Published in final edited form as:

Magn Reson Imaging. 2013 February ; 31(2): 247–261. doi:10.1016/j.mri.2012.07.010.

Spatially Aggregated Multi-Class Pattern Classification in Functional MRI using Optimally Selected Functional Brain Areas

Weili Zheng^{a,*}, Elena S. Ackley^a, Manel Martínez-Ramón^b, and Stefan Posse^{a,c,d}

^aDept. of Neurology, University of New Mexico, School of Medicine, Albuquerque, NM, USA

^bDept. of Signal Theory and Communications, Universidad Carlos III de Madrid, Leganés, Madrid, Spain

^cDept. of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA

^dDept. of Physics and Astronomy, University of New Mexico, Albuquerque, NM, USA

Abstract

In previous works, boosting aggregation of classifier outputs from discrete brain areas has been demonstrated to reduce dimensionality, and improve the robustness and accuracy of fMRI classification. However, dimensionality reduction and classification of mixed activation patterns of multiple classes remain challenging. In the present study, the goals were (a) to reduce dimensionality by combining feature reduction at the voxel level and backward elimination of optimally aggregated classifiers at the region level, (b) to compare region selection for spatially aggregated classification using boosting and partial least squares regression methods and (c) to resolve mixed activation patterns using probabilistic prediction of individual tasks. Brain activation maps from interleaved visual, motor, auditory and cognitive tasks were segmented into 144 functional regions. Feature selection reduced the number of feature voxels by more than 50%, leaving 95 regions. The two aggregation approaches further reduced the number of regions to 30, resulting in more than 75% reduction of classification time and misclassification rates of less than 3%. Boosting and partial least squares (PLS) were compared to select the most discriminative and the most task correlated regions, respectively. Successful task prediction in mixed activation patterns was feasible within the first block of task activation in real time fMRI experiments. This methodology is suitable for sparsifying activation patterns in real-time fMRI and for neurofeedback from distributed networks of brain activation.

Keywords

Pattern classification; fMRI; boosting; partial least squares; real time

1. Introduction

Pattern classification in functional Magnetic Resonance Imaging (fMRI), often based on support vector machine (SVM) [1–5], has led to a number of important discoveries in

© 2012 Elsevier Inc. All rights reserved.

*Corresponding author: Weili Zheng, PhD, HUH-MR Research/Radiology, Wayne State University, 3990 John R Street, Detroit, MI 48201, Fax: +1 313-745-9182, zhengwl@gmail.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

cognitive neuroscience [6–8]. It can predict subject's behaviors, characterize cognitive tasks or mental status, and it has been suggested that it can be used for lie-detection [9]. Some of these studies have demonstrated feasibility of performing pattern classification in real-time during the ongoing scan to predict subject behavior [4,10,11]. Nevertheless, fast and reliable classification remains challenging due to the high dimensionality of fMRI data, the small number of available data sets for training and subject variability in task performance. Moreover, characterization of distributed activation networks [12] and mixed activation patterns during multi-task in real-time fMRI experiments demands automated tools for dynamic interpretation [13].

Feature presentation and selection is a notoriously difficult step in classification problem due to the “curse of dimensionality” [14]. Among extensive techniques [12,15,16] for dimensionality reduction, the most well-know approach is the decision feedback equalizers (DFE-SVM) introduced by Guyon et al. [15]. It iteratively optimizes the weight vector by removing voxels that are not relevant. The DFE-SVM method, which imposes a large computational burden, has been compared to more computationally efficient methods that force voxel sparsity through L1 or L1-L2 regularization of the estimator weight vector [17,18]. However, DFE-SVM seems to be restricted to linear structures due to its computational burden that is exacerbated when it is applied to a nonlinear kernel SVM. Other sparse methods as the one presented in [19] are strictly linear in their formulation, and thus they are not compatible with structures formulated in Reproducing Kernel Hilbert Spaces (RKHS) [20] where the patterns are projected into a higher (often infinite) dimension. In these cases, the relation to patterns in the anatomy-based input space is lost and the translation of feature selection in RKHS into the input space may not be possible. Nevertheless, an intrinsically linear feature selection method is a reasonable approach in fMRI where the number of features is always much higher than the number of available patterns. Feature dimensions can be reduced by choosing only the common spatial features shared among different information sources or different classification methods. Mourão-Miranda et al. [21] used principal component analysis (PCA) of whole brain fMRI to reduce the feature space for classification in attention experiments. Fan et al. [22] increased the accuracy by jointly evaluating both anatomical and functional features and selected the most discriminative features using nested subset selection in a forward manner through a linear classifier. Mitchell et al. [16] defined the most representative feature space as the common region of interest, chosen by a wide range of classifiers at their optimal performance. Feature reduction can also be achieved by using smaller brain areas and ad hoc techniques. In our previous work, we reduced the feature space by segmenting the whole functional activation map into 14 smaller regions based on an anatomical atlas [23]. The local classifier decisions from all the brain regions were then aggregated through an assemble model to provide a final prediction. The weights of the local classifiers were learnt from a boosting procedure [24,25]. This spatially aggregated classifier not only reduced the feature space of the local classifiers, but also improved both the speed and the accuracy compared with a single classifier operating on the entire data set. However, in that study, the rather coarse brain segmentation with only 14 masks extending across multiple functional areas limited specificity of the boosting maps and the generalization of performance. The overlap of activation patterns belonging to different classes was confirmed as cross-correlations of predictions among four classes [23], however, the possibility of a measured activation pattern belonging to multiple classes remains unsolved [23].

The dimensionality problem addressed in our previous study [23] can be further resolved by the aggregation of only part, not all, of the region classifiers in the assemble model. In this approach, the feature space is the classifier output from the regions classifiers. Pros and cons of using the most discriminative features verses the most correlated features have been discussed in [12,15,16].

When the interest is to find the most discriminative regions for a particular classification problem [26], boosting techniques [24,25,27] are among the favored approaches to aggregate classifiers. A boosting algorithm sequentially trains a group of classifiers to form a more powerful ensemble model. The weights of the training data are initially equally assigned. The aggregation coefficients are computed based on the misclassification rate of classifiers on the testing data such that a lower misclassification rate leads to a higher coefficient of a classifier and vice versa. The ensemble coefficients are normalized. In the next iteration, prior to the training procedure, the weights of the training data are updated by increasing the weights of misclassified training data while the weights of correctly classified data are decreased. This forces the subsequent training iteration to focus on the data sets that are difficult to classify by the previous ensemble model and enables the most discriminative classifiers to have the highest priorities in making the final decision. The first polynomial time boosting algorithm was introduced by Schapire [27], followed by the Adaboost algorithm for binary classification introduced by Freund and Schapire [24], and later a multiclass classification version was developed by Allwein et al. [28]. A general framework for boosting based on the minimization of an arbitrary convex loss function was introduced by Mason et al. [29]. The theory of probably approximately correct (PAC) learning [30] shows that given enough training data, weak classifiers (performing slightly better than random guessing) can be aggregated to form an arbitrarily good classifier satisfying any desired (classification or generalization) error tolerance. Since boosting is rooted in PAC learning, where the probabilities of training data belonging to the assigned classes are maximized, it can be applied to the problem of decoupling mixed activation patterns by providing probabilistic predictions of brain activation (testing data) belonging to multiple tasks (classes).

When the interest is to find the most task correlated regions [31], partial least squares (PLS) [32–35] is an alternative approach to weight the regions in the ensemble model. PLS is a class of methods that can be implemented through several different algorithms that find the relationships between two sets of variables (i.e., predictor variables as brain activity and observation variables as behavior or task design) by means of the so-called latent variables, or variables that are not directly measurable. These latent variables or score vectors are estimated through the maximization of the covariance between predictor and observation variables with respect to these latent variables. The PLS method, first applied to econometrics and chemometrics by Wold [36], has been successfully used in functional neuroimaging by McIntosh et al. [37,38] and many others (see a review in [31]). PLS correlation analysis [37] was used to extract the underlying patterns related to task behaviors or experimental design [38]. PLS regression [35,39,40] has been used for prediction behavior [41], classification [42–44], and for modeling [45]. When used for classification purposes, it was found to have performance similar to optimization techniques [43]. In PLS, the probabilistic predictions can be generated in a similar format as in boosting to decouple the mixed activation patterns. However, since PLS minimizes the correlation between tasks (observation variables) and brain activation maps (predictor variables) instead of the probability of brain activations belonging to tasks, they will not be optimized.

The contributions of the present work consist of a procedure that resolves the dimensionality problem by feature selection, a parcellation of brain activation maps into 144 smaller anatomical regions and comparison of two optimal selections of a sparse subset of these regions to construct an ensemble model for aggregation. Boosting selected the most discriminative regions and PLS selected the regions that were most correlated to the activation maps of the tasks. We also developed probabilistic predictions in multi-class classification to resolve mixed tasks using the raw data from our previous study [23], which were acquired at 1.5 and 4 Tesla with interleaved visual, motor, auditory and cognitive tasks. Furthermore, we used the pretrained classifier to demonstrate dynamic decoupling of

mixed tasks on a TR-by-TR basis using the same interleaved tasks to assess the feasibility of real-time fMRI with this pattern classification technique.

2. Materials and Methods

2.1 Functional MRI experiments

2.1.1 Subjects and tasks—Informed consent according to institutionally reviewed guidelines was obtained prior to participation in the study. Data were collected from ten healthy subjects on a 1.5 T Siemens Sonata scanner and on a 4.0 T Brucker MedSpec Scanner equipped with 8 channel head coils as described in [23] (details of the data sets are provided in Table 1).

The paradigm, which has been validated in our previous study [23], consisted of four interleaved tasks: visual (8 Hz checkerboard stimulation), motor (2 Hz right index finger tapping), auditory (syllable discrimination), and cognitive (mental calculation). These tasks were arranged in a randomized block design (8 s per block), with a crosshair serving as baseline for a total of 132 s per scan. The total duration for each condition was thus approximately 27 s. Visual stimulation consisted of 8 Hz reversing black and white checker boards. Finger tapping in the motor task was paced with an auditory tone (1 kHz). Subjects were asked to tap with maximum extension of the finger onto a button-response pad (Cedrus corp., San Pedro, CA). During the auditory task, subjects listened to recorded syllables (i.e., “Ah, Ba, Ha, Ka, Ra, ...”) and pressed a button when they heard the target syllable “Ta” (25% of syllables). The cognitive task consisted of mental calculations. Subjects were asked to sum three aurally presented numbers and divide the sum by three, responding with a button press when the sum was divisible by three (50% of trials). Subjects were instructed to attend to each task with a constant effort across scans and field strengths.

2.1.2 Data acquisition—Functional MRI data at 1.5 and 4 Tesla were acquired using single-shot echo-planar imaging (EPI) with TR: 2 s, TE: 50 ms, flip angle: 90°, matrix size: 64 x 64 or 32 x 32 pixels, FOV: 200 mm or using multi-echo EPI with: TR: 2 s, 4 TEs ranging from 18 to 93 ms for 64×64 matrix at 1.5T, 8 TEs ranging from 11 to 137 ms for 32×32 matrix at 1.5T, 3 TEs ranging from 18 to 68 ms for 64×64 matrix at 4 T, 7 TEs ranging from 11 to 119 ms for 32×32 matrix at 4 T, flip angle: 90°, FOV: 192 mm (200 mm for 64×64 matrix at 4 Tesla). Data with the 32 x 32 matrix were acquired with different bandwidth, either with 1200 Hz/pixel (LBW) or with 2400 Hz/pixel (HBW), which changes the degree of geometrical distortion and the signal-to-noise ratio (SNR). Slices were 6 mm thick, with 10% gap, 66 volumes were collected for a total measurement time of 132 s per run.

2.1.3 Retrospective fMRI data analysis—The data were preprocessed retrospectively using a custom designed TurboFIRE (version v.5.2.8) real-time fMRI analysis software package [46]. Preprocessing steps included weighted combination of multi-echo EPI data [47], and motion [48] and slice time correction. Preprocessed images were spatially normalized to Montreal Neurological Institute (MNI) template [49] using an affine transformation and nonlinear deformations as in SPM5 (with 7×8×7 nonlinear basis functions, medium nonlinear regularization, and 2×2×4 voxel size), followed by isotropic Gaussian filtering using an 8 mm kernel [50]. Cumulative correlation analysis, which was, implemented computationally efficiently as described in [51], was performed concurrently for each of the four reference vectors and the final correlation maps of each scan were used for pattern classification.

2.2 Design of the spatially aggregated classifier

The structure of the spatially aggregated classifier is shown in Fig. 1. In this study, correlation maps from fMRI analysis were used for pattern classification. Maps were normalized in amplitude and, then, a feature selection procedure described below was applied to select those voxels that contributed most to the classification performance based on their weights in a whole brain linear SVM. After that, a segmentation of the brain into 144 regions was applied and each region was vectorized. Vectors belonging to each region were input to local classifiers. The local classifier decisions were then aggregated through an assembly model for each task, leading to a probabilistic prediction. The task with the highest probability was the winner, as detailed below.

The available data set consisted of 312 activation maps taken from 10 different subjects, two different field strengths and four tasks (details in Table 1). The classification was thus across subjects, field strengths and tasks, which has been characterized for each of these subsets of data in our previous study [23]. The pipeline for preprocessing of the data and training of the classifier is depicted in Fig. 2 and consists of the following five steps: (a) Splitting the 312 available correlation maps into training and testing sets, and amplitude normalization; (b) feature selection using a linear SVM [52] on the whole brain voxels and then segmentation of the selected voxels into 144 anatomical regions based on the Talairach Daemon database [53]; (c) optimizing hyper-parameters for local nonlinear multi-class SVM classifiers (one for each region) with Gaussian radial basis function (RBF) kernels using code from the open source library LIBSVM (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>) [52]; (d) selecting an optimal combination of regions via iterative estimation of the aggregation coefficients of local classifiers using boosting and backward elimination; and (e) final training and testing of the aggregated classifier on the test dataset. Details of each step are described in the subsections below.

2.2.1 Normalization of the correlation maps—In our current implementation, the inputs of the spatially aggregated classifier, i.e. features, are the Spearman correlation maps [51]. Maps were split using a 2:1 ratio into a training set D_{train} and a testing set D_{test} with the same distributions of subjects, field strengths and tasks. Variation in correlation maps among subjects and across tasks due to inter-individual differences in task performance and functional neuroanatomy was reduced using amplitude normalization. Since the data was acquired across subjects, field strengths and tasks, the histogram of correlation maps did not follow a simple form of Gaussian distribution. We chose a Gaussian mixture model including three Gaussian distributions and one uniform distribution as a general model that most accurately represents all types of correlation map histogram distributions. The solution for the mixture model of each activation map was obtained through Estimation Maximization (EM) fitting [54]. Subsequently, the data was normalized by subtracting the mean and by dividing by the standard deviation of the most dominant Gaussian distribution, which represented thermal and physiological noise.

2.2.2 Feature selection—To reduce the dimension for local classifiers, the most discriminative voxels were selected by thresholding the weights of the voxels in a whole brain classification. These weights were calculated as a weighted average of the absolute value of all the support vectors obtained from training a four-class linear SVM using $\frac{1}{3}$ of the training set D_{train} . Half of the remaining $\frac{2}{3} D_{\text{train}}$ data was used for training a single SVM (i.e., to determine the support vectors and their weights) and the other half for testing its performance (i.e., to determine the classification rate across subjects, field strengths and tasks). Voxel selection for both aggregation methods (see below) was performed for a wide range of thresholds by increasing the threshold with a step size of 0.01 to test the performance of the SVM. The best-featured voxels were selected by maximizing

performance with the smallest number of voxels. To gain a better statistical selection of the featured voxels, the above procedure was repeated 200 times using bootstrapping of the training set D_{train} to optimize the confidence interval of the selection [55]. The average number of featured voxels was estimated and the featured voxels were selected as those with the highest selection probability in the bootstrapping procedure.

The featured voxels were then segmented in the following way. There are 144 brain regions at the 5th level labeled in the Talairach Daemon database [53] (<http://www.talairach.org/daemon.html>). The voxel coordinates of each region in Talairach space were transformed into MNI space [49] using Matthew Brett's formula [56] to create a vectorized mask for that region. Then the featured voxels from each activation map, which were already in MNI space and vectorized, were segmented into 144 regions using the regions masks defined in MNI space. The masks were pruned since some voxels in the masks were not selected as featured voxels. If no voxel in a region was selected during feature selection, that region was not considered in subsequent boosting or PLS aggregation. The final vector dimensions for each region (i.e., local classifier) varied depending on the volume of the region.

2.2.3 Model parametric optimization of local classifiers—Nonlinear SVM classifiers (using the ‘one versus the rest’ strategy implemented in LIBSVM) with Gaussian radial basis function (RBF) kernels for four-class classification were applied to the individual regions and the hyper-parameters of this kernel type SVM model (i.e. the trade-off factor C of the SVM primal objective function [2] and the kernel width δ) were optimized with 10-fold cross validation using the entire training set D_{train} .

2.2.4 Distributed classifier using boosting aggregation of local classifier decisions—Boosting is a learning algorithm that combines the performance of classifiers to produce a dramatically improved classifier. The boosting procedure used in this paper is detailed in our previous study [23]. In the present implementation, the probability of each task being represented in the activation pattern was also computed. The multi-class problem was dealt with in the manner of ‘one versus the rest’ for each class due to the possibility that a given activation pattern may belong to more than one class as in text classification. Let $h_1(x), h_2(x), \dots, h_N(x)$ be a set of local classifier outputs, and denote their coefficients. The ensemble classifier output was formulated as:

$$H(x) = \sum_{i=1}^N \lambda_i h_i(x) \quad (1)$$

The ensemble classifier model for each class was:

$$H_k(x) = \sum_{i=1}^N \lambda_{i,k} h_{i,k}(x) \quad (2)$$

Minimizing the total cost function $\sum_{k=1}^K J(H_k(x))$ regarding the ensemble classifier output $H_k(x)$ of class k , $1 \leq k \leq K$, by minimizing its cost function:

$$J(H_k) = E(e^{-y H_k(x)}) \quad (3)$$

where,

$$H_k(x) = \log p_k(x) - \frac{1}{K} \sum_{j=1}^K \log p_j(x) \quad (4)$$

was the one-against-the-rest classifier output for input pattern x , this led to the following expression of the probability that a pattern x belonged to class k :

$$p_k(x) = \frac{e^{H_k(x)}}{\sum_{j=1}^K e^{H_j(x)}} \quad (5)$$

The classifier with the highest probability was the winner. Our boosting procedure used the Quasi-Newton method to update the coefficients and weights to minimize each cost function as described in our previous study [23].

2.2.5 Distributed classifier based on PLS regression—PLS finds the relationships between two sets of variables by means of the latent variables and can be implemented through several different algorithms (see the detailed review in [31]). In this paper, PLS regression was used as a means to solve the previously defined assembly classifier model in Eqs. (1) and (2).

Assuming a set of predictor variables $h_k = \{h_{i,k}(x)\}$ and a set of labels of classes $y_k = \{y_{i,k}(x)\}$, the PLS procedure constructed the following representation of the data:

$$\begin{cases} H = TP + E \\ Y = UQ + F \end{cases} \quad (6)$$

where H and Y were matrices containing all available instances of vectors h_k and y_k respectively. Matrices T and U contained a set of orthogonal latent variables corresponding to the two sets, and P and Q were the corresponding projections of the data over the latent variables. Thus, expressions TP and UQ were an approximate representation of the data with residuals or errors E and F . A nonlinear partial least squares (NIPALS) [36,57] procedure was used to find the latent variables in matrix T and U in Eq. (6).

The regression equation $Y = \Lambda H + F^*$ can be expressed with regression parameters:

$$\Lambda = H^T U (T^T H H^T U)^{-1} T^T Y \quad (7)$$

and with residuals:

$$F^* = H Q^T + F \quad (8)$$

The regression matrix $\Lambda = (\lambda_{1,k}, \lambda_{2,k}, \dots, \lambda_{N,k})$ was obtained for the aggregation of N local classifier outputs regarding class k . Note that Eq. (2) is still used in our final output to provide a probability prediction, but it does not result from the minimization of the designed cost function and has no relation to it.

2.2.6 Search for optimal combination of regions—Among the several advanced strategies to find a smaller subset of variables (regions in the assembly model in this paper) discussed in [15], the backward elimination and forward selection approaches are most frequently used in classifier-based functional neuroimaging analysis to select the brain regions involved in an activation pattern. Here, we hypothesize that segmentation using a

small number of large functional areas, which we have used previously [23], limits the classifier performance, since many of these large regions may contain significantly activated voxels. Sparsifying the input data using finer brain segmentation resulting in a larger number of functional brain regions may translate into better classification accuracy and increased specificity of the selection of discriminative and task-related regions as a result of the reduction of the structural complexity of the classifier. We chose to use backward elimination instead of forward selection to select regions, since, as argued in¹⁵, backward elimination may outperform forward selection by rejecting a ‘loner’ region which works better by itself than in combination with other regions.

The optimal combination of regions was selected by starting from an aggregation with all regions, iteratively dropping one region at a time and performing a boosting procedure to train the local classifiers and to estimate the aggregation coefficients that provide the best prediction. The iteration was stopped when only four regions remained. The training data D_{train} was bootstrapped 200 times, and each time split into 2/3 D_{train} for training and 1/3 D_{train} for testing to estimate the error of the classifiers across the number of selected regions. The optimal number of regions was defined at the best classification rate average, and the optimal combination was defined as the most frequently selected set of regions.

Finally, the training set D_{train} and testing set D_{test} were used to train and test, respectively, the optimal aggregation of local classifiers. The local SVM models (i.e., the support vectors and their coefficients) were trained with half of the training data D_{train} , and the aggregation parameters were trained using the other half. The aggregated model was tested using the testing data D_{test} , which had not been used in any of the previous processing steps. The trained classifier model was integrated into the TurboFIRE real-time fMRI analysis pipeline for online classification as described below.

2.3 Real-time fMRI experiments

A real time fMRI experiment was performed in order to detect sensorimotor and cognitive tasks online. The boosting and PLS classifiers, which were trained offline as described above, and the resulting parameters were stored for use in real time testing.

Note that none of the data obtained from the real-time experiment was used to train the classifier. Two healthy subjects were scanned using a 3T Siemens Trio scanner equipped with a 12 channel head coil while performing the four interleaved tasks described in the Subjects and tasks section. Data were acquired using a multi-echo EPI pulse sequence to maximize the BOLD sensitivity [47]. Acquisition parameters were TR= 2s, 64×64 matrix and 2-fold GRAPPA acceleration using either: five TEs between 11 and 111 ms with 16 slices and 3×3×6 mm³ voxel size; or, four TEs between 11 and 59 ms with 25 slices and 5×5×5 mm³ voxel size.

Real-time fMRI analysis with online pattern classification was performed using TurboFIRE (version v.5.3.2). Data transfer from the reconstruction computer via the scanner host computer to the external RedHat Linux Dual-Core 2.40GHz processor workstation was implemented using TCP/IP communication and stream sockets. Multi-echo EPI preprocessing was performed to generate TE-averaged images. Automatic pixel-specific weights were based on real-time fitting of the multi-echo EPI data to compute pixel-wise T_2^* values to maximize BOLD sensitivity [47]. Further processing was identical to the steps described above in the offline analysis (see Materials and Methods section). Activation maps obtained with cumulative correlation analysis were classified at each TR. One of the pretrained classifier models (either the boosting or the PLS classifier) was preloaded into TurboFIRE using the parameters obtained in the training phase, and run using a test function included with the LIBSVM (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>) library as part of the

online classifier. Both boosting and PLS structures had identical computational burden. The online classifier produced a probabilistic estimation for each of the four possible tasks. The winner for the selected reference vector was displayed on the GUI main window during the scan.

Steady state data acquisition, image reconstruction, image transfer, image analysis and pattern classification were performed with data acquisition rates of up to 70 2D images/sec and time delay of less than 1 TR from the acquisition of a complete volume to the output of the pattern classification results. Nevertheless, since TurboFIRE supports single model classification, a comparison between both approaches had to be performed retrospectively. Plots of the relative contributions of the classification were produced offline in Gnuplot (<http://www.gnuplot.info/>)

3. Results

3.1 Feature selection

Using thresholding of the voxel weights obtained from a linear four-class SVM, the most discriminative voxels were found at a voxel weight threshold of 0.2 at which the minimum misclassification rate (5.7%) performance was achieved with approximately 3×10^4 voxels. These voxels were fewer than half of the whole brain voxels and located in 95 of the 144 total regions. The activated brain regions detected at 1.5 and 4 Tesla are summarized in Table 2. The activated regions were largely consistent across these two field strengths and for each task they individually corresponded to the brain regions reported in previous studies using similar tasks. The references in Table 2 (motor [58–64], visual [65–68], cognitive [69–76] and auditory [77–82]) reporting the same areas of activation by similar tasks as ours were based on the Brainmap functional database search results using the Sleuth (version 2.0.3) application (www.brainmap.org) [83–85]. The search criteria were constrained by the stimulus condition, auditory tones for the motor task and auditory syllables for the auditory task, and by the experimental paradigm class: finger tapping, flashing checkerboard, and counting/calculation for the motor, visual and cognitive tasks, respectively. The selected references were further refined to include only those studies that best matched our experimental paradigms. Here, the minimum misclassification rate (5.7%) was less than the misclassification rate of more than 9% reported in our previous work [23] using the same tasks and raw fMRI data. This indicates that the feature selection in this study was superior to the thresholding method based on the standard deviation of activation map values across subjects, fields and tasks, which was used in our previous work [23]. Fig. 3 shows the group averaged correlation map of each task (motor, visual, cognitive and auditory) and the feature selected voxels mask. Both motor and visual tasks had the most distinct activation patterns among the tasks, although the motor pattern contained auditory activation due to auditory gating of the motor task. Cognitive and auditory activation patterns had extensive overlap of regions between each other, since both tasks activated frontal regions involved in the executive function. The cognitive activation pattern also contained auditory activation due to the verbal input.

3.2 Optimal combination of regions

Fig. 4 shows that the misclassification rate as a function of the number of regions in the aggregation model exhibited a flat minimum between 30 to 75 regions, which demarcates the range between overfitting beyond 75 regions, and underfitting below 30 regions. Underfitting is particularly pronounced for PLS compared to boosting when fewer than 15 regions were selected, whereas PLS exhibited slightly better performance ($p < 0.05$) than boosting when more than 15 regions were selected.

Among the optimal 30 regions (the least number of regions needed for the best performance) that are listed in Table 2 and 3, the majority (25 regions) were the same for both boosting and PLS selection criteria. The weak correlation ($R=0.49$) between boosting and PLS coefficients indicated that these two aggregation methods assigned different importance to individual regions.

Fig. 5 shows that the performance of individual local classifiers using nonlinear SVM with optimized RBF kernel parameters was limited. The 30 regions picked by both methods had the ability to discriminate the tasks with an accuracy ranging between 48 and 91%. However, the aggregated classifier had accuracy higher than 97%. This demonstrates the power of the aggregation classifier as originally described in [30]. Only the regions with the highest classification accuracy (71~91%) were picked by boosting. This agreed with the strategy of boosting algorithm to reward the most discriminative regions by giving them the highest weights. Among the regions picked only by PLS, Brodmann area 30 [R] (region 25) and Medial Dorsal Nucleus [R] (region 58) had individually weak accuracy of 49% and 51%.

3.3 Interpretations of Spearman correlation, representative features, and aggregation coefficients

Fig. 6 shows the comparison of representative features extracted from a linear SVM applied on the whole brain voxels, and the aggregation coefficients of the ensemble classifier from boosting and PLS regression. Interpretation of the patterns of selected regions strongly depended on the chosen learning method. In Fig. 6, the representative patterns, which had the maximal projection onto the separation planes of the linear SVM, were similar to the patterns in the correlation maps (Fig. 3). For example, the regions with strong correlation with the motor tasks (BA1-5[L], primary motor cortex BA4 being the strongest) were identical to the regions highly weighted (BA4 being most highly weighted) by the linear SVM. The patterns for the other tasks in the SVM weight maps were easily understood in the same way.

However, interpretation of the aggregation coefficient maps of boosting and PLS regression was not straightforward and not necessarily related to the expected activation patterns. For example, some regions were explanatory of a given task if they were deactivated only for this particular task. For boosting, the coefficients indicated the ability of a region to discriminate a task. For example, primary auditory cortex BA41[L] had a strong correlation with the auditory task in the correlation map (Fig. 3), but the information (local classifier output) in this region itself had a low ability ($\lambda=0.007$) to discriminate the auditory task from the others because it was also correlated with the motor task, which was gated using auditory input. In contrast, BA8[L] (frontal cortex, which plays an important role in eye movement) had strong negative correlation with the auditory task, but ten times more capability ($\lambda=0.079$) in discriminating this task. This suggests that negative correlation with a task is also important information for identifying the involvement of a region in a particular task. In the PLS regression coefficient maps, taking the visual task as an example, both the regions that were strongly positively correlated with the visual task (BA 17-19[L,R], primary, secondary and associative visual cortex) and those regions that were strongly negatively correlated with the visual task (BA21 [L,R], middle temporal gyrus) were identified as important (local classifier decision) in the underlying pattern, which best reflected the change from this task to the other tasks.

3.4 Real-Time fMRI with multi-class classification

The SNR, geometrical distortion and overall BOLD sensitivity of the real-time fMRI data acquired at 3 Tesla were between those of the 1.5 and the 4 Tesla data. The spatial

distribution of the activation patterns detected at 3T was consistent with those detected in the offline studies at 1.5 and 4 Tesla described above (Table 2). Fig. 7 shows the comparison of the performance of both the boosting and the PLS classifier in one of the probabilistic prediction experiments. During the first 20 to 30 seconds of the real-time fMRI experiments, the outputs clearly reflected a transient state after which the classifiers stabilized. The transient roughly determined the minimum number of volumes needed to produce a useful correlation map, which coincided with the completion of the first activation block in all tasks (see panels a, b and d) except for the cognitive task (panel c), which took about 80 seconds to stabilize. The presence of auditory activation during the cognitive task, which corresponded to the subject listening to the questions, dominated the classification output between 20 and 80 seconds into the scan. The successful decoupling of both tasks after the initial 80 seconds reflected the increase in contrast-to-noise ratio in functional regions that were specific to the cognitive task. Although auditory and cognitive activation patterns had extensive overlap in auditory cortex and anterior cingulate cortex (Fig. 7c, d), the optimally aggregated classifiers predicted the auditory task with a probability greater than 80% and revealed almost no involvement of the cognitive task in the auditory task and approximately 20 % involvement of the auditory task in the cognitive task after the initial 80 s of the scan, which illustrated the specificity of the prediction.

4. Discussion

Feature selection is an important step in achieving fast, reliable and accurate predictions. Feature selection in this paper using linear SVM on the whole brain voxels provided more robust results than thresholding the standard deviation of the activation map, which was used in our previous work [23]. Compared with the recursive procedure of the RFE-SVM used in [17], the feature selection used here is suboptimal but faster despite the use of bootstrapping. Furthermore, the applied method is more automatic since in RFE-SVM, the number of iterations and the number of features to be discarded at each iteration must be preset, and the most discriminative feature set selected after the optimization requires a heuristic choice. The feature selection in this paper assigned the ranks of voxels, which did not change in later training. This enabled the investigation of feature selection performance as a function of a threshold of voxel weights. Feature selection exhibited an increasing misclassification rate with a decreasing number of selected voxels. Thus, it was possible to construct a convex function, which automatically converged to an optimal number for feature voxels.

Boosting and PLS regression selected 25 common regions and only 5 different regions. Comparing the selection of individual regions, which were expressed in the different aggregation coefficients, revealed the different strategies of boosting and PLS regression in evaluating the roles of regions. Boosting emphasized individual discriminative performance: all of the selected regions had better performance than those regions that were discarded. PLS by contrast evaluated the impact of the presence of a region on the performance of the entire whole assemble model: it included two regions that had poor performance individually, but contributed to the overall prediction of the assemble model. The results obtained with both aggregation methods demonstrated that negative correlation with task specific activation patterns plays an important role in understanding the involvement of a region in a task. However, the involvement of negatively task-correlated regions cannot be judged solely on classification performances. Other classification methods [86–88] and aggregation strategies, for example Monte Carlo methods [89] and Evolutionary algorithms [90], can also be used for aggregation to characterize the involvement of individual brain regions in prediction.

The spatially aggregated classifier provided high classification accuracy of more than 97 % for the retrospective analysis, which was slightly higher than in our previous study [23]. This high classification accuracy was obtained despite the considerable overlap of activated regions between tasks as a result of task mixture (e.g., auditory gating of motor task), and the fact that data were acquired at different field strengths with a range of pulse sequence parameters. In our previous study comparing two different implementations of the Adaboost algorithm for aggregated pattern classification we demonstrated that the effect of field strength on classification performance was minor: training at 1.5 Tesla and testing at 1.5 Tesla resulted in similar error rates as training at 4 Tesla and classification at 1.5 Tesla. Error rates for testing and training with different image resolutions and readout bandwidths showed larger variation, but these were still considerably smaller than for a single whole brain classifier. The low error rates in the present study suggest that the classifier is robust with respect to variability in SNR in the raw data and inter-individual differences in activation patterns. It will thus be of interest to apply spatially aggregated classification to more complex patterns of activation encountered in cognitive tasks that are more difficult to discriminate and thus provide an even more appropriate setting for testing the classification performance of this new methodology.

The considerable reduction of total classification time makes the aggregated classifier suitable for real-time fMRI experiments. In the present study we showed the feasibility of applying a pretrained classifier for dynamic probabilistic prediction during real-time fMRI experiments, which revealed onset differences in activation patterns during the initial phases of the scan. This is not feasible with hard decision classifiers. The classifiers were also shown to be robust with respect to the low contrast to noise ratio of the test data in the initial phase of the scan.

The increasing range of applications of online classification of fMRI data and the importance of real-time fMRI in studying brain disease demands speed, reliability and accuracy [13]. Sensitivity constraints in real-time fMRI dictate paradigms with limited number of task conditions resulting in complex activation patterns with multiple mixed task components that require automated tools for interpretation. Our data show that spatially aggregated pattern classification is suitable for online classification of a single block of task execution. Our real-time fMRI data suggest that a sliding window length of 20 to 30 seconds would be adequate to classify task related activity during visual, motor and auditory tasks. Using sliding window correlation analysis [51] this spatially aggregated classifier is thus expected to enable tracking of dynamically changing task contributions during a real-time fMRI scan, which may enable automated monitoring of shifts in task execution or attention.

The real-time fMRI experiments further substantiated the findings of our previous study [23] that spatially aggregated pattern classification methods are robust with respect to the acquisition of training data at different field strengths, with different fMRI pulse sequences and with different acquisition parameters. Moreover, the classifiers were robust with respect to the significant difference in contrast-to-noise ratio between the training and the noisy test data at the beginning of the real-time scan. The stabilization of the classifier outputs during the initial phase of the scan shown in Figure 7 reflected the increase in the contrast-to-noise ratio of the test data during cumulative analysis, after which the classifier outputs became invariant with respect to the contrast-to-noise ratio, as expected.

We anticipate that the high spatial specificity of this aggregation approach will improve the physiological validity of neurofeedback, by providing a feedback signal from an entire functional network representing a complex task compared to traditional neurofeedback from a single or a few discrete brain regions. This may lead to new approaches of neurofeedback training. Finally, to further improve temporal resolution for neurofeedback studies we are

also investigating the feasibility of spatially aggregated pattern classification in time domain raw data with online detrending of signal drifts.

5. Conclusion

In the present study, a spatially aggregated classifier was developed with automated assessment of the involvement of individual brain regions in multi-class classification across subjects, field strengths, tasks and fMRI data acquisition methods. Effective reduction of dimensionality by means of feature voxel reduction and optimal aggregation of brain regions was shown to make fast classification with high prediction accuracy feasible with this spatially aggregated classifier. Comparing boosting and PLS regression for aggregation showed that these methods enable selection of the most discriminative and the most task correlated regions, respectively. We also demonstrated the speed and classification performance of this aggregated classification approach enables the monitoring of the stabilization of the prediction output during the initial phase of the scan during real-time fMRI, which suggests that monitoring of changes in task prediction during individual blocks of task activation will be feasible. This study also demonstrated dynamic decoupling of mixed activation patterns to provide task specific information from functional networks rather than discrete brain regions, thus providing a new approach towards fMRI-based neurofeedback training.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants 1R01 EB002618-01 and 1R41NS062474-0. We thank Maxim Zaitzev and Jochen Rick (U. Freiburg) and Oliver Speck (U. Magdeburg) for sharing the MEPI pulse sequence and for supporting the implementation of the real-time export interface. We gratefully acknowledge the support of Oliver Hinds (MIT) for his advice on network communication. We thank the Center for Advanced Research Computing at UNM for providing access to its high performance computing facilities.

References

1. Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI) “brain reading” : detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*. 2003; 19(2): 261–270. [PubMed: 12814577]
2. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Mining Knowl Discov*. 1998; 2(2):1–32.
3. Vapnik, V. *The Nature of Statistical Learning Theory*. N.Y: Springer-Verlag; 1995.
4. LaConte SM, Strother S, Cherkassky V, Anderson J, Hu X. Support vector machines for temporal classification of block design fMRI data. *NeuroImage*. 2005; 26(2):317–329. [PubMed: 15907293]
5. LaConte, SM.; Strother, S.; Cherkassky, V.; Hu, X. Predicting motor tasks in fmri data with support vector machines. Toronto, Ontario, Canada: 2003.
6. Wang, X.; Hutchinson, R.; Mitchell, TM., editors. *Training fMRI Classifiers to Detect Cognitive States across Multiple Human Subjects*. Vol. 16. MIT Press; Cambridge, MA: 2004.
7. Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*. 2005; 8(5):679–685.
8. Haynes J-D, Rees G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*. 2005; 8(5):686–691.
9. Davatzikos C, Ruparel K, Fan Y, Shen DG, Acharyya M, Loughead JW, Gur RC, Langleben DD. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *NeuroImage*. 2005; 28(3):663–668. [PubMed: 16169252]

10. LaConte SM, Scott J, Peltier, Hu XP. Real-Time fMRI Using Brain-State Classification. *Human Brain Mapping*. 2007; 28:1033–1–44. [PubMed: 17133383]
11. deCharms CR. Reading and controlling human brain activation using real-time functional magnetic resonance imaging. *Trends Cogn Sci*. 2007; 11(11):473–481. [PubMed: 17988931]
12. Norman K, Polyn SM, Detre G, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends Cogn Sci*. 2006; 10(9):424–430. [PubMed: 16899397]
13. LaConte SM. Decoding fMRI brain states in real-time. *NeuroImage*. 2011; 56:440–454. [PubMed: 20600972]
14. Bellman, R. *Adaptive Control Processes*. Princeton, New Jersey: Princeton University Press; 1961.
15. Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*. 2003; 3:1157–1182.
16. Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X, Just M, Newman S. Learning to decode cognitive states from brain images. *Mach Learn*. 2004; 57:145–175.
17. De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E. Combining multivariate voxel selection and support vector machines for mapping and classification of fmri spatial patterns. *NeuroImage*. 2008; 43(1):44–58. [PubMed: 18672070]
18. Ryali S, Supekar K, Abrams DA, Menon V. Sparse logistic regression for whole-brain classification of fMRI data. *NeuroImage*. 2010; 51(2):752–764. [PubMed: 20188193]
19. Bi J, Bennett KP. Dimensionality Reduction via Sparse Support Vector Machines. *Journal of Machine Learning Research*. 2003; 3:1229–1243.
20. Aizerman A, Braverman EM, Rozoner LI. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*. 1964; 25:821–837.
21. Mourão-Miranda J, Bokde ALW, Born C, Hampel H, Stetter M. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *NeuroImage*. 2005; 28(4):980–995. [PubMed: 16275139]
22. Fan Y, Resnick SM, Wu X, Davatzikos C. Structural and functional biomarkers of prodromal Alzheimer's disease: A high-dimensional pattern classification study. *NeuroImage*. 2008; 41(2): 277–285. [PubMed: 18400519]
23. Martínez-Ramón M, Koltchinskii V, Heileman GL, Posse S. fMRI pattern classification using neuroanatomically constrained boosting. *NeuroImage*. 2006; 31(3):1129–1141. [PubMed: 16529955]
24. Schapire, RE. *A brief introduction to boosting*. San Francisco, CA, USA: Morgan Kauffman Pub. Inc; 1999. p. 1401-1406.
25. Schapire RE, Singer Y. Improved boosting algorithms using confidence-rated predictions. *Mach Learn*. 1999; 37(3):297–336.
26. Williams MA, Dang S, Kanwisher NG. Only some spatial patterns of fMRI response are read out in task performance. *Nature Neuroscience*. 2007; 10:685–686.
27. Freund, Y.; Schapire, R. *Lecture Notes in Computer Science*. Vol. 904. Springer; Berlin: 1995. A decision-theoretic generalization of online learning and an application to boosting. *Computational Learning Theory*; p. 23-37.
28. Allwein EA, Schapire R, Singer Y. Reducing multiclass to binary: a unifying approach for margin classifiers. *J Mach Learn Res*. 2000; 1:113–141.
29. Mason, L.; Baxter, J.; Bartlett, P.; Frean, M. *Functional Gradient Techniques for Combining Hypotheses*. MIT Press; Cambridge: 1999. *Advances in large margin classifiers*. Ch; p. 33-58.
30. Valiant LG. A theory of the learnable. *Communications of the ACM*. 1984; 27(11):1134–1142.
31. Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage*. 2011; 56(2):455–475. [PubMed: 20656037]
32. Frank I, Friedman J. A statistical view of some chemometrics regression tools. *Technometrics*. 1993; 35:109–135.
33. Helland I. On the structure of partial least squares regression. *Commun Statist imul*. 1998; 17:581–607.
34. Helland I. Partial least squares regression and statistical models. *Scand J Statist*. 1990; 17:97–114.

35. Wold, H. Soft modelling, the basic design and some extensions. In: Wold, H.; Jöreskog, K-G., editors. *Systems Under Indirect Observation: Causality-Structure-Prediction Part II*. Amsterdam: North-Holland Publishing Company; 1982. p. 1-54.
36. Wold, H. Estimation of principal components and related models by iterative least squares. In: Krishnaiah, DPR., editor. *Multivariate Analysis*. N.Y: Academic Press; 1966. p. 391-420.
37. McIntosh AR, Bookstein FL, Haxby JV, Grady CL. Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*. 1996; 3:143–157. [PubMed: 9345485]
38. McIntosh AR, Chau WK, Protzner AB. Spatiotemporal analysis of event-related fMRI data using partial least squares. *NeuroImage*. 2004; 23(2):764–775. [PubMed: 15488426]
39. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst*. 2001; 58:109–130.
40. de Jong S. SIMPLS: an alternative approach to partial least squares regression. *Chemom Intell Lab Syst*. 1993; 18:251–263.
41. Giessing C, Fink G, Rosler F, Thiel C. fMRI data predict individual differences of behavioral effects of nicotine: a partial least squares analysis. *J Chemometr*. 2007; 18:208–215.
42. Gottfries J, Blennow K, Wallin A, Gottfries CG. Diagnosis of dementias using partial least squares discriminant analysis. *Dementia*. 1995; 6:83–88. [PubMed: 7606284]
43. Lehmann C, Koenig T, Jelic V, Prichep L, John RE, Wahlund L-O, Dodge Y, Dierks T. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity. *J Neurosci Meth*. 2007; 161:342–350.
44. Chen K, Reiman EM, Huan Z, Caselli RJ, Bandy D, Ayutyanont N, Alexander GE. Linking functional and structural brain images with multivariate network analyses: A novel application of the partial least square method. *NeuroImage*. 2009; 47(2):602–610. [PubMed: 19393744]
45. Esposito-Vinzi, V.; Wynne, W.; Chin, WW.; Henseler, J.; Wang, H. *Handbook of Partial Least Squares: Concepts, Methods and Applications*. New York: Springer Verlag; 2010.
46. Posse S, Binkofski F, Schneider F, Gembris D, Wiese S, Kiselev V, Graf T, Elghawaghi B, Eickermann T. A New Approach to Measure Single Event Related Brain Activity using Real-Time fMRI: Feasibility of sensory, motor, and higher cognitive tasks. *Human Brain Mapping*. 2001; 12(1):25–41. [PubMed: 11198103]
47. Posse S, Wiese S, Gembris D, Mathiak K, Kessler C, Grosse-Ruyken M-L, Elghawagh B, Richards T, Dager SR, Kiselev VG. Enhancement of BOLD-Contrast Sensitivity by Single-Shot Multi-Echo Functional MR Imaging. *Magnetic Resonance Med*. 1999; 42(1):87–97.
48. Mathiak K, Posse S. Evaluation of Motion and Realignment for Functional Magnetic Resonance Imaging in Realtime. *Magn Reson Med*. 2001; 45(1):167–171. [PubMed: 11146500]
49. Evans AC, Collins D, Mills SR, Brown ED, Kelly RL, Peters TM. 3D statistical neuroanatomical models from 305 MRI volumes. 1993:1813–1817.
50. Gao K, Posse S. TurboFire: Real-Time fMRI with Automated Spatial Normalization and Talairach Daemon Database. 2003:838.
51. Gembris D, Taylor JG, Schor S, Frings W, Suter D, Posse S. Functional magnetic resonance imaging in real time (FIRE): Sliding-window correlation analysis and reference-vector optimization. *Magnetic resonance in medicine*. 2000; 43(2):259–268.
52. Chang, C-C.; Lin, C-J. LIBSVM : a library for support vector machines. 2001. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm> Accessed
53. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT. Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*. 2000; 10(3):120–131. [PubMed: 10912591]
54. Dempster AD, Larid NM, Rubin DB. Maximum likelihood from the incomplete data via the EM algorithm. *J R Statist Soc*. 1977; 39:1–37.
55. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability 57)*. Chapman & Hall/CRC; 1994.
56. Brett M, Christoff K, Cusack R, Lancaster J. Using the Talairach atlas with the MNI template. *NeuroImage*. 2001; 13

57. Wold, H. Path models with latent variables: The NIPALS approach. In: al HMBe, editor. Quantitative Sociology: International perspectives on mathematical and statistical model building. New York: Academic Press; 1975. p. 307-357.
58. Jäncke L, Shah NJ, Peters M. Cortical activations in primary and secondary motor areas for complex bimanual movements in professional pianists. *Cognitive Brain Research*. 2000; 10:177–183. [PubMed: 10978706]
59. Mattay VS, Callicott JH, Bertolino A, Santha AKS, Van Horn JD, Tallent KA, Frank JAW, DR. Hemispheric control of motor function: A whole brain echo planar fMRI study. *Psychiatry Research*. 1998; 83:7–22. [PubMed: 9754701]
60. Ramsey NF, Kirkby BS, Gelderen tV, Karen F, Berman, Duy JH, Frank JA, Mattay VS, John D, Van Horn, Esposito G, Moonen CTW, et al. Functional Mapping of Human Sensorimotor Cortex with 3D BOLD tMRI Correlates Highly With H2150 PET rCBF. *Journal of Cerebral Blood Flow and Metabolism*. 1996; 16:755–764. [PubMed: 8784221]
61. Rao SM, Harrington DL, Haaland KY, Bobholz JA, Cox RW, Binder JR. Distributed Neural Systems Underlying the Timing of Movements. *The Journal of Neuroscience*. 1997; 17(14):5528–5535. [PubMed: 9204934]
62. Yoo SS, Wei X, Dickey CC, Guttmann CRG, Panych LP. Long-term reproducibility analysis of fMRI using hand motor task. *International Journal of Neuroscience*. 2005; 115:55–77. [PubMed: 15768852]
63. Sabatini U, Boulanouar K, Fabre N, Martin F, Carel C, Colonnese C, Bozzao L, Berry I, Montastruc JL, Chollet F, et al. Cortical motor reorganization in akinetic patients with Parkinson's disease: A functional MRI study. *Brain*. 2000; 123:394–403. [PubMed: 10648446]
64. Gelnar PA, Krauss BR, Sheehe PR, Szeverenyi NM, Apkarian AV. A Comparative fMRI Study of Cortical Representations for Thermal Painful, Vibrotactile, and Motor Performance Tasks. *NeuroImage*. 1999; 10:460–482. [PubMed: 10493903]
65. Moradi F, Liu LC, Cheng K, Waggoner RA, Tanaka K, Ioannides AA. Consistent and precise localization of brain activity in human primary visual cortex by MEG and fMRI. *NeuroImage*. 2003; 18:595–609. [PubMed: 12667837]
66. Hermann D, Smolka MN, Klein S, Heinz A, Mann K, Braus DF. Reduced fMRI activation of an occipital area in recently detoxified alcohol-dependent patients in a visual and acoustic stimulation paradigm. *Addiction Biology*. 2007; 12:117–121. [PubMed: 17407505]
67. Bristow D, Frith C, Rees G. Two distinct neural effects of blinking on human visual processing. *NeuroImage*. 2005; 27:136–145. [PubMed: 15893941]
68. Cohen L, Martinaud O, Lemer C, Lehericy S, Samson Y, Obadia M, Slachevsky A, Dehaene S. Visual Word Recognition in the Left and Right Hemispheres: Anatomical and Functional Correlates of Peripheral Alexias. *Cerebral Cortex*. 2003; 13:1313–1333. [PubMed: 14615297]
69. Simon O, Kherif F, Flandin G, Poline J-B, Rivière D, Mangin J-Fo, Bihan DL, Dehaene S. Automatized clustering and functional geometry of human parietofrontal networks for language, space, and number. *NeuroImage*. 2004; 23:1192–1202. [PubMed: 15528119]
70. Simon O, Mangin J-Fo, Cohen L, Bihan DL, Dehaene S. Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*. 2002; 33:475–487. [PubMed: 11832233]
71. Cowell SF, Egan GF, Code C, Harasty J, Watson JDG. The functional neuroanatomy of simple calculation and number repetition: a parametric PET activation study. *NeuroImage*. 2000; 12:565–573. [PubMed: 11034863]
72. Ortuno F, Ojeda N, Arbizu J, López P, Martí-Clement JM, uelas IP, Cervera S. Sustained attention in a counting task: normal performance and functional neuroanatomy. *NeuroImage*. 2002; 17:411–420. [PubMed: 12482094]
73. Ghatan PH, Hsieh JC, Petersson KM, Stone-Elander S, Ingvar M. Coexistence of Attention-Based Facilitation and Inhibition in the Human Cortex. *NeuroImage*. 1998; 7:23–29. [PubMed: 9500831]
74. Chochon F, Cohen LG, Van De Moortele PF, Dehaene S. Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of Cognitive Neuroscience*. 1999; 6:617–630. [PubMed: 10601743]

75. Mazoyer B, Zago L, Mellet E, Bricogne S, Etard O, Houdé O, Crivello F, Joliot M, Petit L, Tzourio-Mazoyer N. Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Research Bulletin*. 2001; 54(3):287–298. [PubMed: 11287133]
76. Stanescu-Cosson R, Pinel P, Van De Moortele PF, LeBihan D, Cohen LG, Dehaene S. Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain*. 2000; 123:2240–2255. [PubMed: 11050024]
77. Hutchison ER, Blumstein SE, Myers EB. An event-related fMRI investigation of voice-onset time discrimination. *NeuroImage*. 2008; 40:342–352. [PubMed: 18248740]
78. Benoit MM, Raij T, Lin F-H, Jääskeläinen IP, Stufflebeam S. Primary and multisensory cortical activity is correlated with audiovisual percepts. *Human Brain Mapping*. 2010; 31:526–538. [PubMed: 19780040]
79. Jancke L, Shah NJ, Posse S, Grosse-Ryken M, Muller-Gartner HW. Intensity coding of auditory stimuli: An fMRI study. *Neuropsychologia*. 1998; 36:875–883. [PubMed: 9740361]
80. Riecker A, Wildgruber D, Dogil G, Grodd W, Ackermann H. Hemispheric lateralization effects of rhythm implementation during syllable repetitions: An fMRI study. *NeuroImage*. 2002; 16:169–176. [PubMed: 11969327]
81. Husain FT, Fromm SJ, Pursley RH, Hosey LA, Braun AR, Horwitz B. Neural Bases of Categorization of Simple Speech and Nonspeech Sounds. *Human Brain Mapping*. 2006; 27:636–651. [PubMed: 16281285]
82. Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. *Cereb Cortex*. 2005; 15:1621–1631. [PubMed: 15703256]
83. Laird AR, Lancaster JL, Fox P. BrainMap: The social evolution of a functional neuroimaging database. *Neuroinformatics*. 2005; 3:65–78. [PubMed: 15897617]
84. Fox P, Lancaster J. Mapping context and content: The BrainMap model. *Nature Rev Neurosci*. 2002; 3:319–321. [PubMed: 11967563]
85. Fox P, Laird A, Fox S, Fox M, Uecker A, Crank M, Koenig S, Lancaster J. BrainMap taxonomy of experimental design: Description and evaluation. *Human Brain Mapping*. 2005; 25:185–198. [PubMed: 15846810]
86. Kunchev LI, Rodríguez JJ. Classifier ensembles for fMRI data analysis: an experiment. *Magnetic Resonance Imaging*. 2010; 28:583–593. [PubMed: 20096528]
87. Ku, S-p; Grettton, A.; Macke, J.; Logothetis, NK. Comparison of pattern recognition methods in classifying high-resolution BOLD signals obtained at high magnetic field in monkeys. *Magnetic Resonance Imaging*. 2008; 26:1007–1014. [PubMed: 18691999]
88. Formisano E, Martino FD, Valente G. Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. *Magnetic Resonance Imaging*. 2008; 26:921–934. [PubMed: 18508219]
89. Björnsdóttir M, Rylander K, Wessberg J. A Monte Carlo method for locally multivariate brain mapping. *NeuroImage*. 2011; 56:508–516. [PubMed: 20674749]
90. Björnsdóttir Åberg M, Wessberg J. An evolutionary approach to the identification of informative voxel clusters for brain state discrimination. *IEEE Journal of Selected Topics in Signal Processing*. 2008; 2(6):919–928.

Fig. 1.
The structure of the aggregated classifier

Fig. 2.
Preprocessing and training pipeline for boosting and PLS aggregation

Fig. 3. Group average of correlation map of each of the four tasks task (top four) and voxels selected for pattern classification (bottom in white) at six selected slices ($Z=4, 8, 12, 16, 20, 24$ mm) in MNI space. The color bar on the top corresponds to correlation coefficients ranging from -1 to 1 .

Fig. 4. Misclassification rate as a function of the number of regions. Best performance with minimum number of regions was obtained for 30 regions.

Fig. 5. Individual performance of nonlinear SVM of 95 preselected regions with optimal trade-off factor C of the SVM primal objective function² and optimal kernel width d . Region 25 is Brodmann area 30 [R] and region 58 is Medial Dorsal Nucleus [R]. The names of the selected regions are listed in Table 2.

Fig. 6.

Comparison of (a) representative features from linear SVM analysis, (b) aggregation coefficients for boosting and (c) PLS regression for each task. Maps are displayed at six selected slices in MNI space ($Z = 4, 8, 12, 16, 20, 24$ mm). The color bar on the bottom indicates relative coefficient weights from minimum to maximum.

Fig. 7.

Dynamic predictions of the 4 interleaved tasks: (a) motor task, (b) visual task, (c) cognitive task and (d) auditory task, during a real-time fMRI experiment using boosting and PLS aggregation. For each task the selected slices of the activation maps that correspond to the slices in Fig. 3 are shown in the top row. The boosting output as a function of scan time is shown in the 2nd row. The time course of task execution (red) and the corresponding modeled reference vector (green) are shown in the 3rd row. The fourth row displays the PLS aggregation output as a function of scan time. Prediction results for both classifiers stabilize between scan numbers 20 and 30 (TR= 2s), which corresponds to the completion of the first set of activation blocks, except for the cognitive task.

Table 1

Number of activation maps acquired with different field strengths, spatial matrix sizes, read-out bandwidths (BW) using either low bandwidth (LB) or high bandwidth (HB), and EPI or multi-echo EPI (MEPI) for offline pattern classification.

Field	Matrix Size	BW	Echo	#maps
1.5T	32×32	LB	EPI	16
		HB	EPI	36
			MEPI	20
	64×64	LB	EPI	16
		HB	EPI	24
			MEPI	24
4T	32×32	LB	EPI	28
		HB	EPI	40
			MEPI	28
	64×64	HB	EPI	40
			MEPI	40

\$watermark-text

\$watermark-text

\$watermark-text

Table 2

Comparison of the activated brain regions detected at 1.5 and 4 Tesla and comparison with other studies. Regions with at least 10% of the voxels passing a correlation threshold 0.3 are shown. Z scores were calculated from correlation coefficients using the Fisher transformation. At a glance, bolded regions are common to both field strengths, and underlined regions indicate a selected region in either one of the assemble models (Table 3).

Region	4T (44 subjects)			1.5T (34 subjects)			All (78 subjects)		Region selection		References
	Max Avg Z	Voxel passed (%)	Max Avg Z	Max Avg Z	Voxel passed (%)	Max Avg Z	Max Avg Z	Voxel passed (%)	Boosting	PLS	
MOTOR	<u>BA42 R</u>	0.33	19	0.37	23	0.33	0.33	15	X	X	58
	<u>BA01 L</u>	0.62	66	0.41	28	0.51	0.51	58	---	---	---
	<u>BA02 L</u>	0.68	65	0.41	24	0.54	0.54	50	---	---	58, 59
	<u>BA03 L</u>	0.78	65	0.42	34	0.59	0.59	59	---	---	58, 60
	<u>BA04 L</u>	0.77	55	0.43	29	0.59	0.59	50	X	X	60, 61, 62
	<u>BA24 L</u>	0.40	11	---	---	---	---	---	X	X	63
	<u>BA40 L</u>	0.64	19	0.42	11	0.52	0.52	15	X	X	58, 64
	<u>BA41 L</u>	0.40	62	0.36	29	0.38	0.38	49	X	---	61, 62
	<u>BA42 L</u>	0.40	24	0.36	14	0.37	0.37	19	X	X	58,
VISUAL	<u>BA17 R</u>	0.59	56	0.80	67	0.61%	0.61%	61	X	X	65
	<u>BA18 R</u>	0.59	31	0.77	45	35%	35%	35	X	X	65, 66, 67
	<u>BA19 R</u>	---	---	0.63	24	---	---	---	---	---	68
	<u>BA17 L</u>	0.50	31	0.71	62	46%	46%	46	X	X	---
	<u>BA18 L</u>	0.52	19	0.73	37	26%	26%	26	X	X	---
	<u>BA19 L</u>	---	---	0.58	---	---	---	---	X	X	67, 68
COGNITIVE	<u>BA24 R</u>	0.42	10	---	---	---	---	---	---	---	---
	<u>BA32 R</u>	0.47	28	0.40	20	0.43	0.43	24	---	---	Ref. 69
	<u>BA33 R</u>	0.31	10	---	---	---	---	---	---	---	---
	<u>BA47 R</u>	0.39	14	0.42	11	0.40	0.40	13	---	---	---
	<u>L_GPR</u>	0.35	12	---	---	---	---	---	---	---	---
	<u>M_GPR</u>	0.35	26	---	---	0.33	0.33	10	---	---	---
	<u>BA06 L</u>	0.52	25	---	---	0.47	0.47	13	X	X	70, 71
	<u>BA24 L</u>	0.44	14	---	---	---	---	---	X	X	72, 73
	<u>BA32 L</u>	0.54	38	0.44	19	0.47	0.47	31	X	X	69, 74, 75

\$watermark-text

\$watermark-text

\$watermark-text

4T (44 subjects)			1.5T (34 subjects)			All (78 subjects)			Region selection		References
Region	Max Avg Z	Voxel passed (%)	Max Avg Z	Voxel passed (%)	Max Avg Z	Voxel passed (%)	Max Avg Z	Voxel passed (%)	Boosting	PLS	
BA44 L	0.52	59	0.45	32	0.49	45	0.49	45	---	X	74
BA45 L	0.49	32	0.45	25	0.46	23	0.46	23	---	---	69
BA46 L	---	---	0.42	21	---	---	---	---	---	---	75, 76
BA13 R	0.39	12	---	---	---	---	---	---	X	X	77, 78
BA21 R	0.59	60	0.68	57	0.62	59	0.62	59	X	X	78
BA22 R	0.51	29	0.65	49	0.56	35	0.56	35	---	---	---
BA32 R	0.45	27	---	---	0.35	12	0.35	12	---	---	78
BA38 R	0.39	17	---	---	---	---	---	---	---	---	---
BA41 R	---	---	0.45	32	---	---	---	---	---	---	79
BA42 R	0.42	12	0.53	41	0.46	21	0.46	21	X	X	---
BA44 R	0.38	17	---	---	---	---	---	---	X	---	---
BA45 R	0.38	16	---	---	---	---	---	---	---	---	---
BA47 R	0.41	27	---	---	0.39	19	0.39	19	---	---	---
BA06 L	0.55	13	---	---	---	---	---	---	X	X	78, 80, 81
BA21 L	0.58	53	0.64	49	0.58	51	0.58	51	X	X	---
BA22 L	0.49	28	0.59	55	0.52	37	0.52	37	X	X	82
BA24 L	0.49	22	---	---	---	---	---	---	X	---	---
BA32 L	0.52	35	---	---	0.44	18	0.44	18	X	X	78
BA38 L	---	---	0.38	11	---	---	---	---	---	---	78
BA41 L	---	---	0.46	53	0.38	19	0.38	19	X	---	79, 80
BA42 L	---	---	0.53	33	0.43	15	0.43	15	---	X	---
BA44 L	0.38	23	---	---	---	---	---	---	---	X	80
BA47 L	0.39	13	---	---	---	---	---	---	X	---	---

AUDITORY

Abbreviations: BA * L: Brodmann Area * in left hemisphere; BA * R: Brodmann Area * in right hemisphere; L GP R: Lateral Globus Pallidus in right hemisphere; M GP R: Medial Globus Pallidus in right hemisphere.

Table 3

Regions selected by boosting and PLS in the assemble model. BA: Brodmann area.

Boosting and PLS	Boosting only	PLS only
BA 5 [R], BA 11 [R], BA 13 [R], BA 17 [R], BA 18 [R], BA 20 [R], BA 21 [R], BA 37 [R], BA 40 [R], BA 42 [R], BA 4 [L], BA 6 [L], BA 7 [L], BA 8 [L], BA 13 [L], BA 17 [L], BA 18 [L], BA 19 [L], BA 21 [L], BA 22 [L], BA 24 [L], BA 32 [L], BA 40 [L], Caudate Tail [L], Putamen [L].	BA 44 [R], BA 9 [L], BA 39 [L], BA 41 [L], BA 47 [L]	BA 30 [R], Medial Dorsal Nucleus [R], BA 42 [L], BA 44 [L], Lateral Globus Pallidus [L]