Identification of OCD-Relevant Brain Areas through Multivariate Feature Selection

Emilio Parrado-Hernández¹, Vanessa Gómez-Verdejo¹, Manel Martinez-Ramon¹, Pino Alonso²,³, Jesús Pujol⁴, José M. Menchón²,³, Narcís Cardoner²,³, and Carles Soriano-Mas²,³

¹ Signal Processing and Communications, Universidad Carlos III de Madrid, Leganés, 28911, Spain
{emipar, vanessa, manel}@tsc.uc3m.es
² Bellvitge Biomedical Research Institute-IDIBELL, Psychiatry Service, Bellvitge University Hospital, Barcelona, Spain
³ Carlos III Health Institute, Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Spain
{mpalonso, jmenchon, ncardoner}@bellvitgehospital.cat, csoriano@idibell.cat
⁴ CRC Hospital del Mar, Barcelona, Spain
jpujol@crccorp.es

Abstract. In this work we apply multivariate feature selection methods to construct a classifier that is able to differentiate among control subjects and OCD patients, with the purpose of bringing out regions of the brain that are relevant for the detection of the disease. Results show a discovery of regions that present great agreement with traditional methods used in OCD problems, but with the advantage of showing which ones are representative of control subjects or patients and providing cleaner and more accurate region maps.

1 Introduction

Multivariate pattern analysis is a useful tool to explore the information in Magnetic Resonance Imaging (MRI) data. Many approaches have been recently applied to functional MRI data to predict subjects’ behavior detecting changes in the sensorimotor or cognitive status [1–6]. In this context, it is generally accepted that optimal results can be achieved upon feature extraction, allowing the machine to operate in a reduced feature space and providing useful neuroanatomical and clinical information.

The above mentioned works are intended to introduce new multivariate strategies and provide experimental evidence about the possibility of automatically detecting spatially specific activations in the brain, and they apply the methods

* This work was supported by grants Spain CICYT grant TEC2008-02473, Madrid Regional Government grant CCG10-UC3M/TIC-5511 and PASCAL2 Network of Excellence IST-2007-216886, and the Carlos III Health Institute (PI09/01331 and CP10/00604). Dr. Soriano-Mas is funded by a Miguel Servet contract from the Carlos III Health Institute (CP10/00604).
Identification of OCD-Relevant Brain Areas 61

only to fMRI, but structural MRI is not considered. Besides, there are little or no references to their possible clinical applications to disease detection and characterization. In other works [7–10], authors introduce methods potentially useful for schizophrenia detection and characterization. All these methods apply dimensionality reduction or feature extraction methods in order to reduce the dimensionality of the patterns to be processed.

Interestingly, some studies have also used structural MRI data for disease detection or characterization. Structural MRI data have been typically considered to be more reliable than functional findings, being probably better suited for longitudinal and multicentric comparisons. Ecker et al. [11], for instance, used RFE and SVM to discriminate autistic subjects from healthy controls, while Costafreda et al. [12] used a feature extraction step based on mass-univariate statistical analyses followed by an SVM classification of depressive patients versus healthy controls. While structural abnormalities have been consistently described in other psychiatric conditions, such as Obsessive-Compulsive Disorder (OCD) [13, 14], multivariate classification approaches have not been extensively explored, although results using less sophisticated approaches [15] seem to indicate that there is potential for the classification and characterization of OCD patients using anatomical data.

In this work we propose to apply an SVM based Bagging strategy able to automatically detect what regions of the brain are of interest in the characterization of OCD. Experimental results show that, unlike standard feature selection techniques, this approach is also able to identify whether a region is associated to either an OCD patient or a control subject, which provides a great advance in the automatic characterization of OCD disease. Additionally, combining this feature selection strategy with a linear SVM, we will show that the proposed method is able to provide similar classification rates than standard feature selection methods.

2 SVM Bagging for Voxel Selection in MRI

An MRI brain scan is a vector whose components are always greater or equal than zero (the probability of each voxel being gray matter). Therefore, linear classifiers in such input space admit a pretty straightforward manner to interpret the role of each voxel in the discriminative task. A linear classifier assigns each brain scan of \( D \) voxels \( \mathbf{x} = [x_1, \ldots, x_D]^T \) to a possible output class \( \hat{y}(\mathbf{x}) \) using

\[
\hat{y}(\mathbf{x}) = \text{sign}\left\{ \mathbf{w}^T \mathbf{x} + b \right\} = \text{sign}\left\{ \sum_{d=1}^{D} w_d x_d + b \right\}
\]  

(1)

where \( \mathbf{w} = [w_1, \ldots, w_D]^T \) and \( b \) are the weight vector and the bias term of the classifier, respectively. The feature selection method used in this work builds on the starplots method of [16]. Basically, we carry out a bagging procedure consisting in sampling a subset of \( M \) instances of the training set and learn a linear SVM [17, 18] with them. We repeat this procedure a significant number