11-15-2005

Classifying spatial patterns of brain activity with machine learning methods: application to lie detection

Christos Davatzikos
University of Pennsylvania, christos@rad.upenn.edu

Kosha Ruparel
University of Pennsylvania, ruparel@bbl.med.upenn.edu

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

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

M. Acharyya
University of Pennsylvania

See next page for additional authors

Recommended Citation

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D., Acharyya, M., Loughead, J., Gur, R., & Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. Retrieved from https://repository.upenn.edu/neuroethics_pubs/10

Postprint version. Published in *NeuroImage*, Volume 28, Issue 3, 2005, pages 663-668.
Publisher URL: http://dx.doi.org/10.1016/j.neuroimage.2005.08.009

Classifying spatial patterns of brain activity with machine learning methods: application to lie detection

Abstract

Patterns of brain activity during deception have recently been characterized with fMRI on the multi-subject average group level. The clinical value of fMRI in lie detection will be determined by the ability to detect deception in individual subjects, rather than group averages. High-dimensional non-linear pattern classification methods applied to functional magnetic resonance (fMRI) images were used to discriminate between the spatial patterns of brain activity associated with lie and truth. In 22 participants performing a forced-choice deception task, 99% of the true and false responses were discriminated correctly. Predictive accuracy, assessed by cross-validation in participants not included in training, was 88%. The results demonstrate the potential of non-linear machine learning techniques in lie detection and other possible clinical applications of fMRI in individual subjects, and indicate that accurate clinical tests could be based on measurements of brain function with fMRI.

Keywords

FUNCTIONAL MAGNETIC-RESONANCE, DECEPTION, FMRI, CLASSIFICATION, RESPONSES, CORTEX

Disciplines

Other Medicine and Health Sciences

Comments

Postprint version. Published in *NeuroImage*, Volume 28, Issue 3, 2005, pages 663-668.
Publisher URL: http://dx.doi.org/10.1016/j.neuroimage.2005.08.009

Author(s)

Christos Davatzikos, Kosha Ruparel, Yong Fan, Dinggang Shen, M. Acharyya, James Loughead, Ruben Gur, and Daniel D. Langleben

Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection

C. Davatzikos[a], K. Ruparel[b], Y. Fan[a], D.G. Shen[a], M. Acharyya[a], J.W. Loughead[b], R.C. Gur[b] and D.D. Langleben[b, c]

[a]Department of Radiology, University of Pennsylvania, 3600 Market Street, Suite 380, Philadelphia, PA 19104, USA
[b]Department of Psychiatry, University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104, USA
[c]Treatment Research Center, University of Pennsylvania, 3900 Chestnut Street, Philadelphia, PA 19104, USA

Received 1 March 2005; revised 15 June 2005; accepted 4 August 2005. Available online 5 October 2005.

Introduction

A large body of functional neuroimaging literature has elucidated relationships between structure and function, as well as functional activity patterns during a variety of functional activation paradigms. Statistical parametric mapping (SPM) (Friston et al., 1995) has played a fundamental role in these studies, by departing from the conventional biased ROI- and hypothesis-based methods of data analysis and enabling unbiased voxel-by-voxel examination of all brain regions. While a great deal of knowledge has been gained during the past decade regarding brain regions that are activated during various tasks using voxel-based SPM analysis, the quantitative characterization of entire spatio-temporal patterns of brain activity, as opposed to voxel by voxel examination, has received much less attention, especially as a means for deducing “the state of the mind” from functional imaging data. The important distinction between a voxel-based analysis and the analysis of a spatio-temporal pattern is the same as the distinction between (mass) uni-variate and multi-variate analysis (Davatzikos, 2004). Specifically, a pattern of brain activity is not only a collection of active voxels, but carries with it correlations among different voxels. Notable efforts towards the functional activity pattern analysis have been made (Strother et al., 1995 and McIntosh et al., 1996), some of which, attempt to use these methods to classify complex activation patterns using machine learning methods (Cox and Savoy, 2003 and LaConte et al., 2005).

In this paper, we present an approach to the problem of identifying patterns of functional activity, by using a high-dimensional non-linear pattern classification method. We apply this approach to one of the long-standing challenges in applied psychophysiology, namely lie detection. Deception is a socially and legally important behavior. The limitations of the specificity of the currently available physiological methods of lie detection prompted the exploration of alternative methods based on the correlates of the central nervous system activity, such as EEG and fMRI (Rosenfeld, 2001, Spence et al., 2001 and Langleben et al., 2002). Using SPM-based analyses of multi-subject average group data, several recent fMRI studies demonstrated differences in brain activation between truthful and non-truthful responses in various experimental paradigms (Langleben et al., 2002, Langleben et al., in press, Ganis et al., 2003, Kozel et al., 2004a, Kozel et al., 2004b and Lee et al., 2002). In order to translate these data into a clinically relevant application, discrimination between lie and truth has to be achieved at the level of single participants and single trials (Kozel et al., 2004b), not just via group analysis. The potential of the SPM-based approach to achieve this goal is limited due to the between-subject variability of regional brain activity. In the current work, we have overcome this limitation using a multi-variate non-linear high-dimensionality pattern classification technique (Lao et al., 2004) applied to spatial patterns of brain activation recorded via fMRI. Using data acquired with a previously reported formal deception paradigm (Langleben et al., in press), we have tested the hypothesis that truthful and non-truthful responses could be distinguished via analysis of the pattern of functional brain activity. In addition to classifying a pattern of functional activity to the “truthful” and “non-truthful” categories, we construct spatial maps that display the brain regions that are most distinctive between these two response categories, by following the gradient direction of the decision function of the classifier, thereby allowing for visualization of the aspects of the functional activity patterns that are most different between the two conditions (lying and truth-telling).

Methods

Participants and deception task

Twenty-two right-handed male undergraduate students (M age = 19.36, SD = 0.5). After complete description of the study to the participants, written informed consent was obtained. The experimental procedure has been described in detail elsewhere (Langleben et al., in press).

A pseudorandom sequence of playing cards was composed of five stimulus classes: (1) Lie (5 of clubs or 7 of spades, participant's choice); (2) Truth (5 of clubs or 7 of spades); (3) Recurrent Distracter (2 hearts); (4) Varied distracter (remaining cards 2 through 10, all suits); and (5) Null (back of a card). Stimuli were presented for 2 s followed by a variable ISI (0–16 s) during which the null condition was displayed. Twenty-four Lie, Truth, and recurrent distracter cards and 168 variant distracter cards were shown. Participants were instructed to press a left button (yes) to confirm that a card was in their possession and the right button (no) to deny it. A total of 432 stimuli were presented with a total session length of 14.4 min.

Instruction protocol

Participants were presented an envelope containing 2 cards (5 clubs and 7 spades) and $20. One investigator (DDL) instructed participants to deny possession of one of the cards and acknowledge possession of the other in the imaging phase of the study. Participants were warned that the “fMRI investigator”, naïve to the true intent of the study, would give contradictory instructions. The assignment of “Lie” and “Truth” to the cards contained in the envelope, was left to the participants. Participants were told by DDL that they could keep the $20 only if successful in concealing identity of the lie card during the scan session. Participants were escorted to the scanner by a third party and instructed by the “fMRI investigator” (JWL) to respond to each trial as accurately and truthfully as possible. After the session, participants were debriefed, informed that the compensation was not contingent on performance, queried about the lie card, and asked if they employed any strategy during the task.

Image acquisition

Images were acquired using BOLD imaging (Bandettini et al., 1992) on a clinical 3 T Siemens Trio Scanner (Iselin, NJ). A 5 min magnetization-prepared, rapid acquisition gradient echo image (MPRAGE) was acquired for anatomic overlays of functional data and spatial normalization. BOLD imaging used a 33-slice whole-brain, single-shot gradient-echo (GE) echo-planar (EPI) sequence (TR/TE = 2000/21 ms, FOV = 240 mm, matrix = 64 × 64, slice thickness/gap = 4/0 mm). This sequence delivered a nominal voxel resolution of 3.75 × 3.75 × 4 mm.

Data analysis—parameter estimate images (PEIs)

Functional data were processed and analyzed using statistical parametric mapping (SPM2, Wellcome Department of Cognitive Neurology, London, UK). Data were slice-time corrected and motion-corrected to median image using b-spline interpolation (4 degrees of freedom). Co-registered images were normalized to standard anatomical space (T1 MNI template) using tri-linear interpolation (Ashburner and Friston, 1999) and smoothed with a Gaussian filter (full-width half maximum = 8 mm, isotropic). A high pass frequency filter (to a maximum of 1/100 Hz) and correction for autocorrelation {AR(1)} between scans were applied subsequently to the data. Initial statistical analysis was performed using the general linear model. The expected hemodynamic response was modeled by convolving stimuli onsets with a double gamma function as implemented in SPM2 (Friston et al., 1998). Forty-eight regressors modeled “lie” and “truth” events individually while two additional regressors modeled the variant distracter and recurrent distracter conditions. Parameter Estimate Images (PEIs), i.e. regression coefficients or “beta”, of the HRF regressor for each of the 50 conditions, were calculated from the least mean square fit of the model to the time series. The PEIs were used for the second stage analysis of non-linear pattern classification.

Non-linear pattern classification

To quantify the spatial patterns of brain activity that distinguish truthful from non-truthful responses, we first sampled brain activity uniformly throughout the entire brain, in a way that was unbiased by any a priori hypothesis of specific regions of interest. We then used these brain activity samples in conjunction with a pattern classification method to find patterns that allow us to distinguish between truthful and non-truthful responses. Our methodology is detailed next.

The bounding box of the linearly-registered PEIs was evenly subdivided into 560 cubes, each having a size of 16 mm × 16 mm × 16 mm. The average value of the PEI of each event was calculated within each of these boxes, thereby resulting in 560 measurements from each scan distributed uniformly throughout the entire brain. This collection of measurements constituted a sample of the spatial activation pattern. Customary to the machine learning terminology, we call these 560 values features. These features were subsequently normalized, so that they ranged from 0 to 1. Specifically, all PEIs of each individual, including truthful and non-truthful responses, were pooled together. The maximum and minimum values were determined and used to linearly rescale all PEIs of the respective individual between 0 and 1. This was repeated for all individuals separately. Feature normalization is an important procedure, since we are interested in examining relative levels of activation, i.e. activation patterns. Moreover, this scaling process is often used to reduce the total variance of the measurement across individuals, and thereby help make classes more separable.

These normalized measurements were then used as features into a non-linear support vector machine (SVM) with a Gaussian kernel (Lao et al., 2004, Vapnik, 1999 and Scholkopf and Smola, 2001). SVM is a powerful classification method that finds the hypersurface (high-dimensional analog to a surface separating two classes) that maximizes the margin between two distributions, the truthful and non-truthful responses in our case. One of the most important characteristics of SVM is that it is not calculated from all samples (activation patterns), but only from samples that lie close to the interface between the two groups of interest. In our case, the algorithm focuses only on activation patterns that are difficult to classify into truthful or non-truthful responses, since other patterns are easy to classify, anyway. A second important characteristic of SVM is that it can determine non-linear separating boundaries (hypersurfaces), by mapping the 560-dimensional feature vector into an infinite-dimensional space. Accordingly, complex and non-linear decision boundaries can be formed that separate the two conditions.

Using the aforementioned procedures, classifiers were constructed and tested using cross-validation. In each repetition of the cross-validation procedure, 1% of the PEIs were left out, and a classifier was constructed from the remaining 99% of the PEIs. This classifier was then tested on the left out 1%. This procedure was repeated 30 times, for randomly selected sets of left out PEIs. The cross-validation procedure is important, since SVM is a high-dimensional non-linear classifier (560 features) that was applied to a relatively small sample (24 “lie”, 24 “truth” for 22 participants, yielding 1056 training PEIs), suggesting that it is flexible enough to separate these two conditions. Therefore, how well a trained SVM “explains” the data may not be indicative of how well the same classifier will separate new sets of data.

Although the training and testing procedures provide us with quantitative measures of the ability of the classifier to separate the two sets of responses, they do not directly provide us with an intuitive way to understand which brain structures are most informative in terms of the classification. In Lao et al. (2004), we described a procedure for constructing such visual displays. Briefly, after the classifier has been trained, a hypersurface that lies in-between the set of truthful responses and the set of non-truthful responses has been determined; this hypersurface generates a maximum margin between these two classes of response. Using the training samples that lie on one side of this hypersurface, one can follow the gradient of the decision function from +1 to −1, thereby capturing the most important differences between the truthful and non-truthful conditions. We applied this procedure for all support vectors, i.e. for all example patterns that after training ended up being close to the dividing hypersurface, and recorded the boxes whose values changed the most along the path from truthful to non-truthful response. By overlaying these boxes on the brain images themselves, we obtained 3D displays of the regions whose activation was important in the classification process.

The procedure described above may have some practical limitations since, correct baseline data from an individual being tested have to be collected and used to construct a classifier (these images would be part of the 99% of the sample), before the actual test data can be collected and analyzed. Thus, we also used the following procedure to determine the potential of building a single classifier that could be applied to a number of individuals taking the same test or performing the same experimental deception task. During each repetition of the classification experiment, we left out *all images of one individual*, instead of 1% of all images, then trained a classifier on images of all other individuals and tested on the left out individual. In order to reduce the effects of various types of noise, we formed averages of all PEIs of truthful and non-truthful responses for all individuals, thereby ending up with 44 images in total, 2 for each individual. We then investigated the generalization performance of the classifier by training it on the 42 images of the 21 individuals, then testing it on the 2 images of the left out individual, then repeating this procedure 22 times, each time leaving the PEIs of a different individual out.

Results

Behavioral

A significant difference in accuracy [$F(3,19) = 9.62$, $P < 0.05$] and response time [$F(3,19) = 31.44$, $P < 0.05$] was observed among the four stimuli classes. The response time to Varied Distracter, Repeat Distracter, Lie, and Truth were 634 (SD 17), 654 (20), 717 (24), and 806 (27) ms, respectively, and the % of correct answers was 96 (SD 0.79), 98 (0.26), 97 (0.94), and 93 (1.25), respectively.

Non-linear pattern analysis

As described in Methods, we first examined the individual PEIs. A support vector machine achieved 99.3% separation of the truth/lie conditions, when trained on the PEIs. In the testing stage, during which 30 repetitions of training on 99% and testing on 1% of the PEIs were applied, we obtained classification accuracy equal to 87.9% (90% sensitivity, 85.8% specificity). Fig. 1 shows the training and testing accuracy of the classifier, in these experiments (only a randomly selected representative set of samples is shown, for clarity of the figure).

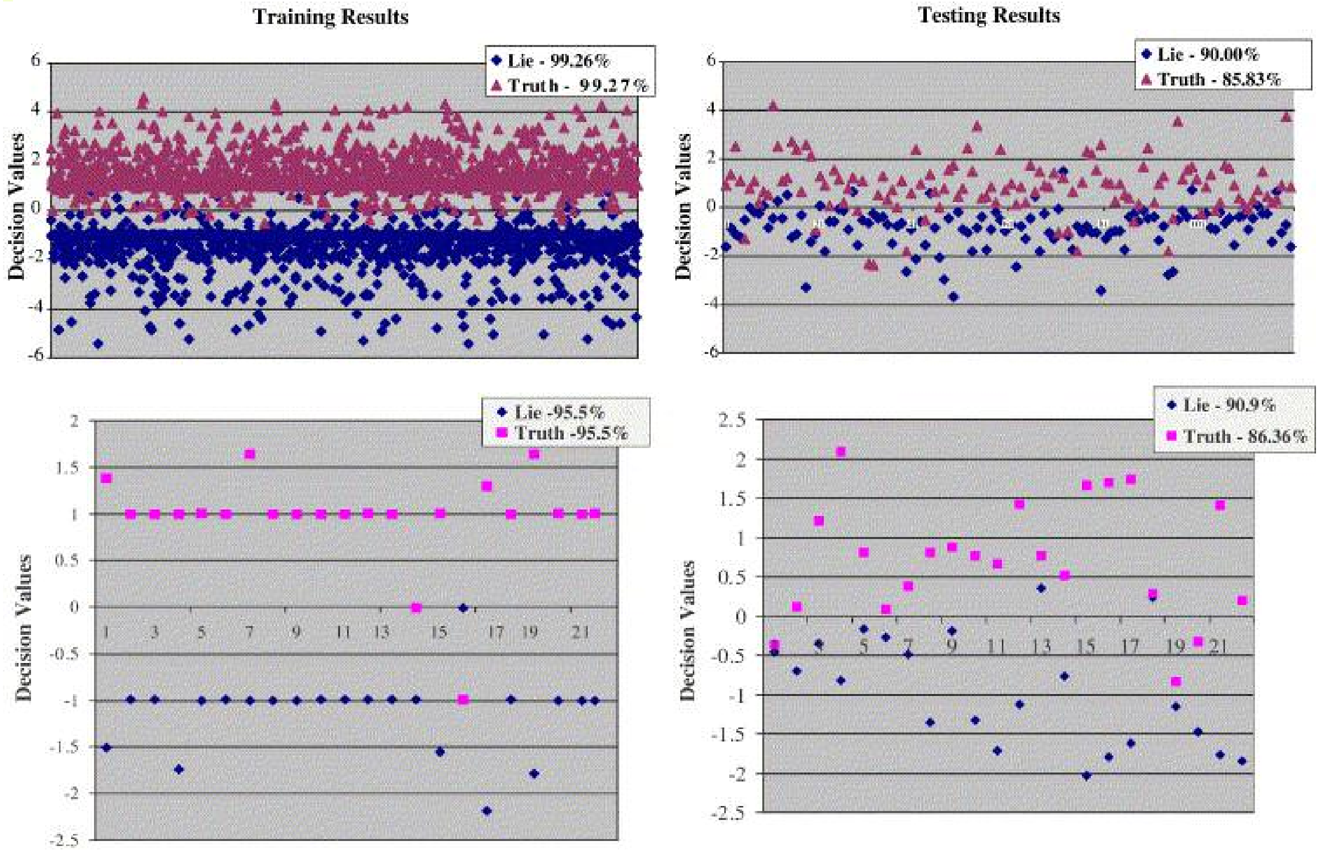

Fig. 1. Plots of the value of the "decision function", which shows whether or not the spatial activation pattern is representative of truth-telling or lying (1 is classified as truth and −1 as lie). Top row is for individual PEIs and bottom row for average PEIs. Training accuracy (i.e. how well the model "explained" the data) was above 99% and 95% (left). Testing accuracy (which reflects the predictive value of the classification scheme) was obtained via cross-validation, i.e. by testing the trained classifiers on previously unseen data sets and repeating for many times (right): sensitivity was 90% and 90.9%, for the two types of experiments, and specificity was 85.83% and 86.36%. The vertical axis shows the value of the decision function, and the horizontal axis shows different subjects (subject number).

For comparison purposes, we applied a linear classification method (Fisher's Linear Discriminant Analysis, FLDA) to the same data, testing the efficacy of linear separators. We obtained classification accuracy equal to 65.6%, which is notably lower than the SVM result. Moreover, we tested voxel-wise *t* statistic maps, which are customary in spm-type of analysis. Since statistical parametric mapping is not meant for classification, but rather for generating voxel-by-voxel statistical parametric maps, we tried to create a comparison that is as fair as possible to the two methods. In particular, we selected the box that displayed the most significant *P* value (highest *t* value), based on *t* tests analysis of the PEIs, and trained a one-variable SVM classifier based on it. The resulting classification rate was 63.1%. This most discriminative box is shown in Fig. 2. These two results strongly indicate that not only there are relationships in the activity patterns of different regions that must be considered in classification, but also that these relationships are non-linear.

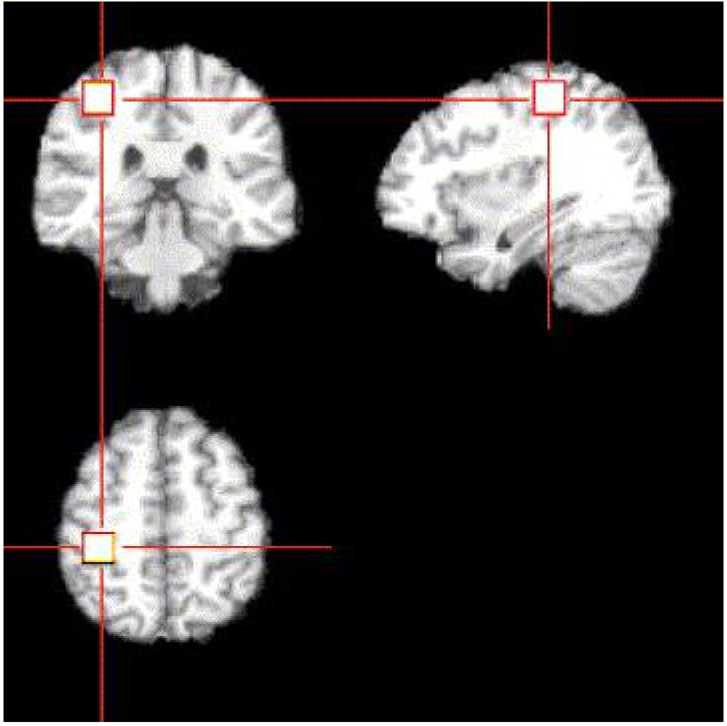

Fig. 2. The area with the lowest *P* value obtained by the *t* test applied to the PEIs of the 512 cubic subregions in which the brain was divided. This area was used for classification, for comparison purposes.

In order to display the regions that played important role in the classification, as described in Methods, we constructed paths from truthful to non-truthful responses, by following the gradient of the decision function. Fig. 3 (right) shows an overlay of the most significant boxes on a 3D rendering of the template brain. We note that some of these boxes are deeper in the brain, hence they appear fuzzed out. In the same figure, on the left, we show regions which had relatively higher (lower) PEI values in truthful responses (green) and in non-truthful responses (red).

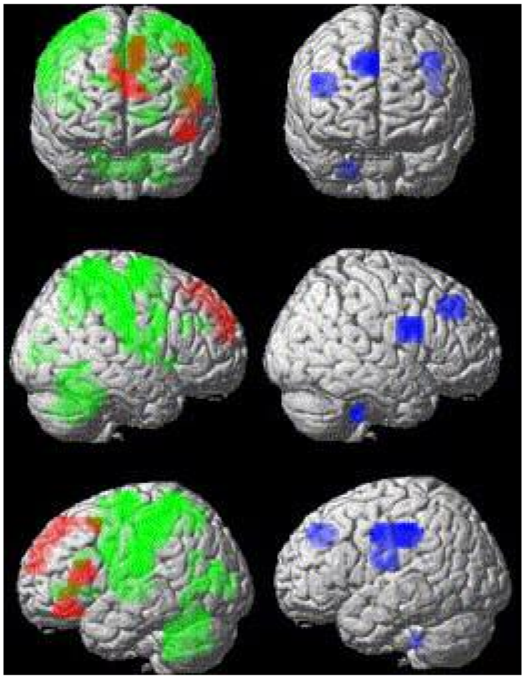

Fig. 3. Left: Different 3D views of regions or relatively higher activity during truth-telling (green) and lying (red). Right: Areas in which functional activity was found to be most informative in terms of pattern classification, in the sense that activities in primarily those regions were jointly used to classify an activity pattern. Some of the blue regions are relatively deeper and therefore appear fuzzed out in this rendition.

In the second set of analyses, described in Methods, we tested the classification method on average PEIs, one for each subject and separately formed for truthful and non-truthful responses. The classification results were nearly the same as in the previous sets of experiments. A 95.5% separation accuracy was obtained in this case. Cross-validation using leave-one-subject-out yielded an average classification rate of 88.6% (90.9% sensitivity, 86.4% specificity), with the most discriminant features (boxes) being almost identical to the ones in Fig. 2.

Discussion

We presented an approach to classification of patterns of brain activity obtained during lying and truth-telling, during a forced choice deception experiment (LaConte et al., 2005, Lykken, 1991, OTA, 1983 and Rosenfeld et al., 1988). The main premise of this paper has been that a high-dimensional non-linear classification analysis can reveal spatial patterns of activity that distinguish lying, from telling the truth. As detailed in our discussion in Lao et al. (2004), although voxel-based analysis has been very useful in identifying regions activated in various cognitive activation paradigms, it could be of limited value when used to detect complex and spatially distributed image patterns, since correlations among different brain regions might be highly distinctive from the brain activity pattern (and they are, in the current study, as the experimental results showed). This is particularly important in lie detection, since as described in the Introduction, voxel-by-voxel group differences do not allow us to reliably classify patterns of brain activity during truth-telling and lying, due to across-subject variability. Our hypothesis herein was that examining the entire brain pattern in a high-dimensional space formed by all brain regions jointly would help better separate truthful from non-truthful responses and potentially yield a methodology of value in applied lie detection. The high classification accuracy of our approach suggests that, with a more extensive training of the classifier, SVM analysis of fMRI data could surpass the accuracy of the polygraph, which is estimated to be between 65 and 95% (Stern, 2004).

We used two complementary approaches to classification, based on the same data. In the first one, we attempted to classify individual responses, having trained on a large number of images. In the second approach, we averaged all responses of the same individual (one average for truth-telling and one for lying), and therefore we trained the classifier on a relatively small number of data, but data that presumably had less noise and variability across individuals. There is an obvious trade-off between these approaches, hence they gave very similar results. However, we believe that the second approach will ultimately prove far more powerful, if more training samples become available. This is because noise in the data can adversely affect high-dimensional classifiers, a limitation that signal averaging should be expected to alleviate to a large extent, without sacrificing separation flexibility, provided that a large enough set of training samples is available. The second approach is also more practical, because it does not require that the classifier be first trained on the person under study, before it can be applied for testing purposes.

We would like to emphasize the importance of the cross-validation experiments we performed, since this is an issue that is often neglected in the literature. In particular, statistical models are often used to "explain the data", which in our case amounts to separating the two conditions. However, a model that explains one set of data well does not necessarily generalize well to other sets of data, even if those are drawn from the same distribution. Accordingly, we used well-established cross-validation methods for estimating the generalization ability of a classifier (Hastie et al., 2001). However, it should be stressed that performance of the same classifier on new data that are not drawn from the same distribution could be significantly worse, since cross-validation guards only against overtraining, but not against training on a data set that is not fully representative of a statistical distribution. There are many reasons why new data could yield significantly different classification performance, including significant differences in scanner characteristics, acquisition protocols, and of course study design. Therefore, the main conclusion from this study is not that a universal classifier can be built that detects truthful from non-truthful responses, but rather that sophisticated classification methods can be trained on specific experimental set-ups to identify certain patterns of brain activity.

The regions that were most informative in terms of classification (Fig. 2) were the right prefrontal regions (inferior and superior frontal gyri, Brodmann Area [BA] 6 and 44), and the bilateral posterior cortex (superior temporal and inferior parietal gyri, BA 38, 40).

Additional regions included bilateral pericentral (BA 2,3 and 4) areas and foci in the right cerebellum and the striatum. Most of the prefrontral regions have been previously implicated in a variety of executive function tasks, including response inhibition and error monitoring, including deception experiments (Langleben et al., in press and Kozel et al., 2004b). The temporoparietal parietal cortex (BA 38 and 40), has been associated with average increased activity during deception in most fMRI studies of deception reported so far (Spence et al., 2004). Remarkably, this somatosensory association area and the primary somatosensory cortex (BA 2,3) have a role in the skin conductance response (Critchley et al., 2000), providing a possible link between central and peripheral physiological markers of deception (Bechara et al., 2000).

Feature normalization is an important aspect of the classification process, since it effectively changes the characteristics of the underlying probability distributions, in a way that hopefully reduces intra-class variation and therefore amplifies inter-class separation. In Methods, we presented one possible way of feature normalization. We also experimented with a variety of other approaches, such as normalizing each feature separately, i.e. by normalizing the activity within each box so that it represents a *z* score across subjects, as well as normalization by the baseline images estimated by the SPM analysis used to determine the PEIs. We also used the baseline activity measurements as separate features. However, all of these methods gave inferior results.

Feature selection is also a key issue in pattern classification. In these experiments, we did not use any feature selection, but rather sampled the entire brain region at the resolution of each cube. This approach has both advantages and disadvantages over a variety of alternatives (Chambers and Hastie, 1991 and Clark and Pregibon, 1993). Its main advantage is that it avoids potential biases that might be introduced by feature selection, and which would necessitate cross-validation of the feature selection process itself, along with the classification. That procedure would be extremely computationally expensive. The main disadvantage of our feature selection method is that it does not attempt to identify in advance the features (brain regions) that are most informative, and therefore might include features that are just noise and tend to deteriorate the performance of the classifier. This bolsters our confidence that classification accuracy can potentially be improved using more sophisticated feature selection methods, such as methods based on wavelets and other hierarchical image representation schemes (Lao et al., 2004).

In conclusion, although spatial patterns of brain activation have been previously shown to present average differences between lying and truth-telling, the specificity of the average pattern to individual events of lie and truth has not been demonstrated and thus the value of these data for clinical lie detection applications was uncertain (Kozel et al., 2004b). Our findings demonstrate that a high-dimensionality non-linear pattern classification method is capable of accurately detecting subtle, spatially distributed, and complex patterns of brain activity associated with lying, thus bridging the gap between average group data and the practical lie detection in individuals. Under the conditions of the present deception paradigm (GKT2), the separation ability was almost 100%. Predictive power was also very high, as determined from responses of individuals that were not part of the SVM training procedure. These cross-validation results are particularly promising since training a classifier on an individual to be tested may be impossible in practical lie detection. Since the number of participants was relatively limited in these experiments, we anticipate that performance will improve significantly with more extensive training. Moreover, since the classification method is not specific to lie detection, it could ultimately be used to a very broad range of applications in which the state of mind is to be inferred from spatio-temporal patterns of brain activity. More sophisticated feature selection approaches (Lao et al., 2004) could further help improve the performance of the classifier by selecting only the brain regions that are most distinctive between the two conditions. Finally, the same approach could potentially be used with other measures of brain activity, such as EEG.

References

Ashburner and Friston, 1999 J. Ashburner and K.J. Friston, Nonlinear spatial normalization using basis functions, *Hum. Brain Mapp.* **7** (1999) (4), pp. 254–266.

Bandettini et al., 1992 P.A. Bandettini *et al.*, Time course EPI of human brain function during task activation, *Magn. Reson. Med.* **25** (1992) (2), pp. 390–397.

Bechara et al., 2000 A. Bechara, H. Damasio and A.R. Damasio, Emotion, decision making and the orbitofrontal cortex, *Cereb. Cortex* **10** (2000) (3), pp. 295–307.

Chambers and Hastie, 1991 J.M. Chambers and T.J. Hastie, Statistical Models. CRC, Chapman and Hall, New York (1991).

Clark and Pregibon, 1993 L.A. Clark and D. Pregibon, Tree-based models. In: J.M. Chambers and T.J. Hastie, Editors, *Statistical Models*, S. Chapman and Hall, New York, NY (1993), pp. 377–419.

Cox and Savoy, 2003 D.D. Cox and R.L. Savoy, Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex, *NeuroImage* **19** (2003) (2), pp. 261–270.

Critchley et al., 2000 H.D. Critchley *et al.*, Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study, *J. Neurosci.* **20** (2000) (8), pp. 3033–3040.

Davatzikos, 2004 C. Davatzikos, Why voxel-based morphometric analysis should be used with great caution when characterizing group differences, *NeuroImage* **23** (2004), pp. 17–20.

Friston et al., 1995 K.J. Friston *et al.*, Statistical parametric maps in functional imaging: a general linear approach, *Hum. Brain Mapp.* **2** (1995), pp. 189–210.

Friston et al., 1998 K.J. Friston *et al.*, Nonlinear event-related responses in fMRI, *Magn. Reson. Med.* **39** (1998) (1), pp. 41–52.

Ganis et al., 2003 G. Ganis *et al.*, Neural correlates of different types of deception: an fMRI investigation, *Cereb. Cortex* **13** (2003), pp. 830–836.

Hastie et al., 2001 T. Hastie, R. Tibshirani and J. Friedman, The elements of statistical learning: data mining, inference and prediction, *Springer Series in Statistics*, Springer-Verlag, New York (2001).

Kozel et al., 2004a F.A. Kozel *et al.*, A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men, *J. Neuropsychiatry Clin. Neurosci.* **16** (2004), pp. 295–305.

Kozel et al., 2004b F.A. Kozel, T.M. Padgett and M.S. George, A replication study of the neural correlates of deception, *Behav. Neurosci.* **118** (2004) (4), pp. 852–856.

LaConte et al., 2005 S. LaConte, S. Strother, V. Cherkassky, J. Anderson and X. Hu, Support vector machines for temporal classification of block design fMRI data, *NeuroImage* **26** (2005), pp. 317–329.

Langleben et al., 2002 D.D. Langleben, L. Schroeder, J.A. Maldjian, R.C. Gur, S. McDonald, J.D. Ragland, C.P. O'Brien and A.R. Childress, Brain activity during simulated deception: an event-related functional magnetic resonance study, *NeuroImage* **15** (2002), pp. 727–732.

Langleben et al., in press Langleben, D.D., Loughead, J.W.L., Bilker, W.B., Ruparel, K., Childress, A.R., Busch, S., Gur, R.C., in press. Telling truth from lie in individual subjects with fast event-related fMRI. Hum. Brain Mapp.

Lao et al., 2004 Z. Lao *et al.*, Morphological classification of brains via high-dimensional shape transformations and machine learning methods, *NeuroImage* **21** (2004) (1), pp. 46–57.

Lee et al., 2002 T.M.C. Lee *et al.*, Lie detection by functional magnetic resonance imaging, *Hum. Brain Mapp.* **15** (2002) (3), pp. 157–164.

Lykken, 1991 D.T. Lykken, Why (some) Americans believe in the lie detector while others believe in the guilty knowledge test, *Integr. Physiol. Behav. Sci.* **26** (1991) (3), pp. 214–222.

McIntosh et al., 1996 A.R. McIntosh *et al.*, Spatial pattern analysis of functional brain images using partial least squares, *NeuroImage* **3** (1996) (3), pp. 143–157.

OTA, 1983 OTA, Scientific Validity of Polygraph Testing: A Research Review and Evaluation—A Technical Memorandum, US Congress: Office of Technology Assessment, Washington, DC (1983).

Rosenfeld, 2001 J.P. Rosenfeld, Event-related potentials in detection of deception. In: M. Kleiner, Editor, *Handbook of Polygraphy*, Academic Press, New York (2001).

Rosenfeld et al., 1988 J.P. Rosenfeld *et al.*, A modified, event-related potential-based guilty knowledge test, *Int. J. Neurosci.* **42** (1988) (1–2), pp. 157–161.

Scholkopf and Smola, 2001 B. Scholkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond (Adaptive Computation and Machine Learning)* (1st ed.), The MIT Press (2001 (December 15)), p. 644.

Spence et al., 2001 S.A. Spence *et al.*, Behavioural and functional anatomical correlates of deception in humans, *NeuroReport* **12** (2001) (13), pp. 2849–2853.

Spence et al., 2004 S.A. Spence *et al.*, *A Cognitive Neurobiological Account of Deception: Evidence from Functional Neuroimaging* **vol. 359**, Royal Society (2004), pp. 1755–1762.

Stern, 2004 P.C. Stern, The polygraph and lie detection. In: T.N.R. Council, Editor, *Report of The National Research Council Committee to Review the Scientific Evidence on the Polygraph*, The National Academies Press, Washington, DC (2004), pp. 340–357.

Strother et al., 1995 S.C. Strother, I. Kanno and D.A. Rottenberg, Principal component analysis, variance partitioning, and functional connectivity, *J. Cereb. Flow Metab.* **15** (1995) (3), pp. 355–360.

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