Medical Image Analysis 15 (2011) 729–737

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Spatial regularization of SVM for the detection of diffusion alterations associated with stroke outcome

Rémi Cuingnet [a,b,c,d,*,1], Charlotte Rosso [a,b,c,e], Marie Chupin [a,b,c], Stéphane Lehéricy [a,b,c,f], Didier Dormont [a,b,c,f], Habib Benali [d], Yves Samson [a,b,c,e], Olivier Colliot [a,b,c]

[a] UPMC Université Paris 6, UMR 7225, UMR_S 975, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (CRICM), Paris F-75013, France
[b] CNRS, UMR 7225, CRICM, Paris F-75013, France
[c] Inserm, UMR_S 975, CRICM, Paris F-75013, France
[d] Inserm, UMR_S 678, LIF, Paris F-75013, France
[e] AP-HP, Urgences Cérébro-Vasculaires, Groupe hospitalier Pitié-Salpêtrière, Paris F-75013, France
[f] Centre for NeuroImaging Research, CENIR, Department of Neuroradiology, Groupe hospitalier Pitié-Salpêtrière, Paris F-75013, France

ARTICLE INFO

Article history:
Available online 6 June 2011

Keywords:
SVM
Regularization
Group analysis
Stroke
DWI

ABSTRACT

In this paper, we propose a new method to detect differences at the group level in brain images based on spatially regularized support vector machines (SVM). We propose to spatially regularize the SVM using a graph Laplacian. This provides a flexible approach to model different types of proximity between voxels. We propose a proximity graph which accounts for tissue types. An efficient computation of the Gram matrix is provided. Then, significant differences between two populations are detected using statistical tests on the outputs of the SVM. The method was first tested on synthetic examples. It was then applied to 72 stroke patients to detect brain areas associated with motor outcome at 90 days, based on diffusion-weighted images acquired at the acute stage (median delay one day). The proposed method showed that poor motor outcome is associated to changes in the corticospinal bundle and white matter tracts originating from the premotor cortex. Standard mass univariate analyses failed to detect any difference on the same population.

1. Introduction

Diffusion-weighted imaging (DWI) is of considerable interest to the clinical evaluation of acute stroke patients (Chalela et al., 2007). The location of the lesions has been suggested to represent a better predictor than their global volume (Crafton et al., 2003). At the subacute or chronic phases, previous studies have shown that damages to the corticospinal tract (CST) (Domi et al., 2009) and lesions to the primary sensorimotor cortex (Crafton et al., 2003; Lo et al., 2010) correlated with poor motor outcome. At the acute stage, regional changes in the apparent diffusion coefficients (ADC) were suggested as early quantitative indices of regional irreversible ischemic damage (Rosso et al., 2009). However, at the acute stage, the spatial pattern of ADC changes associated with motor outcome remains unclear.

Group analyses of differences between populations in brain imaging have widely relied on univariate voxel-wise analyses, such as voxel-based morphometry (VBM) for structural MRI (Ashburner and Friston, 2000) or their equivalent for diffusion imaging (VB-DWI). In such analyses, brain images are first spatially registered to a common stereotaxic space, and then mass univariate statistical tests are performed in each voxel to detect significant group differences. However, the sensitivity of theses approaches is limited when the differences are spatially complex and involve a combination of different voxels or brain structures (Davatzikos, 2004). Recently, there has been a growing interest in support vector machines (SVM) methods (Vapnik, 1995; Schölkopf and Smola, 2001) to overcome the limits of these univariate analyses. Theses approaches allow capturing complex multivariate relationships in the data and have been successfully applied to the individual classification of a variety of neurological conditions (Lao et al., 2004; Fan et al., 2007; Klöppel et al., 2008a; Vemuri et al., 2008). Moreover, the output of the SVM can also be analyzed to localize spatial patterns of discrimination, for example by drawing the coefficients of the optimal margin hyperplane (OMH) – which, in the case of a linear SVM, live in the same space as the MRI data (Klöppel et al., 2008a; Vemuri et al., 2008; Mourão-Miranda et al., 2005). However, one of the problems with analyzing directly the OMH

* Corresponding author at: UPMC Université Paris 6, UMR 7225, UMR_S 975, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (CRICM), Paris F-75013, France.

E-mail addresses: remi.cuingnet@imed.jussieu.fr (R. Cuingnet), marie.chupin@upmc.fr (M. Chupin), habib.benali@imed.jussieu.fr (H. Benali), olivier.colliot@upmc.fr (O. Colliot).

[1] Present address: CRICM – Equipe Cogimage (ex LENA), Hôpital de la Pitié-Salpêtrière, 47, Boulevard de l'Hôpital, 75651 Paris Cedex 13, France.

1361-8415/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2011.05.007

coefficients is that the corresponding maps are noisy and lack spatial coherence. Moreover few of these approaches perform a statistical analysis of the OMH coefficients (Mourão-Miranda et al., 2005).

In this paper, we propose a new method to detect group differences in brain images based on spatially regularized SVM. In particular, we show how spatial consistency can be directly enforced into the SVM by using Laplacian regularization. Such regularization leads to a diffusion kernel on a graph (Kondor and Lafferty, 2002). We then propose a statistical analysis based on the spatially regularized SVM to detect brain regions which are significantly different between two groups of subjects. The proposed framework is tested on 2D synthetic test images and then applied to the detection of differences between stroke patients with good and poor outcome based on DWI acquired at the acute stage.

This paper extends work previously presented at a conference (Cuingnet et al., 2010). It is organized as follows. In Section 2, we show that the regularization operator framework provides a flexible approach to model different types of proximity via the definition of a regularization graph. Section 3 presents a method to to detect group differences based on the spatially regularized SVM. Then, in Section 4, the proposed framework is tested on 2D synthetic images (Section 4.1) and then applied to the detection of differences between stroke patients with good and poor outcome based on DWI acquired at the acute stage (Section 4.2). A discussion of the methods and results is presented in Section 5.

2. Spatially regularized SVM using the graph Laplacian

In this section, we first present some background on SVM, and on the framework of regularization operators (Section 2.1). We then propose a regularization operator based on the graph Laplacian (Section 2.2). This provides a flexible approach to model different types of proximity between voxels. A proximity graph which accounts for tissue types is finally presented.

2.1. Regularization and priors in SVM

2.1.1. Linear SVM

In this contribution, we consider any feature computed at each voxel of a 3D brain image. These images can be any characteristics extracted from MRI, such as gray matter concentration maps (in VBM) or ADC maps (in diffusion MRI). We further assume that images were spatially normalized to a common stereotaxic space (e.g. Shen and Davatzikos, 2002; Ashburner, 2007) as in many group studies or classification methods (Lao et al., 2004; Fan et al., 2007; Klöppel et al., 2008a; Vemuri et al., 2008; Querbes et al., 2009; Cuingnet et al., 2011). Let $\mathcal{X} = \mathbb{R}^d$. Let $(\mathbf{x}_s)_{s\in[1,N]} \in \mathcal{X}^N$ be the images of N subjects and $(y_s)_{s\in[1,N]} \in \{-1,1\}^N$ their group labels (e.g. diagnosis or outcome).

Support vector machines search for the hyperplane for which the margin between groups is maximal, the OMH. The standard linear SVM solves the following optimization problem (Vapnik, 1995; Schölkopf and Smola, 2001; Shawe-Taylor and Cristianini, 2004):

$$(\mathbf{w}^{\text{opt}}, b^{\text{opt}}) = \operatorname{argmin}_{\mathbf{w}\in\mathcal{X}, b\in\mathbb{R}} \times \underbrace{\frac{1}{N}\sum_{s=1}^{N} \ell_{\text{hinge}}(y_s[\langle \mathbf{w}, \mathbf{x}_s\rangle + b])}_{\text{Empirical Loss}} + \underbrace{\lambda\|\mathbf{w}\|^2}_{\text{Classical Tikhonov Regularization}} \quad (1)$$

where $\lambda \in \mathbb{R}^+$ is the *regularization parameter* and ℓ_{hinge} the *hinge loss function* defined as: $\ell_{\text{hinge}} : u \in \mathbb{R} \mapsto \max(0, 1-u)$. The ordered pair $(\mathbf{w}^{\text{opt}}, b^{\text{opt}}) \in \mathcal{X} \times \mathbb{R}$ denotes the minimal point solution to optimization problem (1) (Fig. 1).

With a linear SVM, the *feature space* is the same as the *input space*. Thus, when the input features are some characteristics defined at the voxel level (e.g. voxel intensities), each component of $\mathbf{w}^{\text{opt}}$ also corresponds to a voxel. One can therefore represent the values of $\mathbf{w}^{\text{opt}}$ in the image space, and use this map to localize differences. However, the map $\mathbf{w}^{\text{opt}}$ can be noisy and scattered (as for example in Klöppel et al. (2008a)). This is due to the fact that the regularization term of the standard linear SVM *is not a spatial regularization*. Moreover, voxel-based comparisons are subject to registration errors and interindividual variability. Gaussian smoothing is therefore often used as a preprocessing step. However, some image information is lost during the smoothing step which, for example, mixes white matter with gray matter voxels. Tissue probability maps could be used to overcome this limitation. More generally, if voxels are connected, meaning for example spatially, anatomically or functionally close, we would like the SVM to consider them as similar.

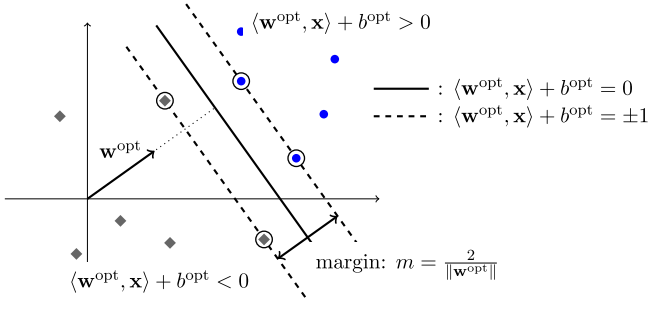

Fig. 1. Illustration of an optimal margin hyperplane obtained with a linear SVM. The optimal separating hyperplane is defined by the ordered pair $(\mathbf{w}^{\text{opt}}, b^{\text{opt}})$, the minimal point solution to optimization problem (1). It is the hyperplane for which the margin m between the two groups (gray and blue) is maximal. The support vectors are circled. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.1.2. How to include priors in SVM

To spatially regularize SVM, one has to include some prior knowledge on the proximity of features. In the literature, three main ways have been considered in order to include priors in SVM.

In SVM, all the information used for the classification is encoded in the kernel. Hence, the first way to include prior knowledge is to directly design the kernel function (Schölkopf and Smola, 2001). But this implies knowing a metric or an affinity measure on the input space $\mathcal{X}$ consistent with the prior knowledge. The use of an inappropriate metric or affinity measure would only map the data into a higher dimensional space without adding any information.

Another way is to constrain the classifier function to be locally invariant to some transformations. This can be done: (*i*) by directly engineering a kernel which leads to locally invariant SVM (Schölkopf et al., 1998) to some chosen transformations, (*ii*) by generating artificially transformed examples from the training set to create virtual support vectors (Schölkopf et al., 1996), (*iii*) by using a combination of both these approaches called kernel jittering (Decoste and Schölkopf, 2002). But the main difficulty with these methods is to define the transformations to which we would like the kernel to be invariant.

The last way is to consider SVM from the regularization viewpoint (Smola and Schölkopf, 1998; Schölkopf and Smola, 2001). The idea is to constrain the classifier function to be regular with respect to some criteria. This is the viewpoint which is adopted in this paper.

2.1.3. Regularization operators

Our aim is to introduce a spatial regularization of the classifier function of the SVM. This function can be written as: sign $(f(\mathbf{x}_s) + b)$, where $f \in \mathbb{R}^{\mathcal{X}}$. This is done through the definition of a *regularization operator* on f. Following (Smola and Schölkopf, 1998; Schölkopf and Smola, 2001), a regularization operator P is defined as a linear map from a space $\mathcal{F} \subset \mathbb{R}^{\mathcal{X}}$ into a dot product space $(\mathcal{D}, \langle \cdot, \cdot \rangle_{\mathcal{D}})$.

If there exists a function $G : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ verifying:

$$\forall f \in \mathcal{F}, \forall \mathbf{x} \in \mathcal{X}, f(\mathbf{x}) = \langle P(G(\mathbf{x}, \cdot)), P(f) \rangle_{\mathcal{D}} \tag{2}$$

then G is a positive semi-definite kernel and the minimization problem:

$$(f^{\text{opt}}, b^{\text{opt}}) = \arg\min_{f \in \mathcal{F}, b \in \mathbb{R}} \frac{1}{N} \sum_{s=1}^{N} \ell_{\text{hinge}}(y_s[f(\mathbf{x}_s) + b]) + \lambda \|P(f)\|_{\mathcal{D}}^2 \tag{3}$$

is equivalent to the SVM minimization problem with kernel G.

Since, in linear SVM, the feature space is the input space, f lies in the input space. Therefore, the optimization problem (3) is very convenient to include spatial regularization on f via the definition of P.

Note that, usually, $\mathcal{F}$ is a Reproducing Kernel Hilbert Space (RKHS) with kernel K and $\mathcal{D} = \mathcal{F}$. Hence, if P is bounded, and $P^\dagger P$ is invertible (e.g. when P is injective and compact) where $P^\dagger$ denotes the adjoint of P, the function G defined as:

$$\forall \mathbf{x}_s \in \mathcal{X}, \quad G_{\mathbf{x}_s} = (P^\dagger P)^{-1} K_{\mathbf{x}_s} \in \mathbb{R}^{\mathcal{X}}$$

where $G_{\mathbf{x}_s} = G(\mathbf{x}_s, \cdot) \in \mathbb{R}^{\mathcal{X}}, K_{\mathbf{x}_s} = K(\mathbf{x}_s, \cdot) \in \mathbb{R}^{\mathcal{X}}$, verifies Eq. (2). Note that G can be considered as the Green function of $P^\dagger P$.

One has to define the regularization operator P so as to obtain the suitable regularization for the problem.

2.2. Regularization based on diffusion on graph

2.2.1. Choice of the regularization operator

Weighted graphs are a natural framework to take spatial information into consideration (Peleg, 1980; Geman and Geman, 1984; Haris et al., 1998; Shi and Malik, 2000). Voxels of a brain image can be considered as nodes of a graph which models the voxels' proximity. This graph can be, as detailed below, the voxel connectivity (6, 18 or 26) or a more sophisticated graph.

We chose the following regularization operator:

$$\begin{aligned} P : \mathcal{F} = \mathcal{L}(\mathcal{X}, \mathbb{R}) &\to \mathcal{F} \\ f = \mathbf{w}^* &\mapsto \left(e^{\frac{1}{2}\beta L}\mathbf{w}\right)^* \end{aligned} \tag{4}$$

where L denotes the graph Laplacian (Chung, 1992) and $\mathbf{w}^*$ the dual vector[2] of $\mathbf{w}$. The parameter β controls the size of the regularization. The optimization problem then becomes:

$$\begin{aligned} (\mathbf{w}^{\text{opt}}, b^{\text{opt}}) = \arg\min_{\mathbf{w} \in \mathcal{X}, b \in \mathbb{R}} & \\ \times \underbrace{\frac{1}{N} \sum_{s=1}^{N} \ell_{\text{hinge}}(y_s[\langle \mathbf{w}, \mathbf{x}_s \rangle + b])}_{\text{Empirical Loss}} + & \underbrace{\lambda \|e^{\frac{1}{2}\beta L}\mathbf{w}\|^2}_{\text{Spatial Regularization}} \end{aligned} \tag{5}$$

Such a regularization exponentially penalizes the high-frequency components and thus forces the classifier to consider as similar voxels highly connected according to the graph adjacency matrix. According to the previous section, this new minimization problem (5) is equivalent to an SVM optimization problem. The new kernel K_β is given by:

$$K_\beta(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) = \mathbf{x}_{s_1}^T e^{-\beta L} \mathbf{x}_{s_2} \tag{6}$$

[2] The dual vector $\mathbf{w}^* \in \mathcal{L}(\mathcal{X}, \mathbb{R})$ of a vector $\mathbf{w} \in \mathcal{X}$ denotes the continuous linear function defined by: $\forall \mathbf{x} \in \mathcal{X}, \mathbf{w}^*(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle$.

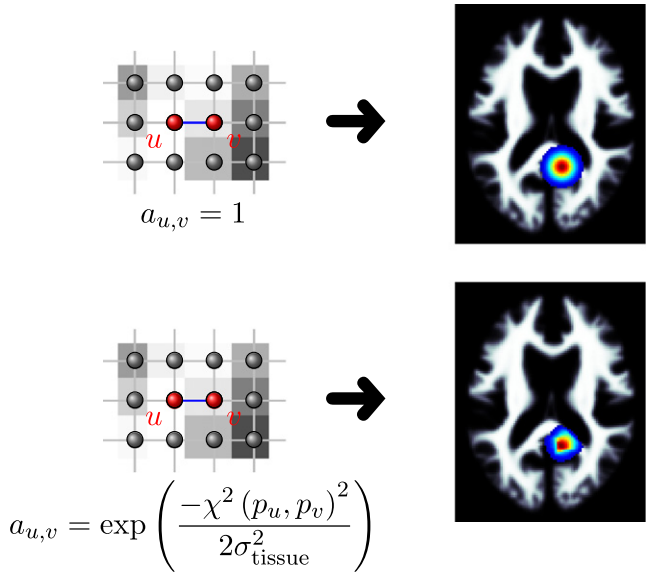

Fig. 2. Regularization graphs and corresponding impulse responses. Upper row: when the regularization graph is the image connectivity, the regularization is equivalent to smoothing the images with a Gaussian smoothing kernel. Lower row: the proposed regularization graph takes into consideration the tissue types (in this example we only used WM and non-WM as tissue types).

for any pair of subjects $(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$. Note that $K_\beta(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) = \langle e^{-\frac{\beta L}{2}}\mathbf{x}_{s_1}, e^{-\frac{\beta L}{2}}\mathbf{x}_{s_2} \rangle$. This is a heat or diffusion kernel (Kondor and Lafferty, 2002).

2.2.2. Examples of regularization graphs

One has now to define the graph depending on the type of spatial proximity one wants to enforce. The simplest option is to use the image connectivity (6, 18 or 26). In this case, the regularized SVM would be equivalent to smoothing the data with a Gaussian kernel with standard deviation $\sigma = \sqrt{\beta}$ (Smola and Kondor, 2003). But this would result in mixing gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Instead, we propose a graph which takes into account both spatial location and tissue types (Fig. 2). Let $\mathcal{T}$ be a set of tissues (e.g. $\mathcal{T} = \{\text{GM}, \text{WM}, \text{CSF}\}$). In each voxel v, we have the probability $p_v^{(t)}$ that this voxel belongs to tissue t. This probability is defined from a template (e.g. SPM template). In the following, $p_v$refers to the tuple $p_v = \left(p_v^{(t)}\right)_{t \in \mathcal{T}}$.

We considered the following graph. Two voxels are connected if and only if they are neighbors in the image (6-connectivity). The weight $a_{u,v}$ of the edge between two connected voxels u and v was defined as:

$$a_{u,v} = \exp\left(\frac{-\chi^2(p_u, p_v)^2}{2\sigma^2_{\text{tissue}}}\right) \tag{7}$$

where χ^2 is the χ^2-*distance* and σ_{tissue} a parameter. It is used to provide a measure of similarity between two tuples (p_u and p_v). Specifically, the χ^2-distance is defined as:

$$\chi^2(p_u, p_v)^2 = \frac{1}{2} \sum_{t \in \mathcal{T}} \frac{\left(p_u^{(t)} - p_v^{(t)}\right)^2}{p_u^{(t)} + p_v^{(t)}} \tag{8}$$

The parameter σ_{tissue} was set beforehand to the estimated standard deviation of $\chi^2(p_u, p_v)$. The graph edge weights were finally normalized by the mean edge weight.

2.2.3. Computation of the Gram matrix

Taylor series expansion. To solve the spatially regularized SVM optimization problem (Eq. (5)), one has to compute the Gram

matrix G, which is defined by: $G = (K(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}))_{s_1,s_2}$. According to Eq. (6), it requires to compute $e^{-\frac{\beta}{2}L}\mathbf{x}_s$ for any subject $s \in [1,N]$. There are various methods to compute the exponential of a matrix (Moler and Loan, 2003). As $L \in \mathbb{R}^{d\times d}$ with $d \sim 10^6$, the direct computation of $e^{-\frac{\beta}{2}L}$ is intractable. Therefore, we used the Taylor series expansion. Considering the terms up to pth order yields the following approximation:

$$e^{\frac{-\beta}{2}L}\mathbf{x}_s \approx \sum_{k=0}^{p} \frac{1}{k!}\left(\frac{-\beta}{2}L\right)^k \mathbf{x}_s \tag{9}$$

Scaling. It should be noted that $\left(\frac{\beta}{2}L\right)^k$ could have very large components, which may lead to numerical problems and in particular to round-off errors (Moler and Loan, 2003). To avoid such errors, we used the scaling method. It consists in first computing $e^{\frac{-\beta}{2n}L}$ where n is an integer chosen such as:

$$\left\|\frac{-\beta}{2n}L\right\|_1 \leqslant 1 \tag{10}$$

The first step *scales* the matrix norm between 0 and 1 by dividing the matrix by n. We chose n as the smallest integer satisfying inequality (10). Then $e^{-\frac{\beta}{2}L}$ is computed from $e^{-\frac{\beta}{2n}L}$ using the following formula:

$$e^{-\frac{\beta}{2}L} = \left(e^{\frac{-\beta}{2n}L}\right)^n \tag{11}$$

Order of the Taylor series. We are working with symmetric matrices; therefore, the spectral norm is lower or equal to the 1-norm. Therefore, the approximation error of $e^{\frac{-\beta}{2n}L}$ for the spectral norm is bounded by $\frac{1}{(p+1)!}$ (e.g. Kondor and Lafferty, 2002). As L is semidefinite positive, $\left\|e^{\frac{-\beta}{2}L}\right\|_2 \leqslant 1$. This yields:

$$\left\|\left(\sum_{k=0}^{p}\frac{1}{k!}\left(\frac{-\beta}{2n}L\right)^k\right)^n - e^{\frac{-\beta}{2}L}\right\|_2 \leqslant \left(1+\frac{1}{(p+1)!}\right)^n - 1 \tag{12}$$

We $p = 10$ for the experiments (Section 4). It yielded approximation errors lower than 10^{-5}.

3. Statistical analysis of the hyperplane coefficients

In this section, we propose a statistical analysis for the detection of brain regions which are significantly different between two groups of subjects, based on the results of the spatially regularized SVM. It is based on significance tests on the SVM outputs.

The classification function, obtained with a linear SVM, is the sign of the inner product of the features with $\mathbf{w}^{\text{opt}}$ (Vapnik, 1995; Shawe-Taylor and Cristianini, 2000, 2004; Schölkopf and Smola, 2001): sign $(\langle \mathbf{w}^{\text{opt}} | \mathbf{x} \rangle + b)$. Therefore, if the absolute value of the ith component of vector $\mathbf{w}^{\text{opt}}$, $|w_i^{\text{opt}}|$, is small compared to the other components $\left(\left|w_j^{\text{opt}}\right|\right)_{j\neq i}$, then the ith feature will have little influence on the classification. Conversely, if $|w_i^{\text{opt}}|$ is relatively large, the ith feature will play an important role in the classifier (Fig. 3).

The coefficients of two different OMH obtained with two different SVM comparisons cannot be compared directly. Let $\mathcal{S}_1, \mathcal{S}'_1, \mathcal{S}_2$ and $\mathcal{S}'_2$ be four groups of subjects. Let $\mathbf{w}^{(1)\text{opt}}$ be the hyperplane coefficients obtained with an SVM analysis of groups $\mathcal{S}_1$ and $\mathcal{S}'_1$. Similarly, we defined $\mathbf{w}^{(2)\text{opt}}$ as the OMH coefficients obtained with an SVM analysis of groups $\mathcal{S}_2$ and $\mathcal{S}'_2$. Note that, if the separation between groups $\mathcal{S}_1$ and $\mathcal{S}'_1$ is larger than the separation between groups $\mathcal{S}_2$ and $\mathcal{S}'_2$, then:

$$\|\mathbf{w}^{(1)\text{opt}}\| \leqslant \|\mathbf{w}^{(2)\text{opt}}\|$$

To sum up, there are two opposite effects:

(i) For a given comparison, the ith feature will have more importance than feature j if and only if: $w_i^{\text{opt}} > w_j^{\text{opt}}$.
(ii) If the separation between groups $\mathcal{S}_1$ and $\mathcal{S}'_1$ is larger than the separation between groups $\mathcal{S}_2$ and $\mathcal{S}'_2$, then: $\|\mathbf{w}^{(1)\text{opt}}\| \leqslant \|\mathbf{w}^{(2)\text{opt}}\|$

Therefore, one cannot compare directly the weights $|w_i^{\text{opt}}|$.

SVM search for the hyperplane for which the margin between groups is maximal (Fig. 1). The margin m is large when there is a large separation between two groups (Fig. 3). Note that, for the standard linear SVM (1), the margin, m, can be written as Schölkopf and Smola (2001):

$$m = \frac{2}{\|\mathbf{w}^{\text{opt}}\|} \tag{13}$$

and for the spatially regularized SVM (5):

$$m = \frac{2}{\left\|e^{\frac{1}{2}\beta L}\mathbf{w}^{\text{opt}}\right\|} \tag{14}$$

Thus, by combining m and $\frac{|w_i^{\text{opt}}|}{\|\mathbf{w}^{\text{opt}}\|}$, one can simultaneously quantify the separation between groups and the relative influence of the different features. Therefore, we propose to analyze the statistic of:

$$\frac{m|w_i^{\text{opt}}|}{\|\mathbf{w}^{\text{opt}}\|} \tag{15}$$

We performed permutation tests on $\frac{m|w_i^{\text{opt}}|}{\|\mathbf{w}^{\text{opt}}\|}$ under the null hypothesis $\mathcal{H}_0$ corresponding to "no relationship between the class labels and the MR scan". By randomly permuting the subjects labels 20,000 times and training the SVM with this permutation of labels, we estimated for each voxel i the probability distribution of $\frac{m|w_i^{\text{opt}}|}{\|\mathbf{w}^{\text{opt}}\|}$ under $\mathcal{H}_0$. Based on these distributions, it is possible to test $\mathcal{H}_0$ at the voxel level. The false discovery rate (FDR) was used to correct for multiple comparisons (Soric, 1989; Benjamini and Hochberg, 1995). To the best of our knowledge, other statistical analyses of the OMH (Mourão-Miranda et al., 2005; Wang et al., 2007; Sato et al., 2009) did not take the margin into account.

4. Experiments and results

We first tested our method on 2D synthetic images. We then applied it on real data for the detection of brain areas associated with three-month stroke outcome based on diffusion weighted MRI acquired at the acute stage.

4.1. Synthetic images

4.1.1. Materials

We first evaluated the ability of the method to detect artificial differences between two groups of twenty 2D synthetic images (Fig. 4) – 116×92 with 1.5 mm isotropic voxels – which were constructed as follows. We considered a slice of a WM template. The WM template was constructed as follows. The images of 509 subjects randomly selected from the ADNI database were segmented into GM, WM and CSF using the SPM5 unified segmentation (Ashburner and Friston, 2005) and spatially normalized using DARTEL (Ashburner, 2007). The normalized WM maps were then averaged. We used the slice corresponding to $z = 15$ mm in the MNI space.

For each of the 40 images, the voxels of the WM were assigned a random number between zero and one, the intensity of the other voxels being set to zero. In each image of the second group, we constructed hyperintense areas (Fig. 5.a). The hyperintensities h_{green} of the green region and h_{red} of the red region were

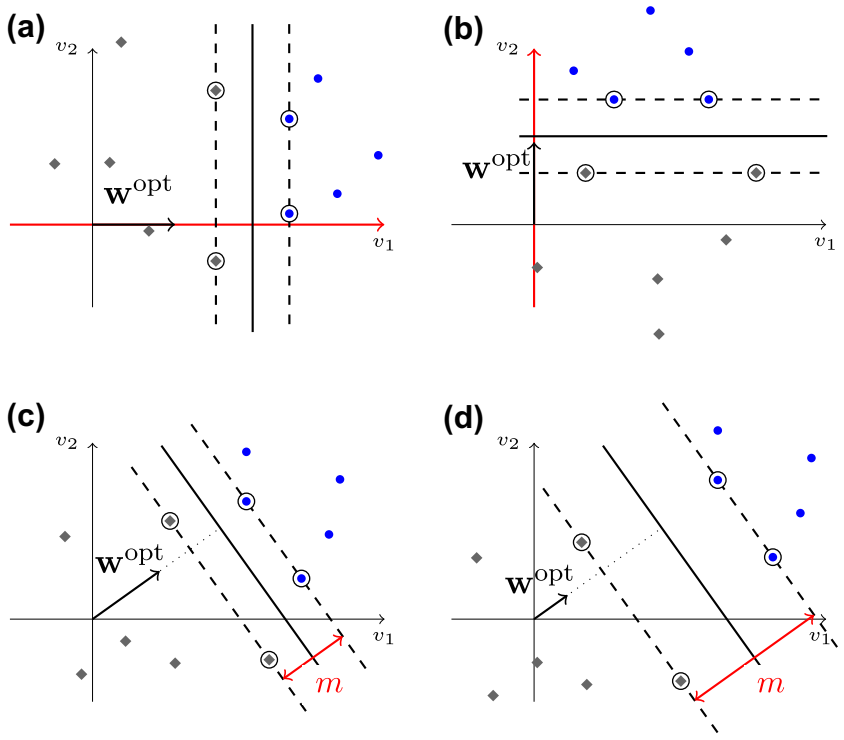

Fig. 3. Illustration of the discriminative information of the OMH and the margin. Upper row (a and b): The OHM coefficients informs about the relative importance of the features. Lower row (c and d): the margin m quantifies the difference between two groups.

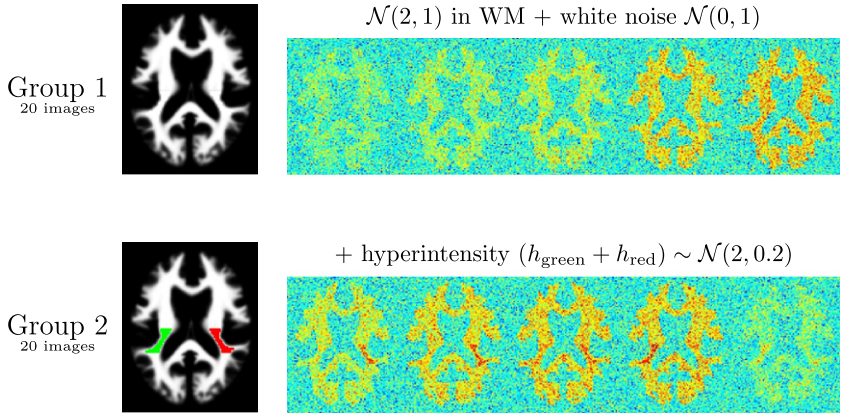

Fig. 4. Synthetic example. Upper row: from left to right, the WM template and five randomly selected images from the first group with no hyperintensities (control). Lower row: from left to right, the WM template with the hyperintense regions to detect (in red and green) and five randomly selected images from the second group with hyperintensities. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

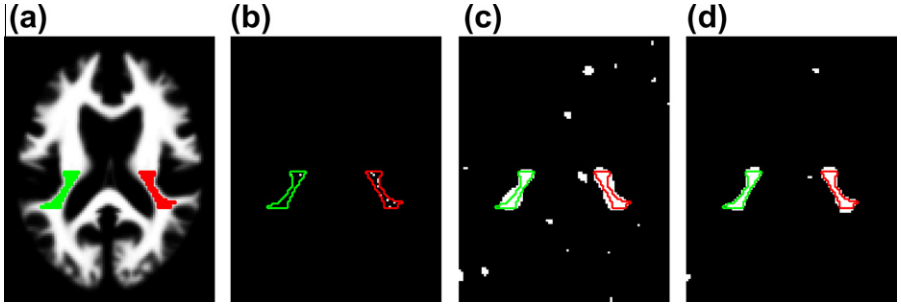

Fig. 5. Synthetic example: (a) WM template and the hyperintense regions to detect (in red and green); (b) detection with a linear SVM on raw images; (c) detection with a linear SVM on smoothed images; (d) detection with a spatially regularized SVM on raw images. All univariate analyses detected no difference. Detected voxels are in white. Contours of the simulated hyperintensities are in red and green. All tests were corrected for multiple comparisons using the false discovery rate ($q = 0.05$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Demographic characteristics of the study population. Values are indicated as mean ± standard-deviation [range]. mRS: modified Rankin Scale. Differences between groups were assessed using Student's *t*-test.

Outcome	*good* mRS 0–2	*poor* mRS 3–5	
Number	39	33	
Age (years)	55.8 ± 15.0 [24–81]	64.9 ± 10.9 [38–80]	$p = 0.002$
mRS (3 months)	1.2 ± 0.7 [0–2]	4.0 ± 0.8 [3–5]	$p < 0.0001$
NIHSS (<6 h)	11.4 ± 5.4 [2–23]	18.5 ± 4.9 [6–30]	$p < 0.0001$
NIHSS (day one)	5.5 ± 4.1 [0–14]	16.8 ± 6.0 [6–35]	$p < 0.0001$
Time to initial MRI (min)	152 ± 66 [66–360]	164 ± 63 [66–341]	$p = 0.2$

constructed so that they follow $(h_{\text{green}} + h_{\text{red}}) \sim \mathcal{N}(2, 0.2)$. Gaussian white noise $(\mathcal{N}(0,1))$ was added to all images.

4.1.2. Analyses

We tested six methods: three univariate methods and three SVM methods. We performed three univariate analyses on the voxel intensities: on the raw images, on the images smoothed with a Gaussian kernel and on the images pre-processed by $e^{-\frac{\beta}{2}L}$ (where L was the Laplacian of the graph used in the spatially regularized SVM).

We tested three SVM methods: the standard linear SVM on the raw images, the standard linear SVM on the smoothed images and the spatially regularized SVM on the raw images.

All tests were corrected for multiple comparisons with a 5% FDR. The C parameter of the SVM was fixed to one $\lambda = \frac{1}{2NC}$ (Schölkopf and Smola, 2001). The full width at half maximum (FWHM) of the Gaussian smoothing kernel was set to three voxels (the red and green regions' widths).

β controls the scale of the spatial regularization. As mentioned in paragraph Subsection 2.2.2, using the image connectivity graph as a regularization graph would be equivalent to smoothing the data with a Gaussian kernel with standard deviation $\sigma = \sqrt{\beta}$.

Therefore, for being able to compare the results obtained with the spatial regularization and with the Gaussian smoothing, we chose β as the square of the standard deviation of the Gaussian smoothing kernel ($\beta = \sigma^2$).

4.1.3. Results

The group differences detected on the 2D synthetic images are shown on Fig. 5. The three univariate analyses detected no difference. The SVM on raw images detected only very few voxels. On smoothed images, it detected both regions, but detected also many scattered clusters outside the hyperintense regions. The spatially regularized SVM also detected both hyperintense regions and substantially decreased the number of scattered clusters.

4.2. Brain areas associated with stroke outcome

We applied our method to the detection of brain areas associated with three-month stroke outcome based on diffusion weighted MRI acquired at the acute stage.

4.2.1. Subjects

Consecutive patients meeting the following criteria were included to participate in the study: (1) ischemic stroke in carotid territory, (2) initial 1.5T MRI with DWI performed within the first six hours following stroke onset, (3) control MRI with DWI performed within the next 3 days, and (4) clinical assessment by the modified Rankin scale (mRS) at three months. Exclusion criteria were symptomatic haemorrhagic transformation or death during follow-up (90 days). As a result, 72 consecutive acute stroke patients (mean age: 60 ± 14 years [24–81]) were included in this study. The patients could receive intravenous rtPA (recombinant tissue plasminogen activator) within a 5-h time window according to the routine clinical procedure at our institution. In this procedure, intravenous rtPA is given based on clinical and MRI criteria including baseline National Institutes of Health Stroke Scale (NIHSS) > 4 without major improvement, acute cerebral ischemia detected from MRI data, an exclusion of hemorrhage, and evidence of an intracranial occlusion. The neurological examination was assessed using NIHSS at admission and at day one. The modified Rankin Scale (mRS) was used to assess outcome at 90 days. Good outcome was defined as independency (mRS 0, 1 or 2; 39 subjects) and poor outcome as severe disability (mRS 3–5; 33 subjects). The demographic characteristics of the study population are reported in Table 1.

All imaging and clinical data were obtained during routine clinical workup of the patients in our stroke center. Therefore, according to the French legislation, explicit informed consent was waived. The study was approved by the La Pitié-Salpêtrière Hospital Ethics Committee.

4.2.2. MRI acquisition and preprocessing

MR imaging was performed using a 1.5 Tesla MR unit (General Electric Signa Horizon Echospeed) with enhanced gradient hardware for echoplanar imaging. The analysis was performed on the control MRI. The median delay between stroke onset and MRI acquisition was 1.2 day (lower and upper quartiles: 1.1–1.8). We performed three sequences in our MRI protocol: DWI, Fluid Attenuated Inversion Recovery (FLAIR), and an intra-cranial time-of-flight MRA. The parameters of the axial DWI spin echo EPI were: 24 slices, 2825 ms repetition time (TR), 98.9 ms echo time (TE), 90° flip angle, field-of-view (FOV) of $280 \times 210\ \text{mm}^2$, 96×64 matrix, 5 mm slice thickness, and 0.5 mm interslice gap. A baseline T2 image and three diffusion-weighted images in the x, y, and z directions using a b-value of $1000\ \text{s mm}^{-2}$ were acquired within 40 s.

The quantitative ADC maps were generated using commercially available software (Functool 2, General Electric, Buc, France). The ADC maps were normalized to the Montreal Neurological Institute (MNI) reference frame using the T2-weighted template from SPM5 (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, London, UK). The spatially normalized images had 2 mm isotropic voxels. The T2-weighted template was made symmetric by averaging the template with its mirrored image. To put all the lesions on the same side, ADC maps with the infarct lesion in the left hemisphere were flipped with respect to the interhemispheric plane for the analyses.

4.2.3. Analyses

Group analyses between patients with good outcome and patients with poor outcome were performed on ADC maps acquired one day after the stroke onset. Group differences were assessed using the spatially regularized SVM with a tissue dependent regularization on raw images. For comparison, we also performed univariate analyses with both a permutation test and a parametric Student's *T*-test on smoothed images (8-mm FWHM Gaussian filter). To construct the graph, we used the gray matter, white matter and CSF templates provided with SPM5. As in the

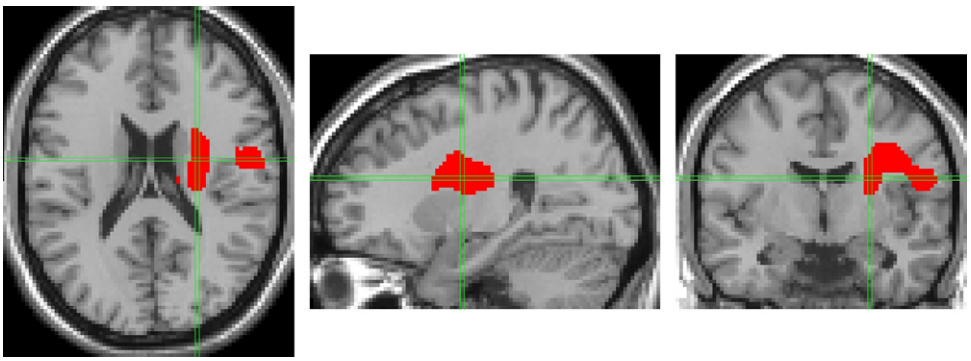

Fig. 6. Statistically significant differences of the ADC maps obtained one day after the stroke onset between poor and good motor outcome with a spatially regularized SVM superimposed on the SPM T_1 single subject template (z = 20 mm, x = 28 mm and y = −8 mm in the MNI-space). Tests were corrected for multiple comparisons using the false discovery rate (q = 0.05).

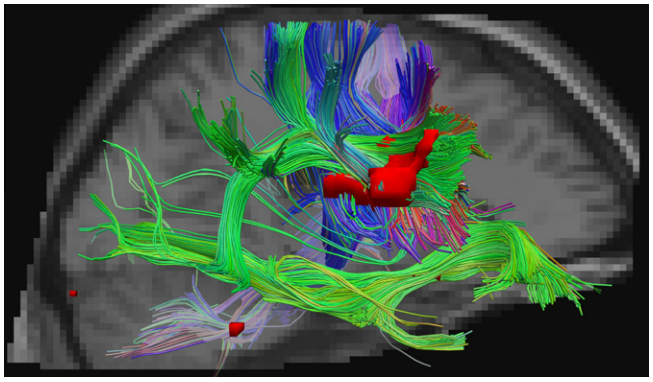

Fig. 7. In red: regions detected superimposed on white matter tracts of a single control subject computed from diffusion spectral imaging (DSI) data using Diffusion Toolkit (http://www.trackvis.org/dtk/). The colors of the fiber track code for the orientation.

previous section, the β parameter was chosen to correspond to the FWHM of the univariate analyses. All tests were corrected for multiple comparisons using the false discovery rate (q = 0.05).

4.2.4. Results

The group analyses between stroke patients with good and poor outcome based on ADC maps acquired at the acute stage (one day after the stroke onset) led to the following results. No difference was detected by the univariate analyses. The spatially regularized SVM detected significant ADC changes between patients with good and poor outcome (Fig. 6).

The changes detected one day after the onset involved a large cluster (1265 voxels) that included the periventricular white matter, the posterior part of the internal capsule, the posterior part of the putamen and the lower part of the primary motor cortex (Figs. 6 and 7). Changes were also detected in some small clusters located in the insula (37 voxels), the uncinate fasciculus (21 voxels) and the contralateral medulla (25 voxels).

While the individual classification of patients was not the objective of this paper, we compared the classification accuracy of the spatially regularized SVM to that of a standard linear SVM. The spatially regularized SVM was as accurate as the linear SVM (76% accuracy).

5. Discussion

In this paper, we proposed a new method based on spatially regularized SVM to detect group differences in brain images. Spatial consistency was directly included in the SVM by using the graph Laplacian. The proposed approach was applied to the detection of early ADC changes associated with three-month stroke outcome. The ADC changes were localized in the corticospinal tract, the periventricular white matter and in the lower motor cortex that were undetected by univariate group analyses.

Group analyses of brain images have widely relied on the classical mass-univariate approach. However, sensitivity of univariate approaches is limited when the differences are spatially complex and involve a combination of different voxels or brain structures (Davatzikos, 2004). In our study, in both the synthetic tests and the application to stroke patients, the spatially regularized SVM allowed unveilling differences that were undetected by univariate analyses. More specifically, in the synthetic example, the lesions or hyperintensities constructed in the image of the second group were distributed between two different regions, the red and the green ones. As a result, the differences between the first (without any hyperintensity) and the second group (with hyperintensity) were too subtle in each voxel, for the differences to be detected by a univariate analysis, considering the sample size. Only the combination of both regions was discriminative. Such group differences were only detected using SVM-based group analyses.

In such approaches, the hyperplane coefficients are analyzed to detect significant group differences between populations. However, like other voxel-based comparisons, SVM-based analyses are subject to noise, registration errors and interindividual variability. This is due to the fact that the regularization term of the standard linear SVM is not a spatial regularization term. Thus, the map $\mathbf{w}^{\text{opt}}$ obtained with a standard linear SVM can be noisy and scattered. In the synthetic example white noise was added to images. As a consequence, only very few voxel were detected by the standard linear SVM. Gaussian smoothing is therefore often used as a preprocessing step. In the synthetic exemple, since the noise was unstructured, the smoothing step partially removed it, which lead to a better detection with a linear SVM. However such preprocessing step does not fully take into account anatomical information about the data. For instance, it mixes white matter with gray matter voxels. In this contribution, we proposed instead to add to the SVM a spatial regularization which accounts for tissue types.

The regularization was based on the graph Laplacian. The graph framework has often been used in image processing (Geman and Geman, 1984; Shi and Malik, 2000) to encode spatial priors. More generally, weighted graphs allow modeling the voxels' proximity. In this paper, this prior information was introduced into the SVM using a spatial regularization. This regularization exponentially penalized the high frequency components of the OMH with respect to the graph. In other words, it constrained the classifier to

consider as similar voxels highly connected according to the graph adjacency matrix. Such regularization leads to a diffusion kernel on a graph. Graph diffusion kernels for SVM were introduced by Kondor and Lafferty (2002). They belong to the family of spectral regularization kernels (Smola and Kondor, 2003). Spectral regularization kernels have been used in satellite imaging (Gómez-Chova et al., 2008) and bioinformatics (Vert and Kanehisa, 2003; Lanckriet et al., 2004; Tsuda and Noble, 2004). However, our approach differs from these diffusion kernels: in our case, the nodes of the graph are the features, here the voxels, whereas in these other papers, the nodes were the objects to classify. To our knowledge, kernels similar to those used in this study have not been used in neuroimaging but only for the classification of microarray data (Rapaport et al., 2007.)

The penalization used in this study was exponential, which led to the diffusion kernel. Many other penalization functions, such as the regularized Laplacian, the p-Step Random Walk or the Inverse Cosine (Smola and Kondor, 2003) for instance, could have been used instead of the diffusion process. Nevertheless using the diffusion process as a penalization function extends the widely used framework which consists in smoothing the data with a Gaussian kernel as a preprocessing step.

The proposed approach was applied to the detection of brain areas associated with stroke outcome based on DWI acquired at the acute stages. It allowed detecting ADC changes that were localized in the corticospinal tract (CST), the periventricular white matter, the posterior part of the putamen and part of the primary motor cortex. Univariate analyses failed to detect any difference.

The implication of the CST, obtained at the acute stage using ADC measures, was in line with previous DTI studies carried out at the subacute (Konishi et al., 2005; Thomalla et al., 2005; Cho et al., 2007; Kunimatsu et al., 2007; Nelles et al., 2007; Jang et al., 2008; Domi et al., 2009; Yu et al., 2009) or chronic phase (Feydy et al., 2002; Wenzelburger et al., 2005; Newton et al., 2006; Stinear et al., 2007; Schaechter et al., 2008). Previous functional imaging (fMRI) studies have demonstrated that recovery is mainly related to the reorganization of the preserved cortical motor network of peri-infarcted, motor-related areas (Jaillard et al., 2005) and homologous areas in the intact contralateral hemisphere (Rijntjes, 2006). In this study, significant ADC changes were found mainly in sub-cortical regions in addition to a smaller part of the primary motor cortex. Damages in subcortical regions, especially in the CST and the periventricular white matter, could be more directly related to stroke outcome than cortical regions for ischemic stroke in carotid territory (Rosso et al., 2011; DeVetten et al., 2010). Indeed, we may hypothesize that, when the main motor outflow tract is interrupted, motor commands cannot be conveyed to the spinal cord and therefore reorganization at the cortical level remains insufficient to the patient recovery. This hypothesis is supported by Seitz et al. (2009), where periventricular white matter damages were associated with a poor patient outcome. The authors have underlined that periventricular white matter is the junction where many white matter tracts are crossed over. Additionally, a number of previous studies using either standard MRI or CT-scans (Feydy et al., 2002; Wenzelburger et al., 2005) or measures of other diffusion variables (Konishi et al., 2005; Thomalla et al., 2005; Nelles et al., 2007; Jang et al., 2008; Domi et al., 2009) at the subacute or chronic stages have shown that CST lesions correlate to motor outcome. We have extended these results at the very acute stage using ADC changes at day one and using a global measure of outcome. These results could provide useful information to guide therapies and rehabilitation programs at an early stage. They could also allow giving more precise information to relatives. Finally, they could be used in stratification or design of future clinical trials.

This study has the following limitations. In addition to that large cluster involving the CST, smaller clusters were also detected in the regions of the insula, the uncinate fasciculus and the contralateral medulla. These smaller clusters may result from an artifact of the method, in particular given that the false discovery rate was used to correct for multiple comparisons. Moreover, it should be noted that the location of these clusters can only be approximately described because DWI data acquired at the acute stage is prone to misregistrations and has a large voxel size (5 mm). External validations on other groups of patients are needed to assess whether these smaller clusters are artefactual or not. Besides, the method in its present form does not include covariates in the analysis. For this reason, age and sex were not included as covariates. Nevertheless, detected differences were strongly lateralized on the side of the lesion. This suggests pathology-related changes rather than age-related changes. Future work will focus on the inclusion of covariates in the spatially regularized SVM.

In conclusion, we proposed a method based on spatially regularized SVM to study differences between populations. When applied to the analysis of brain areas associated with stroke outcome, it detected early ADC changes mainly localized in the corticospinal tract and the periventricular white matter that could not be detected using standard univariate analyses. The proposed approach is not specific to diffusion MRI or stroke patients, and can be applied to other types of data (e.g. anatomical MRI) and other pathologies (e.g. neurodegenerative disorders). It has the potential to overcome the limits of traditional mass univariate voxelwise analyses by detecting complex spatial patterns of alterations.

Acknowledgments

This work was partially supported by the "Programme Hospitalier de Recherche Clinique EVAL-USINV" (No. AOM 03 008).

This work was partially supported by ANR (Project HM-TC, No. ANR-09-EMER-006).

References

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. NeuroImage 38, 95–113.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry – the methods. NeuroImage 11, 805–821.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–851.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57, 289–300.

Chalela, J.A., Kidwell, C.S., Nentwich, L.M., Luby, M., Butman, J.A., Demchuk, A.M., Hill, M.D., Patronas, N., Latour, L., Warach, S., 2007. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. The Lancet 369, 293–298.

Cho, S.H., Kim, D.G., Kim, D.S., Kim, Y.H., Lee, C.H., Jang, S.H., 2007. Motor outcome according to the integrity of the corticospinal tract determined by diffusion tensor tractography in the early stage of corona radiata infarct. Neuroscience Letters 426, 123–127.

Chung, F.R.K., 1992. Spectral Graph Theory 1992, AMS.

Crafton, K.R., Mark, A.N., Cramer, S.C., 2003. Improved understanding of cortical injury by incorporating measures of functional anatomy. Brain 126, 1650–1659.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., 2011. The Alzheimer's disease neuroimaging initiative, Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. NeuroImage 56, 766–781.

Cuingnet, R., Rosso, C., Lehéricy, S., Dormont, D., Benali, H., Samson, Y., Colliot, O., 2010. Spatially regularized SVM for the detection of brain areas associated with stroke outcome. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010. Lecture Notes in Computer Science, vol. 6361. Springer, Berlin/Heidelberg, pp. 316–323.

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

Decoste, D., Schölkopf, B., 2002. Training invariant support vector machines. Machine Learning 46, 161–190.

DeVetten, G., Coutts, S.B., Hill, M.D., Goyal, M., Eesa, M., O'Brien, B., Demchuk, A.M., Kirton, A.for the MONITOR and VISION study groups, 2010. Acute corticospinal tract Wallerian degeneration is associated with stroke outcome. Stroke 41, 751–756.

Domi, T., deVeber, G., Shroff, M., Kouzmitcheva, E., MacGregor, D.L., Kirton, A., 2009. Corticospinal tract pre-Wallerian degeneration: a novel outcome predictor for pediatric stroke on acute MRI. Stroke 40, 780–787.

Fan, Y., Shen, D., Gur, R., Gur, R., Davatzikos, C., 2007. COMPARE: classification of morphological patterns using adaptive regional elements. IEEE Transactions on Medical Imaging 26, 93–105.

Feydy, A., Carlier, R., Roby-Brami, A., Bussel, B., Cazalis, F., Pierot, L., Burnod, Y., Maier, M., 2002. Longitudinal study of motor recovery after stroke: recruitment and focusing of brain activation. Stroke 33, 1610–1617.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Gómez-Chova, L., Camps-Valls, G., Muñoz-Marí, J., Calpe, J., 2008. Semi-supervised image classification with Laplacian support vector machines. IEEE Geoscience and Remote Sensing Letters 5, 336–340.

Haris, K., Efstratiadis, S., Maglaveras, N., Katsaggelos, A., 1998. Hybrid image segmentation using watersheds and fast region merging. IEEE Transactions on Image Processing 7, 1684–1699.

Jaillard, A., Martin, C.D., Garambois, K., Lebas, J.F., Hommel, M., 2005. Vicarious function within the human primary motor cortex?: a longitudinal fMRI stroke study. Brain 128, 1122–1138.

Jang, S.H., Bai, D., Son, S.M., Lee, J., Kim, D.S., Sakong, J., Kim, D.G., Yang, D.S., 2008. Motor outcome prediction using diffusion tensor tractography in pontine infarct. Annals of Neurology 64, 460–465.

Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S.J., 2008a. Automatic classification of MR scans in Alzheimer's disease. Brain 131, 681–689.

Kondor, R.I., Lafferty, J.D., 2002. Diffusion kernels on graphs and other discrete input spaces. In: Proceedings of the International Conference on Machine Learning, pp. 315–322.

Konishi, J., Yamada, K., Kizu, O., Ito, H., Sugimura, K., Yoshikawa, K., Nakagawa, M., Nishimura, T., 2005. MR tractography for the evaluation of functional recovery from lenticulostriate infarcts. Neurology 64, 108–113.

Kunimatsu, A., Itoh, D., Nakata, Y., Kunimatsu, N., Aoki, S., Masutani, Y., Abe, O., Yoshida, M., Minami, M., Ohtomo, K., 2007. Utilization of diffusion tensor tractography in combination with spatial normalization to assess involvement of the corticospinal tract in capsular/pericapsular stroke: feasibility and clinical implications. Journal of Magnetic Resonance Imaging 26, 1399–1404.

Lanckriet, G., Deng, M., Cristianini, N., Jordan, M., Noble, W., 2004. Kernel-based data fusion and its application to protein function prediction in yeast. In: Proceedings of the Pacific Symposium on Biocomputing, pp. 300–311.

Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C., 2004. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. NeuroImage 21, 46–57.

Lo, R., Gitelman, D., Levy, R., Hulvershorn, J., Parrish, T., 2010. Identification of critical areas for motor function recovery in chronic stroke subjects using voxel-based lesion symptom mapping. NeuroImage 49, 9–18.

Moler, C., Loan, C.V., 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Review 45, 3–49.

Mourão-Miranda, J., Bokde, A.L., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. NeuroImage 28, 980–995, Special Section: Social Cognitive Neuroscience.

Nelles, M., Gieseke, J., Flacke, S., Lachenmayer, L., Schild, H., Urbach, H., 2007. Diffusion tensor pyramidal tractography in patients with anterior choroidal artery infarcts. AJNR: American Journal of Neuroradiology 29, 488–493.

Newton, J.M., Ward, N.S., Parker, G.J.M., Deichmann, R., Alexander, D.C., Friston, K.J., Frackowiak, R.S.J., 2006. Non-invasive mapping of corticofugal fibres from multiple motor areas–relevance to stroke recovery. Brain 129, 1844–1858.

Peleg, S., 1980. A new probabilistic relaxation scheme. IEEE Transactions on Pattern Analysis and Machine Intelligence 2, 362–369.

Querbes, O., Aubry, F., Pariente, J., Lotterie, J.A., Demonet, J.F., Duret, V., Puel, M., Berry, I., Fort, J.C., Celsis, P.the Alzheimer's Disease Neuroimaging Initiative, 2009. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain 132, 2036–2047.

Rapaport, F., Zinovyev, A., Dutreix, M., Barillot, E., Vert, J., 2007. Classification of microarray data using gene networks. BMC Bioinformatics 8, 35.

Rijntjes, M., 2006. Mechanisms of recovery in stroke patients with hemiparesis or aphasia: new insights, old questions and the meaning of therapies. Current Opinion in Neurology 19, 76.

Rosso, C., Colliot, O., Pires, C., Delmaire, C., Valabrègue, R., Crozier, S., Dormont, D., Baillet, S., Samson, Y., Lehéricy, S., 2011. Early ADC changes in motor structures predict outcome of acute stroke better than lesion volume. Journal of Neuroradiology 38, 105–112.

Rosso, C., Hevia-Montiel, N., Deltour, S., Bardinet, E., Dormont, D., Crozier, S., Baillet, S., Samson, Y., 2009. Prediction of infarct growth based on apparent diffusion coefficients: penumbral assessment without intravenous contrast material. Radiology 250, 184–192.

Sato, J.R., Fujita, A., Thomaz, C.E., da Graca Morais Martin, M., Mourão-Miranda, J., Brammer, M.J., Junior, E.A., 2009. Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction. NeuroImage 46, 105–114.

Schaechter, J.D., Perdue, K.L., Wang, R., 2008. Structural damage to the corticospinal tract correlates with bilateral sensorimotor cortex reorganization in stroke patients. NeuroImage 39, 1370–1382.

Schölkopf, B., Burges, C., Vapnik, V., 1996. Incorporating invariances in support vector learning machines. In: Proceedings of the 1996 International Conference on Artificial Neural Networks. Springer Verlag, p. 47.

Schölkopf, B., Simard, P., Smola, A., Vapnik, V., 1998. Prior knowledge in support vector kernels. In: Proceedings of Conference on Advances in Neural Information Processing Systems'97. MIT Press, pp. 640–646.

Schölkopf, B., Smola, A.J., 2001. Learning with Kernels. MIT Press.

Seitz, R., Sondermann, V., Wittsack, H., Siebler, M., 2009. Lesion patterns in successful and failed thrombolysis in middle cerebral artery stroke. Neuroradiology 51, 865–871.

Shawe-Taylor, J., Cristianini, N., 2000. Support Vector Machines and other Kernel-based Learning Methods. Cambridge University Press.

Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press.

Shen, D., Davatzikos, C., 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Transaction on Medical Imaging 21, 1421–1439.

Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 888–905.

Smola, A., Kondor, R., 2003. Kernels and regularization on graphs. In: Proceedings of the 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel. Springer Verlag, p. 144.

Smola, A.J., Schölkopf, B., 1998. On a kernel-based method for pattern recognition, regression, approximation, and operator inversion. Algorithmica 22, 211–231.

Soric, B., 1989. Statistical "discoveries" and effect-size estimation. Journal of the American Statistical Association 84, 608–610.

Stinear, C.M., Barber, P.A., Smale, P.R., Coxon, J.P., Fleming, M.K., Byblow, W.D., 2007. Functional potential in chronic stroke patients depends on corticospinal tract integrity. Brain 130, 170–180.

Thomalla, G., Glauche, V., Weiller, C., Röther, J., 2005. Time course of Wallerian degeneration after ischaemic stroke revealed by diffusion tensor imaging. Journal of Neurology, Neurosurgery & Psychiatry 76, 266–268.

Tsuda, K., Noble, W.S., 2004. Learning kernels from biological networks by maximizing entropy. Bioinformatics 20, i326–333.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer-Verlag.

Vemuri, P., Gunter, J.L., Senjem, M.L., Whitwell, J.L., Kantarci, K., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack, C.R., 2008. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. NeuroImage 39, 1186–1197.

Vert, J., Kanehisa, M., 2003. Graph-driven feature extraction from microarray data using diffusion kernels and kernel CCA. Advances in Neural Information Processing Systems, 1449–1456.

Wang, Z., Childress, A.R., Wang, J., Detre, J.A., 2007. Support vector machine learning-based fMRI data group analysis. NeuroImage 36, 1139–1151.

Wenzelburger, R., Kopper, F., Frenzel, A., Stolze, H., Klebe, S., Brossmann, A., Kuhtz-Buschbeck, J., Golge, M., Illert, M., Deuschl, G., 2005. Hand coordination following capsular stroke. Brain 128, 64–74.

Yu, C., Zhu, C., Zhang, Y., Chen, H., Qin, W., Wang, M., Li, K., 2009. A longitudinal diffusion tensor imaging study on Wallerian degeneration of corticospinal tract after motor pathway stroke. NeuroImage 47, 451–458.