Structured Outlier Detection in Neuroimaging Studies with Minimal Convex Polytopes

Erdem Varol$^{(\boxtimes)}$, Aristeidis Sotiras, and Christos Davatzikos

Center for Biomedical Image Computing and Analytics, University of Pennsylvania,
Philadelphia, PA 19104, USA
`erdem.varol@uphs.upenn.edu`

Abstract. Computer assisted imaging aims to characterize disease processes by contrasting healthy and pathological populations. The sensitivity of these analyses is hindered by the variability in the neuroanatomy of the normal population. To alleviate this shortcoming, it is necessary to define a normative range of controls. Moreover, elucidating the structure in outliers may be important in understanding diverging individuals and characterizing prodromal disease states. To address these issues, we propose a novel geometric concept called minimal convex polytope (MCP). The proposed approach is used to simultaneously capture high probability regions in datasets consisting of normal subjects, and delineate outliers, thus characterizing the main directions of deviation from the normative range. We validated our method using simulated datasets before applying it to an imaging study of elderly subjects consisting of 177 controls, 123 Alzheimer's disease (AD) and 285 mild cognitive impairment (MCI) patients. We show that cerebellar degeneration is a major type of deviation among the controls. Furthermore, our findings suggest that a subset of AD patients may be following an accelerated type of deviation that is observed among the normal population.

1 Introduction

Mass-univariate and multivariate pattern analysis techniques aim to reveal disease effects by comparing a patient group to the control population [1,9]. The latter is commonly assumed to be homogeneous. However, as noted in recent works [6,13], controls may often consist of subjects that are outside a normative range, and this may confound the actual pathological effect when comparing against the patient group. The confounding effect may be remedied by identifying a normative range and removing outliers that lie outside this range.

There have been two main directions of outlier detection in the context of neuroimaging. The first class of methods include parametric models that aim to select a subset of samples such that the determinant of the covariance matrix is minimized. This is in contrast to non-parametric methods such as the one-class support vector machine (OC-SVM) [7,13,14] which attempt to separate a subset of samples from the origin with maximum margin in the Gaussian radial basis function (GRBF) kernel space. Another complementary non-parametric

approach is the support vector data description (SVDD) [15] whose objective is to solve for the smallest radius hypersphere that encloses a subset of the samples (Fig. 1b). All of the aforementioned outlier detection methods effectively capture the main probability mass of a dataset and delineate samples outside this region as outliers. However, they do not provide further information about whether there are different types of outliers. In this work, we posit that there may be a structure by which outliers deviate from the normal population. Capturing this structure may be instrumental in characterizing and understanding how pathogenesis originates from those who are healthy. Thus, the overall aim of our approach is to learn the organization by which samples deviate from the main probability mass.

We resolve the limitation of prior methods regarding learning the structure of outliers by containing the high probability region of a dataset using convex polytopes [16]. The geometry of our formulation allows to simultaneously enclose the normative samples within the convex polytope while excluding outliers with maximum margin. The assignment of outliers to unique faces of the convex polytope permits our formulation to be posed as a clustering problem. This clustering allows to subtype the directions of deviation from the normal.

The remainder of this paper is organized as follows. In Sect. 2 we detail the proposed approach, while experimental validation follows in Sect. 3. Section 4 concludes the paper with our final remarks.

2   Method

To learn the organization by which samples deviate from the main probability mass, we aim to find the minimal convex polytope (MCP) that excludes ρ percent of the samples with maximum margin. The convex polytope is minimal in the sense that the radius of the largest hypersphere that is circumscribed within the polytope is the minimum possible. Furthermore, the convex polytope is maximum margin in the sense that the margin between samples within the polytope and the outliers surrounding the polytope is maximized (Fig. 1c).

(a)   (b)   (c)
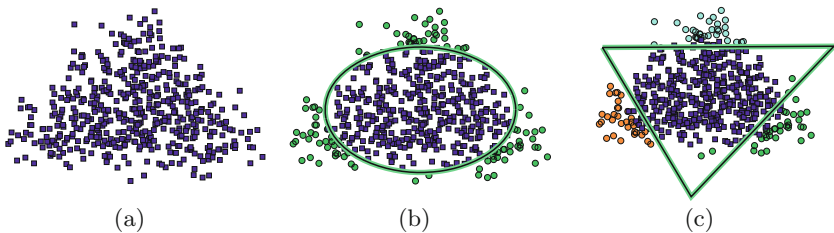
Fig. 1. (a) A simulated dataset with three deviations from normal; (b) the minimum hypersphere that excludes ρ percent of samples; (c) Proposed solution: minimum convex polytope (MCP) that excludes ρ percent of samples. Note that the MCP characterizes the types of deviations by associating outliers to different faces (indicated by colors orange, green and blue).

The previous problem involves two steps. The first step is to find the minimal hypersphere that excludes ρ percent of samples and the second is to find the convex polytope that circumscribes this hypersphere. Let $\mathbf{x}_i \in \mathbb{R}^d$ for $i = 1, \ldots, n$ denote the ith d-dimensional sample in the dataset. The minimal hypersphere that excludes ρ percent of samples can be cast as the following optimization problem:

$$\underset{R, \mathbf{x}_c}{\text{minimize}} \; R^2 + \frac{1}{n\rho} \sum_{i=1}^{n} \max\{0, R^2 - \|\mathbf{x}_i - \mathbf{x}_c\|_2^2\}, \tag{1}$$

where R describes the radius and $\mathbf{x}_c$ denotes the center of the hypersphere. This problem is convex [15] and can be solved using LIBSVM[1].

Once the dichotomy between the outliers and normative samples has been established, the maximum margin convex polytope [16] that separates the outliers from the normative samples can be cast as the following objective:

$$\underset{\substack{\{\mathbf{w}_j, b_j\}_{j=1}^{K} \\ \{a_{i,j}\}_{j=1,i=1}^{K,n} \\ \sum_j a_{i,j}=1 \\ a_{i,j}\geq 0}}{\text{minimize}} \quad \underbrace{\sum_{j=1}^{K} \|\mathbf{w}_j\|_1}_{\text{regularization/margin}} \quad +C \left[\underbrace{\sum_{\substack{i:\|\mathbf{x}_i-\mathbf{x}_c\|_2 \leq R \\ j=1,\ldots,K}} \frac{1}{K} \max\{0, 1 + \mathbf{w}_j^T \mathbf{x}_i + b_j\}}_{\text{loss for normative samples}} \right.$$

$$\left. + \underbrace{\sum_{\substack{i:\|\mathbf{x}_i-\mathbf{x}_c\|_2 > R \\ j=1,\ldots,K}} a_{i,j} \max\{0, 1 - \mathbf{w}_j^T \mathbf{x}_i - b_j\}}_{\text{assignment \& loss for outliers}} \right].$$

$$\tag{2}$$

This objective bears resemblance to standard large margin classifiers such as SVM. The first term encourages sparsity to capture focal directions of deviation which are often encountered in neuroimaging studies. The loss term is broken into one for normative samples and another for outliers. Specifically, the normative samples are constrained to be in the negative halfspace of all faces of the polytope while the outliers are constrained to be in the positive halfspace for at least one of the faces. This leads to an assignment problem which is encoded by the $a_{i,j}$ entries of the matrix $\mathbf{A}$ that inform us whether ith sample belongs to the jth face of polytope or not. The resulting formulation is non-convex and an iterative optimization between solutions of the faces, $\mathbf{W}, \mathbf{b}$ and assignments, $\mathbf{A}$ is necessary.

When fixing the assignments, the problem can be solved by K applications of weighted LIBSVM[2]. On the other hand, when fixing the convex polytope, the outliers can be assigned to the face that yields the maximum value of $\mathbf{w}_j^T \mathbf{x}_i + b_j$. The overall optimization scheme is summarized in Algorithm 1.

---

[1] https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/svdd/.
[2] https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/weights/.

Algorithm 1. Minimal (enclosing) Convex Polytope (MCP)

Input: $\mathbf{X} \in \mathbb{R}^{n \times d}$, C (loss penalty), ρ (outlier percentage), K (number of outlier subtypes)

Output: $\mathbf{W} \in \mathbb{R}^{d \times K}$, $\mathbf{b} \in \mathbb{R}^{1 \times K}$ (Outlier excluding convex polytope); $\mathbf{A} \in [0,1]^{n \times K}$ (Outlier subtype assignment)

Outlier delineation: Solve for $R, \mathbf{x}_c$ in Eq. (1b) using LIBSVM-SVDD[1]

Initialization: Initialize outlier assignments $\mathbf{A}$ randomly

Loop: Repeat until convergence (or a fixed number of iterations)

• Fix $\mathbf{A}$ — Solve for $\mathbf{W}, \mathbf{b}$ with weighted LIBSVM with weights[2]:

$$w_{i,j} = \begin{cases} \frac{C}{K} & \text{if } \|\mathbf{x}_i - \mathbf{x}_c\|_2 \leq R \\ C a_{i,j} & \text{if } \|\mathbf{x}_i - \mathbf{x}_c\|_2 > R \end{cases}$$

• Fix $\mathbf{W}, \mathbf{b}$ — Solve for $\mathbf{A}$: $\begin{cases} a_{i,j} = \frac{1}{K} & \text{if } \|\mathbf{x}_i - \mathbf{x}_c\|_2 \leq R \\ a_{i,j} = 1 & \text{if } \|\mathbf{x}_i - \mathbf{x}_c\|_2 > R \text{ and } j = \arg\max_j \mathbf{w}_j^T \mathbf{x}_i + b_j \\ a_{i,j} = 0 & \text{otherwise} \end{cases}$

2.1 Model Selection

The proposed MCP model is ultimately a clustering method whose performance depends on the selection of the following three parameters: (1) K, the number of deviation subtypes; (2) ρ, the outlier amount; (3) C, the loss penalty for violating margin. We choose the parameter combination that yields the most stable clustering [2]. To measure stability, we compute the average pairwise adjusted Rand index (ARI) [8] in a 10-fold cross-validation setting. The considered parameter space is: $K \in \{1, \ldots, 9\}$, $\rho \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and $C \in \{10^{-3}, \ldots, 10^1\}$.

3 Experimental Validation

3.1 Simulated Data

Due to lack of ground truth in clinical datasets and the need to quantitatively evaluate performance, we validated our method on two simulated datasets where the number of directions of deviations from the normal was a priori determined. Both datasets composed of 1000 samples and 150 features. 130 out of 150 of the features were drawn from a zero mean, unit variance, multivariate Gaussian distribution. For the first dataset, the remaining 20 features were replicates of the univariate random variable that is uniformly distributed within a unit side length equilateral triangle (as in Fig. 1a). Thus, the number of simulated deviations from the spherical white noise was three for this dataset. The second dataset was analogously generated except that the 20 signal-carrying features were replicates of the univariate random variable that is uniformly distributed within a unit side length square. Hence, this dataset was designed to yield four types of outliers.

For the triangular dataset, the parameter selection revealed that the most stable clustering occurs at $K = 3, \rho = 0.1, C = 0.01$ (Fig. 2a), while for the square dataset, the most stable clustering occurred at $K = 4, \rho = 0.5, C = 0.01$

(a) Triangular simulated dataset   (b) Square simulated dataset
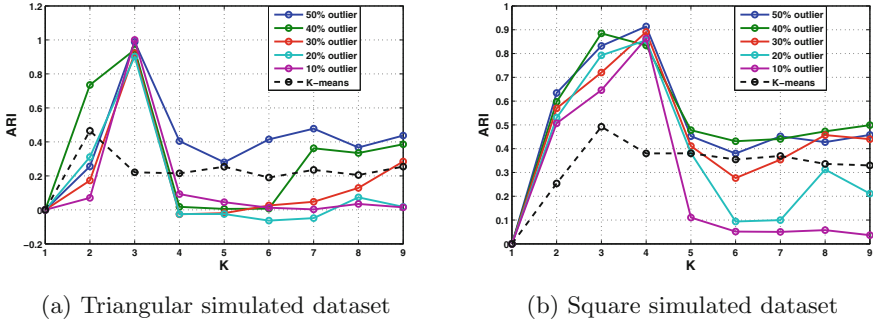
Fig. 2. The parameter selection for (a) triangular simulated dataset, and (b) square simulated dataset. (a) $K = 3, \rho = 0.1, C = 0.01$ were selected, (b) $K = 4, \rho = 0.5, C = 0.01$ were selected. Different solid lines indicate the ARI of MCP at different values of ρ at the maximum ARI yielding C parameter. Black dashed lines indicate the ARI of K-means for comparison. Note that MCP yields more stable clusterings that align with the ground truth.
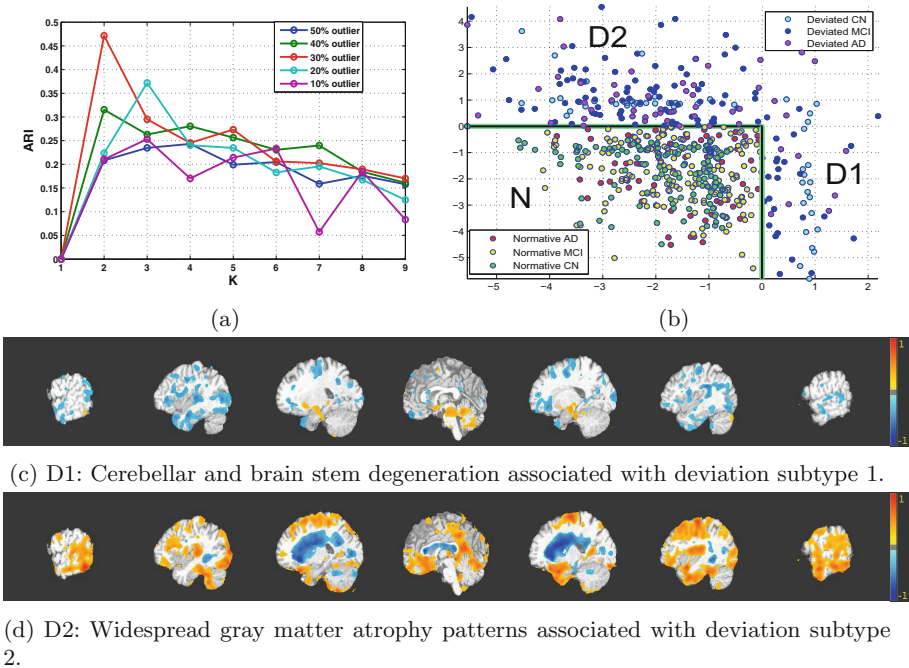
(Fig. 2b). For both of these datasets, the ARI values for the optimal K were comparable across varying ρ and C, which indicates that the most important directions of deviation were captured regardless of the amount of outliers searched. These results demonstrate the ability of MCP to capture the underlying directions of deviation.

For comparison, K-means clustering was applied to the same datasets (see Fig. 2a, b, dashed lines). For the triangular and square datasets, $K = 2$ and $K = 3$ yielded the most stable clusterings, respectively. This demonstrates that K-means was not able to accurately capture the main directions of deviation, but was most likely grouping outliers with the normative samples.

3.2 Application to a Study of Alzheimer's Disease

The proposed method was applied to a subset of the ADNI study[3] which is composed of magnetic resonance imaging (MRI) scans of 177 controls (CN), 123 Alzheimer's disease (AD) patients and 285 mild cognitive impairment (MCI) patients. T1-weighted MRI volumetric scans were obtained at 1.5 Tesla. The images were pre-processed through a pipeline consisting of (1)alignment to the Anterior and Posterior Commissures plane; (2) skull-stripping; (3) N3 bias correction; (4) deformable mapping to a standardized template space. Following these steps, a low-level representation of the tissue volumes was extracted by automatically partitioning the MRI volumes of all participants into 153 volumetric regions of interest (ROI) spanning the entire brain. The ROI segmentation was performed by applying a multi-atlas label fusion method [4]. The derived ROIs were used as the input features for our method. Before training the model, all ROIs were linearly residualized to remove the effect of age and sex [5].

---

[3] http://adni.loni.usc.edu/data-samples/mri/.

(a)   (b)

(c) D1: Cerebellar and brain stem degeneration associated with deviation subtype 1.

(d) D2: Widespread gray matter atrophy patterns associated with deviation subtype 2.

Fig. 3. (a) The parameter selection for ADNI control group, $K = 2, \rho = 0.3, C = 1$ yielded the highest clustering stability. (b) The projections of all ADNI subjects along the two faces of the MCP. Normative samples (N) are in the negative orthant while deviated subtypes are on the upper left (subtype 2) and lower right (subtype 1). (c, d) The voxel-based group differences between all normative samples and deviation subtype 1 (c), and deviation subtype 2 (d) are shown. Warmer colors indicate that the normative group volume is greater, while colder colors indicate that the deviated group volume is greater.

The method was applied only to the control group. The parameter selection revealed that $K = 2$ subtypes, and 30 % outliers with $C = 1$ yielded the highest clustering stability (Fig. 3a). Once the MCP that captured the normative controls was found, it was used to subtype the rest of the ADNI dataset consisting of AD and MCI subjects into three groups denoted by normative (N), deviation subtype 1 (D1) and deviation subtype 2 (D2).

The distribution of the entire ADNI dataset with respect to the MCP is illustrated in Fig. 3b. Furthermore, the demographic and clinical biomarker information of CN, MCI and AD subjects within their respective subgroup is summarized in Table 1. 56 % of AD and 62 % of MCI patients were categorized into the normative group. This indicated that the main type of AD and MCI neuropathology was dissimilar to the deviations exhibited by the normal population. However, a non-negligible portion, 37 % of AD and 28 % of MCI was found to deviate

Table 1. Demographic and clinical characteristics of CN, AD, MCI subjects and their grouping into the normative (N) or deviated subtypes (D1, D2). [a] – Mini mental state exam. [b] – Presence of at least one APOE $\varepsilon 4$ allele. [c] – Cerebrospinal fluid (CSF) concentrations of Amyloid-beta (Aβ), total tau (t-tau), and phosphorylated tau (p-tau). [d] – p-values using ANOVA between three subgroups

Normative and deviated subjects in ADNI study							
Group	AD				MCI		
Subtype	AD-N	AD-D1	AD-D2	p-value[d]	MCI-N	MCI-D1	MCI-D2	p-value[d]
n (%)	69 (56.0)	8 (6.5)	46 (37.3)		178 (62.4)	26 (9.1)	81 (28.4)	
Age (years)	76.4 ± 6.7	74.4 ± 6.6	72.0 ± 7.8	0.007	75.4 ± 6.7	77.6 ± 6.3	71.9 ± 7.4	0.01
Sex (female), n (%)	36 (52.1)	3 (37.5)	23 (50.0)	0.73	59 (33.1)	6 (23.0)	33 (40.7)	0.15
MMSE[a]	23.6 ± 1.9	23.7 ± 1.9	23.5 ± 1.7	0.92	27.0 ± 1.7	27.0 ± 1.8	26.9 ± 1.7	0.89
APOE $\varepsilon 4^{b}$, n (%)	45 (65.2)	6 (75.0)	31 (67.3)	0.85	95 (53.3)	12 (46.1)	50 (61.7)	0.47
CSF Aβ (pg/mL)[c]	153.8 ± 49.2	146.2 ± 39.7	126.9 ± 23.3	0.03	158.8 ± 51.9	187.7 ± 59.2	164.9± 54.5	0.24
CSF t-tau (pg/mL)[c]	118.5 ± 59.0	104.2 ± 44.7	132.3 ± 59.0	0.49	90.0 ± 38.4	129.2 ± 71.0	106.2 ± 53.0	0.0029
CSF p-tau (pg/mL)[c]	37.9 ± 18.3	38 ± 14	45.5 ± 20.6	0.27	34.2 ± 16.2	38.1 ± 20.3	36.6 ± 16.5	0.05

Group	CN			
Subtype	CN-N	CN-D1	CN-D2	p-value[d]
n (%)	125 (70.6)	19 (10.7)	33 (18.6)	
Age (years)	75.4 ± 5.3	78.1 ± 4.2	76.2 ± 4.7	0.10
Sex (female), n (%)	58 (46.4)	10 (52.6)	19 (57.5)	0.49
MMSE[a]	29.0 ± 1.1	29.0 ± 0.7	29.4 ± 0.8	0.19
APOE $\varepsilon 4^{b}$, n (%)	34 (27.2)	6 (31.5)	8 (24.2)	0.85
CSF Aβ (pg/mL)[c]	203.4 ± 55.7	233.7 ± 35.2	218.5 ± 52.3	0.21
CSF t-tau (pg/mL)[c]	67.4 ± 23.8	63.7 ± 22.6	73.9 ± 29.2	0.54
CSF p-tau (pg/mL)[c]	24.5 ± 14.5	22.4 ± 11.5	24.8 ± 11.4	0.90

along the second subtype direction along with 18 % of CN. This suggested that a sizeable portion of the normal population might have the propensity to deviate towards AD-like pathology.

To better understand and interpret the neuroanatomical directions of these deviations from the normative range, voxel-based analysis was performed on all subjects in the normative group versus either of the two subtypes of deviations using gray matter tissue density maps. The group differences are visualized in Fig. 3.

There has been a substantial amount of research in the past that has demonstrated that the normal pattern of aging consists of prefrontal and motor cortex thinning along with increased ventricle size [11,12]. Corresponding manifestations of these patterns can be observed in group D2 (Fig. 3d). The significantly younger ages of AD and MCI subjects (Table 1) that fall into this subtype may indicate that the cognitive decline they exhibit may be caused by early and accelerated aging that follows this pattern. Furthermore, the relatively lower CSF amyloid-β and t-tau concentrations (Table 1) of these patients is another strong indicator of AD [3].

On the other hand, the patterns seen in group D1 (Fig. 3c) indicate cerebellar degeneration which is usually accompanied by brain stem atrophy [10]. Although cerebellar thinning has been demonstrated to be part of normal aging, our findings suggest that the increased rate of this degenerative pattern may be a type of deviation. Lastly, it should be mentioned that the majority of the AD and MCI subjects were not designated to be moving along either of the directions of

deviations of normal subjects. A possible explanation is that for these particular subjects, the deviation towards AD may have begun at an earlier time point, which was not represented by the control subjects present in the study.

4   Conclusion

In summary, we have introduced a method that can simultaneously detect a homogeneous normative group and define subtypes of outliers. This allows a better understanding of the structure of deviations in control groups in neuroimaging cohorts. This, in turn, aids in the better interpretation of the pathological processes, which occur when subjects diverge from the normative region.

References

1. Ashburner, J., Friston, K.J.: Voxel-based morphometry-the methods. Neuroimage **11**(6), 805–821 (2000)
2. Ben-Hur, A., et al.: A stability based method for discovering structure in clustered data. In: Pacific Symposium on Biocomputing, vol. 7, pp. 6–17 (2001)
3. Blennow, K.: Cerebrospinal fluid protein biomarkers for Alzheimer's disease. NeuroRx **1**(2), 213–225 (2004)
4. Doshi, J., et al.: MUSE: multi-atlas region segmentation utilizing ensembles of registration algorithms and parameters, and locally optimal atlas selection. NeuroImage **127**, 186–195 (2015)
5. Dukart, J., Schroeter, M.L., Mueller, K., Initiative, A.D.N., et al.: Age correction in dementia-matching to a healthy brain. PloS one **6**(7), e22193 (2011)
6. Fritsch, V., et al.: Detecting outliers in high-dimensional neuroimaging datasets with robust covariance estimators. Med. Image Anal. **16**(7), 1359–1370 (2012)
7. Gardner, A.B., et al.: One-class novelty detection for seizure analysis from intracranial EEG. J. Mach. Learn. Res. **7**, 1025–1044 (2006)
8. Hubert, L., Arabie, P.: Comparing partitions. J. Classif. **2**, 193–218 (1985)
9. Kawasaki, Y., et al.: Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. Neuroimage **34**(1), 235–242 (2007)
10. Luft, A.R., et al.: Patterns of age-related shrinkage in cerebellum and brainstem observed in vivo using three-dimensional MRI volumetry. Cereb. Cortex **9**(7), 712–721 (1999)
11. Raz, N., Rodrigue, K.M.: Differential aging of the brain: patterns, cognitive correlates and modifiers. Neurosci. Biobehav. Rev. **30**(6), 730–748 (2006)
12. Salat, D.H., et al.: Thinning of the cerebral cortex in aging. Cereb. Cortex **14**(7), 721–730 (2004)
13. Sato, J.R., et al.: An fmRI normative database for connectivity networks using one-class support vector machines. Hum. Brain Mapp. **30**(4), 1068–1076 (2009)
14. Schölkopf, B., et al.: Estimating the support of a high-dimensional distribution. Neural Comput. **13**(7), 1443–1471 (2001)
15. Tax, D.M., Duin, R.P.: Support vector data description. Machine Learn. **54**(1), 45–66 (2004)
16. Varol, E., Sotiras, A., Davatzikos, C.: Hydra: revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. NeuroImage (2016)