Learning to Classify Psychiatric Disorders based on fMR Images: Autism vs Healthy and ADHD vs Healthy

Sina Ghiassian[1], Russell Greiner[1], Ping Jin[1], and Matthew R. G. Brown[2]

[1] Department of Computing Science, University of Alberta
{ghiassia,rgreiner,pjin1}@ualberta.ca
[2] Department of Psychiatry, University of Alberta
mbrown2@ualberta.ca

Abstract. A clinical tool that can diagnose psychiatric illness using functional magnetic resonance brain images would greatly assist physicians. Here, we propose a learning algorithm that uses the histogram of oriented gradients (HOG) features of fMRI images as the input to support vector machines, then show that this system can produce such classifiers when run on two large public datasets: able to diagnose ADHD with hold-out accuracy of 0.626 (over baseline = 0.550) when trained on the ADHD-200 global competition dataset, and to diagnose autism with hold-out accuracy of 0.619 (over baseline = 0.516) when trained on the Autism Brain Imaging Data Exchange (ABIDE) dataset. While these results are not yet to clinical relevance, they outperform all previously presented methods on both datasets. These results suggest that our learning approach may lead to diagnostic classifiers (from functional images) for yet other psychiatric disorders.

Keywords: FMRI, Machine Learning, Classification, ADHD, Autism, HOG Features

1 Introduction

Functional magnetic resonance imaging (fMRI) is a neuroimaging technique that uses strong magnetic fields from standard MRI scanners to produce images of biological tissues, which can be used to investigate brain changes over a few minutes [1]. While the person is inside the scanner, the instrument will produce a complete scan of the brain every 1 to 3 second [1]. This produces, in effect, a waveform signal for each voxel of the brain.

FMRI can be used to analyze the brain in different ways. Many researchers compare groups of people and/or compare brain states within the same individuals. For example, Wolf *et al.* applied independent component analysis (ICA) to a dataset of 12 healthy and 12 ADHD adults during a working memory task and observed that ADHD patients had significantly less activation in the left ventrolateral prefrontal cortex (VLPFC), cerebellar and occipital regions compared with healthy controls [2]. (However, they did not try to build a diagnostic system.) Such "association studies" aggregate over individuals in different

groups or over brain states repeatedly sampled within given individuals. This is in contrast to machine learning classification studies that attempt to diagnose individual people.

FMRI is most commonly based on the blood oxygenation level dependant (BOLD) signal. In task-based fMRI, the subject is asked to perform a task while in the scanner. By contrast, in resting-state fMRI (RS-fMRI), the subject is asked to lie in the magnet and rest quietly. RS-fMRI is often used to investigate the functional architecture of the brain [3]. Methods used to analyze resting state data include seed-based approaches, independent component analysis, graph methods, clustering algorithms, neural networks, and pattern classifiers [3].

In the ADHD-200 global competition [4], many teams tried to learn a classifier that could determine if a subject in the dataset is healthy or has ADHD, using RS-fMRI data. The results show that it is possible to classify the ADHD disorder using only functional images, with the caveat that it is not ready for clinical purposes since the accuracy is far from perfect [5]. Other researchers have tried to learn to classify autism, using the ABIDE dataset [6,7]. We describe their results below.

Our work makes 3 contributions: *a*) We show that histogram of oriented gradient (HOG) feature descriptors can be useful for diagnosing psychiatric disorders. *b*) Our method outperforms any previously published classification results on two large resting state fMRI datasets. *c*) For these datasets, we found that a learner that uses time-averaged fMRI signal descriptors performs better than a learner that uses $T1$-weighted images.

Below, section 2 will describe our learning method and section 3 will present our experimental results on those two datasets.

2 Methodology

The ADHD-200 global competition dataset originally included 776 resting state fMRI and anatomical scans from 8 different imaging sites, 491 of which were obtained from "controls" (typically developing subjects) while 285 were cases, with ADHD. We also included the ADHD-200 global competition test data, which originally included 197 instances from both types of healthy and ADHD cases. We had to remove 33 subjects (6 had no resting state scan, 1 could not be preprocessed using SPM8 [8] and 26 from the Brown site had no labels), leaving us with 94 healthy subjects and 77 ADHD cases in the test set and 490 healthy controls and 279 ADHD cases in the training set. The ADHD-200 dataset also included other non-imaging data including gender, age etc; see [4]. The ABIDE dataset originally included 1112 subjects, but we had to remove one (as it could not be preprocessed using SPM8) leaving us with a dataset of size 1111 individuals, including 573 healthy controls and 538 autistic cases. The ABIDE dataset also provided an extensive array of phenotypic information; see [6].

Due to issues like the movement of subjects in the scanner, it is necessary to preprocess fMRI data before analysis. Researchers use different preprocessing steps to reduce the variability of data that is not associated with the experimental task, and to prepare the data for statistical testing [1]. For preprocessing we have used statistical parametric mapping (SPM8) [8] software and our in-house MATLAB code. Our preprocessing included 6 steps: *a*) 6-parameter rigid body motion correction to remove the head motions in the magnet, *b*) coregistration of functional scans to subject-specific anatomical scans to guide the normalization step, *c*) non-linear spatial normalization to match all of the images to the MNI T1 template [9–11], *d*) spatial smoothing with $8mm$ FWHM [1] Gaussian kernel to increase the signal-to-noise ratio, *e*) temporal preprocessing to make all the images equal in terms of their volume times (time to scan the whole brain once) and temporal duration and *f*) z-normalizing each image to make the intensities of images from different sites comparable. For more information about the preprocessing steps, see Huettel *et al.* [1].

The fMRIs are $4D$ images. We first reduced the number of dimensions to 3 by averaging over time for each voxel of the image. This produced what we call a $3D$ functional MR image for each individual. We used the mean value because the values of the waveform signal for the voxels appear to follow a Gaussian distribution. (We also repeated our experiments using the median values of the voxels over time and found similar results.)

We then computed $3D$ HOG (histogram of oriented gradients) features of the images. This, in our case, describes each subject with $116,480$ features. HOG is a powerful feature descriptor for visual object recognition [12] that computes spatial gradient information among pixels. To be specific, HOG divides the whole image into blocks and each block into several cells. In each cell, it utilizes the normalized local histograms of gradient orientations as a new feature. It then normalizes each cell within the same block. A thorough explanation of different kinds of normalizations including gamma/colour normalization and different block normalization schemes can be found in [12]. The HOG method relies on the assumption that local histogram of gradient information can characterize local objects or shapes very well, even when discarding the position information of corresponding gradients [12]. HOG has been successfully applied to $2D$ images for different kinds of tasks related to object recognition [12].

Figure 1 shows an example where 8 orientation bins are evenly spaced over 0° - 360° [12]. Fig 3[1] shows how the HOG features (right) of a bike (left) – showing the major direction of each region: eg, horizontal at the top and bottom of the wheel and vertical on the left and right sides (in green), and horizontal along the crossbar (in blue). For $3D$ images, such as magnetic resonance images, first we divide the whole image into $3D$ blocks instead of $2D$ ones, and then we divide these $3D$ blocks into smaller $3D$ cells. Then we generalize the $2D$ orientation bins into the $3D$ space using 26 orientation bins; (see Fig. 2).

Using the HOG features as input, we learned binary classifiers to diagnose either ADHD versus control (using ADHD-200 data) or autism versus control (us-

[1] image from www.mathworks.com

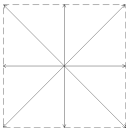

Fig. 1: $2D$ bins

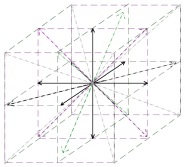

Fig. 2: $3D$ bins

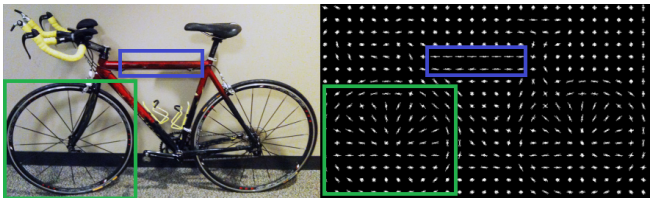

Fig. 3: Input and output of HOG

ing ABIDE data). We considered several base classifiers including Naive Bayes, K-Nearest Neighbours with different number of neighbours and RBF SVM with different sigma values. As each dataset had only about 1000 individuals, there was a high chance the learning algorithms will overfit to the training data if we used all of the features. So we used MRMR (maximum relevance minimum redundancy) [13,14] to select the most relevant features. For notation, let acc(L, D, FS) be the 5 values computed by using 5-fold cross-validation when running the learning L on the dataset D, using only the feature set FS; and let Eacc(L,D,FS) be the mean of these 5 values. Also let FS_k(D) be the top k MRMR features over the dataset D. For each learner L, on dataset D, our L2CPD system (Algorithm 1) sequentially considered using the top k $= \{1, 11, 21, ...\}$ MRMR features, stopping when the Eacc(L,D, FS_k) decreased or reached a plateau. We then identified each learner L with the mean of the accuracies for the best FS, and the range of these accuracy values. We identified each base learner with both the mean accuracy achieved using the best feature set, and also the range of accuracy values. We found the three base learners with the top 3 mean accuracy values; these appear in Table 1. (Note all 3, for each dataset, were SVM with RBF kernels. They differ only by the Sigma values.) Since the top 3 mean accuracies are very close, L2CPD chose the learner from these 3 with the smallest range of accuracy values and returned that learner. In the ADHD-200 dataset, for learning algorithms with top accuracy values, we saw accuracy decrease when the number of features went over 211 and in the ABIDE dataset we almost reached a plateau when we went over 101 features.

Algorithm 1 L2CPD algorithm

```
1: procedure L2CPD(D: training data, 𝓛: set of base_learners)
2:     for each base_learner L in 𝓛 do
3:         FS ← top 1 MRMR feature (on D)
4:         bsf_m ← 0 // best-so-far mean
5:         bsf_r ← 0
6:         while True do
7:             vals[L; 1 : 5] ← acc( L, D, FS ) // 5-fold c-v accuracy of L on D, using FS
8:             mean ← ave {vals[i]}
9:             range ← max_i{vals[i]} − min_j{vals[j]} // range of accuracy over 5 folds
10:            AveAcc[L, |FS|] ← [mean, range]
11:            if mean ≥ bsf_m then
12:                bsf_m ← mean
13:                bsf_r ← range
14:                FS ← FS + top 10 new MRMRSelectedFeatures from D
15:            else
16:                Break
17:    SortAveAcc ← sort AveAcc (corresponding to first element), large to small
18:    BestAcc[1 : 3] ← SortAveAcc[1: 3]
19:    BestL ← learner in BestAcc[1 : 3] with smallest 2nd entry
20: return BestL
```

3 Experimental Results

The ADHD-200 Global Competition divided the ADHD-200 dataset into training and test sets. In the ABIDE dataset, we randomly selected 4/5 of the data as the training data and left the remaining 1/5 as the testing data. We ran our L2CPD system on these two datasets. For ADHD-200, L2CPD determined the best classifier was SVM with Sigma = 8 on 121 MRMR features. When run on the hold-out set, its accuracy was 0.6257, which was better than the best imaging-based diagnostic performance, 0.6154, achieved in the ADHD-200 global competition [5]. (Note that our accuracy scores for the ADHD-200 test set did not include the 26 subjects from the Brown site, as their diagnostic labels have not been released). For the ABIDE dataset, L2CPD determined that SVM with Sigma = 9, on 241 features, was the best. When it was run on its hold-out set, its accuracy was 0.6188 (note on this hold-out set, the baseline was 0.5157). This is better than the result of Nielsen *et al.*, who achieved 0.60 accuracy against their baseline of 0.5363 [15]. The difference in baseline accuracies was because Nielsen *et al.* omitted 148 of the individuals due to preprocessing problems.

We then explored the accuracy of various base learners, as a function of the number of features. In this paper we only show the 3 learners that had the best overall accuracy. The dashed lines in Figures 4 and 5 show the average 5 fold cross validation accuracy using SVM with RBF kernel with different sigma values, as a function of features selected. (This corresponds to the values of AveAcc[L, k].mean, for each learner L, for the different feature set sizes, k in $\{1, 11, ...\}$.

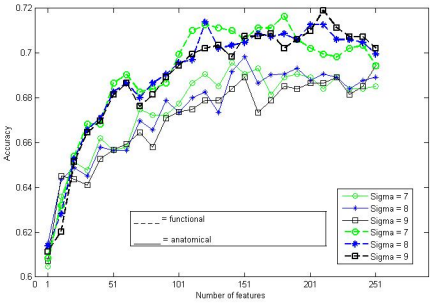

Fig. 4: ADHD train data, baseline: 0.6372

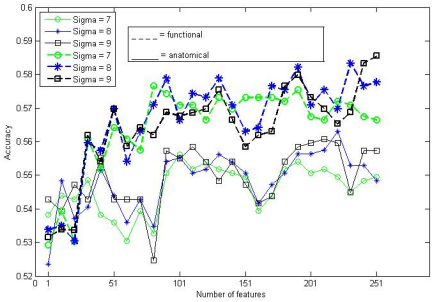

Fig. 5: ABIDE train data, baseline: 0.5158

Table 1: Top 3 base learners

	ADHD			ABIDE		
	Sigma = 9	Sigma = 7	Sigma = 8	Sigma = 9	Sigma = 9	Sigma = 8
Accuracy	0.71907	0.71644	0.71387	0.58556	0.58334	0.58330
Number of features	211	181	121	251	241	231
Range	0.07996	0.106103	**0.073538**	0.075509	**0.044944**	0.06244

We then considered whether we could get similar performance using structural MRI – that is, just using $T1$-weighted images. We therefore ran the same processing, including HOG feature extraction, MRMR feature selection, and using the same range of base learners; these appear as the solid lines in Figures 4 and 5. The figures show that functional-based features are almost always doing better than the structural based features, over the entire range of feature set sizes.

4 Discussion

We anticipate our approach may be used to learn classifiers for other diseases. We also think our classification accuracy will improve if we also include personal characteristic data, in addition to fMRI-based information, as was done in [16].

To summarize, our results showed that both autism and ADHD can be diagnosed using the functional images and also that HOG features, which are well known for object detection, can also be useful for classification of psychiatric diseases using brain images. Since we successfully applied our method to learn two diagnosis methods from two large multisite datasets, we expect that our approach might be appropriate for other datasets. Further research will be needed to address this question.

References

1. Scott Huettel et al. Functional magnetic resonance imaging. 2004.
2. Robert C Wolf et al. Regional brain activation changes and abnormal functional connectivity of the ventrolateral prefrontal cortex during working memory processing in adults with attention-deficit/hyperactivity disorder. *Human brain mapping*, 30(7):2252–2266, 2009.
3. MH Lee et al. Resting-state fmri: A review of methods and clinical applications. *American Journal of Neuroradiology*, 2012.
4. ADHD-200 global competition. `http://fcon_1000.projects.nitrc.org/indi/adhd200/`, 2011.
5. ADHD-200 global competition results. `http://fcon_1000.projects.nitrc.org/indi/adhd200/results.html`, 2011.
6. ABIDE dataset. `http://fcon_1000.projects.nitrc.org/indi/abide/`, 2011.
7. A Di Martino et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 2013.
8. Statistical parametric mapping, version 8. `http://www.fil.ion.ucl.ac.uk/spm/`, 2009.
9. D Louis Collins et al. Animal+ insect: improved cortical structure segmentation. In *information processing in medical imaging*, pages 210–223. Springer, 1999.
10. VS Fonov et al. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47:S102, 2009.
11. Vladimir Fonov et al. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1):313–327, 2011.
12. Navneet Dalal et al. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.
13. Hanchuan Peng et al. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005.
14. Chris Ding et al. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02):185–205, 2005.
15. Jared A Nielsen et al. Multisite functional connectivity mri classification of autism: Abide results. *Frontiers in human neuroscience*, 7, 2013.
16. Matthew RG Brown et al. Adhd-200 global competition: diagnosing adhd using personal characteristic data can outperform resting state fmri measurements. *Frontiers in systems neuroscience*, 6, 2012.