Machine Learning for Clinical Diagnosis from Functional Magnetic Resonance Imaging

Lei Zhang, Dimitris Samaras
Department of Computer Science
SUNY at Stony Brook
Stony Brook, NY 11790
{lzhang,samaras}@cs.sunysb.edu

Dardo Tomasi, Nora Volkow, Rita Goldstein
Department of Medical Image
Brookhaven National Laboratory
Upton, NY
{tomasi,volkow,rgoldstein}@bnl.gov

Abstract

Functional Magnetic Resonance Imaging (fMRI) has enabled scientists to look into the active human brain. FMRI provides a sequence of 3D brain images with intensities representing brain activations. Standard techniques for fMRI analysis traditionally focused on finding the area of most significant brain activation for different sensations or activities. In this paper, we explore a new application of machine learning methods to a more challenging problem: classifying subjects into groups based on the observed 3D brain images when the subjects are performing the same task. Here we address the separation of drug-addicted subjects from healthy non-drug-using controls. In this paper, we explore a number of classification approaches. We introduce a novel algorithm that integrates side information into the use of boosting. Our algorithm clearly outperformed well-established classifiers as documented in extensive experimental results. This is the first time that machine learning techniques based on 3D brain images are applied to a clinical diagnosis that currently is only performed through patient self-report. Our tools can therefore provide information not addressed by traditional analysis methods and substantially improve diagnosis. [1]

1. Introduction

Functional Magnetic Resonance Imaging (fMRI) has enabled scientists to look into the active human brain. FMRI provides a sequence of 3D brain images with intensities representing blood oxygenation level dependent (BOLD) brain activations. This has revealed exciting insights into the spatial and temporal changes underlying a broad range of brain functions, such as how we see, feel, move, understand each other and lay down memories. In this paper, we explore a new application of machine learning methods to classify drug-addicted subjects from controls based on the observed 3D brain images. Drug addiction diagnosis is unique because it's not externally validated. By applying machine learning methods to the 3D brain images, we can find the hidden pattern differentiating the drug addicted subjects from healthy controls, thus perform classification for diagnosis. To our knowledge, this is the first time that machine learning techniques are applied to clinical diagnosis, which today is performed only through patient self-report.

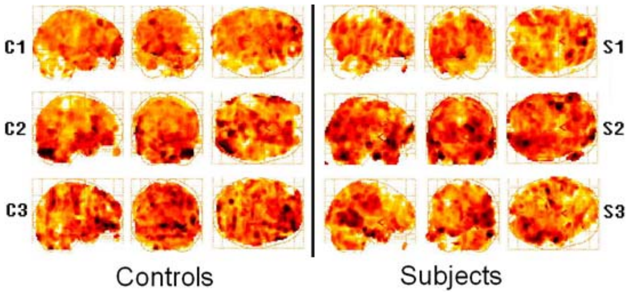

Figure 1. Can we find the hidden pattern in these 3D brain images to differentiate the drug addicted subjects from control normals? In the image, left columns show the brain images of controls and right columns show those of subjects. Each column shows three slides of the 3D images in different views.

The analyses and interpretation of fMRI data that are most commonly employed by cognitive-behavioral and emotional neuroscientists depend on the behavioral probes that are developed to tap regional brain functions. In this traditional neuroscience framework, the brain responses are a-priori labeled based on the putative underlying task condition (e.g., regions involved in reward vs. regions involved in punishment) and are then used to separate a priori defined

[1] We thank Steve Berry, B.A., for help with preliminary data analyses, L.A. Cottone, N Alia-Klein, A.C. Leskovjan, F. Telang, E.C. Caparelli, L. Chang, T. Ernst and N.K. Squires for helpful discussions; This study was supported by grants from the National Institute on Drug Abuse (to NDV: DA06891-06; and to RZG: 1K23 DA15517-01), Laboratory Directed Research and Development from U.S. Department of Energy (OBER), NARSAD Young Investigator Award, SB/BNL seed grant (79/1025459), National Institute on Alcohol Abuse and Alcoholism (AA/ODO9481-04), ONDCP, and General Clinical Research Center (5-MO1-RR-10710).

groups of subjects. Most such studies provide the results in the form "fMRI activity in brain region R is on average greater when performing task T than when in control condition C."[15] In this paper, we consider a different pattern recognition problem (Figure 1): training classifiers to automatically separate different groups of human subjects based on the observed 3D fMRI BOLD images. Solving this problem is essential because patterns of variability in brain states may be unique to a certain psychopathology and can be therefore used for improving diagnosis (e.g. diagnosis of drug addiction, relapse or craving). In addition, the development of this "*clinical machine learning framework*" can be applied to further our understanding of other human disorders and states such as those impacting insight and awareness, that similarly to drug addiction are currently identified based mostly on subjective criteria.

This classification problem is particularly challenging owing to the following factors: 1) undersized data space: limited data size due to the difficulties inherent in human subjects research; 2) oversized dimensionality of the fMRI BOLD data, e.g., in our experiment, the dimension of one 3D fMRI scan is about $53 \times 63 \times 46$ and one task contains 87 scans; 3) increased variability: inter-subject variability (i.e. different brain activation patterns are associated with different individuals) and intra-subject variability: even for the same person, the human brain activations are different from trial to trial even under the same experimental environment due to the brain complexity. 4) decreased group heterogeneity: because our goal is to separate healthy control subjects from individuals with subtle or preclinical brain changes, the more pronounced task vs. baseline activations cannot be used (the traditional fMRI analysis indicated that both drug addicted subjects and healthy controls had similar task related general brain activation patterns[8]).

In this paper, we contribute a comprehensive framework for the exploration of fMRI BOLD data sets for clinical diagnostic applications through the extensive and exhaustive comparison of three methodologies that have been successfully applied in other classification problems: i) PCA based dimensionality reduction and classification; ii) Voxel-based feature selection and classification and iii) AdaBoost. The first two methodologies differ in the feature selection step (indirect vs. direct selection). Once the features have been selected, we applied a number of classifier training methods (*Gaussian Naive Bayes (GNB)* [14], *Support Vector Machine (SVM)* [3], *k Nearest Neighbor (kNN)* [14]). Most of the above methods performed adequately well with AdaBoost being the best on data that was collected under identical conditions. However, when there was variability in the sequence of the stimuli, performance dropped significantly. One of the difficulties of our classification problem is that even for the same participant, the brain activations are different from trial to trial even under exactly the same experimental setting due to brain-behavior complexity. We propose a new boosting algorithm with side information[16] on subject identity to remove the intrasubject variability in order to improve classification. Our experiments show that the new algorithm allows for less restrictive data collection conditions with a significantly reduced performance penalty. This algorithm can work on combined data sets of different tasks effectively, tripling the amount of training data, which is significant given the labor intensiveness in data collection.

In Sec. 2, we discuss related work and in Sec. 3, we describe the acquisition and pre-processing of the fMRI BOLD data. In Sec. 4, we describe the exploration steps and the design of machine learning approaches. The experimental results and comparison of all three categories of methods is described in Sec. 5. Finally Sec. 6 presents the conclusions and future work directions.

2. Related Work

Earlier studies demonstrated that post-analysis is feasible on brain activation maps derived with Positron Emission Tomography (PET) data [13] where the PET scans of HIV positive patients were successfully separated from healthy controls. Recently [5], fMRI contrast images and significance maps were cpmpared for patient classification using a Fisher linear discriminant (FLD) classifier to differentiate patients from controls accurately for Alzheimer's disease, schizophrenia, and mild traumatic brain injury. For these types of psychopathologies, there commonly are other validation methods that aid in diagnosis (e.g. marked neuropsychological deterioration over time or from a prodromal baseline). In drug addiction the cognitive deficits are not as markedly pronounced [8] and frequently they go unrecognized; their attribution to "non-cognitive" factors (e.g., dysthymia during withdrawal, lack of motivation) further complicates identification and prompt delivery of adequate interventions. Indeed, in contrast to other neuropsychiatric disorders, drug addiction is only now being recognized as a disorder of the brain. The relatively moderate level of cognitive deficits in addiction and the difficulty in diagnosing addiction as a separate entity led us to apply more sophisticated computer learning algorithms since methods similar to [5] proved to be inadequate for our learning task.

In recent work [19][15], Mitchell et al. have demonstrated the feasibility of training classifiers which automatically decoded the subject's cognitive state (e.g., looking at a picture or reading a sentence). More specifically, they trained both single subject and cross subjects classifiers that distinguished among a set of predefined cognitive states, based on a single fMRI image or a sequence of fMRI images of activations to the presentation of a particular stimulus. Thus, based on the cognitive states decoded from the brain data, they separated the stimuli that activated distinct regions of the brain. However, our goal in the current study was to separate drug-addicted subjects from controls, while using the same stimuli for both groups. Hence, our data set included activations in the same brain regions in response to the same cognitive-behavior paradigm in all subjects, thus complicating the classification task as described above.

3. Acquisition of fMRI data

Functional Magnetic Resonance Imaging: functional MRI [11][2] is based on the increase in blood flow to the local vasculature that accompanies neural activity in the brain. Using an appropriate imaging sequence, human cortical functions can be observed without the use of exogenous contrast agents.

In our experiments, the data were collected to study the neuropsychological problem of loss of sensitivity to the saliency of money in cocaine users[8]. The MRI studies were performed on the 4T Varian scanner at Brookhaven National Laboratory and all the stimuli were presented using LCD-goggles connected to a PC. The human participants pressed a button or not based on a picture shown to them. They received a monetary reward if they performed correctly. Specifically, three runs were repeated twice (T1, T2, T3; and T1R, T2R, T3R) and in each run, there were three monetary conditions (high money, low money, no money) and a baseline condition where a fixation cross was shown on the screen; the order of monetary conditions was pseudo-randomized and identical for all participants. Participants were informed about the monetary condition by a 3-sec instruction slide, which visually presented the stimuli: or $0.45, $0.01 or $0.00. The feedback for correct responses in each condition consisted of the respective numeral designating the amount of money the subject has earned if correct. The symbol (X) followed incorrect trials in all conditions. To simulate real-life motivational salience, subjects could gain up to $50 depending on their performance on this task. In our experiments, drug addicted subjects were 16 cocaine dependent individuals, 18-45 years of age, in good health, matched with 13 non-drug-using controls on sex, race, education and general intellectual functioning.

Statistical Parametric Mapping (SPM)[7] was used for fMRI data preprocessing (realignment, normalization/registration and smoothing) and statistical analyses. SPM refers to the construction and assessment of spatially statistical processes that are used to test hypotheses about [neuro] imaging data from SPECT/PET and fMRI. The time series were analyzed independently at each normalized, resampled voxel ($3 \times 3 \times 3$ mm) using regression analysis and creating 3D contrast maps for pairs of conditions. Contrast values are estimates of the difference in activation between two different conditions: a positive contrast value for a voxel is interpreted as an increase in brain activation for the first condition compared to the second, while a negative value is often assumed to reflect a decrease [5][1][10]. In our work, we applied a t-test to determine the probability that the means of the two groups with Gaussian distributions are significantly different between the task conditions as determined by thresholding the activation values. Thus, we created a data set of six contrast maps (CM) for each subject for each run ($45 > Baseline$, $1 > Baseline$, $0 > Baseline$, $45 > 0$, $45 > 1$ and $1 > 0$). Figure 1 shows examples of the created 3D contrast maps. These specific contrasts were created based on previous observations that drug addiction has at its core a deficit in the processing of relative reward [8]; the activation differences between the monetary condition pairs were therefore assumed crucial to our classification problem.

4. Machine Learning for Diagnosis

In this section, we will describe our exploration of machine learning methods for classification of drug-addicted subjects from controls. We aim to approximate the classification function:

$$f : \langle fMRI \quad data \rangle \rightarrow [DrugAddicted|Control] \qquad (1)$$

The format of this function is similar to the classification functions estimated successfully in [19][15]. We first performed similar learning experiments: we selected features (*Average, ActiveAvg(n) and Active(n)*) and explored a number of classifier training methods (*Gaussian Naive Bayes (GNB)* [14], *Support Vector Machine (SVM)* [3], *k-th Nearest Neighbor (kNN)* [14]) on the preprocessed fMRI sequences (Please see [19] for more on the feature selection and learning). In our experiments, the problem of data registration for multiple subjects has been solved by using SPM for data preprocessing (normalization step). We found that all these learning methods resulted in poor classification rates. This result could have been attributed to the similar fMRI BOLD activation patterns for both subject groups as previously described [8]. Thus, it was not possible to achieve acceptable rates of classification by simply using the general task related brain activations across all monetary conditions. Guided by a prior hypothesis and previous results (i.e. the loss of sensitivity to relative saliency of money in cocaine users, [8]), we therefore performed classifications on the activation differences between monetary conditions pairs. For this purpose, we use the contrast map data set created by SPM as described in Section 3.

4.1. Diagnosis with Standard Learning Methods

We group the classifiers that were trained using different feature selection methods in three categories: i) PCA based dimensionality reduction and classification; ii) Voxel-based feature selection and classification and iii) Adaboost:

4.1.1 PCA-Based Dimensionality Reduction and Classification

Principal Components Analysis (PCA) is a standard method for creating uncorrelated variables by fitting linear combinations of the variables to the raw data and selecting the best fits. PCA is also a standard method for dimensionality reduction that eliminates redundancies in the data and reduces the number of dimensions needed to model the available data. Intriguingly, a PCA+Fisher's Linear Discriminant [9] classification method has been reported in [5] to classify patients from controls accurately for Alzheimer's disease, schizophrenia and mild traumatic brain injury. In our experiments, we first performed dimensionality reduction by using PCA. We then applied a number of other learning methods *KNN*, *GNB*, and *SVM* in addition to using *FLD*.

4.1.2 Voxel-Based feature selection and classification

Several feature selection methods have been successfully used in [19] to perform classification analyses. We performed the experiments using two such methods, using the computed contrast maps as input feature vectors instead of the raw fMRI scans used in [19].

• *ActiveROI(n):* We divided the whole brain into 8 Region of Interest(ROI), and for each ROI, we selected the n most active voxels.

• *Active(n):* We selected the n most active voxels over the entire brain.

Again, we considered the following learning methods: *KNN*, *GNB* and *SVM* for classification.

4.1.3 Adaboost

Many classification problems have been successfully addressed by Boosting [18][4]. A variant of Adaboost [6] has been used successfully both to select the features and to train the classifier in a face detection system [18]. Boosting produces a strong classifier by computing the weights with which to combine a number of weak classifiers. In our experiments, the weak learning algorithm is designed to select the single voxel that best separates the positive and negative examples. For each voxel, the weak learner determines the optimal threshold classification function, such that a minimum weighted error rate is acquired. A weak classifier $h(x, f, p, \theta)$ thus consists of a feature (f), a threshold (θ) and a polarity (p) indicating the direction of the inequality:

$$h\left(x, f, p, \theta\right) = \begin{cases} 1 & if\ pf(x) < p\theta \\ 0 & otherwise \end{cases} \tag{2}$$

Here x is one voxel of the contrast map.

4.2. Boosting with Side Information

As we described in the Section 1, one of the difficulties of our learning problem is that brain activations are different from trial to trial even for the same person under exactly the same experimental settings, due to complex brain behaviors. Previous work [12] has shown that when only a small number of data are available, feature selection is essential to achieve accurate rates of classification. In the current group classification study, the desired features should depend on inter-subject (and not in intra-subject) brain activation. Shashua et al [16] have shown the use of side information in the context of a hard feature selection problem. Traditionally, the notion of side information is to provide auxiliary data in the form of an additional dataset containing only the feature space that is irrelevant to the classification task and thus undesirable. Stated differently, using side information allows for the feature selection process to select only those features that enhance the relevant dimensions in the main dataset while inhibiting the irrelevant dimensions in the auxiliary dataset. Here we propose a novel boosting algorithm enhanced by side information to remove the intra-subject variations. This is essential in our study because our goal is to classify subjects into two groups, hence our desired features should perform classification consistently for training data of the same subject. In this paper, side information is integrated into the boosting algorithm by adjusting the weak classifier selection and weight updating steps.

The weak classifier $h(x, f, p, \theta)$ is the same as in Eq. 2. Table 1 shows the details of the learning algorithm. In our algorithm, we keep the same weight w_i^j for all data instances of the same participant i. In the weak classifier selection step, we use a set of parameters ρ to enhance the inter-subject variability. The basic idea in selecting ρ is that given training data that are weighted equally, we prefer to select those features that miss a smaller number of training data and a smaller number of participants. For example, assume two features A and B; feature A misclassifies n pieces of data for one subject and feature B misclassifies $n/2$ pieces of data for 2 subjects (for a total of n misses as well). In this case, we prefer feature A whose performance is more consistent w.r.t. each subject. More formally, since for each participant, we have 6 pieces of training data for each monetary reward under the same task, we propose to select the set of $\rho_m, 0 <= m <= 6$ according to the following three rules:

$$\begin{aligned} &A : \rho_a \times a < \rho_b \times b \quad if \quad a < b \\ &B : \rho_{a+b} \times (a+b) < \rho_a \times a + \rho_b \times b \\ &C : if \quad a+b = a' + b' \quad and \quad a^2 + b^2 > a'^2 + b'^2 \ , \\ &\quad \rho_{a+b} \times (a+b) < \rho_{a'+b'} \times (a' + b') \end{aligned} \tag{3}$$

Rule A ensures that for the subset of each subject, the features that miss a smaller number of training data will out-

```
Boosting with Side Information
* Given $n$ example training data of $K$ participants $(x_1^1, y_1^1), ..., (x_1^{z_1}, y_1^{z_1}), (x_2^1, y_2^1), ..., (x_2^{z_2}, y_2^{z_2}), ......., (x_K^1, y_K^1), ..., (x_K^{z_K}, y_K^{z_K})$ where $z_1, z_2, ..., z_K$ are the number of training data of each participant with $\sum_{i=1}^{K} z_i = n$ and $y_i^j = 0, 1$ for negative and positive examples respectively. (in the following, $i = 1, ..., K$ and $j = 1, ..., z_K$ )
* Initialize weights $w_{1,i}^j = \frac{1}{2m}, \frac{1}{2l}$ for $y_i^j = 0, 1$ respectively, where $m$ and $l$ are the number of negative and positive examples respectively and $m + l = n$.
* For $t = 1, ..., T$:
    1. Normalize the weights, $w_{t,i}^j \leftarrow \frac{w_{t,i}^j}{\sum_{i,j} w_{t,i}^j}$
    2. Select the best weak classifier $h(x, f, p, \theta)$ with respect to the weighted error:
       $$\varepsilon_t = \min_{f,p,\theta} \sum_{i=1}^{K} \rho_{m_i} \varepsilon_t^i$$
       where:
       $m_i = \sum_{j=1}^{z_j} \left| h(x_i^j, f, p, \theta) - y_i^j \right|$ is the number of misclassified instances for subject $i$,
       $\rho_{m_i}$ is a pre-computed parameter and $\varepsilon_t^i = \sum_{j=1}^{z_j} w_{t,i}^j \left| h(x_i^j, f, p, \theta) - y_i^j \right|$
    3. Define $h_t(x) = h(x, f_t, p_t, \theta_t)$ where $f_t$,$p_t$ and $\theta_t$ are the minimizers of $\epsilon_t$.
    4. Update the weights: $w_{t+1,i}^{j=1..zi} = w_{t,i}^j \beta_t^{1-e_i}$ where $e_i = \frac{\sum_{j=1}^{z_i} \left| h_t(x_i^j) - y_i^j \right|}{z_j}$, and $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$.
* The final strong classifier is:
  $$C\,(x) = \left\{ \begin{array}{ll} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{array} \right.$$
  where $\alpha_t = \log \frac{1}{\beta_t}$.
```

Table 1. The boosting algorithm with side information for feature selection and training of the classifier. The final output is a weighted linear combination of the T weak classifiers where the weights are inversely propotional to the training error

put a smaller error number. Rule B ensures that if two features misclassify the same number of training data, we prefer the feature whose misses are in the subset of one subject. Rule C implies that we prefer the features whose misses are not evenly distributed, e.g. we prefer the feature that misses 1 piece of training data for subject i and 5 pieces of data for subject j to the feature that misses 3 pieces for each of these two subjects. Our weak classifier selection reduces to the standard Adaboost if all ρ_m are set to 1.

In our experiments, we used exponential functions to compute the set of ρ_m: let $\rho_m = (1/m)^{2/k}$, where k is a constant to be computed according to Eq. 3. In our case, for 0 $<= m <=$ 6, $k = 20$ satisfies all three rules.

5. Experiments and Results

After we trained the classifiers as described above, we evaluated these classifiers using a "leave-one-out" cross validation procedure. Each of the K human subjects was used as a test subject and each fMRI contrast map of each subject was used as a test input while training on the contrast maps of the remaining $K - 1$ subjects, and the mean accuracy over these held out subjects was then calculated. In the following section, we will report the experimental results and comparison of these learning methods.

In our experiments, there are totally 6 runs: T1, T2, T3 and T1R, T2R, T3R. For each run, we created 6 contrast maps as described in Section 3. Due to the head motion, for some participants, data of some task have too much displacement to be used. The first set of experiments was conducted to test the notion that it is difficult to classify drug addicted subjects from healthy controls by just looking at brain activation under each monetary condition individually. This will also explain why the methods proposed in [19] cannot be applied directly to our learning problem. Table 2 shows the classification results on single monetary contrast maps and verifies previous observations.

In the following, we will focus on the classification based on the brain activation differences between pairs of mone-

	$45 > B$				$1 > B$				$0 > B$			
	T1+1R	T2+2R	T3+3R	ALL	T1+1R	T2+2R	T3+3R	ALL	T1+1R	T2+2R	T3+3R	ALL
GNB	55.8%	56.6%	55.1%	52%	55.8%	54.7%	55.1%	51.3%	57.7%	56.5%	53.1%	50.8%
5NN	59.2%	58.5%	55.1%	52.6%	59.2%	60.4%	57.1%	52.6%	57.7%	58.5%	55.1%	52%

Table 2. Classification results on the data of individual monetary condition and the results validate our observation that it is very hard to get good classification by simply looking at the brain activation under each monetary condition individually since both drug addicted group and control group have the similar brain activation patterns

	$45 > 1$				$45 > 0$				$1 > 0$			
	T1+1R	T2+2R	T3+3R	ALL	T1+1R	T2+2R	T3+3R	ALL	T1+1R	T2+2R	T3+3R	ALL
GNB	84.6%	79.2%	81.6%	71.4%	76.9%	75.4%	72.7%	67.5%	76.9%	77.3%	75.5%	69.5%
SVM	82.7%	79.2%	79.6%	70.8%	75%	73.6%	73.5%	67.5%	73.1%	73.6%	73.5%	66.9%
FLD	82.7%	81.1%	81.6%	71.4%	78.8%	75.4%	75.5%	69.5%	78.8%	77.3%	73.5%	68.8%
5NN	88.5%	86.8%	85.7%	74.7%	82.7%	81.1%	81.6%	71.4%	82.7%	83.0%	79.6%	70.1%

Table 3. Classification results of PCA based methods. 5NN performs the best among these methods. The classification based on $45 > 1$ contrast maps give the best experimental results which validates the traditional observation that relative reward processing is at the core of the brain deficit in drug addiction.

tary conditions. Table 3 shows the classification results of PCA based methods, where we found that 5NN performed the best among these learning methods. The classification based on $45 > 1$ contrast maps provided the best experimental results. These results validated the previous observation using traditional fMRI analysis that relative reward processing is at the core of the brain deficit in drug addiction [8]. In the following experiments, we therefore focused on the experiments using $45 > 1$ contrast maps. Table 4 shows the classification results of voxel-based methods. Table 5 shows the classification results of AdaBoost with different numbers of selected features. Among all those learning methods, AdaBoost performed best. Our experimental results indicate that we can successfully classify drug addicted-subjects from healthy controls by using the data of each run separately. However, for datasets containing all the contrast maps, classification performance dropped due to the intra-subject variability. Table 6 compares the classification results of our novel boosting algorithm with side information to the results of standard Adaboost. From Table 6, we see that boosting with side information outperformed standard AdaBoost on the data set containing the contrast maps from different runs.

	ActiveROI(40)			
	T1+1R	T2+2R	T3+3R	ALL
GNB	80.8%	81.1%	81.6%	72.7%
SVM	78.8%	79.2%	79.6%	70.8%
5NN	84.6%	81.1%	81.6%	74%
	Active(40)			
	T1+1R	T2+2R	T3+3R	ALL
GNB	78.8%	81.1%	79.6%	72.1%
SVM	80.8%	75.5%	77.6%	69.5%
5NN	82.7%	79.2%	79.5%	73.4%

Table 4. Classification results of voxel-based methods on $45 > 1$ data. We see these methods have similar classification performance as PCA-based methods.

	T1+1R	T2+2R	T3+3R	ALL
100	90.4%	88.7%	89.8%	80.5%
200	92.3%	90.6%	91.8%	82.4%

Table 5. Classification results of Adaboost on $45 > 1$ data by selecting different number of features and we found AdaBoost outperforms the methods reported in Table 3 and 4.

In Table 6, it is interesting to note that the classification on the ”T1+T2+T3” data is not as good as the classification on the ”T1R+T2R+T3R” data. From a neuroscience point of view, we employ a task that evokes motivation and the effects we are measuring are susceptible to subject habituation. We expected the habituation effect (over the 6 repetitions) to decrease the intensity of brain activation to the task, especially for the drug-addicted subjects. The results from Table 6 imply that drug addicted subjects may indeed have different habituation speeds than healthy controls. Results further suggest that task repetitions offer better rates of classification between controls and subjects with psychopathology than the initial task runs.

Our experiments show that by applying machine learning methods to fMRI brain data, we can separate drug-addicted subjects from the normal controls successfully. Such classification provides both theoretical and clinical benefits. From a theoretical point of view, we show that the experimental results validate related neuropsychological theories from an alternate view of the brain data:
1) we observed that we cannot separate the drug addicted subjects from the controls by simply looking at the data activation of each monetary condition individually, while we can classify the two groups accurately based on the brain difference of pairs of monetary conditions.
2) We also found that the classification results of $45 > 1$ contrast maps are better than $45 > 0$ and $1 > 0$ data, vali-

	T1+T2+T3	T1R+T2R+T3R	ALL
Adaboost 100	81.6%	85.9%	80.5%
Adaboost 200	82.9%	87.2%	82.4%
Boost-SI 100	85.5%	87.2%	85.7%
Boost-SI 200	86.8%	89.7%	87.7%

Table 6. The comparison of Adaboost and Boost-SI methods on the mixture data set of $45 > 1$ contrast maps. Boost-SI improves the classification performance on the data sets containing contrast maps from different runs.

dating the previous observation that a core of the deficit in drug addiction pertains to relative award processing.
3) the classification on the mixture data set ALL was not as good as on each run separately due to the intra-subject variability and by using boosting algorithm with side information, we improved the classification.
4) the dataset ”T1R+T2R+T3R” was easier to be classified than the dataset ”T1+T2+T3” which implies that drug addicted subjects may have different habituation speeds from the controls.

From a clinical point of view, the trained classifiers can be used for clinical drug addiction diagnosis. Finally, our results call for further exploration of applying similar machine learning methods to other situations where the diagnosis can only be made using patient self-report (e.g. emotion identification) or diagnosis can only be made using patient self-report (e.g. emotion identification) or diagnosis with states and disorders of insufficient development of insight and awareness (e.g. children, anger and aggression).

6. Conclusions and Future Work

We have shown that we can successfully separate the drug addicted subjects from controls by using the 3D brain images obtained with fMRI BOLD, despite the difficulties pertaining to the subtlety of neuro-cognitive deficits in drug addiction and activation variability. This new application provides an alternate view of brain data and validates related neuropsychological theories. Our exploration of applying machine learning methods to 3D brain images allows diagnosis based on derived data, in cases that today are diagnosed only through self-report and thus can be extended to other applications.

Feature selection is the key for pattern recognition problems. We were able to extend similar feature selection and classification methods [20][17] successfully applied in 2D visual images to 3D brain images. After further validation with other data sets (additional subjects with addiction or other psychopathology), we will explore combining temporal and spatial information to find better features. We will thus explore the dynamic nature of the interactive brain regions; our analyses to date focused on static activations (i.e. at a certain time during the task) while neural networks interact in a dynamic way. This would allow the demarcation of the causal relationships between different regions within the functioning human brain.

References

[1] G. Aguirre, E. Zarahn, and M. D’Esposito. The inferential impact of global signal covariates in functional neuroimaging analyses. *NeuroImage*, 8(3):302–306, Oct. 1998.

[2] P.A. Bandettini, A. Jesmanowicz, E.C. Wong, and J.S. Hyde. Processing strategies for time-course data sets in functional mri of the human brain. *Magn Reson Med*, (30), 1993.

[3] Burges C. A tutorial on support vector machines for pattern recognition. *Journal of Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[4] M. Collins. Ranking algorithms for named-entity extraction: Boosting and the voted perceptron. In *ACL*, 2002.

[5] J. Ford, H. Farid, F. Makedon, L.A. Flashman, T.W. McAllister, V. Megalooikonomou, and A.J. Saykin. Patient classification of fmri activation maps. In *MICCAI03*, 2003.

[6] Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Computational Learning Theory: Eurocolt 95*, 1995.

[7] K. Friston, A. Holmes, K. Worsley, and et al. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, pages 2:189–210, 1995.

[8] R.Z. Goldstein et al. A modified role for the orbitofrontal cortex in attribution of salience to monetary reward in cocaine addiction: an fmri study at 4t. In *Human Brain Mapping Conference*, 2004.

[9] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Predictions*. Springer, 2001.

[10] M. Hutchinson, W. Schiffer, S. Joseffer, and et al. Task-specific deactivation patterns in functional magnetic resonance imaging. *Mag Res Imag*, 17(10):1427–36, Dec 1999.

[11] K.K. Kwonget et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci USA*, (89):5675–5679, 1992.

[12] K Levi and Y. Weiss. Learning object detection from a small number of examples: The importance of good features. In *CVPR*, 2004.

[13] J. Liow, K. Rehm, and S. Strother. Comparison of voxel- and volume-of-interest-based analyses in fdg pet scans of hiv positive and healthy individuals. *J. Nucl Med*, 41(4):612–621, April 2000.

[14] T.M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[15] T.M. Mitchell, R. Hutchinson, R. Niculescu, F. Pereira, X. Wang, M. Just, and S. Newman. Learning to decode cognitive states from brain images. *Machine Learning: Special Issue on Data Mining Lessons Learned*, 2003.

[16] A. Shushua and L. Wolf. Kernel feature selection with side data using a spectral approach. In *ECCV*, 2004.

[17] R. Vidal, Y. Ma, and J. Piazzi. A new gpca algorithm for clustering subspaces by fitting, differentiating and dividing polynomials. In *CVPR*, pages 510–517, 2004.

[18] P. Viola and M.J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 2004.

[19] X. Wang, R. Hutchinson, and T.M. Mitchell. Training fmri classifiers to detect cognitive states across multiple human subjects. In *NIPS03*, Dec 2003.

[20] Y. Wu and A. Zhang. Feature selection for classifying high-dimensional numerical data. In *CVPR*, 2004.