ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 28 (2005) 980 – 995

Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data

Janaina Mourão-Miranda,[a] Arun L.W. Bokde,[b] Christine Born,[c] Harald Hampel,[b] and Martin Stetter [a,*]

[a]*Siemens Corporate Technology, Information and Communications, Munich, Germany*
[b]*Dementia and Neuroimaging Research Section, Alzheimer Memorial Center and Geriatric Psychiatric Branch, Department of Psychiatry, Ludwig-Maximilian University, Munich, Germany*
[c]*Institute for Clinical Radiology, Ludwig-Maximilian University, Munich, Germany*

Received 23 February 2005; revised 17 June 2005; accepted 23 June 2005
Available online 4 November 2005

In the present study, we applied the Support Vector Machine (SVM) algorithm to perform multivariate classification of brain states from whole functional magnetic resonance imaging (fMRI) volumes without prior selection of spatial features. In addition, we did a comparative analysis between the SVM and the Fisher Linear Discriminant (FLD) classifier. We applied the methods to two multisubject attention experiments: a face matching and a location matching task. We demonstrate that SVM outperforms FLD in classification performance as well as in robustness of the spatial maps obtained (i.e. discriminating volumes). In addition, the SVM discrimination maps had greater overlap with the general linear model (GLM) analysis compared to the FLD. The analysis presents two phases: during the training, the classifier algorithm finds the set of regions by which the two brain states can be best distinguished from each other. In the next phase, the test phase, given an fMRI volume from a new subject, the classifier predicts the subject's instantaneous brain state.

Keywords: Machine learning methods; Support Vector Machine; Classifiers; Functional magnetic resonance imaging data analysis

Introduction

Functional brain imaging has been used in a large number of experiments that identify which regions of the brain are associated with which cognitive processes. Researchers have addressed the study of human cognitive process by investigating which areas activated when performing a task in comparison to another task. Thus, an experiment was based on a priori hypothesis, and the design of the experiment was optimized to answer this question. This approach is a natural way to search for functional localization. The analysis of the functional imaging data is done through a univariate approach, that is, each voxel is treated individually. The standard approach for fMRI data analysis is the GLM (Friston et al., 1995a). The GLM models treat the observed time series at each voxel as a linear combination of explanatory functions (descriptors that govern the overall condition of the experiment) plus an error term. The outputs from this approach are statistical parametric maps showing differences in brain activation between tasks at a specific P value.

The fMRI data are multivariate in nature since each fMRI volume contains information about brain activation at thousands of measured locations (voxels). Multivariate techniques have been applied to neuroimage data in many studies (for review, see Friston and Büchel, 1997). For example, Friston et al. (1995b) introduced a multivariate approach using standard multivariate statistic (MANCOVA) and GLM to make inferences about effects of interest and canonical variates analysis (CVA) to describe the important feature of these effects. McIntosh et al. (1996) introduced partial least squares (PLS) as a multivariate method. In this work, they performed singular value decomposition (SVD) analysis on the cross-correlation matrix between brain images and experimental design matrices to obtain spatial patterns of brain activity representing the optimal association between the brain images and either of the experimental conditions. McKeown et al. (1998) described a multivariate method for analyzing fMRI data based on independent components analysis (ICA) (Bell and Sejnowski, 1995), which can be used to distinguish between non-task-related signal components, movements and other artifacts, as well as consistently or transiently task-related fMRI activations. Kherif et al. (2002) presented a general method for specifying a model to analyze neuroimaging data based on multivariate linear model on a training data set. These techniques have mostly been

* Corresponding author. Neural Computation, Siemens AG, CT IC 4, Otto-Hahn-Ring 6, 81739 Munich, Germany. Fax: +49 89 636 49767.
E-mail address: stetter@siemens.com. (M. Stetter).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2005.06.070

used to detect brain activations, to reduce noise artifacts, to describe the variability of the data in a concise manner, and to specify a model to be applied on the data.

Multivariate analysis has also been applied to fMRI data analysis with focus on reproducibility of fMRI. Tegeler et al. (1999) studied the reproducibility of activation patterns in the whole brain by using scatter plots of statistical values and calculating the correlation coefficient between pairs of activation maps. They used a *t* test between control and task periods as a univariate method and FLD as a multivariate approach.

Another interesting aspect to be investigated with multivariate analysis is the mapping from observed fMRI data to a subject's instantaneous brain state, instead of mapping from task to brain locations. The mapping of activities in the brain to cognitive states can be treated as a pattern recognition problem. In this approach, the fMRI data can be treated as a spatial pattern, and a statistical pattern recognition method (e.g., a machine learning algorithm) is used to map these patterns to instantaneous brain states. The algorithm is designed to learn and later predict or classify multivariate data based on statistical properties of the data set.

The previous use of classifiers for fMRI data analysis can be divided in two groups. The first group applied classifiers after preprocessing using a feature selection methods based on prior hypotheses (Mitchell et al., 2004; Wang et al., 2003; Cox and Savoy, 2003; Ford et al., 2003). In these studies, the data were encoded as a vector of features, one feature for each voxel (hundred of thousands of features). Because of the high dimensionality of this feature vector, a feature selection method based on prior hypotheses was applied to the data set to reduce the dimensionality, and afterwards the selected features were used as inputs to a classifier, that is, the discriminating regions were chosen a priori and given as input to the classifiers. The second group used principal components analysis (PCA)/singular value decomposition (SVD) analysis as dimensionality reduction method and applied the classifier on PCA/SVD basis without prior selection of spatial features (e.g., Mørch et al., 1997; Kjems et al., 2002; LaConte et al., 2003; Carlson et al., 2003; Strother et al., 2004). Mørch et al. (1997) introduced the concept of models of functional activation for classification. They compared the performance of two models, a linear and a nonlinear based on artificial neural networks, by measuring the generalization error as a function of the number of training examples. One important advantage of these studies is that the algorithm finds the discriminating regions, and this information can be presented as "spatial activation maps". However, the focus of most of these works was on evaluation of the performance of models for neuroimaging data analysis (Kjems et al., 2002) and on evaluation of preprocessing choices for neuroimaging data analysis (LaConte et al., 2003; Strother et al., 2004).

The first study that focused on detecting global brain activation pattern to discriminate between two or more cognitive states was by Carlson et al. (2003) who applied FLD on PCA basis to investigate patterns of activity in the categorical representation of objects. By using this approach, they seek to find which voxels contribute to the pattern of activity that is indicative of the stimulus category the observer is viewing. The FLD classifies by linearly projecting the training set on the axis that is defined by the difference between the center of mass for both classes (tasks) corrected by the within-class covariance. One drawback of the FLD is the high susceptibility of the mean and the covariance estimates to contamination by outliers.

In the present paper, we applied the SVM algorithm to map whole fMRI volumes from different subjects to brain states without prior selection of features. In addition, we did a comparative analysis between the SVM and the FLD. The SVM is based in statistical learning theory (Vapnik, 1995). It has emerged as a powerful tool for statistical pattern recognition (Boser et al., 1992). It selects from many possible solutions the optimal one. The optimal solution is determined by the most informative training examples, i.e. the support vectors. We demonstrate that SVM clearly outperforms FLD in classification performance as well as in the robustness of the spatial maps obtained. Our novel contribution takes advantage of the good scaling in high dimensions and the large margin properties of SVM for the fMRI brain state classification problem.

Materials and methods

Outline of study

The overall scheme for classification, as shown in Fig. 1, starts by training the classifier using fMRI volumes from different subjects (training data set) acquired during two brain states (e.g., task 1 and task 2). Each fMRI volume is treated as a feature vector in a high-dimensional space, where each feature is the fMRI signal at a specific voxel at a specific time. PCA is applied for dimensionality reduction by projecting each training volume onto the principal components obtained from all training volumes. During the training phase, the learning algorithm (SVM or FLD) finds the most discriminating features (or voxels) between the two brain states. During the test phase, after the training, an fMRI volume from a new subject (test data set) acquired during one of two predefined brain states is projected onto the principal components and provided as input to the system. Finally, the classifier predicts as output in which brain state for each volume the subject was.

We tested the performance of the SVM and FLD classifiers for their ability to distinguish different brain states from fMRI single volumes independent of the variability between subjects. We applied the methods to two different visual attention experiments: a face matching and a location matching. Previous neuroimaging experiments have shown selective activation of the ventral visual pathway when attending to faces stimuli and the dorsal visual pathway when attending to spatial information (Corbetta et al., 1991a,b; Haxby et al., 1991, 1994). In these tasks, we expected the most discriminate regions to be in the ventral and dorsal visual path for the face and location task, respectively.

Additionally, to test the robustness of both methods, we compare their performance by using the training sets with and without spatial filter. It was shown previously by using CVA (the multivariate extension of FLD) (LaConte et al., 2003; Strother et al., 2004) that the spatial filter has a high impact on prediction accuracy. Gaussian spatial filter is a standard preprocessing step for fMRI data analysis, especially for multisubject comparisons where it compensates for the anatomical differences between the subjects and increases the signal to noise ratio. The use of a multisubject training set without spatial filter represents a challenge for the machine learning methods. Even in this situation, the SVM algorithm was able to find between-subject regularities and successfully predict instantaneous brain states.

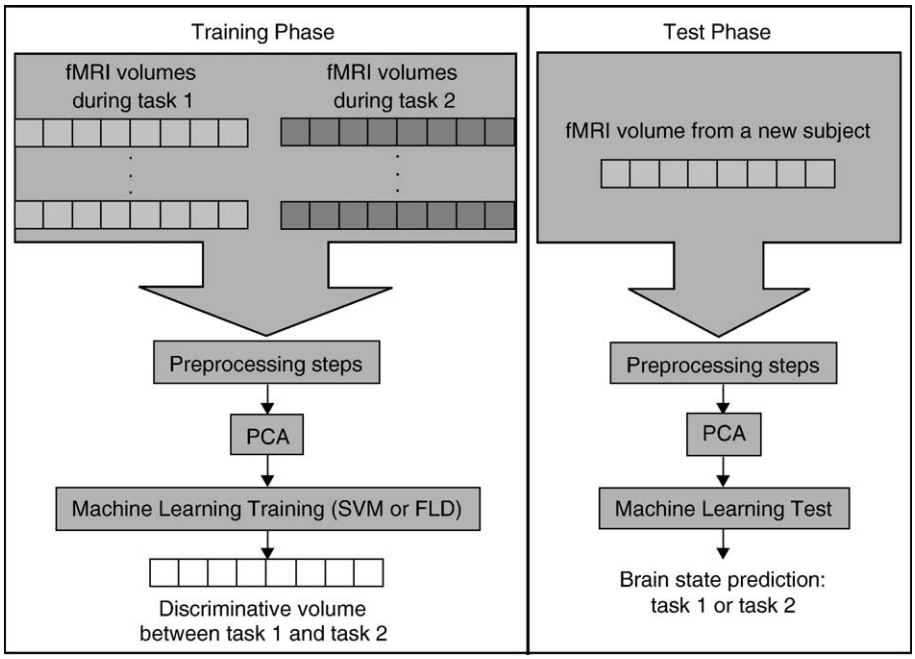

Fig. 1. Schematic illustration of the steps for the classification. The analysis presents two phases: during the training, the classifier algorithm finds the set of regions by which the two brain states can be best distinguished from each other, i.e. the discriminating volume. In the next phase, the test phase, given an fMRI volume from a new subject, the classifier predicts the subject's instantaneous brain state. For both phases, the data are preprocessed, and principal components analysis (PCA) is applied for dimensionality reduction.

Subjects

We used fMRI data from 10 female and 6 male right-handed healthy subjects with average age 66.3 (standard deviation = 4.3) years. Participants did not have any history of neurological and psychiatric illness. All subjects had normal vision or corrected by use of MR-compatible eyeglasses to normal vision. All subjects gave written informed consent to participate in the study after the study was explained to them. The study was performed in accordance with the Declaration of Helsinki and the Ethics Committee of the Medical Faculty of Ludwig Maximilian University approved the study.

Data acquisition

The data were acquired with a 1.5 T Siemens Vision scanner (Siemens, Erlangen, Germany). The imaging sequence was an interleaved T2*-weighted echo planar sequence with 28 axial slices with interleaved echo planar sequence (slice thickness = 4 mm, gap between slices = 1 mm, repetition time (TR) = 3.6 s, repetition time per slice = 0.60, echo time (TE) = 60 ms, flip angle = 90°, field of view (FOV) = 240 mm and matrix = 64 × 64). In each run, 69 functional volumes were acquired. For anatomical reference in each subject, a 28-slice T1-weighted structural image was acquired in the same orientation as the EPI sequence (TR = 620 s, TE = 12 ms, flip angle = 90°, field of view (FOV) = 240 mm, matrix = 224 × 256, rect. FOV = 7/8, effective thickness = 1.25 mm) and a high resolution T1-weighted 3D Magnetization Prepared rapid Gradient Echo (MPRAGE) structural image was acquired (TR = 11.4 ms, TE = 4.4 ms, flip angle = 8°, FOV = 270 mm, matrix = 224 × 256, rect. FOV 7/8, effective thickness = 1.25 mm).

Stimuli and tasks

Stimuli were presented in a blocked fashion. There were three different experimental conditions: instruction period, control task and attentional task. In each run, there were 3 blocks of the attentional task (each with 7 scans) and 4 blocks of the control task (with 7 scans in the first three and 9 scans in the last one). Each task block was preceded by an instruction period (each with 2 scans). There were two runs per subject, in one, the attentional task was a face matching task and, in the other, was a location matching task. The face matching task consisted of two faces presented simultaneously, and participants were asked to decide on each trial if a pair of faces was identical or not (Fig. 2A). If they were, the subject would respond by pressing a button in the right hand. No response was required if the faces were dissimilar. There were 8 trials of the attentional task per block. The faces were from the Max Planck Institute for Biological Cybernetics database (Blanz and Vetter, 1999). Each run also included 4 volumes at the beginning of the run to factor out T1 magnetic transients.

The location matching task, as shown in Fig. 2B, consisted of two abstract images located within a smaller square. The subject had to decide if the relative location of the small square relative to the larger one was the same. The subject would press a button if the relative locations were identical.

In the control task for the face and location matching tasks, the subject had to press the button every time an image appeared. The image was identical to the location matching task, with the squares located always in the center. The parameters for the presentation of the images were identical to the attentional task. The different conditions were counter-balanced across subjects.

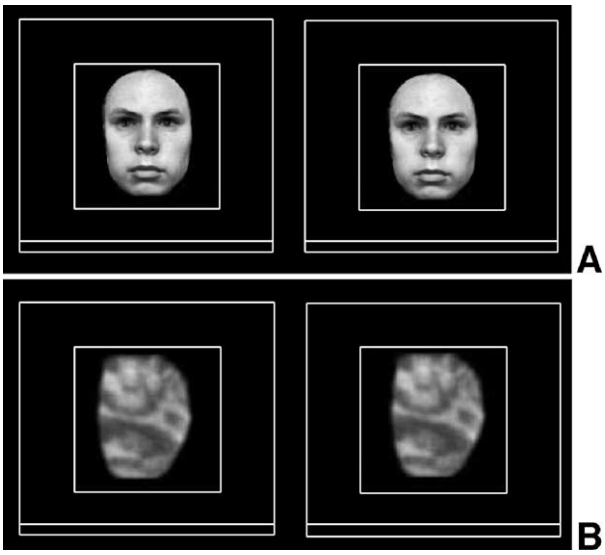

Fig. 2. Illustration of the stimuli during the attentional tasks: (A) Face matching task. (B) Location matching task.

During the task, performance was monitored, and accuracy and response times were measured.

Data preprocessing

The initial step was to delete the first 4 volumes of each scan to remove the initial T1 magnetic transients in the data. The remaining data were corrected for the timing differences between each slice using Fourier interpolation. Then, the data were corrected for motion effects (6-parameter rigid body), where the reference volume was in the center of the run. The motion-corrected data were registered to the T1 structural image (6-parameter rigid body transformation), the T1 structural image was registered to the MPRAGE image (6 parameter rigid body) and the MPRAGE image was normalized using a 12-parameter affine transformation to the Montreal Neurological Institute/International Consortium for Brain Mapping 152 standard (MNI/ICBM), as contained within the FSL software package. The fMRI data were then normalized to the MNI/ICBM standard utilizing the transformations described previously. The data were spatially smoothed using an isotropic Gaussian filter (FWHM = 8 mm). The baseline and the low frequency components were removed by applying a regression model for each voxel. The low frequency components were modeled by a set of discrete cosine functions as described in Holmes and Friston (1997). The cut-off period chosen was 128 s. Finally, a mask was applied to select voxels which contain brain tissue for all subjects (filtered preprocessed data set 1).

For testing the classifiers performance in an extreme situation, we used the data treated with all preprocessing steps except spatial filtering (unfiltered preprocessed data set 2).

Dimensionality reduction

The dimensionality reduction of the high-dimensional fMRI data is important because it decreases the computational complexity of the classification algorithms. In cases where the data dimensionality M (voxels) exceeds the number of data points N (volumes or scans), i.e. ill-posed data sets, one can represent the data in a space of smaller dimensionality without loss of information (Mørch et al., 1997). This PCA/SVD formulation has been previously applied as a dimensionality reduction method in neuroimaging field (e.g., Friston et al., 1995b; Weaver, 1995; Strother et al., 2004; Kjems et al., 2002; LaConte et al., 2003).

We used PCA to find the bases of reduced dimensionality. PCA finds an orthogonal basis (principal components—PCs) that explains the greatest amount of variance of the data (Jackson, 1991). By projecting each data point of the original data (in our case an fMRI volume) onto the principal components, a set of correlated variables (voxels) is transformed into a set of uncorrelated variables (projections of the fMRI volume on the principal components). The uncorrelated variables can be expressed as linear combinations of the original variables.

We can define a centered data matrix $M \times N$ with one volume per column and one voxel per row as $\mathbf{D}_c$. PCA consists of finding the eigenvalues and eigenvectors of the $M \times M$ covariance matrix of the data matrix $\mathbf{D}_c$, where $M = 10^5$. Thus, determining M eigenvectors and eigenvalues is of prohibitive computational expenses. However, when the number of dimensions (M) exceeds the number of samples or scans (N), there will be only $N - 1$, rather M, meaningful eigenvectors (the remaining eigenvectors will have associated eigenvectors of zero). It is possible to compute the $N - 1$ PCs of $\mathbf{D}_c$ using singular value decomposition (SVD). For completeness, we provide a description of PCA/SVD in the Appendix A.

In the present work, we did not exclude any PC in the analysis, that is, the PCA step is loss–less dimension reduction and represents only a change of the coordinate system to the subspace spanned by the measured brain volumes.

Classifier

When treating each projected volume as a point in a high-dimensional space (space dimension = number of principal components), the task of linearly classifying fMRI volumes in two classes can be viewed as a task of finding a separating hyperplane. The separating hyperplane is a linear function that is capable of separating the training data (projected volumes) without error. It is described in terms of the training set by

$$\mathbf{H} : (\boldsymbol{w}^p)^T \boldsymbol{v}_i^p + b = 0 \quad (1)$$

where $\boldsymbol{w}^p$ is a learning weight vector, b an offset and $\boldsymbol{v}_i^p$ are the projected volumes of both classes onto the principal components, the superscript p indicates vector/matrix after projection. To illustrate this idea, one can consider the case of only two principal components. In this case, each projected volume can be plotted as a point in a two-dimensional space (first principal component versus second principal component). In Fig. 3A is displayed a hypothetical example with six volumes. The circles represent volumes acquired during task 1 and the squares volumes acquired during task 2. The dashed lines represent possible separating hyperplanes.

Once the separating hyperplane is learned from the training data (fMRI data from group of subjects), it corresponds to a decision function that can be used for classifying a test example (fMRI volume from a new subject). This decision function is able to map fMRI volumes to brain states. Depending on the learning method applied, there are many possible solutions or hyperplanes. For example, for the FLD, the separating hyperplane is perpendicular to the axis that is defined by the difference between the center of mass for both classes (tasks) corrected by

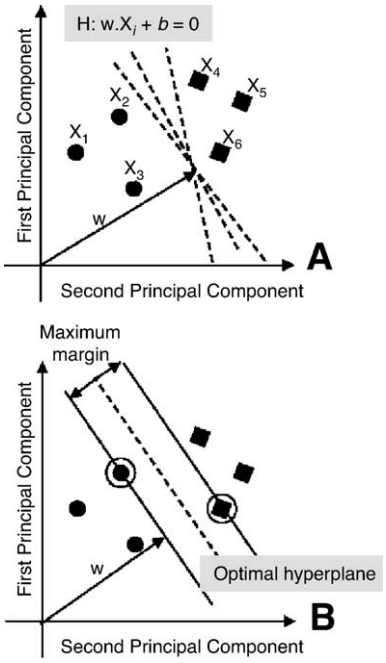

Fig. 3. (A) Illustration of a classification problem between task 1 and task 2 for the simplified case of only two principal components. Each axis represents one principal component. Each symbol represents a brain volume. The circles represent the volumes acquired during task 1, and the squares the volumes acquired during task 2. The dashed lines represent hyperplanes that correctly classify the tasks. The separating hyperplanes (**H**) are described by a learning weight vector ($\boldsymbol{w}$) and an offset (b). (B) Illustration of a hyperplane determined by the Support Vector Machine algorithm. The optimal hyperplane (dashed line) is the one with a maximal margin of separation between the two classes. The symbols at the margin (circled) are the support vectors.

the within-class covariance. However, a classifier that correctly classifies a training set can fail for unseen examples and therefore generalizes badly. Based on generalization performance, one can choose between different learning methods or classifiers.

Support Vector Machine

The SVM algorithm proposed by Boser et al. (1992) finds the largest margin hyperplane. The minimal distance from the separating hyperplane to the closest training example is called margin. They showed that the optimal hyperplane is the one with maximal margin (i.e. more separation between the classes). A larger margin corresponds to the better generalization.

The training examples that lie on the margin are called support vectors, they are conceptually the most difficult data points to classify and therefore they define the location of the separating hyperplane. In Fig. 3B is illustrated an SVM classifier for a two-dimensional problem. Advantages of this method are the selection of training examples that are most informative for the classification and a good scaling for high dimensions.

By using a kernel trick, the SVM can be extended to find nonlinear boundaries. In addition, if the training set is not separable, slack variables can be used to allow misclassification of difficult or noise examples. However, in high dimensions with few data points (the situation here), a linear classifier always shatters the problem, hence, a linear kernel is sufficient by definition. For the same reason, slack variables are unnecessary here and would be only overhead. Instead, we face the problems that a high-dimensional space as the fMRI volume space is essentially empty, and we have too many solutions that heavily overfit and generalize badly. Exactly, this problem is solved by the SVM.

We used a linear kernel Support Vector Machine that allows direct extraction of the weight vector as an image (i.e. the discriminating volume). The SVM toolbox for Matlab was used to perform the classifications: http://www.cis.tugraz.at/igi/aschwaig/software.html. In Appendix B is the mathematical formulation of the method; for a detailed description of the SVM, see Schölkopf and Smola (2002).

Additionally, the FLD was used as a benchmark approach. The FLD classifies by linearly projecting the training set on the axis that is defined by the difference between the center of mass for both classes, corrected for within-class correlations. It is equivalent to a two-class Canonical Variates Analysis (CVA), which was introduced in neuroimaging field by Friston et al. (1995b) and has been used to address prediction accuracy (e.g., LaConte et al., 2003; Strother et al., 2004).

Discriminating volume

The separating hyperplane is determined to be orthogonal to the direction along which the training examples of both classes differ most, i.e. the weight vector. If the examples ($\boldsymbol{v}_i$) are in the voxel space (one voxel per dimension), the weight vector ($\boldsymbol{w}$) will be the direction along which the volumes of either tasks or cognitive states differ most. Hence, it represents a volume with the most discriminating regions, i.e. the discriminating volume. Given two classes, task 1 and task 2, with the label y_i equal to +1 and −1 (for $i = 1$ and $i = 2$, respectively), a positive value in a position of the discriminating volume (i.e. in a voxel) means that these voxels presented higher activity during tasks 1 than during task 2 in most of the training volumes, and a negative value means lower activity during task 1 than during task 2 in most of the training volumes. Because the classifier is multivariate by nature, the combination of all discriminating voxels as a whole is identified as global activation patterns by which the brain states differ.

In fact, as we used the reduced representation of the data ($\mathbf{D}^p$) as input to the classifiers, we will get the weight vector ($\boldsymbol{w}^p$) in a reduced representation. To recover the volume with the most discriminating regions in the original space, we need to map back the weight vector to the high-dimensional space (i.e. voxel space). The weight vector or discriminating volume in the voxel space is given by

$$\boldsymbol{w} = \mathbf{E}\boldsymbol{w}^p \tag{2}$$

where **E** is a matrix containing one eigenvector or PC per column.

Classifier performance

We evaluated the performance of the classifier using the leave-one-subject-out cross validation test. In each trial, we used data from all but one subject ($S - 1$ of the S subjects) to train the classifier. Subsequently, the class assignment of the remaining subject, which was so far unseen by the algorithm, was calculated during the test or application phase. This procedure was repeated S times, each time leaving out a different subject. The general classification procedure was to use 42 volumes (21 of each task) of each training subject to train the classifiers. After the training, the classifier was tested in classifying 42 volumes (21 of each task) of

a test subject outside the training set. To quantify the results, we measured the error rate, the ratio of the number of test volumes wrongly classified to the total of tested volumes. We also quantified the sensitivity and specificity of each test defined as

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (4)$$

where TP is the number of true positives: number of volumes of the task 1 correctly classified; TN is the number of true negatives: number of volumes of the task 2 correctly classified; FP is the number of false positives: number of volumes of the task 2 classified as task 1; FN is the number of false negatives: number of volumes of the task 1 classified as task 2.

Permutation test

The absolute magnitude of each element of the discriminating volume (i.e., of each voxel) determines its importance in discriminating the brain states. The discriminating volume can be thresholded so that the most important voxels for discriminating between cognitive states are selected. The threshold can be set arbitrarily or a probability map can be computed and used to set the threshold level. It is possible to obtain the probability map by testing the null hypothesis that there are no differences between the brain states during the two different tasks, that is, the labels have no contribution for the classification. One possible approach is to utilize a parametric statistical test. In this case, we need to know the probability distribution of the selected statistic (values of the elements of the discriminating volume) under the null hypothesis. However, deriving a parametric distribution for a particular statistic requires making strong assumptions on the generative model of the data. A second possible approach is to apply nonparametric techniques, such as permutation tests, which estimate empirically the distribution of the statistic under a null hypothesis. Nonparametric tests have been previously applied to fMRI data analysis (e.g., Holmes et al., 1996; Bullmore et al., 1996; Nichols and Holmes, 2002). By permuting the class labels 2000 times randomly and training the SVM with this permutation of labels, we estimated a probability distribution for each voxel of the discriminating volume under the null hypothesis of no relationship between the class labels and the global structure of the fMRI volumes. Based on these probability distributions, it is possible to test the null hypothesis at the voxel level: if a voxel value of the discriminating volume – when learnt from the original data set – lies far outside the major mass of the distribution under null hypothesis – indicated by a small P value – it is likely to be significantly predictive for the class label. The P value is calculated as the proportion of values in the distribution under null hypothesis that is greater or equal to the value obtained by using the original (i.e. nonpermutated) training data. The probability maps of discriminating volumes maps shown in the results section display all significant voxels under P values < 0.05 and P values < 0.001.

Classifier vs. GLM

In order to compare the results from the classification methods with the standard general linear model (Friston et al., 1995a), we evaluated the overlap between the results of a Statistic Parametric t-Map (SPMt) and of the discriminating volume using the training set with spatial filter. The GLM analysis was performed using SPM2 (Wellcome Department of Cognitive Neurology, London, UK). Each task on the voxel time series was modeled as box-car function smoothed with a canonical hemodynamic response function. The statistical model included global and low frequency components as described in Holmes and Friston (1997). The contrast between the tasks effects was assessed at each voxel using a fixed-effect model.

To quantify the overlap between the SPMt map and the discriminating volume, we kept the SPMt threshold fixed (at a level corresponding to a corrected P value of 0.05 for face matching vs. control task and at a level corresponding to an uncorrected P value of 0.001 for face matching vs. location matching task) and changed the discriminating volume threshold from 5% to 60% of its maximum value while computing the following three different measurements:

- the total number of voxels that are above the threshold in common for both maps, i.e. the number of voxels in the overlaps;
- the percentage of voxels (above the threshold) from the SPMt results that lie within the above-threshold regions of the discriminating volume;
- the percentage of voxels (above the threshold) from the discriminating volume that lie inside the above-threshold regions of the SPMt.

Summary of the method

Our method can be summarized by the following algorithm steps:

Preprocessing steps

1. Slice time correction;
2. Motion correction;
3. Normalization to standard space (MNI/ICBM);
4. Spatial filter for data set 1 and no spatial filter for data set 2;
5. Removal of the base line and low frequency components of each voxels;
6. Use a mask to select voxels which contain brain tissue.

Dimensionality reduction phase

Project each scan of the training set $\mathbf{D}_{M \times N}$ onto the principal components to get $\mathbf{D}^{p}{}_{N \times N}$.

Learning phase

- Use the projected data $\mathbf{D}^{p}{}_{N \times N}$ to train the classifier. The classifier finds a hyperplane that separates the data. The separating hyperplane is described by a weight vector ($\boldsymbol{w}^{p}{}_{N \times 1}$) for each classifier;
- Map the weight vector ($\boldsymbol{w}^{p}{}_{N \times 1}$) back to the high-dimensional or voxel space ($\boldsymbol{w}_{M \times 1}$) by using the principal components information. It represents a volume with the discriminating regions, i.e. the discriminating volume.

Table 1
Sensitivity and specificity for classifying between face matching vs. control task (sensitivity = probability of correctly predicting face matching task; specificity = probability of correctly predicting control task) and between face matching vs. location matching (sensitivity = probability of correctly predicting face matching task; specificity = probability of correctly predicting location matching task)

Test	Data without spatial filter				Data with spatial filter			
	Sensitivity		Specificity		Sensitivity		Specificity	
	FLD	SVM	FLD	SVM	FLD	SVM	FLD	SVM
Face matching vs. control task	0.72	0.85	0.76	0.90	0.74	0.89	0.78	0.89
Face matching vs. location matching	0.38	0.75	0.69	0.70	0.67	0.81	0.79	0.76

Test phase

- Project each volume from the test subject ($\boldsymbol{v}_{M \times 1}$) onto the principal components;
- Use the projected test data ($\boldsymbol{v}^{p}{}_{N \times 1}$) as input to the classifier and compute the class assignment by using the hyperplane decision function. The output of the test phase is the prediction of a brain state.

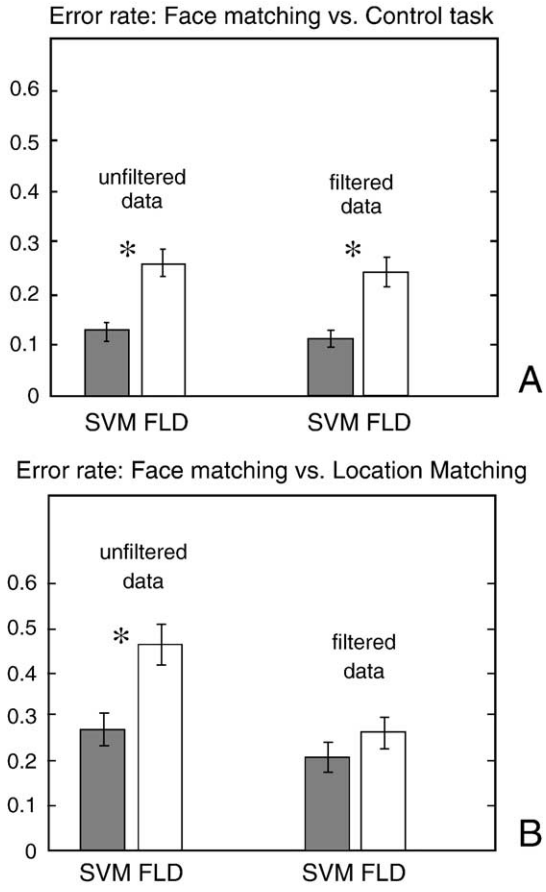

Fig. 4. Average error rates for both methods (SVM and FLD) for classifying between face matching and control task (A) and between face matching vs. location matching task (B). Left bars correspond to unfiltered data set 1 and right bars to filtered data set 2. Error bars indicate the standard error across 16 leave-one-out cross validation tests (for each test, the classifier was trained by using data of 15 subjects and tested with a new subject). For the columns marked by an asterisk, a paired t test indicated significant difference between the means.

Results

Task performance

The average response accuracy was 91.4% (7.6%) and 92.9% (10.7%) for the face and location matching tasks, respectively. The average response time for the face matching task was 1.49 (0.31) s and for the location matching task 1.30 (0.30) s. There was no statistically significant difference in accuracy rate and response time between both tasks.

Prediction of a brain state

We trained and tested the classifiers to distinguish between two brain states: (a) face matching (task 1) vs. control task (task 2) and (b) face matching (task 1) vs. location matching (task 2). The summary of the classifiers' performance is shown in Table 1 and in Fig. 4. The error rate is the mean between misclassifications of fMRI volumes from task 1 and from task 2. The error bars indicate the standard error across 16 leave-one-out tests (for each test, the classifier was trained by using data of 15 subjects and tested with a new subject). If the classifiers were guessing at random, the expected error rate would be 0.5. The results show that the SVM presents a good performance (error rates 0.1–0.27 across subjects) for all tests and for both training sets, without and with spatial filter. In addition, it clearly outperforms the FLD. To show this, we performed a paired t test on the error rates, and we found that the SVM error rates are significantly lower (P value < 0.05) than the FLD error rates for 3 out of 4 tests (columns marked by an asterisk in Fig. 4).

Discriminating volume

In the discriminating volume, the value of each voxel indicates the importance of such voxel in differentiating between two brain states. The value at each voxel is a function of the difference in activation in the voxel between the tasks.

Case 1: discriminating between face matching vs. control task

We present slices of the discriminating volume according to the FLD and the SVM using a training set without spatial filter (Figs. 5A and B) and using a training set filtered using an isotropic Gaussian filter (FWHM = 8 mm) (Figs. 5C and D). In Fig. 5, all voxels with value above 30% of the maximum absolute value of the discriminating volume are shown in color scale (blue scale for negative values, and red scale for positive values). By using this threshold on the positive values of the discriminating volumes (face matching > control task) for the filtered data, there were 4.7% of voxels above the threshold in discriminating volume for the SVM and 4.23% in the discriminating volume for the FLD.

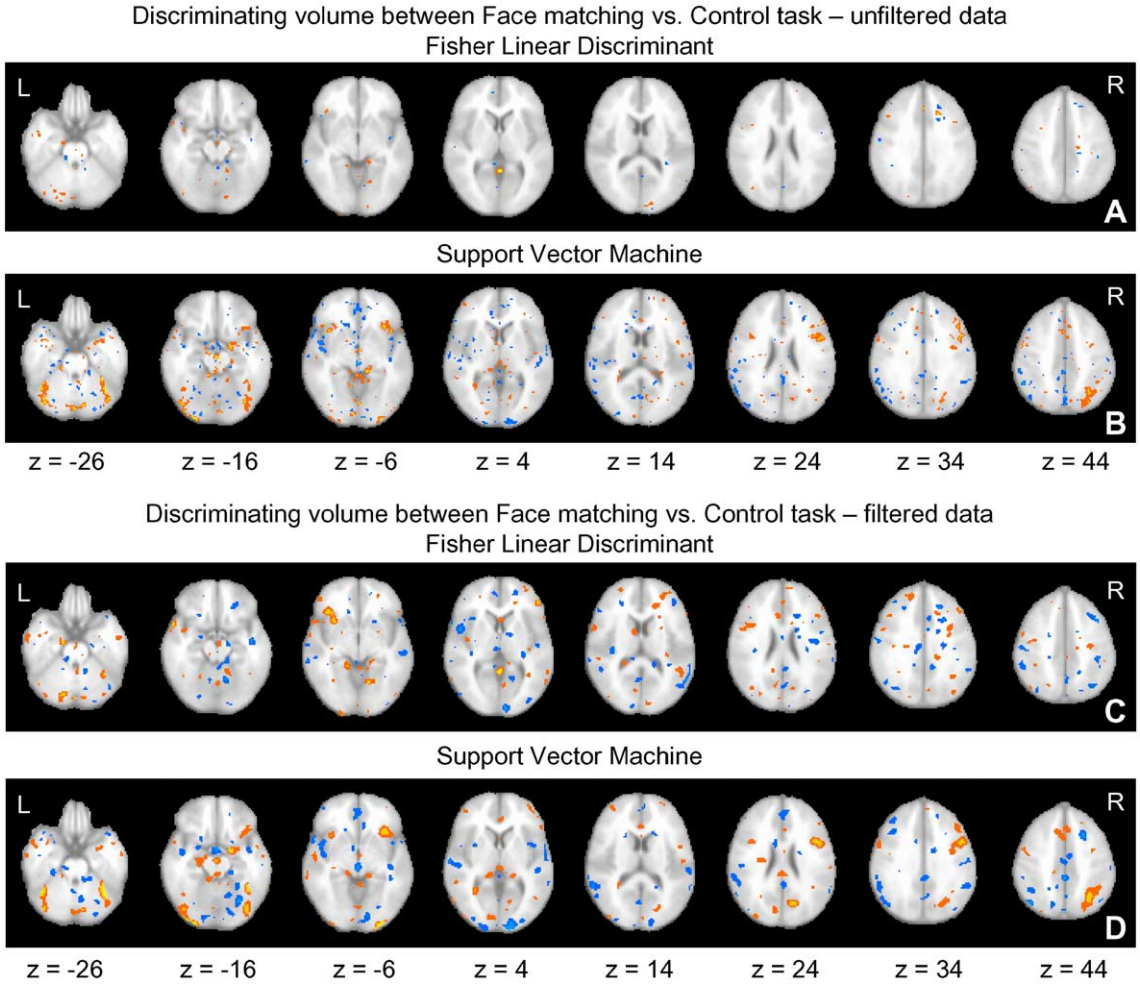

Fig. 5. Discriminating volumes for both methods (FLD and SVM) for classifying between face matching and control task, using the trainings set without and with spatial filter, respectively. All voxels with value above of 30% of the maximum value of the discriminating volume are shown in color scale (blue scale for negative values and red scale for positive values).

According to the SVM, the most discriminating regions, with higher positive values in the discriminating volume (representing higher activations during the face matching task in relation to the control task), were mainly within the ventral visual pathway (Fig. 5D). The most discriminating regions were bilaterally the cerebellum, the fusiform gyrus, the lingual gyrus and the inferior frontal gyrus, and in the right hemisphere, the inferior occipital gyrus, the middle frontal gyrus, the inferior and superior parietal lobe, the precuneus and the anterior cingulate gyrus.

According to the FLD, the most discriminating regions for the face matching task (Fig. 5C) were bilaterally the cerebellum and anterior cingulate, and in the right hemisphere, the precuneus, the lingual gyrus and the middle temporal gyrus and the left inferior frontal gyrus.

Case 2: discriminating between face matching vs. location matching task

The slices of the discriminating volume according to the FLD and the SVM are shown in Fig. 6 for a training set without spatial filter (Figs. 6A and B) and for a training set filtered using an isotropic Gaussian filter (FWHM = 8 mm) (Figs. 5D and 6C). In Fig. 6, all voxels with value above 30% of the maximum absolute value of the discriminating volume are shown in color scale (blue scale for negative values, and red scale for positive values). By using this threshold on the positive values of the discriminating volume (face matching > location matching) for the filtered data, there were 2.11% of voxels above the threshold in discriminating volume for the SVM and 6.47% in the discriminating volume for the FLD. By using this threshold on the negative values of the discriminating volume (location matching > face matching) for the filtered data, there were 2.33% of voxels above the threshold in discriminating volume for the SVM and 6.79% in the discriminating volume for the FLD.

According to the SVM (Fig. 6D), the most discriminating regions, with positive values in the discriminating volume (representing higher activations during the face matching task in relation to the location matching task), were bilaterally the cerebellum and the lingual gyrus, and in the right hemisphere, the fusiform gyrus, the middle occipital gyrus and the middle frontal gyrus. The most discriminating regions, with negative values in the discriminating volume, were bilaterally the lingual gyrus (more anterior than the region found for the positive values) and the cuneus, the left middle occipital gyrus and the right superior parietal lobe.

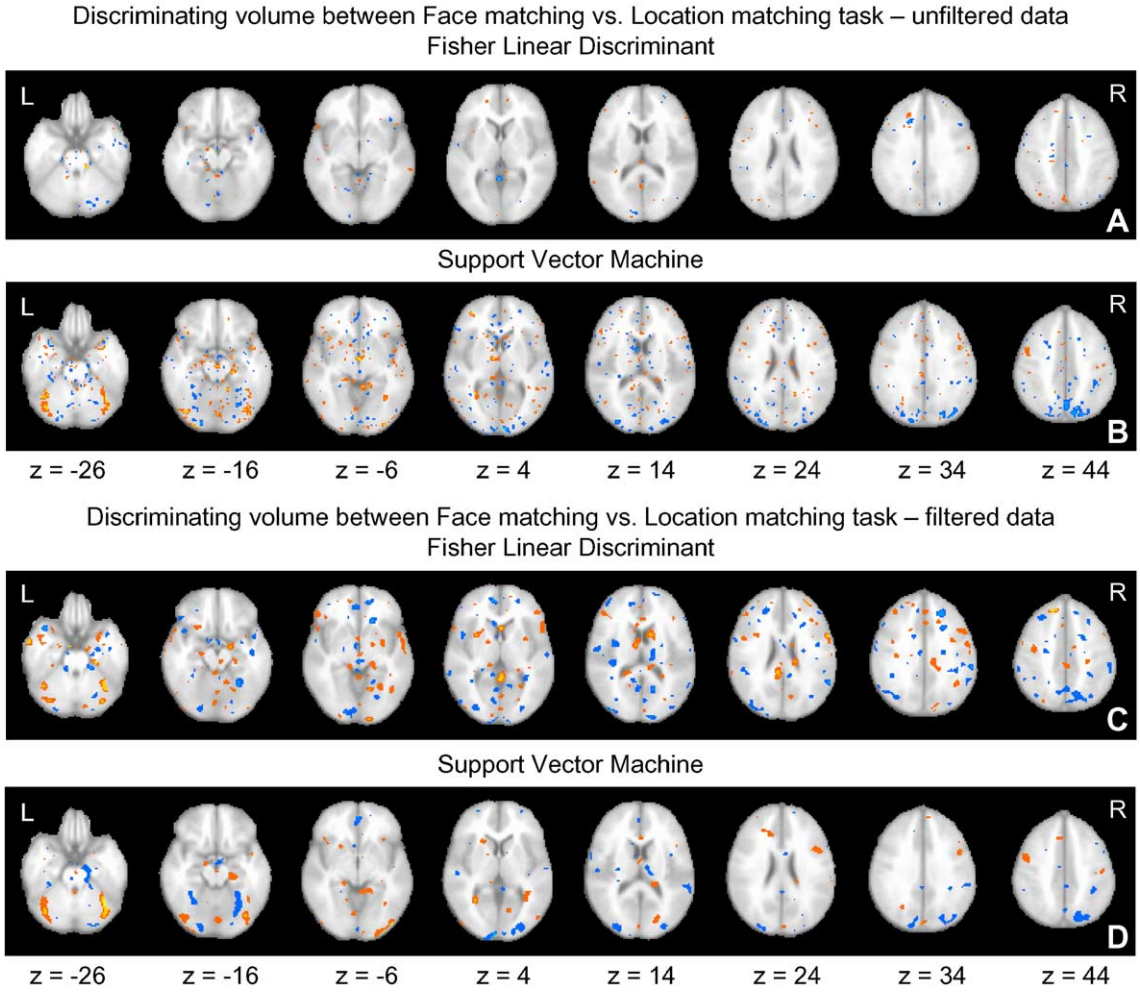

Fig. 6. Discriminating volumes for both methods (FLD and SVM) for classifying between face matching and location matching task, using the training sets without and with spatial filter, respectively. All voxels with value above of 30% of the maximum value of the discriminating volume are shown in color scale (blue scale for negative values and red scale for positive values).

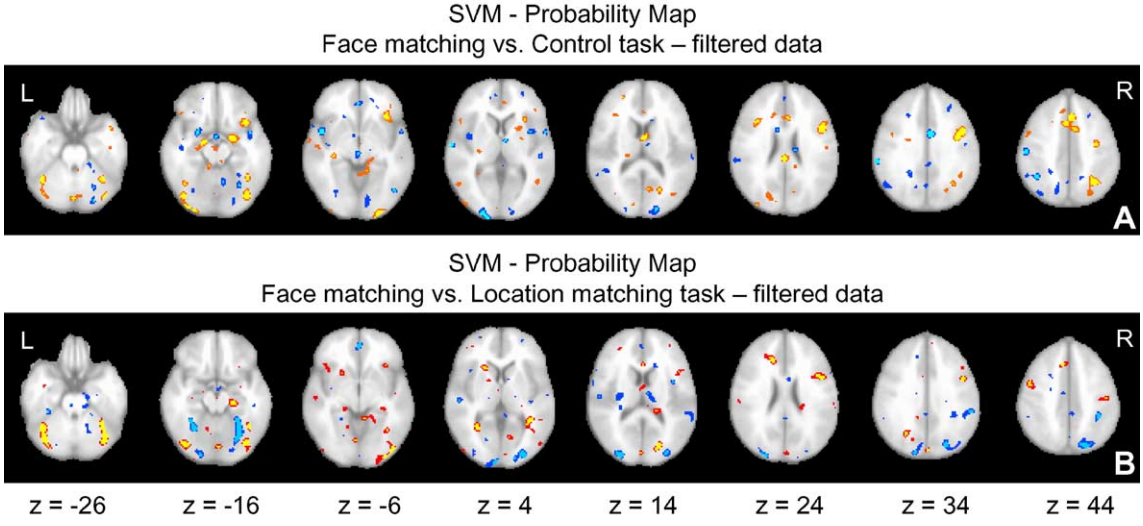

Fig. 7. Probability maps for the SVM discriminating volume under the face matching vs. control task (A) and under the face matching vs. location matching task (B). Voxels with positive values and P value < 0.05 are shown in red, and P value < 0.001 is shown in yellow. Voxels with negative values and P value < 0.05 are shown in dark blue, and P value < 0.001 is shown in light blue.

According to the FLD (Fig. 6C), the most discriminating regions with positive values in the discriminating volume were bilaterally the cerebellum, the right fusiform gyrus and the right middle frontal gyrus. The most discriminative regions with negative values were bilaterally the precuneus, the left middle occipital gyrus and the right superior parietal lobe.

By visual inspection of the discrimination volumes and the error rates, the SVM results are rendered more robust than the FLD results. In particular, comparing the discriminating volumes obtained from the training sets with and without spatial filter, it is possible to see that the spatial filter is a preprocessing step that affects the spatial maps (i.e. discriminating volumes) obtained for both methods. However, the results of the SVM seem less affected than the results of the FLD.

Probability map for SVM

We computed probability maps for the SVM results using the filtered data. Fig. 7A shows the probability map for face matching vs. control task and Fig. 7B for face matching vs. location matching task. Voxels with positive values and P value < 0.05 are shown in red, and P value < 0.001 is shown in yellow. Voxels with negative values and P value < 0.05 are shown in dark blue, and P value < 0.001 is shown in light blue. It is interesting to observe from visual inspection of Figs. 5D, 6D, 7A and B that voxels with value above 30% of the maximum absolute value of the discriminating volume seem to be very similar to the voxels with P value < 0.05 in the probability map.

Classifier and GLM

Figs. 8 and 9 display the overlap analysis of voxels above the threshold in the SPM*t* and in the discriminating volume, according to both methods and based on the training set with spatial filter. The SPM*t* threshold was fixed to a corresponding corrected P value of 0.05 for the face matching vs. control task and to a corresponding uncorrected P value of 0.001 for face

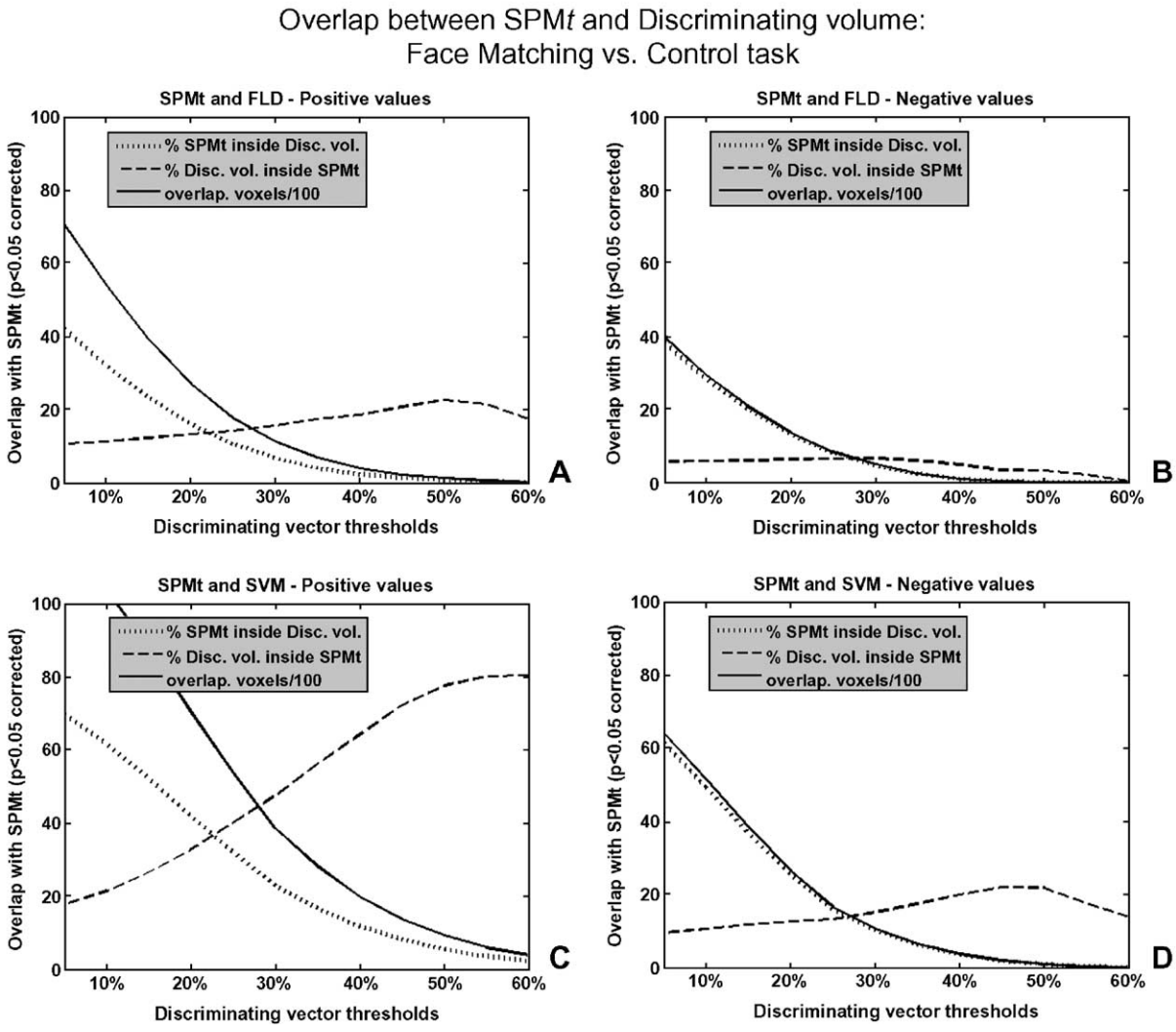

Fig. 8. Overlap analysis between SPM*t* results and discriminating volume results comparing face matching vs. control task. The SPM*t* threshold was fixed at a corrected P value of 0.05, the threshold of the discriminating volume was varied from 5% to 60% of its maximum value, and the total number of overlapping voxels was computed (solid line), the percentage of voxels from the SPM*t* results that lie within the weight vector (dotted line), and the percentage of voxels from the weight vector the lie inside the SPM*t* (dashed line). In panels A and C are shown the comparisons between the positive values of the SPM*t* and the discriminating volume for the FLD and the SVM, respectively. In panels B and D are shown the comparisons between the negative values of the SPM*t* and the discriminating volume for the FLD and the SVM, respectively.

matching vs. location matching. The threshold of the discriminating volume was increased from 5% to 60% of its maximum value, and we computed the total number of overlapping voxels (solid lines), the percentage of voxels from the SPM*t* results that lie within the above-threshold regions of the discriminating volume (dotted lines) and the percentage of voxels from the discriminating volume that lie inside the above-threshold regions of the SPM*t* (dashed lines).

Case 1: face matching vs. control task

The overlap of voxels with values above the threshold between the SPM*t* and the discriminating volume for the SVM and for the FLD for positive values is displayed in Figs. 8A and C, respectively, and for negative values in Figs. 8B and D, respectively. The positive values represent higher activations during face matching task, and the negative values represent higher activations during the control task or deactivation during the face matching task.

As the threshold for the positive values of the discriminating volume was increased, the overlap between the results from the SVM and the results from the GLM analysis increases to a peak of approximately 80% (dashed line in Fig. 8C). This demonstrates that most of the low-intensity voxels of the discriminating volume lie outside the SPM*t* map and that up to 80% high intensity voxels coincide with SPM*t*. In contrast, the highest intensity FLD voxels coincide only by up to 20% (dashed line in Fig. 8A).

For negative values, increasing the threshold of the discriminating volume results in a small increasing of the overlap between the results of the SVM and the results of the GLM (dashed line in Fig. 8D), reaching a maximum of 20%. Additionally, changes in the threshold of the discriminating volume almost did not change the overlap between results from the FLD and from the GLM, which remained around 5% (dashed line in Fig. 8B).

Case 2: face matching vs. location matching task

The overlap of voxels above the threshold between the SPM*t* and the discriminating volume for FLD and for the SVM for positive values is displayed in Figs. 9A and C, respectively, and for negative values in Figs. 9B and D, respectively. In this case, the positive values represent higher activations during the face matching task, and the negative values represent higher activations during the location matching task, both are attentional task with

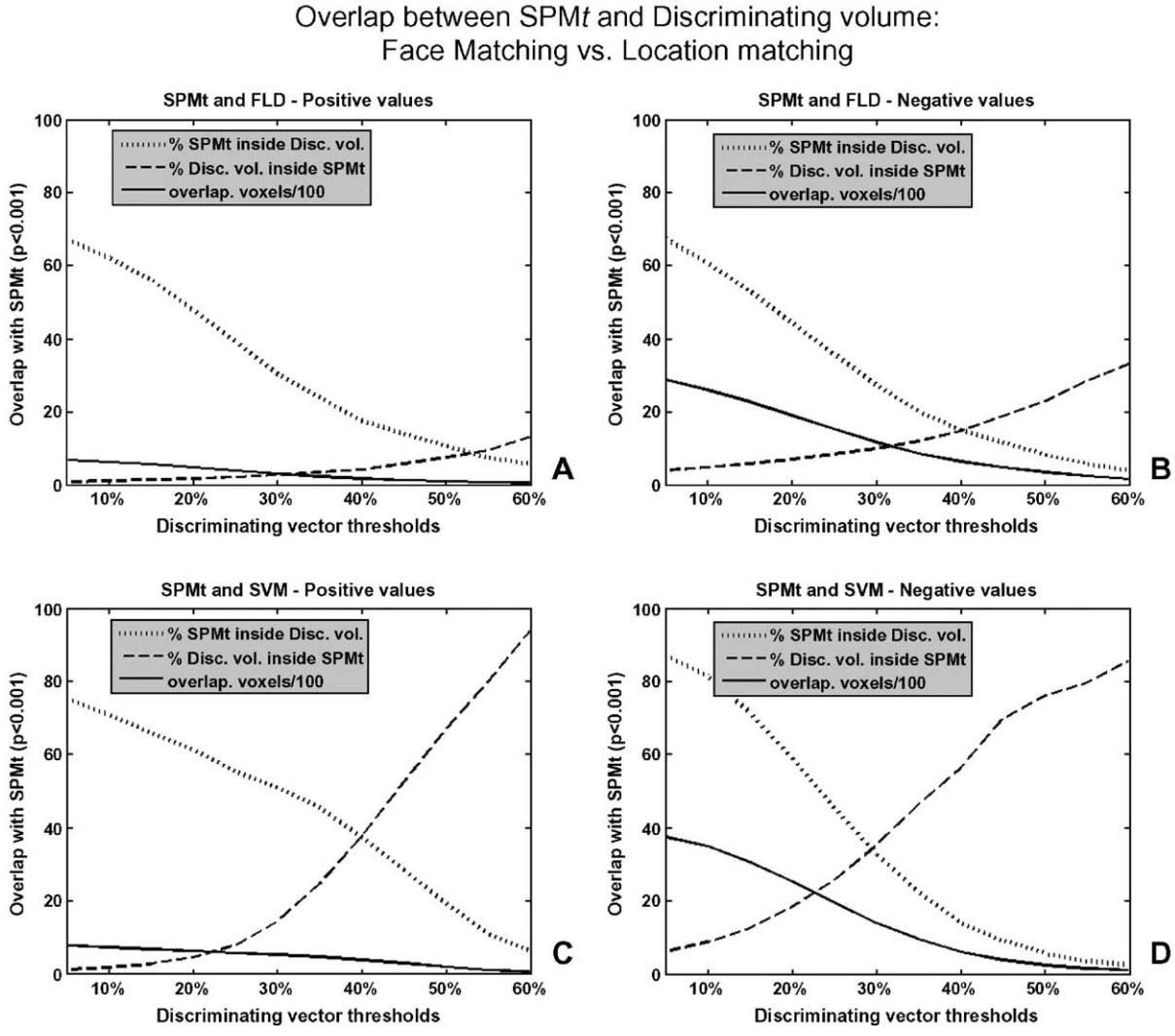

Fig. 9. Overlap analysis between SPM*t* results and discriminating volume results comparing face matching vs. location matching task (displayed as described in Fig. 8). In this case, the SPM*t* threshold was fixed in a corresponding *P* value of 0.001 (uncorrected).

similar cognitive demand. As the threshold of the positive and the negative values of the discriminating volume was increased, the overlap between the results from the SVM and the results from the GLM analysis increases for both positive and negative values, reaching more than 80% (dashed line in Figs. 9C and D).

In contrast, the overlap between results from the FLD and from the GLM reaches a maximum of approximately 12% for positive values and 32% for negative values (dashed line in Figs. 9A and B).

In summary, these results demonstrate that the voxels belonging to the strongest discriminating regions according to the SVM tend to be closely related to those with highest SPM*t* significance, which is not the case for FLD. This again indicates considerable robustness of the SVM compared to FLD.

Overlap between the SVM discriminating volume and SPMt maps

We present overlap maps between the SVM and the SPM*t* results using the filtered data to allow an evaluation about the location of the similarities and the differences between their results. The overlap map between the SPM*t* and the discriminating volume for the comparison between face matching vs. control task are displayed in Figs. 10A (positive values) and B (negative values) and for the comparison between face matching vs. location matching task are displayed in Figs. 10C (positive values) and D (negative values). In the overlap maps, the voxels that were significant in the SPM*t* (corrected P value < 0.05 for face matching vs. control task and uncorrected P value < 0.001 for face matching vs. location matching task) and above the threshold in the discriminating volume (P value < 0.05) were colored yellow; the voxels that were significant only in the SPM*t* map were colored blue, and the voxels above the threshold only in the discriminating volume were colored red.

Discussion

The present study demonstrates greater accuracy of the Support Vector Machine algorithm compared to the Fisher Linear Discriminant in predicting the instantaneous brain states as well as in robustness of the spatial maps. The performance differences

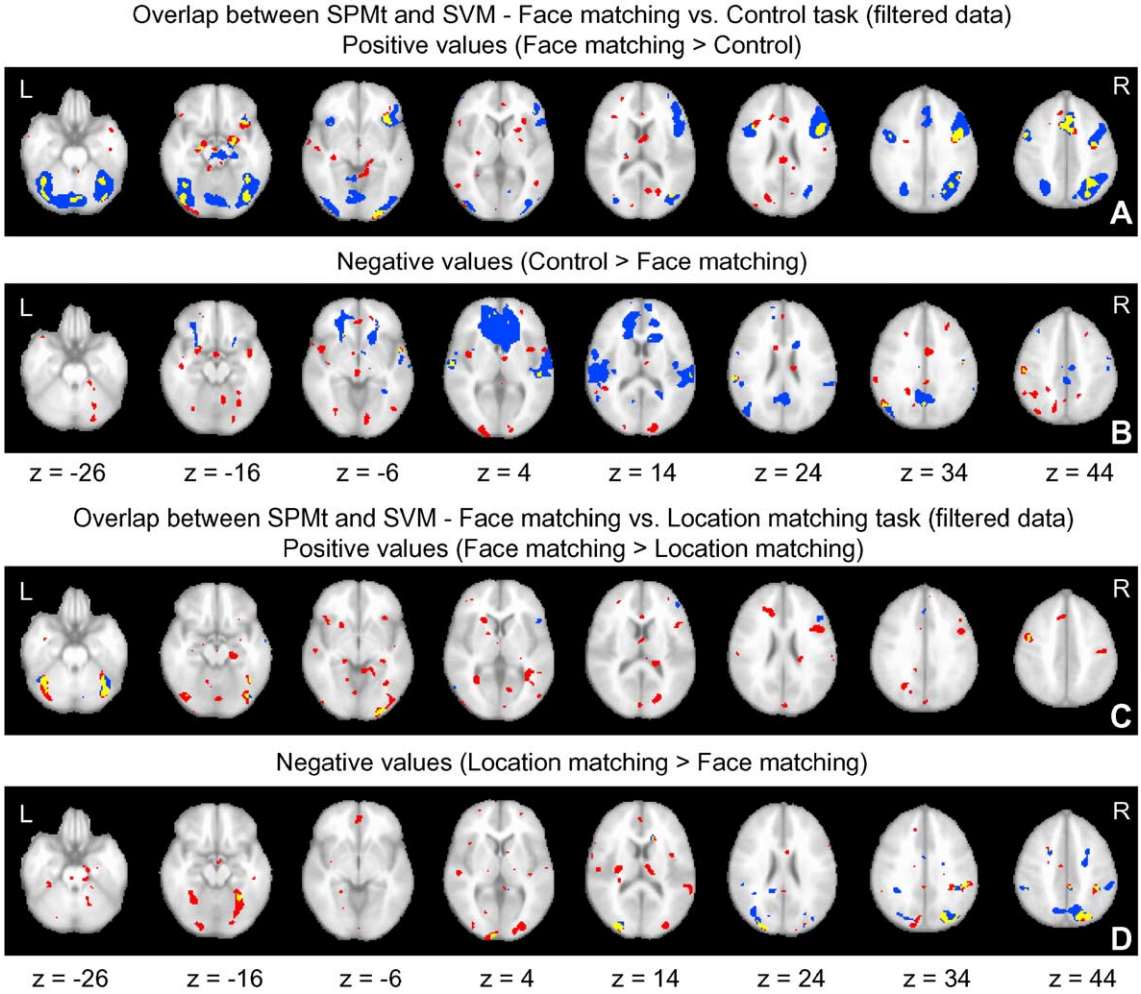

Fig. 10. In panels A and B are shown the overlap of positive values and negative values between the SPM*t* (corrected P value < 0.05) and the SVM discriminating volume (P value < 0.05) for the comparison between face matching and control task respectively, displayed on an average group structural scan in axial slices at different z levels. In blue are shown the voxels above the threshold within the SPM*t* results, in red are shown voxels above the threshold within the SVM results and the overlapping voxels are in yellow. In panels C and D are shown the overlap of positive values and negative values between the SPM*t* (P value < 0.001) and the SVM discriminating volume (P value < 0.05) for the comparison between face matching and location matching task respectively (displayed as described for panels A and B).

between both methods in distinguishing between instantaneous brain states were similar with and without spatial filtering. The most discriminating regions between face matching and location matching task defined by the classifiers were along the expected ventral and dorsal visual pathways. The SVM discrimination maps had greater overlap with the GLM analysis compared to the FLD methods, in particular, with the data sets where no spatial filtering was included.

Prediction accuracy in neuroimaging has been previously addressed utilizing two different approaches. The first approach was to apply the classifier after feature selection method based on prior hypotheses (e.g., Mitchell et al., 2004; Wang et al., 2003; Cox and Savoy, 2003; Ford et al., 2003), and the second approach consisted of use PCA/SVD analysis as dimensionality reduction method and apply the classifier on PCA basis without prior selection of spatial features (e.g., Mørch et al., 1997; Kjems et al., 2002; LaConte et al., 2003; Carlson et al., 2003; Strother et al., 2004). However, none of these studies utilized SVM on raw high-dimensional fMRI volumes without prior selection of spatial features with the objective to predict brain states and find discriminating regions between two brain states.

Some studies demonstrated the feasibility of training classifiers to distinguish a variety of predefined cognitive states, based on features selected from fMRI data. For example, Mitchell et al. (2004) used machine learning algorithm to detect instantaneous cognitive states, and Wang et al. (2003) used machine learning algorithm to detect cognitive state across multiple subjects. Cox and Savoy (2003) trained classifiers to detect and classify distributed patterns of fMRI activity in human visual cortex. In these studies, a feature selection method was used to reduce the dimensionality of the data before training the classifier. Features were selected by averaging the fMRI activations over several chosen voxels or by selecting a subset of the available voxels and times. This approach presents two limitations: first, the machine learning algorithm is not using the whole fMRI spatial information but only selected information; second, the selected information depends on the user, usually based on some a priori hypothesis. Another approach used fMRI brain activation maps generated by GLM to train classifiers to differentiate patients from controls for Alzheimer's disease, schizophrenia, and mild traumatic brain injury (Ford et al., 2003). In this case, only the information previously selected by the univariate GLM approach was used to train the classifier.

In contrast, other studies applied classifiers on PCA/SVD basis without prior selection of spatial features (e.g., Mørch et al., 1997; Kjems et al., 2002; LaConte et al., 2003; Carlson et al., 2003; Strother et al., 2004). One advantage of this approach is that the discriminating regions are an output of the method and not an input, that is, the algorithm finds the discriminating regions, and this information can be presented as spatial maps. However, most of these studies used the prediction metric to measure data-analytic performance in functional neuroimaging (Mørch et al., 1997; Kjems et al., 2002; LaConte et al., 2003; Strother et al., 2004). For example, Kjems et al. (2002) focused on evaluation of the performance of models for neuroimaging data analysis. They presented learning curves as an unbiased means for evaluating the performance of models for neuroimaging data analysis. They demonstrated how prediction error can be expressed as mutual information between scan and the scan label, measured in units of bits and used the mutual information learning curve to evaluate the impact of different methodological choices (e.g., classification label schemes, preprocessing choices). LaConte et al. (2003) used a two-class Canonical Variates Analysis on PCA basis to evaluate how accuracy vs. reproducibility depends on preprocessing choices (alignment, temporal detrending and spatial smoothing). Strother et al. (2004) proposed an approach based on prediction and reproducibility metrics to optimize the preprocessing of fMRI data and used a multivariate linear discriminant analysis (CVA) to detect large-scale brain network. In addition, Carlson et al. (2003) applied also FLD on PCA basis to investigate patterns of activity in the categorical representation of objects. These studies have mainly used FLD or CVA, which is the multivariate extension of FLD, on PCA basis. The FLD is a traditional form of statistical classification analysis. In the FLD classifier, the weight vector or discriminating volume, which defines the separating hyperplane, is determined by the difference between the center of mass for both classes corrected for within-class covariance. One disadvantage of this approach is the high susceptibility of the mean and the covariance estimates to contamination by outliers. In relation to these works, our novel contribution is to take advantage of the good scaling and the large margin properties of SVM for the fMRI brain state classification problem.

The brain state classification from fMRI data volumes corresponds to the classification of few points (scans) in a high-dimensional space (dimension = number of voxels). In this situation, there are many linear classifiers (i.e. hyperplanes) that separate the training data (Schölkopf and Smola, 2002), which heavily overfit and generalize badly. The SVM algorithm can solve this problem (Boser et al., 1992). It finds the optimal hyperplane, i.e. the separating hyperplane that generalizes better. This property makes the linear SVM an optimal tool to address the problem of finding a common brain network between subjects and use this information to classify data from a new subject. For all tests, the training error for the SVM (i.e. the error rate for classifying the training set) was zero, this means that the training data were linearly separable and the SVM algorithm found the optimal separating hyperplane. This reflects the fact that extensions of the SVM as nonlinear kernels or soft-margin SVM with slack variables are unnecessary here and would be counterproductive.

Our results demonstrated that SVM clearly outperforms FLD in classification performance as well as in robustness of the discriminating volumes obtained. This not only demonstrates the clear benefit of use of SVM but at the same time challenges the results of earlier approaches based on FLD. However, these are issues to be covered in future work.

The most discriminating regions in the SVM map were determined by applying a threshold to the SVM map. The threshold was calculated using a nonparametric permutation test. We tested the null hypothesis that there were no differences between the brain states during the two different tasks. If this is true, then the labeling has no contribution for the classification, and, if the labels are randomly permuted, we would observe the same result. To test this hypothesis, we permuted the class labels 2000 times and trained the SVM using the permuted labeling, and, for each voxel, we estimated the probability distribution. By testing the null hypothesis at each voxel, we obtained a map of probabilities. The P value or probability of obtaining a value in a voxel of the discriminating volume by chance equal or greater than the one obtained with the original labeling is the proportion of values in the permutation distribution greater or equal to the value obtained by using the original (i.e. nonpermutated) training data. By comparing the SVM discriminating volumes (Figs. 5D and 6D)

and the probability maps (Figs. 7A and B), it is possible to observe that voxels with value above 30% of the maximum absolute value of the discriminating volume are essentially in the same regions of voxels with P value < 0.05 in the probability maps.

In addition, the results of the overlap analysis showing that voxels with higher values in the discriminating volumes for the SVM can reach greater than 80% overlap with voxels in the above-threshold regions of the SPMt (dashed line in Figs. 8C, 9C and D. By observing the overlap map between the results of the SVM and SPMt for the comparison face matching vs. control (Figs. 10A and B) and face matching vs. location matching task (Figs. 10C and D), it is possible to observe that the overlap voxels are in areas of likely physiological plausibility of the activations according to the tasks, however, there are a few non-overlapping clusters. A visual inspection of the SVM and SPMt maps suggests that SVM might be able to find a globally consistent pattern of restricted clusters by a global multivariate analysis. On the other hand, visual inspection of the results using both discrimination methods (Figs. 5 and 6) suggests little overlap between FLD and SVM results. The distribution in the brain of the small clusters from the FLD suggests that the clusters are located in part along plausible regions of the brain and in part in regions where one would not expect activation. Given the superior performance of the SVM method, we calculated the overlap between discrimination volumes and SPMt only for the results from the SVM method.

The results obtained by using training sets without spatial filter showed that even in this extreme case the error rate was below chance in all cases for the SVM (classifying between face matching vs. control task and between face matching vs. location matching). The FLD showed an error rate below chance level for the comparison between face matching vs. control task but not for the comparison between face matching vs. location matching. Although the classifiers can predict above chance level the brain state using training sets without spatial filter, according to visual inspection, the discriminating volumes obtained in these cases (Figs. 5A, B, 6A and B) are strongly contaminated by a noise patterns, especially for the FLD. This represents evidence that the algorithms are sensitive to the low signal to noise ratio typical of fMRI data that is not spatially filtered. Strother et al. (2004) observed a trade-off between reproducibility and prediction as function of linear discriminant parameterization (i.e. number of PCs passed to the CVA), and they suggested that this uncoupling should be addressed by comparing different type of models, e.g., linear discriminant, Support Vector Machine. Our results suggest that the SVM can at least in part achieve this uncoupling between reproducibility and prediction performance.

We tested the performance of two different methods, the SVM and the FLD, for their ability to distinguish between the following brains states: face matching vs. control task and face matching vs. location matching task, using training sets with and without spatial filters. The error rates obtained and the standard errors are shown in Fig. 4. In all cases, the SVM outperforms FLD, and a paired t test showed that in 3 out of 4 tests the error rates for both methods were statistically different.

The regions selected as being the ones best able to differentiate between tasks were located in the expected areas. When comparing face matching to control task, the discriminating regions, according to SVM (data with spatial filter), were in the ventral visual pathway (early visual areas and fusiform gyrus), right dorsolateral prefrontal area and right parietal lobe, an area associated with attentional control. In the FLD discriminating volume, using the data with spatial filter, there were clusters in the ventral visual pathway, but they were smaller and additionally there were many small clusters spread throughout the whole brain. In the other comparison, face matching versus location matching, the most discriminative regions, according to both methods (SVM and FLD) and using data with spatial filter, with the positive values (face matching > location matching) in the discriminating volume were along the ventral visual pathway. The most discriminative regions with negative values (location matching > face matching) in the discriminating volume were located along the dorsal visual pathway. In addition, the results from the FLD method also led to many small clusters spread throughout the whole brain. The discriminative regions were consistent with previous studies showing selective activation of the ventral visual pathway when attending to faces and selective activation of the dorsal visual pathway when attending to spatial information (Corbetta et al., 1991a,b; Haxby et al., 1991, 1994).

When comparing the overlap between the results of the SVM and the results of the GLM analysis, we observed that, as the threshold of the discriminating volume was increased, the curves showing the percentage of voxels from the discriminating volume lying inside the SPMt results present an interesting behavior: it increases, reaching a maximum of around 80% of overlap in 3 out 4 comparisons: positive values face matching vs. control task (higher activations for face matching task) (Fig. 8C), positive values for the face matching vs. location matching (higher activations for face matching task) (Fig. 9C) and negative values for face matching vs. location matching (higher activations for location matching task) (Fig. 9D). The only comparison where the overlap reached a maximum of only 20% was the negative values for face matching vs. control matching (higher activation during control task or deactivation during face matching task) (Fig. 8D).

The positive values of the SPMt reflect the positive BOLD response (PBR) and represent stimulus-related activation. The negative values of the SPMt reflect the negative BOLD response (NBR). However, both of them only indirectly reflect the neuronal activity (Almeida and Stetter, 2002). Many studies have investigated the functional meaning of the NBR (Shmuel et al., 2002; Smith et al., 2004). The findings of Shmuel et al. (2002) support the contribution to NBR of a significant component of reduction in neuronal activity and possible a component of hemodynamic changes independent of local changes in the neuronal activity. Huettel et al. (2001) observed deactivation in parietal and frontal regions during a visual attention task, they suggested that these regions may underlie baseline (non-task related) semantic processing that is interrupted when the subject begins the perceptual task. Raichle et al. (2001) suggested a network supporting a default mode of brain function. This network would include a set of regions that consistently show greater activity during the rest states than during cognitive tasks, such as posterior cingulate cortex and ventral anterior cingulate cortex (Greicius et al., 2003).

Another possible explanation for the mismatching in our results between regions with negative SPMt and regions with negative values in the discriminating volume for the SVM (in this case, the negative values represent higher activation during the control task) can result from the fact that the GLM is a univariate approach and tries to fit a linear model to the voxel time series, in contrast, the SVM approach is a multivariate method and treats each fMRI volume as a spatial pattern. The two approaches deal differently with the nonlinear components

of the BOLD signal, and the nonlinear components might be higher in the NBR.

To summarize, the novelty of the present work was to demonstrate that the Support Vector Machine algorithm outperforms the Fisher Linear Discriminant: (1) in classifying instantaneous brain states of a new subjects based on multisubject training sets; (2) in finding cross-subjects regularities, i.e. the most discriminating regions, between the brain states. Our approach leads to two outputs: the classification of a brain state and the discriminating regions upon which the classification is based on.

Acknowledgments

The authors would like to thank Drs. T. Meindl and G. Leinsinger for their assistance in acquiring the data. J.M.M. and M.S. gratefully acknowledge support through the German Ministry for Science (BMBF) grant 01IBC01A. A.L.W.B. and H.H. gratefully acknowledge support through the Volkswagen Foundation (Hanover, Germany) and by a grant from the German Competency Network on Dementias (Kompetenznetz Demenzen) funded by the German Ministry for Science (BMBF).

Appendix A. Singular value decomposition (SVD) and principal components analysis (PCA)

Let the fMRI volumes in the original data set to be $\boldsymbol{v}_1$, $\boldsymbol{v}_2$,..., $\boldsymbol{v}_N$. Each volume $\boldsymbol{v}_i$ ($i = 1,\ldots,N$) is treated as a vector which contains M dimensions (voxels). We can define a data matrix $M \times N$ with one volume per column and one voxel per row as

$$\mathbf{D} = [\boldsymbol{v}_1 \ldots \boldsymbol{v}_i \ldots \boldsymbol{v}_N]. \tag{A.1}$$

Let $\mathbf{D}_c$ be $\mathbf{D}$ with the average volume of the data set subtracted from each column. The SVD of $\mathbf{D}_c$ is:

$$\mathbf{D}_c = \mathbf{E}\mathbf{L}^{1/2}\mathbf{e}^T \tag{A.2}$$

where $\mathbf{E}$ and $\mathbf{e}^T$ are the matrices of eigenvectors of $\mathbf{D}_c\mathbf{D}_c^T$ and $\mathbf{D}_c^T\mathbf{D}_c$ respectively. $\mathbf{L}$ is a diagonal matrix, and its nonzero elements are the eigenvalues of both $\mathbf{D}_c\mathbf{D}_c^T$ and $\mathbf{D}_c^T\mathbf{D}_c$. The PCs of $\mathbf{D}_c$ can be determined by

$$\mathbf{E} = \mathbf{D}_c\mathbf{e}\mathbf{L}^{-1/2}. \tag{A.3}$$

The projection of the volumes onto the principal components was carried out as

$$\mathbf{D}^p = \mathbf{E}^T\mathbf{D} \tag{A.4}$$

$$\mathbf{D}^p = \left[\boldsymbol{v}_1^p \ldots \boldsymbol{v}_i^p \ldots \boldsymbol{v}_N^p\right], \tag{A.5}$$

where $\mathbf{E}$ is a matrix $M \times N$ containing one eigenvector or PC per column, and $\boldsymbol{v}_i^p$ is the projection of the volume $\boldsymbol{v}_i$ onto the principal components.

Appendix B. Support Vector Machine (SVM)

The SVM algorithm is described in detail in Schölkopf and Smola (2002) and will be briefly summarized here. It has been shown that the optimal hyperplane is defined as the one with the maximal margin of separation between the two classes (Fig. 4B). There is a weight vector $\boldsymbol{w}^p$ and an offset b such that

$$y_i\left((\boldsymbol{w}^p)^T\boldsymbol{v}_i^p + b\right) > 0 \tag{B.1}$$

where y_i is the class label (+1 for the class 1 and −1 for the class 2), and $\boldsymbol{v}_i^p$ are the training examples (projected volumes onto the principal components).

Rescaling $\boldsymbol{w}^p$ and b such that the point(s) closest to the hyperplane satisfy

$$|(\boldsymbol{w}^p)\boldsymbol{v}_i^p + b| = 1 \tag{B.2}$$

one obtains the canonical form of the hyperplane, given by

$$y_i\left((\boldsymbol{w}^p)^T\boldsymbol{v}_i^p + b\right) \geq 1 \tag{B.3}$$

The margin, the distance to a separating hyperplane from the point closer to it, measured perpendicularly to the hyperplane is $1/\|\boldsymbol{w}^p\|^2$. To maximize the margin, one has to minimize $\|\boldsymbol{w}^p\|$ subject to Eq. (B.3). The solution $\boldsymbol{w}^p$ is constructed by solving a constrained quadratic optimization problem, and it has an expansion in terms of a subset of training examples that lie on the margin (support vectors), given by

$$\boldsymbol{w}^p = \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{v}_i^p \tag{B.4}$$

The training examples $\boldsymbol{v}_i^p$ with coefficients α_i nonzero, called support vectors, carry all information relevant about the classification problem.

The class label of a test example $\boldsymbol{v}^p$ is computed by the hyperplane decision function, given by

$$f(\boldsymbol{v}^p) = \mathrm{sgn}\left(\sum_{i=1}^{N} y_i\alpha_i\left((\boldsymbol{v}^p)^T\boldsymbol{v}_i^p\right) + b\right) \tag{B.5}$$

and the offset b is computed from

$$\alpha_i\left[y_i\left((\boldsymbol{v}_i^p)^T\boldsymbol{w}^p + b\right) - 1\right] = 0 \tag{B.6}$$

References

Almeida, R., Stetter, M., 2002. Modeling the link between functional imaging and neuronal activity: synaptic metabolic demand and spike rates. NeuroImage 17, 1065–1079.

Bell, A.J., Sejnowski, T.J., 1995. An information maximisation approach to blind separation and blind deconvolution. Neural Comput. 7 (6), 1129–1159.

Blanz, V., Vetter, T., 1999. A morphable model for the synthesis of 3D faces. 26th International Conference on Computer Graphics and Interactive Techniques. ACM, Los Angeles, CA, USA.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. D. Proc. Fifth Ann. Workshop on Computational Learning Theory, ACM, pp. 144–152.

Bullmore, E., Brammer, M., Williams, S.C.R., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., Sham, P., 1996. Statistical methods of estimation and inference for functional MR image analysis. Magn. Reson. Med. 35, 261–277.

Carlson, T.A., Schrater, P., He, S., 2003. Patterns of activity in the categorical representations of objects. J. Cogn. Neurosci. 15 (5), 704–717.

Corbetta, M., Miezin, F.M., Dobmeyer, S.M., Shulman, G.L., Petersen, S.E., 1991a. Selective and divided attention during visual discrim-

inations of shape, color, and speed: functional anatomy by positron emission tomography. J. Neurosci. 11 (8), 2383–2402.
Corbetta, M., Miezin, F.M., Shulman, G.L., Petersen, S.E., 1991b. Selective attention modulates extrastriate visual regions in humans during visual feature discrimination and recognition. Exploring brain functional anatomy with positron emission tomography, Ciba Found. Symp, vol. 163, pp. 165–180.
Cox, D.D., Savoy, 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.
Ford, J., Farid, H., Makedon, F., Flashman, L.A., McAllister, T.W., Megalooikonomou, V., Saykin, A.J., 2003. Patient classification from fMRI activation maps. 6th Annual International Conference on Medical Image Computing and Computer Assisted Intervention.
Friston, K.J., Büchel, C., 1997. Functional connectivity: eigenimages and multivariate analysis. In: Ashburner, J., Friston, K., Penny W. (Eds.), Human Brain Function. http://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf2/pdfs/Ch19.pdf.
Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.
Friston, K.J., Frith, C.D., Frackowiak, R.S., Turner, R., 1995. Characterizing dynamic brain responses with fMRI: a multivariate approach. NeuroImage 2 (2), 166–172.
Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc. Natl. Acad. Sci. 100 (1), 253–258.
Haxby, J.V., Grady, C.L., Horwitz, B., Ungerleider, L.G., Mishkin, M., Carson, R.E., Herscovitch, P., Schapiro, M.B., Rapoport, S.I., 1991. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. Proc. Natl. Acad. Sci. 88, 1621–1625.
Haxby, J.V., Horwitz, B., Ungerleider, L.G., Maisog, J.Ma., Pietrini, P., Grady, C.L., 1994. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J. Neurosci. 14, 6336–6353.
Holmes, A.P., Friston, K.J., 1997. Statistical Models and Experiment Design. http://www.fil.ion.ucl.ac.uk/spm/course/notes.html.
Holmes, A.P., Blair, R.C., Watson, J.D.G., Ford, I., 1996. Nonparametric analysis of statistic images from functional mapping experiments. J. Cereb. Blood Flow Metab. 16 (1), 7–22.
Huettel, S.A., Güzeldere, G., McCarthy, G., 2001. Dissociating the neural mechanisms of visual attention in change detection using functional MRI. J. Cogn. Neurosci. 13 (7), 1006–1018.
Jackson, J.E., 1991. A User's Guide to Principal Components. JohnWiley and Sons, Inc, New York.
Kherif, F., Poline, J.B., Flandin, G., Benali, H., Simon, O., Dehaene, S., Worsley, K., 2002. Multivariate model specification for fMRI data. NeuroImage 16, 1068–1083.
Kjems, U., Hansen, L.K., Anderson, J., Frutiger, S., Muley, S., Sidtis, J., Rottenberg, D., Strother, S.C., 2002. The quantitative evaluation of functional neuroimaging experiments: mutual information learning curves. NeuroImage 15 (4), 772–786.
LaConte, S., Anderson, J., Muley, S., Ashe, J., Frutiger, S., Rehm, K., Hansen, L.K., Yacoub, E., Hu, X., Rottenberg, D., Strother, S., 2003. The evaluation of preprocessing choices in single-subject BOLD fMRI using NPAIRS performance metrics. NeuroImage 18 (1), 10–27.
McIntosh, A.R., Bookstein, F.L., Haxby, J.V., Grady, C.L., 1996. Spatial pattern analysis of functional brain images using partial least squares. NeuroImage 3, 143–157.
McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. Hum. Brain Mapp. 6, 160–188.
Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M., Newman, S., 2004. Learning to decode cognitive states from brain images. Mach. Learn. 57, 145–175.
Mørch, N., Hansen, L.K., Strother, S.C., Svarer, C., Rottenberg, D.A., Lautrup, B., Savoy, R., Paulson, O.B., 1997. Nonlinear versus linear models in functional neuroimaging: learning curves and generalization crossover. Proceedings of the 15th International Conference on Information Processing in Medical Imaging.
Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15 (1), 1–25.
Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. 98 (2), 676–682.
Schölkopf, B., Smola, A., 2002. Learning with Kernels. MIT Press.
Shmuel, A., Yacoub, E., Pfeuffer, J., Van de Moortele, P.F., Adriany, G., Hu, X., Ugurbil, K., 2002. Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. Neuron 36, 1195–1210.
Smith, A.T., Williams, A.L., Singh, K.D., 2004. Negative BOLD in the visual cortex: evidence against blood stealing. Hum. Brain Mapp. 21, 213–220.
Strother, S., La Conte, S., Kai Hansen, L., Anderson, J., Zhang, J., Pulapura, S., Rottenberg, D., 2004. Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. NeuroImage 23 (Suppl. 1), S196–S207.
Tegeler, C., Strother, S.C., Anderson, J.R., Kim, S.G., 1999. Reproducibility of BOLD-based functional MRI obtained at 4 T. Hum. Brain Mapp. 7 (4), 267–283.
Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.
Wang, X., Hutchinson, R., Mitchell, T.M., 2003. Training fMRI classifiers to discriminate cognitive states across multiple subjects. The 17th Annual Conference on Neural Information Processing Systems.
Weaver, J.B., 1995. Efficient calculation of the principal components of imaging data. J. Cereb Blood Flow Metab. 15, 892–894.