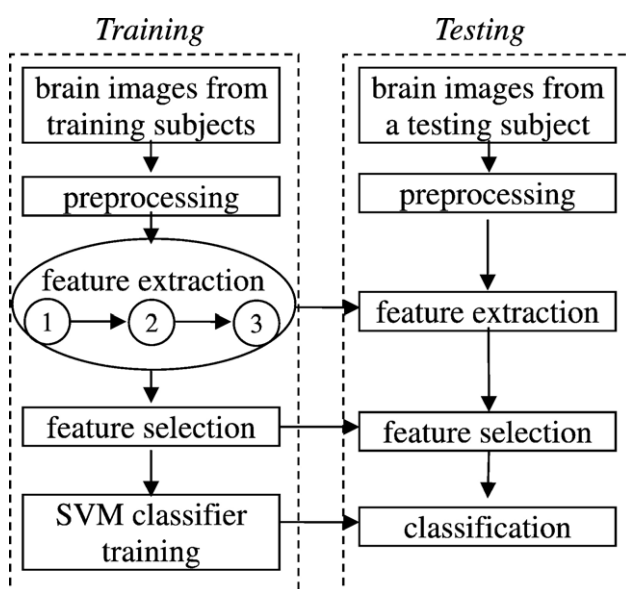


Fig. 1. A framework for brain image classification. In particular, there are three components in the feature extraction step, ① brain template space partition, ② regional feature extraction, and ③ statistical and compact representation of regional features. The finally constructed classifier in the training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

The functional representation of each subject can be obtained from any imaging modality that is sensitive to brain physiology, metabolism and function. In this study, we used each subject's cerebral blood flow (CBF) image as functional feature map, which was warped into the template space (following the deformation matrix determined from structural MRI) with the total CBF value preserved in any arbitrarily defined region during the warping procedure, denoted as functional feature map ( $m_3$ ).

Therefore, three local feature maps ( $m_k$ ,  $k=1, 2, 3$ ) called WM tissue density map  $m_1$ , GM tissue density map  $m_2$ , and functional feature map  $m_3$ , respectively, are adopted in our study. It is worth noting that other feature maps derived from other image modalities can also be used as additional features for multivariate brain analysis. That is, the proposed method is able to integrate the features extracted from different modality images and use them together for multivariate brain analysis.

#### Feature extraction

For each individual brain, the features for classification will be respectively extracted from the adaptively grouped regions in each feature map as defined above, and all extracted regional features from three feature maps are concatenated into a long feature vector to completely represent the structural and functional profile of this individual brain in the template space. In order to extract regional features, each feature map  $m_k$  ( $k=1, 2, 3$ ) will be first independently partitioned into a number of adaptive regions with similar local features, according to the available training samples, using the same procedure of (Fan et al., 2005, 2007). Then, statistical regional features will be extracted from each generated brain region (Fan et al., 2006a,b). These two steps are detailed in the next, and will be finally applied to each of the three feature maps,  $m_k$ ,  $k=1, 2, 3$ .

#### Adaptive partition of brain regions

The brain template space is adaptively partitioned into a number of separate regions by performing a watershed segmentation algorithm (Vincent and Soille, 1991) on a score map of classification power that is defined for each voxel in the template space and estimated from the corresponding feature maps of all training samples (Fan et al., 2005, 2007). The classification power of each voxel-wise feature is related to the *relevance* of this voxel-wise feature to the specific problem under study and also the *generalization ability* of this voxel-wise feature in classifying unseen subjects. Moreover, the *reliability* of each voxel-wise feature, with respect to various registration error and inter-subject variation, is another important issue in classification. The particular techniques for defining the *relevance*, *generalization ability*, and *reliability* of each voxel-wise feature are detailed in the next three paragraphs.

The *relevance* of each voxel-wise feature to the specific problem can be measured by the correlation between this feature and the corresponding class label (patient or normal control) in the training samples. In the field of machine learning and statistical analysis, the correlation measures can be broadly divided into linear correlation and nonlinear correlation. Most nonlinear correlation measures are based on the information-theoretical concept of entropy, such as mutual information, computed by probability estimation. For continuous features, probability density estimation is a hard task especially when the number of available samples is limited. On the other hand, linear correlation measures are easier to compute even for continuous features and are robust to

over-fitting, thus they are widely used for feature selection in machine learning (Guyon and Elisseeff, 2003). Here, we used the Pearson correlation coefficient to measure the relevance of each feature to classification. For feature selection, the Pearson correlation is closely related to the  $t$ -test as indicated in (Guyon and Elisseeff, 2003). In particular, given a voxel-wise feature  $f$ , the Pearson correlation coefficient between this feature and the class label  $y$  is defined as

$$\rho = \frac{\sum_j (f_j - \bar{f})(y_j - \bar{y})}{\sqrt{\sum_j (f_j - \bar{f})^2 \sum_j (y_j - \bar{y})^2}}, \quad (1)$$

where  $j$  denotes the  $j$ -th sample in the training data set. Thus,  $f_j$  is the feature value of the  $j$ -th sample, and  $\bar{f}$  is the mean of  $f_j$  over all samples. Similarly,  $y_j$  is the class label (normal  $-1$  or abnormal  $+1$ ) of the  $j$ -th sample, and  $\bar{y}$  is the mean of  $y_j$  over all samples. Intuitively, the larger the absolute value of Pearson correlation coefficient is, the more relevant to classification this feature is.

The *generalization ability* of this voxel-wise feature can be evaluated via a leave-one-out cross-validation strategy when measuring the overall correlation of a feature to class label by the Pearson correlation coefficient. That is, given  $n$  training samples, the worst absolute Pearson correlation coefficient resulting from  $n$  leave-one-out correlation measurements is selected as the overall correlation coefficient of this feature to class label. We can formulate this conservative definition as the *generalization ability* of a voxel-wise feature  $f$ :

$$g(f) = \operatorname{argmin}_{\{\rho_j(f) | 1 \leq j \leq n\}} |\rho_j(f)|, \quad (2)$$

where  $\rho_j(f)$  is a Pearson correlation coefficient between the feature  $f$  and the class label  $y$ , from the  $j$ -th leave-one-out case where the  $j$ -th sample is excluded. This definition of the overall Pearson correlation for each voxel-wise feature is particularly important when examining a very large number of features; otherwise, outlier features can be found just by a chance. Other robust statistics measurements, such as robust correlation (Wilcox, 1997), can be also used if these measurements can be computed very efficiently and are applicable to the small sample size problems like ours.

The *reliability* of each voxel-wise feature can be measured by the *spatial consistency* of this feature with others in its spatial neighborhood, since intuitively if a voxel-wise feature is similar to other features in its spatial neighborhood, it is reliable and small registration errors or inter-subject variations will not significantly change its value. In particular, as illustrated in Fig. 2 (see Appendix for detail), given a voxel-wise feature  $f$ , we measure its spatial consistency  $c(f)$  by an intraclass correlation coefficient among all features in its spatial neighborhood and across all samples in the training data set (McGraw and Wong, 1996). Thus, the classification power measure  $P(f)$  of a voxel-wise feature  $f$  can be finally obtained by multiplying the intra-class correlation coefficient  $c(f)$  with the overall Pearson correlation coefficient  $g(f)$  defined in Eq. (2).

By calculating a gradient map of the classification power map of each feature map  $m_k$ , and using it in conjunction with a watershed segmentation algorithm, we can adaptively partition the space of each feature map  $m_k$  into  $R$  different regions, i.e.,  $\{r^i, i=1, \dots, R\}$ , where  $R$  can be different for different feature maps. Notably, in order to obtain the relatively large brain regions, Gaussian smoothing is applied to the score map before computing

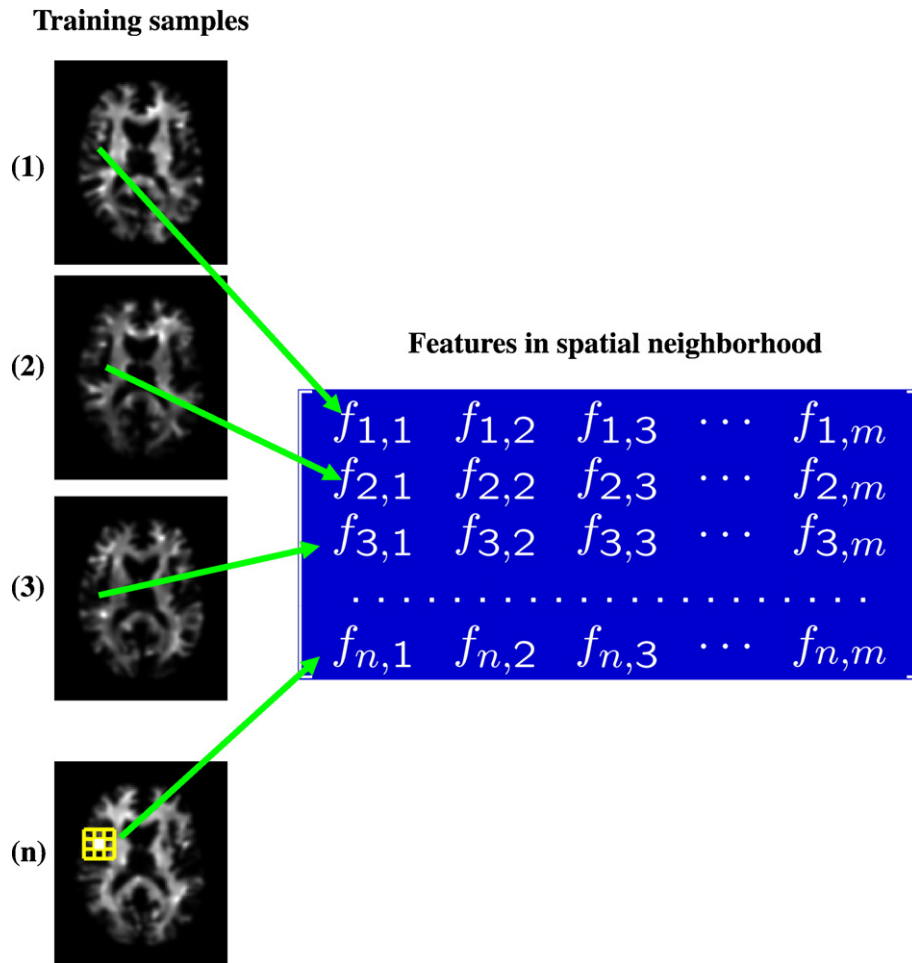


Fig. 2. Illustration of the estimation of the spatial consistency of a feature  $f$  located at  $u$  (white dot) in the feature map. The spatial consistency of a feature is estimated from training samples by the intraclass correlation coefficient among all features in its spatial neighborhood. The feature values within the given feature's spatial neighborhood constitute the rows of a data matrix, whereas the feature values of same spatial location of all training samples constitute the columns of the data matrix. Based on the data matrix, the intraclass correlation coefficient can be computed by the method in McGraw and Wong (1996). The mathematical formulation is detailed in the Appendix.

its gradient map. We typically used a smooth kernel of width equal to three voxels, which was determined empirically.

#### Extraction of statistical regional features

Statistical regional features will be extracted from each region generated above, by computing the probability distribution of voxel-wise features within the region, which is invariant to the inter-subject variation of structure or function within the region. Although the regional features can also be calculated by simply averaging the voxel-wise features within each region, the obtained average feature might be not sufficiently discriminative, compared to the probability estimation of voxel-wise features as we propose (Fan et al., 2006a,b).

To estimate the probability distribution of voxel-wise features within a region, the histogram of features within the region is calculated. Compared to other probability estimation methods, i.e., parametric estimation and nonparametric kernel-based estimation (Duda et al., 2001), the histogram-based regional feature representation is computationally efficient and robust, thereby it has been successfully used in the computer vision field for object

recognition and image retrieval (Schiele and Crowley, 2000). Using this method, each region can be represented by a regional feature vector  $h(r)$ , whose size is equal to the empirically determined number of same width bins used to generate histograms. Therefore, the  $j$ -th sample can be represented by  $R$  feature vectors,  $\{h_j(r^i), i=1, \dots, R\}$ , where  $R$  is the number of regions generated above for a given feature map, i.e., WM or GM tissue density maps, or functional feature map.

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. For these purposes, 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. Particularly, PCA is applied to each region  $r^i$ , to estimate the eigen-space from the covariance matrix of the feature vector,  $h(r^i)$ , based on  $n$  training samples,  $\{h_j(r^i), j=1, \dots, n\}$ . Since only a few coefficients, corresponding to the eigenvectors with largest eigenvalues, are the most relevant to classification, we choose the first five coefficients

to statistically represent each region, i.e.,  $F_j(r^i)$ ,  $i=1, \dots, R$ , for the  $j$ -th sample. In our application, the first five coefficients cover more than 80% variance (energy) of original histogram distributions, and also remove a significant amount of noise encoded in a large number of trailing eigenvectors. Thus, this representation is more compact and also takes into account the statistical variation of regional features across different samples. Notice that the first five PCA coefficients from each brain region of morphological and functional representations are concatenated into a long feature vector to represent the individual brain.

*Hybrid feature selection and SVM classification*

By extracting a compact set of regional features from each automatically generated brain region, we can efficiently represent the brain images of each subject, i.e., by all regional features calculated from three feature maps,  $m_1$ ,  $m_2$ , and  $m_3$ . However, some regional features are less effective, irrelevant or 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 (Fan et al., 2005) is used to select the most discriminative regional features for classification. In particular, 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 of features by a SVM-based subset feature selection algorithm (Rakotomamonjy, 2003).

Based on the regional features selected above, a nonlinear support vector machine (SVM) (Vapnik, 1999) is employed for classification, since SVM has been demonstrated in several studies to have superior performance on small sample problems. In this study, we used the Gaussian radial basis function kernel defined by

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (3)$$

where  $x_1$  and  $x_2$  are two feature vectors, and  $\sigma$  is the width of the Gaussian kernel.

To examine the performance of our proposed method in classifying brain abnormalities, we use a full leave-one-out cross-validation. In each leave-one-out validation case, one subject is

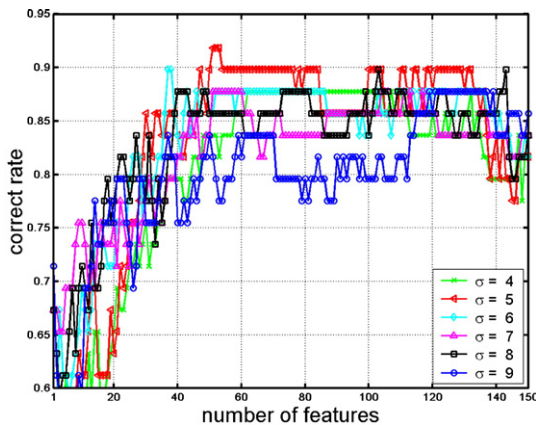


Fig. 3. The change of classification rate with respect to the different number of both structural and functional features used for classification and the different size of kernel used in SVM.

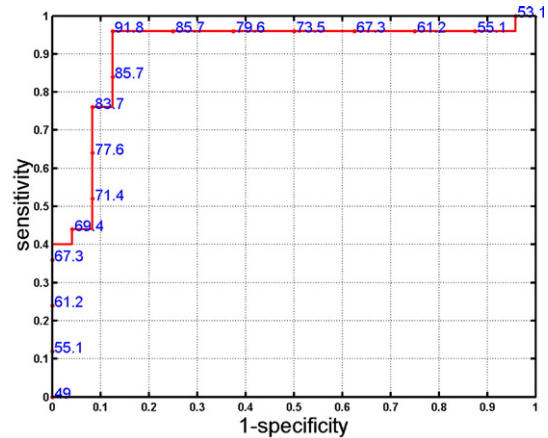


Fig. 4. ROC curve of the classifier that yields the best classification rate. Numbers on the curve are the correct classification rates (%). The area under the ROC curve is 0.91.

first selected as a testing subject, and the remaining subjects are used for adaptive regional feature extraction, feature selection, and classifier training, as described above. By repeatedly leaving each subject out as a testing subject, we can obtain the average classification rate from all of the leave-one-out cases.

*Group difference*

Besides using the pattern analysis method to classify individual brains, we also use it to detect group differences in two brain sets by means of the discriminative direction method (Golland et al., 2001; Fan et al., 2005). The group differences are estimated by averaging all group differences computed from all leave-one-out experiments, thus representing the frequency and significance of group difference in the detected brain regions. In particular, for each leave-one-out case, by following the gradient direction of SVM classification function, each support vector is first projected onto the group separation hypersurface determined by SVM classification function. Then, the difference between each support vector and its projection vector is estimated, which reflects changes of the support vector on the selected regional features when a normal brain changes to the respective configuration in the abnormal group, or vice versa. By summing up all regional differences calculated from all support vectors, an overall group difference vector can be obtained for the current leave-one-out case under study. Finally, the value of each element in the group difference vector is mapped to its corresponding brain region in the template space, and this value is subsequently added by other

Table 1  
Comparison on different feature extraction methods in brain classification, using the proposed hybrid feature selection and nonlinear SVM

Methods	Features		
	Structural features (%)	Functional features (%)	Both features (%)
PCA	63.3	77.6	75.5
Moment	73.5	85.7	87.8
Proposed	85.7	87.8	91.8

The best classification rates obtained by the methods are reported, respectively.

leave-one-out repetitions. Since the regions generated in the leave-one-out experiments can be slightly different, the overall averaged group difference map can be blurred. It is also worth noting that the group difference was examined in “regional probability distribution” feature space. So it is only regionally meaningful, not directly related to voxel-wise group difference such as generated by Statistical Parametric Mapping (SPM) methods (<http://www.fil.ion.ucl.ac.uk/spm>). For better visualization, the insignificant (not clustered) regions in our generated group difference map can be empirically excluded by a threshold.

### Application

As a validation study, our method was applied to study the long-term neurophysiologic effects of prenatal cocaine exposure on children. This study was approved by the local Ethics Committee (IRB). After giving informed consent, 49 subjects participated in this study, including 25 prenatal cocaine-exposed participants (11 females, mean age  $14.4 \pm 1.0$  years) and 24 socioeconomically matched normal controls (13 females, mean age  $13.9 \pm 0.9$  years).

### Data description and preprocessing

Both structural and functional scans of all the subjects were obtained from a Siemens 3.0 T Trio scanner. The high-resolution structural scans were obtained by a 3D MPRAGE sequence with TR=1620 ms, TI=950 ms, TE=3 ms, flip angle=15°, 160 contiguous slices of 1.0 mm thickness, FOV=192×256 mm<sup>2</sup>, matrix=192×256. The functional scans of resting brain were obtained by an amplitude-modulated continuous arterial spin labeling (CASL) perfusion technique (16 slices, 6 mm thk/1.5 mm sp, TR: 4 s, labeling time: 2 s, delay time: 1.2 s, TE: 17 ms, FOV: 22×22 cm<sup>2</sup>, matrix: 64×64). Eighty perfusion-weighted images for each subject were obtained in a 320-s scan.

The Statistical Parametric Mapping (SPM) software package was used to correct MR image series for head movements. Perfusion-weighted image series were then generated by pair-wise subtraction of the label and control images, followed by conversion to absolute cerebral blood flow (CBF) image series based on a single compartment CASL perfusion model (Wang et al., 2005).

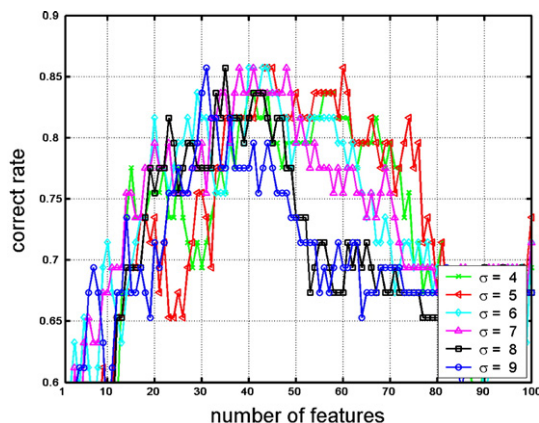


Fig. 5. The change of classification rate with respect to the different number of *structural* features used for classification and the different size of kernel used in SVM.

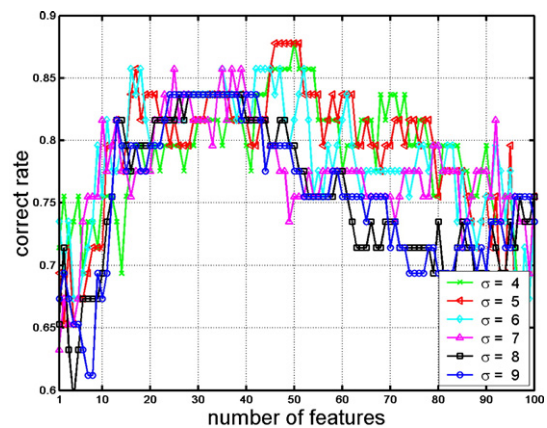


Fig. 6. The change of classification rate with respect to the different number of *functional* features used for classification and the different size of kernel used in SVM.

For each subject, one single mean image of all absolute CBF images, called functional image, was generated and further co-registered with its high-resolution structural image. These functional images were normalized to have same total CBF values.

### Results

#### Classification performance

The best classification accuracy achieved by our method via a leave-one-out cross-validation is 91.8%, using features simultaneously extracted and selected from both structural and functional MR images of the 49 teenagers. For a particular number of selected features, the cross-validation accuracy is simply obtained by averaging the classification accuracies across different leave-one-out cases, although different leave-one-out cases might select different features.

To show the classification stability of our method, the classification rates with respect to the number of features used for classification and the different sizes of kernel used in SVM are shown in Fig. 3. These plots indicate that the described classification method is relatively robust with respect to the size of Gaussian kernel used in SVM and the number of features used in classification. In addition, the average correct classification rate is about 89.9% when using 50 to 80 features and the kernel size of 5 in SVM. To evaluate the overall performance of the classification method, the receiver operating characteristic (ROC) curve of the classifier that yields the best classification result is shown in Fig. 4.

Table 2

Comparison on different feature extraction methods in brain classification, using the ranking-based feature selection and linear SVM

Methods	Features		
	Structural features (%)	Functional features (%)	Both features (%)
PCA	59.2	77.6	69.4
Moment	69.4	77.6	77.6
Proposed	77.6	87.8	87.8

The best classification rates obtained by the methods are reported, respectively.

The area under the ROC curve is 0.91, indicating good classification performance.

Although a promising classification performance is achieved within this leave-one-out cross-validation framework, it might be optimistic to generalize the performance to other data sets. A more rigorous way to validate the classification performance is to use above cross-validation as a training step, and test the obtained classifier's performance with a separate testing data set. However, the high computation cost of the proposed method and the limited number of available samples prevent us from such a validation. Instead, we performed random permutation tests to evaluate the statistical significance of the classification result obtained within this leave-one-out cross-validation framework (Golland and Fischl, 2003). In particular, labels (as prenatally cocaine-exposed or control) were assigned randomly to samples, and then the same training and testing procedures were performed for our method. Totally, 10 permutation tests were performed. The mean  $\pm$ SD of the classification accuracies of these ten tests using best sets of features was  $58.7\% \pm 10.3\%$  and median was 61.8%. These relatively high classification rates on the permutation tests indicate that the proposed method might have a trend to overfitting. On the other

hand, although these results are statistically higher than random guess using this limited number of permutation tests, they are much lower than the accuracy achieved on correctly labeled samples. Therefore, we still have reasonable confidence on our reported classification accuracy, according to (Golland and Fischl, 2003).

#### Comparison with other methods

For comparison purpose, the performance of brain classification was evaluated with respect to (1) the use of different features, i.e., using features from structural images, or features from functional images, or features from both structural and functional images, and (2) the use of different feature extraction methods, i.e., our method, a moment-based regional feature extraction method, or a popular PCA-based classification method. In particular, the moment-based regional feature extraction method extracts the mean, variance, skewness, and kurtosis of voxel-wise measures within each generated brain region (Press et al., 1992). PCA-based feature extraction method extracts the PCA coefficients of a new brain in the eigen-space of a long vector of all voxel-wise morphological

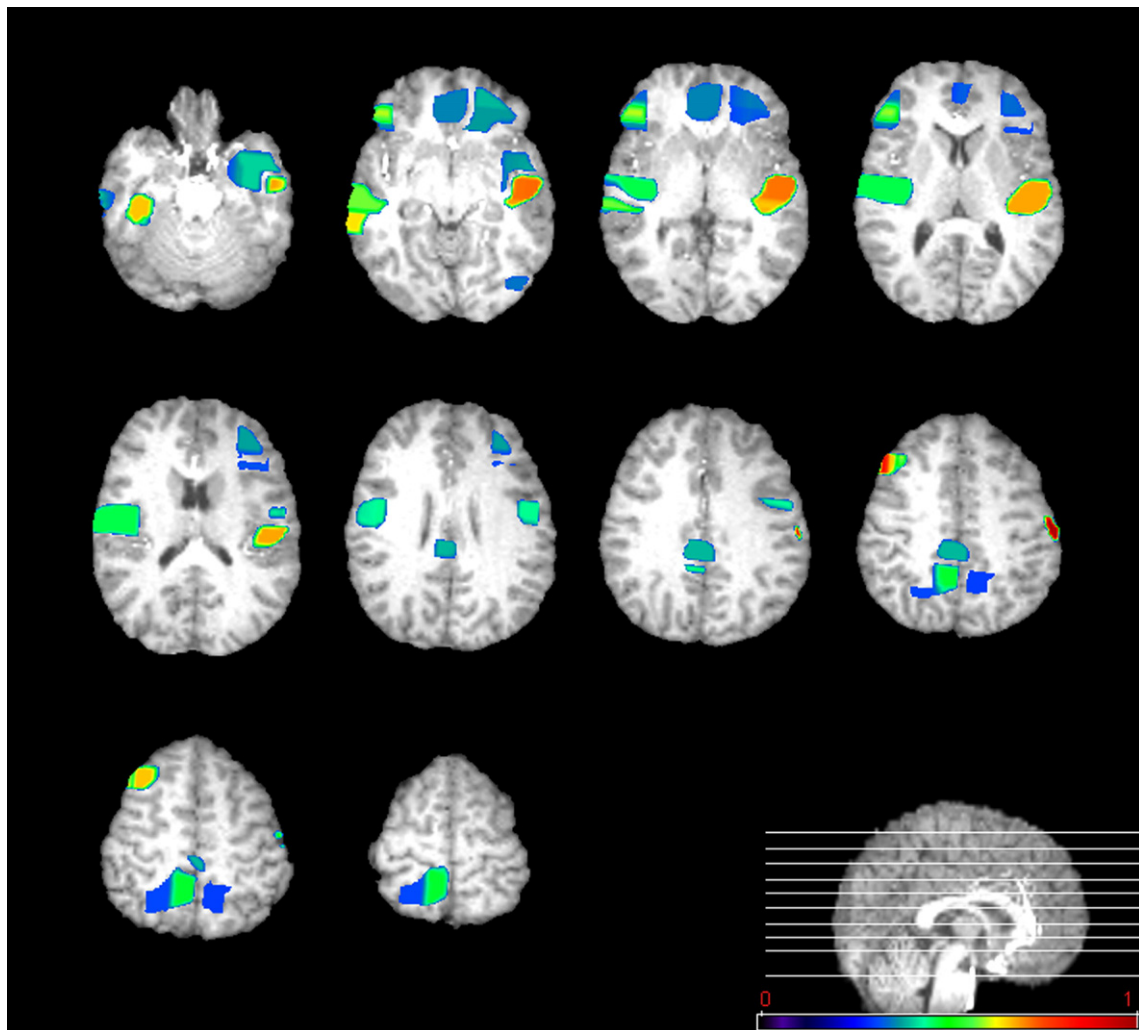


Fig. 7. Group differences identified by our pattern classification method for functional feature map (left is left). The significance of group difference in each region is color-coded according to the color bar shown in the bottom, with 1 as relatively the most important for classification.

and functional measures, trained by a set of training samples (Mourão-Miranda et al., 2005). For fair comparison, the proposed hybrid feature selection method and the nonlinear SVM with Gaussian radial basis function kernel are both used to build the classifiers for all experiments.

The best classification results achieved by these methods are summarized in Table 1. Overall, the classification results by our method are better than the respective ones by moment-based or PCA-based classifications, for any of the cases using features from different single modality images or from both modality images. This should be attributed to the capability of our regional feature extraction method, which can capture the relatively reliable and effective features for classification. In addition, the results by the moment-based feature extraction method are also better than those by the PCA-based feature extraction method. This indicates that the global PCA-based feature extraction method is not effective for the small sample problem. As shown in Figs. 5 and 6, when using only structural features or functional features for classification, the classification performance is relatively unstable with respect to different sizes of kernel used in SVM and different numbers of

features used for classification. On the other hand, as shown in Table 1, the simultaneous use of both structural and functional MR images produced the best result by our method, indicating the importance of simultaneously using multimodality images for examining brain abnormality.

To better understand the performances of these different feature extraction methods, the ranking-based feature selection and the linear SVM-based classification are adopted and further tested with the same leave-one-out cross-validation procedure as described above. The best classification result for each of these feature extraction methods is summarized in Table 2. These results, obtained by linear SVM, indicate again the importance of adaptively extracting histogram-based regional features as well as both morphological and functional measures for brain classification. In particular, the histogram-based regional features generally have better performance than the moment-based features in brain classification, which should be attributed to the use of PCA on the histogram-based features before classification. Of course, the moment-based feature extraction method is simpler and seems more objective, compared to the histogram-based method which

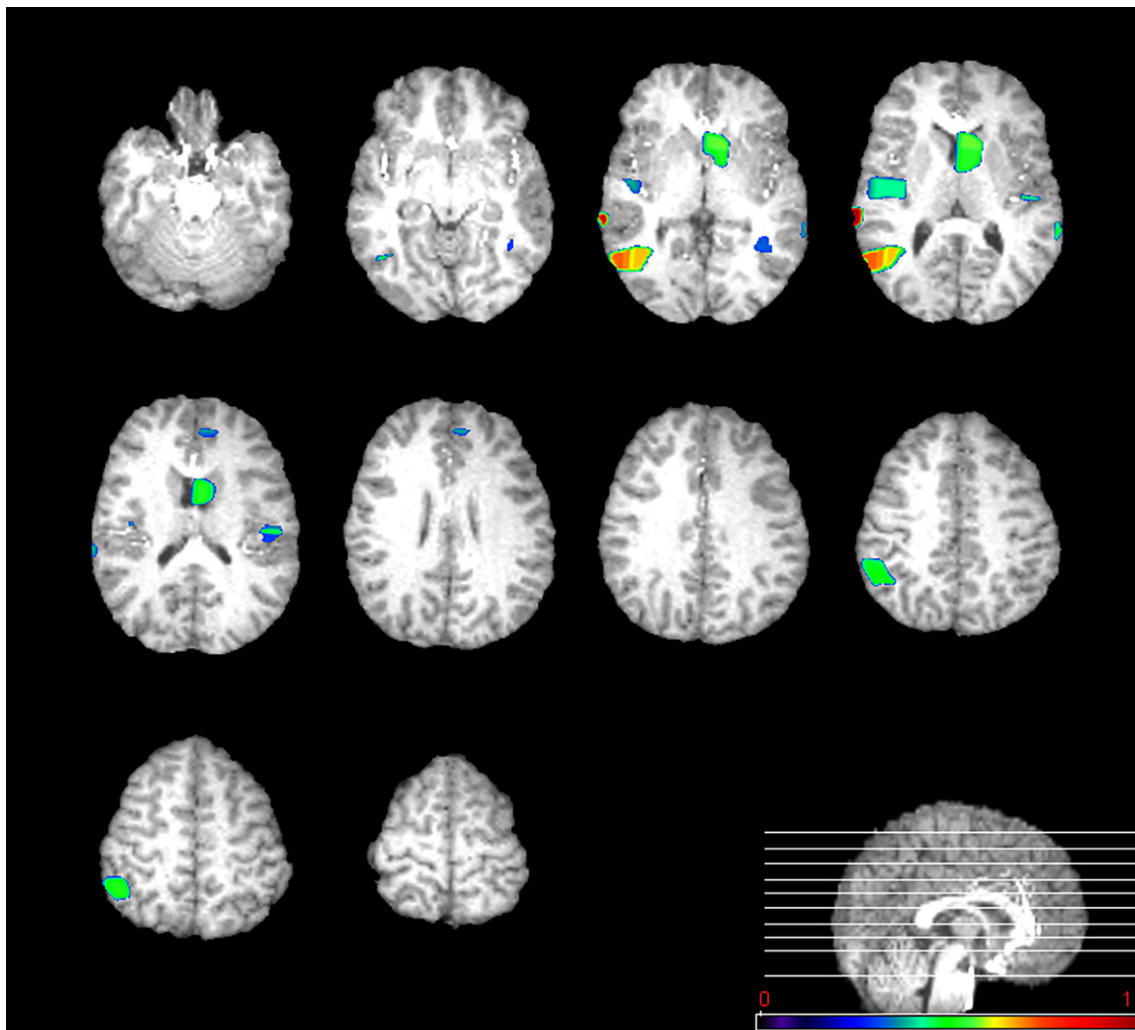


Fig. 8. Group differences identified by our pattern classification method for GM feature map (left is left). The significance of group difference in each region is color-coded according to the color bar shown in the bottom, with 1 as relatively the most important for classification.

requires setting of several parameters, including the number of bins and the number of principal components. On the other hand, PCA-based feature extraction method seems not working well in our small sample size problem, even when we considered using all the possible number of features, from one to the total number of features corresponding to all nonzero eigenvalues, for brain classification. It becomes even worse when the dimensionality of features is extremely large for the cases of structural image-based classification and multimodality image-based classification, as indicated in Table 2. However, the histogram-based regional feature extraction method is more complex than the PCA approach.

#### Group difference

Since three feature maps, i.e., functional feature map, GM and WM tissue density maps, were used, the group differences found in these three maps are overlaid on the template brain by using *MRICro* (<http://www.sph.sc.edu/comd/rorden/mricro.html>), respectively, as shown in Figs. 7–9. For better visualization, the group difference maps are thresholded empirically to make the final group difference maps contain only the clustered regions. It can be

observed that the group differences found in our study confirmed the previous findings, for example, significant group differences located in cingulate/medial frontal in functional imaging, and caudate nucleus in structural imaging (Avants et al., in press; Rao et al., in press). Therefore, the proposed method can be used complementarily with voxel-wise group comparison to detect group difference. It is worth noting that, due to Gaussian spatial smoothing in various stages of our method, the GM (or WM) group difference might slightly display on WM (or GM).

#### Discussion and conclusion

We present a neuroimage-based classification method to examine the brain abnormality using structural and functional MR brain images simultaneously and apply this method to detect prenatal cocaine exposure in adolescents. In the study of separating the prenatally cocaine-exposed adolescents from the socioeconomically matched normal controls, a promising classification performance has been achieved with leave-one-out cross-validation, suggesting that the prenatal cocaine exposure has a long-term effect on brain development, detectable even in adolescents. The

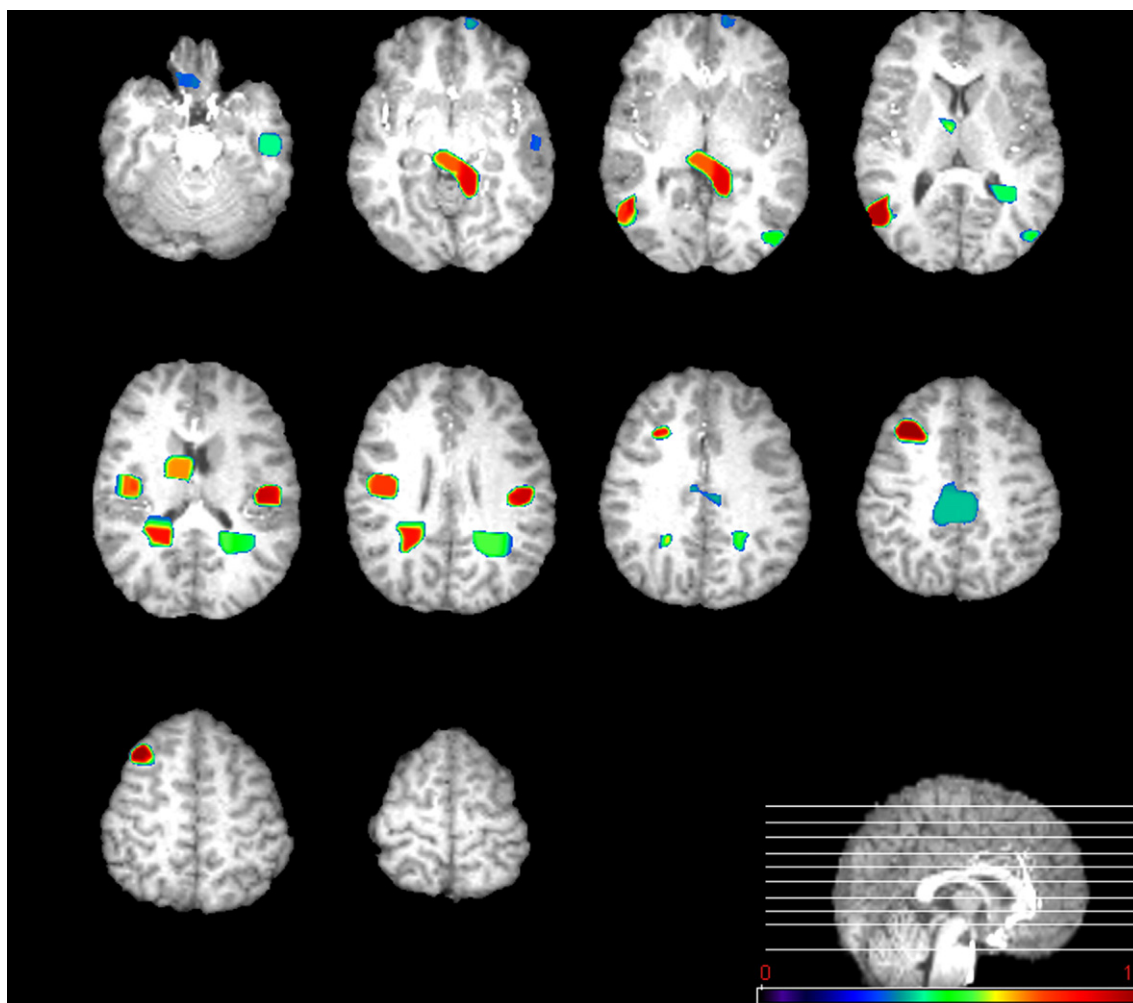


Fig. 9. Group differences identified by our pattern classification method for WM feature map (left is left). The significance of group difference in each region is color-coded according to the color bar shown in the bottom, with 1 as relatively the most important for classification.

group-difference patterns detected by the proposed method are highly consistent with studies using conventional statistical analysis of the structural and functional images, respectively (Avants et al., in press; Rao et al., 2006).

The experimental results have shown that the simultaneous usage of structural and functional images improves the classification accuracy, compared to any of the cases of using only single modality images, i.e., structural MR images or functional MR images. According to the common theme in biology, brain structure and function are interdependent of each other. However, our results show that there exist large differences in structural and functional images at the group level, as indicated in Figs. 7–9. Also, the brain abnormality classifications using only structural or functional images have different performances, as indicated by the classification results. It is worth noting that the classification performance is not always improved when more features are used.

In our method, a statistical regional feature extraction method is used in conjunction with a hybrid feature selection method and a nonlinear SVM classifier (Vapnik, 1999; Rakotomamonjy, 2003; Fan et al., 2005, 2006a,b, 2007). The experimental results show that the PCA-based method has relatively inferior classification performance, indicating that the PCA method is unable to fully capture informative features from a limited number of training samples with images at this dimensionality (Davatzikos et al., 2003). On the other hand, the statistical regional feature-based method performs much better, which should be attributed to the proposed feature extraction method that captures the regional statistical features from the automatically generated brain regions; such method is robust to registration error and inter-subject variation and have good generalization to unseen samples. Although a relatively high correct classification rate is achieved within this leave-one-out cross-validation framework, this result might be optimistic due to the limitation of this cross-validation strategy. We will validate this method in large cohorts in the future.

In conclusion, although several studies have demonstrated that the long-term effects of prenatal drug exposure on the neurocognitive development of children are associated with potential brain structural and functional alterations, the diagnostic value of these findings has never been studied and thus was uncertain. Our study demonstrates that a high-dimensionality nonlinear pattern classification method is capable of accurately detecting spatially distributed and complex patterns of brain alterations associated with prenatal cocaine exposure. The predictive power of these patterns was also very high, as determined from the classification of individuals that were not included in the classifier training procedure. These promising cross-validation results show that the proposed method may provide an alternative and complementary approach to the clinical diagnosis of alterations in brain structure and function associated with prenatal drug exposure as well as other brain disorders.

## Appendix A

Given a voxel-wise feature  $f$  located at  $u$  of the feature map, its spatial consistency  $c(f)$  can be measured by an intraclass correlation coefficient among all features in its spatial neighborhood and across all samples in the training data set. For example, as illustrated in Fig. 2, let us assume that we have  $n$  training samples, and consider the spatial consistency among  $m$  neighbors of feature  $f$  (if the immediate neighborhood is considered, the total

number of neighbors is  $m=27$ ). We construct a  $n \times m$  feature matrix, i.e.,  $\{f_{j,l}(u)\}$ ,  $j=1, \dots, n$ ,  $l=1, \dots, m$ , where  $f_{j,l}(u)$  is the feature value of the  $j$ -th training sample at the location of the  $l$ -th neighbor of  $u$ . In order to measure the agreement amount of features, a two-way random effect model is specified for this feature matrix, i.e.,

$$f_{j,l}(u) = \mu + r_j + c_l + e_{j,l}, \quad j = 1, \dots, n, l = 1, \dots, m, \quad (A1)$$

where  $\mu$  is the grand mean for all density values in the matrix. In the equation above,  $r_j$ ,  $j=1, \dots, n$ , are the row effect-independent random variables, with mean 0 and variance  $\sigma_r^2$ .  $c_l$ ,  $l=1, \dots, m$ , are the column effect independent random variables, with mean 0 and variance  $\sigma_c^2$ .  $e_{j,l}$ ,  $j=1, \dots, n$ ,  $l=1, \dots, m$ , are the residual effect independent random variables, with mean 0 and variance  $\sigma_e^2$ . Based on above model, the spatial consistency  $c(f)$  of a feature  $f$  can be theoretically computed by (McGraw and Wong, 1996)

$$C(f) = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_c^2 + \sigma_e^2}. \quad (A2)$$

In practice, the spatial consistency  $c(f)$  can be estimated from the feature matrix  $\{f_{j,l}(u)\}$ ,  $j=1, \dots, n$ ,  $l=1, \dots, m$ , by calculating a mean square for rows ( $MS_R$ ), a mean square for columns ( $MS_C$ ), and a mean square error for residuals ( $MS_E$ ). In our applications, the value of  $c(f)$  is constrained to be between 0 and 1.

$$c(f) = \frac{MS_R - MS_E}{MS_R + (m-1)MS_E + \frac{m}{n}(MS_C - MS_E)},$$

$$MS_R = m \left( \sum_{j=1}^n r_j - g \right)^2 / (n-1),$$

$$MS_C = n \left( \sum_{l=1}^m c_l - g \right)^2 / (m-1),$$

$$MS_E = \left( \sum_{j=1}^n \sum_{l=1}^m f_{j,l}(u)^2 - (n-1)MS_R - (m-1)MS_C - m \cdot n \cdot g^2 \right) / ((m-1)(n-1)), \quad (A3)$$

$$r_j = \sum_{l=1}^m f_{j,l}(u) / m, \quad c_l = \sum_{j=1}^n f_{j,l}(u) / n, \quad g = \sum_{j=1}^n \sum_{l=1}^m f_{j,l}(u) / (mn).$$

## References

- Avants, B., Hurt, H., Giannetta, J., Epstein, C.L., Shera, D., Rao, H., Wang, J., Gee, J., in press. Effects of Heavy In Utero Cocaine Exposure on Adolescent Caudate Nucleus: A Structural MRI Study. *Pediatr. Neurol.*
- Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G., 1995. Automated multi-modality image registration based on information theory. In: Paola, R.D. (Ed.), *Information Processing in Medical Imaging*. Kluwer, Dordrecht, The Netherlands, pp. 263–274.
- Cox, D.D., Savoya, R.L., 2003. Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* 19, 261–270.
- Davatzikos, C., Genc, A., Xu, D., Resnick, S.M., 2001. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. *NeuroImage* 14, 1361–1369.
- Davatzikos, C., Tao, X., Shen, D.G., 2003. Hierarchical active shape models using the wavelet transform. *IEEE Trans. Med. Imag.* 22, 414–423.

- Davatzikos, C., Ruparel, K., Fan, Y., Shen, D., Acharyya, M., Loughhead, J., Gur, R.C., Langleben, D., 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *NeuroImage* 28, 663–668.
- Dehmeshki, J., Barker, G.J., Tofts, P.S., 2002. Classification of disease subgroup and correlation with disease severity using magnetic resonance imaging whole-brain histograms: application to magnetization transfer ratios and multiple sclerosis. *IEEE Trans. Med. Imag.* 21, 320–331.
- Duda, R.O., Hart, P.E., Stork, D.G., 2001. *Pattern Classification*. John Wiley and Sons, Inc.
- Fan, Y., Shen, D., Davatzikos, C., 2005. Classification of Structural Images via High-Dimensional Image Warping, Robust Feature Extraction, and SVM. In: Duncan, J.S., Gerig, G., Duncan, J.S., Gerig, G., Duncan, J.S., Gerig, G.s. (Eds.), *MICCAI*. Palm Springs, California, USA: Springer Berlin/Heidelberg, pp. 1–8.
- Fan, Y., Rao, H., Giannetta, J., Hurt, H., Wang, J., Davatzikos, C., Shen, D., 2006a. Diagnosis of brain abnormality using both structural and functional MR images. *IEEE 2006 International Conference of the Engineering in Medicine and Biology Society*. New York City, NY, USA.
- Fan, Y., Shen, D., Davatzikos, C., 2006b. Decoding cognitive states from fMRI images of subjects by machine learning and multivariate classification. *IEEE Workshop on Mathematical Methods in Biomedical Image (MMBIA 2006)*. New York.
- Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C., 2007. COMPARE: Classification Of Morphological Patterns using Adaptive Regional Elements. *IEEE Trans. Med. Imag.* 26 (1), 95–105.
- Ford, J., Frarid, H., Makedon, F., Flashman, L.A., McAllister, T.W., Megalooikonomou, V., Saykin, A.J., 2003. Patient classification of fMRI activation maps. *MICCAI03*.
- Frackowiak, R.S.J., Friston, K.J., Frith, C., Dolan, R., Price, C.J., Zeki, S., Ashburner, J., Penny, W.D., 2003. *Human Brain Function*. Academic Press.
- Friston, K.J., 1997. Testing for anatomically specified regional effects. *Hum. Brain Mapp.* 5, 133–136.
- Golland, P., Fischl, B., 2003. Permutation tests for classification: towards statistical significance in image-based studies. *International Conference on Information Processing and Medical Imaging*, pp. 330–341.
- Golland, P., Grimson, W.E.L., Shenton, M.E., Kikinis, R., 2001. Deformation analysis for shape classification. *The 17th International Conference on Information Processing in Medical Imaging*, pp. 517–530.
- Golland, P., Fischl, B., Spiridon, M., Kanwisher, N., Buckner, R.L., Shenton, M.E., Kikinis, R., Dale, A., Grimson, W.E.L., 2002. Discriminative analysis for image-based studies. In: Dohi RK, T., Dohi RKs, T. (Eds.), *Fifth International Conference on Medical Image Computing and Computer Assisted Intervention*. Springer-Verlag GmbH, Tokyo, Japan, pp. 508–515.
- Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182.
- Hayasaka, S., Nichols, T.E., 2004. Combining voxel intensity and cluster extent with permutation test framework. *NeuroImage* 23, 54–63.
- LaConte, S., Strother, S., Cherkassky, V., Anderson, J., Hu, X., 2005. Support vector machines for temporal classification of block design fMRI data. *NeuroImage* 26, 317–329.
- Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C., 2004. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage* 21, 46–57.
- Liu, Y., Teverovskiy, L., Carmichael, O., Kikinis, R., Shenton, M., Carter, C.S., Stenger, V.A., Davis, S., Aizenstein, H., Becker, J., Lopez, O., Meltzer, C., 2004. Discriminative MR image feature analysis for automatic schizophrenia and Alzheimer's disease classification. In: Barillot, C., Haynor, D.R., Hellier, P. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004: 7th International Conference*. Springer-Verlag GmbH, Saint-Malo, France, pp. 393–401.
- McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1, 30–46.
- Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., 2004. Learning to decode cognitive states from brain images. *Mach. Learn.* 57, 145–175.
- Mourão-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *NeuroImage* 28, 980–995.
- Pham, D.L., Prince, J.L., 1999. Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Trans. Med. Imag.* 18, 737–752.
- Poline, J.B., Worsley, K.J., Evans, A.C., Friston, K.J., 1997. Combining spatial extent and peak intensity to test for activations in functional imaging. *NeuroImage* 5, 83–96.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge Univ. Press, pp. 610–614.
- Rakotomamonjy, A., 2003. Variable selection using SVM based criteria. *J. Mach. Learn. Res.* 3, 1357–1370.
- Rao, H., Wang, J., Korkczykowski, M., Giannetta, J., Shera, D., Avants, B., Gee, J., Detre, J.A., Hurt, H., in press. Altered Resting Cerebral Blood Flow in Adolescents with In-utero Cocaine Exposure Revealed by Perfusion Functional MRI. *Pediatrics*.
- Schiele, B., Crowley, J.L., 2000. Recognition without correspondence using multidimensional receptive field histograms. *Int. J. Comput. Vis.* 36, 31–50.
- Shen, D., Davatzikos, C., 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imag.* 21, 1421–1439.
- Thirion, B., Pinel, P., Poline, J.-B., 2005. Finding landmarks in the functional brain: detection and use for group characterization. *MICCAI 2005*.
- Thomaz, C.E., Boardman, J.P., Hill, D.L.G., Hajnal, J.V., Edwards, D.D., Rutherford, M.A., Gillies, D.F., Rueckert, D., 2004. Using a maximum uncertainty LDA-based approach to classify and analyse MR brain images. In: Barillot, C., Haynor, D.R., Hellier, P. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI'02*. Springer-Verlag, Tokyo, Japan, pp. 355–362.
- Timoner, S.J., Golland, P., Kikinis, R., Shenton, M.E., Grimson, W.E.L., Wells, W.M., 2002. Performance issues in shape classification. *Proceedings of the Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI'02*. Springer-Verlag, Tokyo, Japan, pp. 355–362.
- Vapnik, V.N., 1999. *The Nature of Statistical Learning Theory (Statistics for Engineering and Information Science)*. Springer-Verlag.
- Vincent, L., Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* 13, 583–589.
- Wang, J., Zhang, Y., Wolf, R.L., Roc, A.C., Alsop, D.C., Detre, J.A., 2005. Amplitude modulated continuous arterial spin labeling perfusion MR with single coil at 3 T-feasibility. *Radiology* 235, 218–228.
- Wilcox, R., 1997. *Introduction to Robust Estimation and Hypothesis Testing*. Academic Press, New York.
- Yushkevich, P., Joshi, S., Pizer, S.M., Csernansky, J.G., Wang, L.E., 2003. Feature selection for shape-based classification of biological objects. In: C. Taylor, A.N., C. Taylor, A.Ns (Eds.), *Information Processing in Medical Imaging*. Springer-Verlag, Ambleside, UK, pp. 114–125.
- Zhang, L., Samaras, D., Tomasi, D., Alia-Klein, N., Cottone, L., Leskovic, A., Volkow, N., Goldstein, R., 2005. Exploring temporal information in functional magnetic resonance imaging brain data. *MICCAI 2005*.