

# SVM WITH FEATURE SELECTION AND SMOOTH PREDICTION IN IMAGES: APPLICATION TO CAD OF PROSTATE CANCER.

Émilie Niaf<sup>1,2,3</sup>, Rémi Flamary<sup>4</sup>, Alain Rakotomamonjy<sup>5</sup>, Olivier Rouvière<sup>2</sup>, Carole Lartizien<sup>1</sup>

<sup>1</sup>Université de Lyon, CREATIS; CNRS UMR5220; INSERM U1044; INSA-Lyon; Université Lyon 1, France

<sup>2</sup>INSERM, U1032, LabTau, Lyon, F-69003, France; Université de Lyon, Lyon, F-69003, France

<sup>3</sup>A\*STAR-NUS Clinical Imaging Research Centre, Singapore

<sup>4</sup>Laboratoire Lagrange, UMR 7293, Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, Nice, France

<sup>5</sup>LITIS, EA 4108, Université de Rouen, 76801 Saint Etienne du Rouvray, France

## ABSTRACT

We propose a new computer-aided detection scheme for prostate cancer screening on multiparametric magnetic resonance (mp-MR) images. Based on an annotated training database of mp-MR images from thirty patients, we train a novel support vector machine (SVM)-inspired classifier which simultaneously learns an optimal linear discriminant and a subset of predictor variables (or features) that are most relevant to the classification task, while promoting spatial smoothness of the malignancy prediction maps. The approach uses a  $\ell_1$ -norm in the regularization term of the optimization problem that rewards sparsity. Spatial smoothness is promoted via an additional cost term that encodes the spatial neighborhood of the voxels, to avoid noisy prediction maps. Experimental comparisons of the proposed  $\ell_1$ -Smooth SVM scheme to the regular  $\ell_2$ -SVM scheme demonstrate a clear visual and numerical gain on our clinical dataset.

**Index Terms**— Support vector machine, Spatial regularization,  $\ell_1$ -norm, Computer-aided diagnostic, MRI.

## 1. INTRODUCTION

Prostate cancer (PCa) is one of the leading causes of cancer-related death for men worldwide [1]. Until now, random biopsies remain the gold standard technique to detect PCa, but is invasive and imprecise. Radiologists are therefore exploring the performance of multiparametric magnetic resonance (mp-MR) imaging combining various MR sequences to target biopsies towards suspicious areas. However, integrating such a large amount of visual information is a complex task, all the more challenging as PCa and benign tissues may look different in one MR sequence and similar in another.

Over the last decade, computer-aided diagnosis (CAD) systems have been proposed to assist radiologists in screening and diagnosing by highlighting suspicious areas and providing an objective and reproducible confidence index. The prototyped CAD systems [2–10] produce probability maps of malignancy based on a set of numerical features extracted from mp-MR images. Most rely on supervised classification strategies that often outperform unsupervised ones [6]. Various classifiers have been introduced and compared, such

as linear discriminant analysis [2, 10], k-nearest neighbours [3, 10], logistic regression [4], random forest [5], regression vector machine [6], Bayes classifier [3, 5, 10] and boosting [3, 8], but most studies rely on the support vector machine (SVM) classifier [2, 6–11].

One of the main weaknesses of these studies is that the classification procedures are performed on a per-voxel basis and do not take into account the spatial *a priori* inherent to image data and to the spatial coherence of the biological tissue structures - often larger than a single voxel element in size. As shown in [5] and [6], this leads to noisy prediction maps even if good overall performances are achieved. To improve readability, these predicted maps need to be post-processed using, for instance, morphology operations, filtering or a Markov random field approach as proposed in [12] for hyperspectral data.

Moreover, in the literature, a large set of structural [2, 5, 8, 10, 13] and functional [4, 6, 7, 9, 10] features have been used in combination. However, it is difficult to discriminate the features relevant in the classification task from those which may introduce errors or lead to overfitting. Reducing the feature set to its more discriminative elements would decrease the overall computational time. Few studies have tackled this issue by, for example, incorporating a pre-selection step in the classification scheme using filtering criteria [10] or introducing an embedding step to reduce the feature space without clearly discriminating between features [5].

We propose a new classification scheme for PCa discrimination, incorporating an efficient feature selection strategy and promoting spatial smoothness of the predicted map. Following our preliminary work [14], we reformulate the regular SVM optimization problem by introducing both a  $\ell_1$ -norm into the regularization term and an additional cost term to penalize neighbouring prediction discrepancies. Including the spatial *a priori* that neighbouring voxels have similar class do not require any additional annotation of the data. This even allows the use of structural information held by unlabeled voxels (or images) that would, in a regular supervised approach, simply be discarded. This is particularly interesting in a clinical context where the radiologist may not exhaustively delineate PCa lesions over the whole set of images to

define the ground truth needed to train a classifier. Instead, so as to save time, the radiologist outlines only some of the signal abnormalities and roughly delineates the hotspot of the lesions. Section 2 describes the proposed algorithm, which is then evaluated on a clinical database of PCa mp-MR images in Section 3.

## 2. SVM WITH FEATURE SELECTION AND SPATIAL SMOOTHNESS

### 2.1. Problem description and Support Vector Machine

The goal, in a binary supervised learning problem such as SVM, is to learn how to discriminate between  $N$  training examples  $(\mathbf{x}_i)_{i=1\dots N}$  from two classes (herein labeled, without loss of generality, as  $y_i=-1$  or  $y_i=1$ ) on the basis of  $d$  observed predictor variables (or *features*)  $\mathbf{x}_i \in \mathbb{R}^d$ . To do so, we are given a training dataset  $\mathcal{L} := \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^d, y_i \in \{-1, 1\}\}_{i=1\dots n}$ , of  $n$  labeled examples, where  $y_i$  is the label of example  $\mathbf{x}_i$ . In the following, we first recall the basics of regular SVM as defined by Vapnik [15] before introducing our proposed classification scheme joining feature selection and spatial regularization.

SVM aims at constructing a linear discrimination function of the form  $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$  with  $\mathbf{w} = [w_1, \dots, w_d] \in \mathbb{R}^d$  the normal vector to the separating hyperplane  $f(\mathbf{x}) = 0$  and  $b \in \mathbb{R}$  is a bias term. The associated pattern recognition problem is defined as :

$$\min_f \sum_{i \in \mathcal{L}} H(y_i, f(\mathbf{x}_i)) + \lambda_r \Omega_r(f) \quad (1)$$

where  $\Omega_r(f)$  is the regularization term,  $\sum_i H(y_i, f(\mathbf{x}_i))$  the misclassification cost and  $\lambda_r$  the regularization coefficient which allows balancing the influence of the two terms. The regularization term is usually chosen as  $\Omega_r(f) = \|\mathbf{w}\|^2$ , that is the squared  $\ell_2$ -norm penalty, leading to the maximum margin (minimum norm) objective function. The misclassification cost term  $H(y_i, f(\mathbf{x}_i))$  can be defined as the squared hinge loss function  $\max(0, 1 - y_i f(\mathbf{x}_i))^2$  [16].

### 2.2. Spatial regularization on the prediction function

SVM has been proposed for datasets where the samples are considered Independent and Identically Distributed (IID) [15]. This is clearly not the case when the dataset consists in a set of spatially organized samples (such as voxels). Suppose we are given a set of full images (not necessarily of the same dimension), with a total of  $N$  voxels with  $d$  features per voxel. All of these  $N$  voxels  $\{\mathbf{x}_i\}_{i=1\dots N}$  are stored in a matrix  $\mathbf{X} \in \mathbb{R}^{N \times d}$ . As introduced in section 2.1, we assume that a subset  $\{\mathbf{x}_i\}_{i \in \mathcal{L}}$  of  $n$  voxels, where  $n \leq N$ , are labelled with  $\{y_i\}_{i \in \mathcal{L}} \in \{-1, +1\}$ .

We introduce a spatial *a priori* into the learning procedure, by enforcing spatial coherence for the output class predictions. In other words, we want neighbouring voxels of an image  $(\mathbf{x}_i, \mathbf{x}_j)$  to have close prediction scores  $(f(\mathbf{x}_i), f(\mathbf{x}_j))$  in order to promote spatial smoothness in the output prediction function. We introduce an additional penalty term in the

cost function, of the form :

$$\Omega_s(f) = \sum_{i,j=1}^N S_{i,j} \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|^2, \quad (2)$$

The  $N \times N$  matrix  $\mathbf{S}$  is constructed to penalize the distance between the classifier output of two spatially related (connected) voxels  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . To achieve this, we propose to construct  $\mathbf{S}$  such that  $S_{i,j} = 0$  everywhere except when voxels  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are connected according to some adjacency rules ( $S_{i,j} = 1$ ). Note that the general form of this regularization term is commonly used for semi-supervised learning [17] and the specific graph  $\mathbf{S}$  taking into account voxels neighborhood has been proposed by [18]. The computation of this matrix does not require labeled data and can be easily computed for any image, even if all or part of its voxels are unlabelled.

For a linear classifier, the spatial regularization (2) can be rewritten in matrix form as :

$$\Omega_s(f) = \mathbf{w}^\top \sum_{i,j=1}^N \mathbf{S}_{i,j} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{w} = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$$

where  $\boldsymbol{\Sigma} = \mathbf{X}^\top (\mathbf{D} - \mathbf{S}) \mathbf{X}$  with  $\mathbf{D}$  the diagonal matrix such that  $D_{i,i} = \sum_{j=1}^N S_{i,j}$ . When we add the spatial regularization term (2) to the SVM formulation in (1), the optimization problem becomes :

$$\min_f \sum_{i \in \mathcal{L}} H(y_i, f(\mathbf{x}_i)) + \lambda_r \Omega_r(f) + \lambda_s \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \quad (3)$$

where parameter  $\lambda_s$  weights the influence of the spatial smoothness cost functions.

While this problem cannot be solved using classical SVM solvers due to the spatial quadratic regularization, it can be easily transformed. One approach when using the  $\ell_2$  regularization consists of introducing a change of variable such as  $\tilde{\mathbf{w}} = (\boldsymbol{\Sigma} + \frac{\lambda_s}{\lambda_r} \mathbf{I})^{1/2} \mathbf{w}$  and  $\tilde{\mathbf{x}} = (\boldsymbol{\Sigma} + \frac{\lambda_s}{\lambda_r} \mathbf{I})^{-1/2} \mathbf{x}$ , with  $\mathbf{I}$  the identity matrix. The problem can be solved by estimating  $\tilde{\mathbf{w}}$  with a regular SVM solver [16] in the projected space and then obtaining the original  $\mathbf{w}$  due to the inverse projection.

Note that if  $n < N$ , we are considering a semi-supervised learning problem since only some of the training examples are labelled. Nevertheless, we still use unlabelled data for the learning of the discrimination function by using their spatial localization in the images as an *a priori*.

### 2.3. Feature selection via the $l_1$ -norm

In addition to promoting spatial smoothness in the prediction, we aim to automatically select relevant features. As suggested in [10], feature selection is important for discarding uninformative features to prevent over-fitting, speed up the learning process, as well as improving the model's interpretability. Feature selection can be obtained by using a sparsity inducing norm in the regularized empirical risk in (3). A typical sparsity inducing norm is the  $l_1$ -norm such that :  $\|\mathbf{w}\|_1 = \sum_i^d |w_i|$  which penalizes solutions with many nonzero components.

---

**Algorithm 1** ADMM for feature selection

---

**Initialize**  $\mathbf{w}$ ,  $b$ ,  $\mathbf{v}$  and  $\alpha$ **while** convergence is not reached **do**

$$\mathbf{w}, b = \arg \min_{\mathbf{w}, b} \mathcal{L}(\mathbf{w}, b, \mathbf{v}, \alpha)$$

$$\mathbf{v} = \arg \min_{\mathbf{v}} \mathcal{L}(\mathbf{w}, b, \mathbf{v}, \alpha)$$

$$\alpha = \alpha + \mu(\mathbf{w} - \mathbf{v})$$

**end while**

---

We propose to replace the  $\ell_2$ -norm regularization term with a  $\ell_1$ -norm penalty in the optimization problem (3) such as :

$$\min_f \sum_{i \in \mathcal{L}} H(y_i, f(\mathbf{x}_i)) + \lambda_r \|\mathbf{w}\|_1 + \lambda_s \mathbf{w}^\top \Sigma \mathbf{w} \quad (4)$$

This problem is non-differentiable and cannot be solved with a standard SVM solver. We propose to solve this using a proximal splitting algorithm such as the Alternating Direction Method of Multiplier (ADMM) [19]. The idea is to split the complex problem (4) by minimizing the equivalent problem :

$$\min_f \sum_{i \in \mathcal{L}} H(y_i, f(\mathbf{x}_i)) + \lambda_r \|\mathbf{v}\|_1 + \lambda_s \mathbf{w}^\top \Sigma \mathbf{w}, \quad \text{s.t. } \mathbf{v} = \mathbf{w}$$

A solution of the problem is obtained by looking for a saddle point of the Lagrangian :

$$\mathcal{L}(\mathbf{w}, b, \mathbf{v}, \alpha) = \sum_{i \in \mathcal{L}} H(y_i, \mathbf{w}^\top \mathbf{x}_i + b) + \lambda_s \mathbf{w}^\top \Sigma \mathbf{w} + \lambda_r \|\mathbf{v}\|_1 + \alpha^\top (\mathbf{w} - \mathbf{v}) + \frac{\mu}{2} \|\mathbf{w} - \mathbf{v}\|^2$$

where  $\alpha \in \mathbb{R}^d$  are the Lagrange variables. This saddle point is obtained by iteratively optimizing over the primal variables ( $\mathbf{w}$ ,  $b$ ) and  $\mathbf{v}$  and the Lagrangian multiplier  $\alpha$  (see Algo. 1). Note that this algorithm alternates between the resolution of a  $\ell_2$ -linear SVM similar to (3) and a soft-thresholding on the  $\mathbf{v}$  variable.

### 3. NUMERICAL EXPERIMENTS

We propose to compare the regular  $\ell_2$ -SVM to  $\ell_1$ -SVM, performing feature selection, and  $\ell_1$ -SSVM, promoting spatial regularization, on a set of mp-MR images of the prostate.

#### 3.1. Clinical data and feature extraction

Data from 30 patients (median age: 61 [45-70] years) who underwent T2-weighted (T2-w), Diffusion-Weighted (DW) and Dynamic Contrast Enhanced (DCE) MR imaging prior to prostatectomy were included. Acquisition parameters are detailed in [10]. All MR images were resized to match the T2-w dimension (matrix : 256x256x24, FOV : 200x200x3 mm) to get a direct correspondence between voxels of the different MR sequences. Results of the prostatectomy specimens analysis, used as the gold standard, were reviewed by two radiologists and a histopathologist. This allowed an *a posteriori* delineation on MR images of 42 malignant regions in the peripheral zone.

We extracted a large set of 117 features including normalized image intensity values, local texture (eg. entropy, cluster

**Table 1.** Performances achieved by  $\ell_2$ -SVM,  $\ell_1$ -SVM,  $\ell_2$ -SSVM and  $\ell_1$ -SSVM in terms of ACC, AUC and DSC together with the corresponding variances

	$\ell_2$ -SVM	$\ell_1$ -SVM	$\ell_2$ -SSVM	$\ell_1$ -SSVM
ACC	0.83 ( $\pm 0.12$ )	0.87 ( $\pm 0.11$ )	0.89 ( $\pm 0.09$ )	0.89 ( $\pm 0.09$ )
AUC	0.78 ( $\pm 0.11$ )	0.77 ( $\pm 0.10$ )	0.82 ( $\pm 0.10$ )	0.80 ( $\pm 0.10$ )
DSC	0.37 ( $\pm 0.22$ )	0.33 ( $\pm 0.22$ )	0.39 ( $\pm 0.22$ )	0.39 ( $\pm 0.22$ )

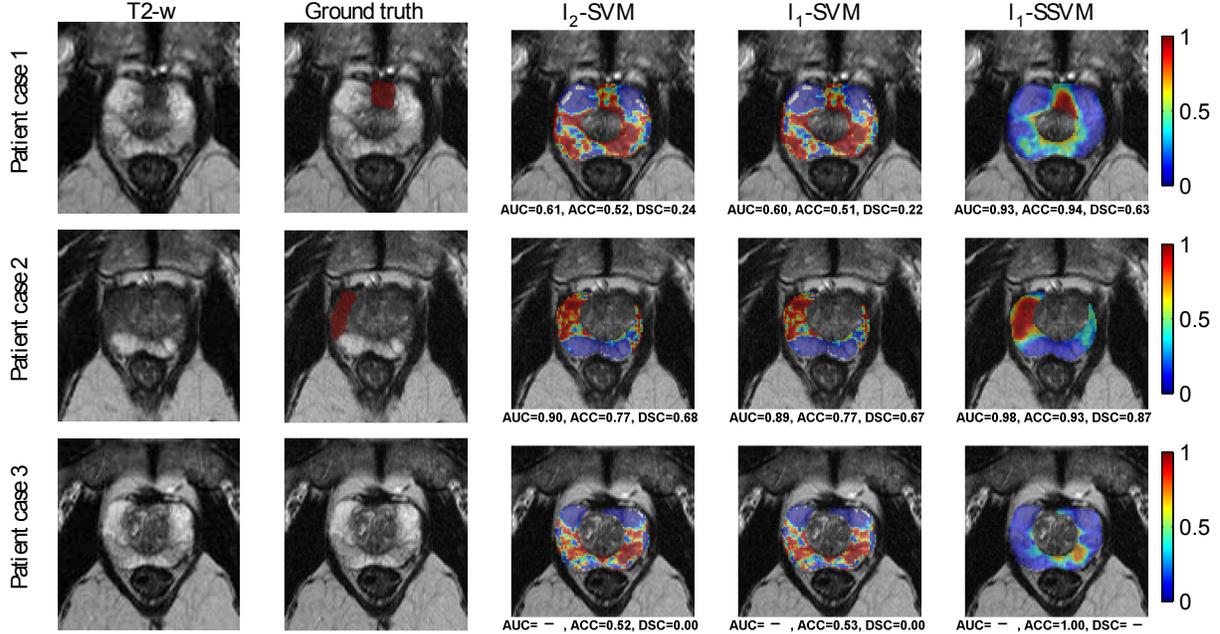
prominence, etc) and gradient (eg. Sobel and Kirsch’s filters, etc) parameters as well as semi-quantitative (eg. wash-in, time-to-peak, etc) and quantitative (eg. forward volume transfer constant Ktrans, etc) pharmacokinetics parameters. Detailed description of these features is given in [10]. All parameters were normalized to zero mean and unit standard deviation.

#### 3.2. Evaluation

For each patient, we deliberately use a restricted subset of only 30 labelled voxels for training (among  $\simeq 10,000$  voxels located within the prostate), mimicking an incompletely labelled dataset. This would occur in clinical practice where the physician may not have time to exhaustively label an entire dataset and would rather outline some points for each patient case instead. The subset of voxels are picked randomly, while still respecting the ratio of malignant/benign voxels per patient; the remaining voxels, for which the class label is artificially considered as unknown, are used in the construction of the spatial *a priori* (2). In this experiment, we arbitrarily defined the vicinity for each voxel  $\mathbf{x}_{k,l}$  as its 4 nearest planar neighbours ( $\mathbf{x}_{k-1,l}$ ,  $\mathbf{x}_{k+1,l}$ ,  $\mathbf{x}_{k,l-1}$ ,  $\mathbf{x}_{k,l+1}$ ). Given the limited number of the patients, classification performances were estimated using a Leave-One-Patient-Out cross-validation approach, to avoid training and testing on the same data. The malignancy score of each voxel of the left-out patient was estimated based on the training data of all the other patients, repeating the procedure for each of the 30 patients. Membership probabilities were estimated from output distances  $f(\mathbf{x})$  using Platt’s algorithm [20]. Performance was evaluated using the area under the ROC curve (AUC) from the output probability maps, as well as the accuracy (ACC) and Dice coefficient (DSC) from the thresholded output distance maps (threshold = 0). Statistical t-tests were performed to compare performances of the different algorithms. Parameters  $\lambda_r$  and  $\lambda_s$  were optimized for values ranging between [0.01 – 10000] using 7 intervals on a log-scale.

#### 3.3. Results

Mean prediction ACC, AUC and DSC performances achieved over the 30 patients are displayed in Table 1. Overall performance was not statistically different between  $\ell_2$ -SVM and  $\ell_1$ -SVM (p-value = 0.2, 0.3 and  $3.10^{-4}$  for ACC, AUC and

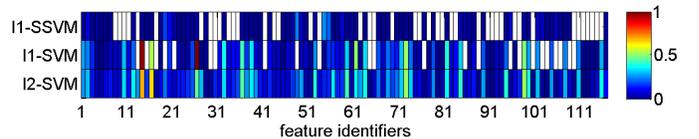


**Fig. 1.** Examples of predicted probability maps obtained through  $l_2$ -SVM,  $l_1$ -SVM and  $l_1$ -SSVM. Predictions performances (AUC, ACC, DSC) corresponding to each map are displayed underneath each image. Note that Patient 3 did not have any PCA lesion on the displayed slice, this aims to show the lack of specificity of the different algorithms.

DSC respectively) nor between  $l_2$ -SSVM and  $l_1$ -SSVM (p-values = 0.4, 0.02 and 0.5 for ACC, AUC and DSC respectively). However, introducing the spatial regularization resulted in significant performance improvements of  $l_2$ -SSVM and  $l_1$ -SSVM compared to  $l_2$ -SVM (p-values < 0.002 for all metrics) and  $l_1$ -SVM (p-values < 0.02 for all metrics) respectively.

Fig. 1 shows the malignancy probability maps of PCA, in medial plane, obtained for 3 patient cases using  $l_2$ -SVM,  $l_1$ -SVM and  $l_1$ -SSVM, together with the T2-w source image and the binary histological ground truth. Numerical evaluation corresponding to each slice is displayed underneath each case. There is no apparent visual difference between  $l_2$ -SVM and  $l_1$ -SVM predicted maps. Similarly,  $l_2$ -SSVM prediction maps are visually similar to  $l_1$ -SSVM ones and are thus not shown. However, the  $l_1$ -SSVM output probability maps are smoother, with lower classification noise. The segmentation of lesions is also improved with regard to the shape of the ground truth.

By setting 39% and 59% of the  $w_i$  coefficients to 0 for  $l_1$ -SVM and  $l_1$ -SSVM respectively (see Fig. 2),  $l_1$ -norm enforces a drastic feature selection, highlighting the relevant features (corresponding to non-zero  $w_i$ ) in the discrimination task. Yet, as seen in Table 1, the classification performances of  $l_1$ -SVM and  $l_1$ -SSVM are equivalent to that of  $l_2$ -SVM and  $l_2$ -SSVM respectively, meaning that using far fewer features did not decrease the discrimination ability, with many features being either uninformative or redundant. Ap-



**Fig. 2.** Absolute values of the output coefficients ( $w_i$ ) $_{i=1\dots 117}$ , depending on the classification method. White color corresponds to non-selected feature ( $w_i=0$ ).  $l_2$ -SVM has only non-zero values (100% features selected). Only 61% and 41% of the features are selected in  $l_1$ -SVM and  $l_1$ -SSVM respectively.

part from the informative interest of knowing which features are actually relevant, being able to discriminate relevant features will avoid the calculation of unused features for future test datasets and thus save computational time.

#### 4. PERSPECTIVES AND CONCLUSIONS

This paper presents a new SVM-based approach for image voxel classification, automatically selecting features and promoting spatial smoothing of the prediction maps. Experimental evaluation on clinical data demonstrates a clear improvement in both numerical performance and readability. New perspectives include extension to non-linear SVM by introducing kernels to the new framework as well as the redefinition of the spatial regularization cost using a weighting scheme, depending on the distance to the neighbours.

## 5. REFERENCES

- [1] G. P Haas, N. Delongchamps, O. W Brawley, C. Y Wang, and G. de la Roza, "The worldwide epidemiology of prostate cancer: perspectives from autopsy studies," *The Canadian Journal of Urology*, vol. 15, no. 1, pp. 3866–3871, 2008.
- [2] I. Chan, W. Wells 3rd, R. V Mulkern, S. Haker, J. Zhang, K. H Zou, S. E Maier, and C. M Tempny, "Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier," *Medical Physics*, vol. 30, no. 9, pp. 2390–2398, 2003.
- [3] A. Madabhushi, J. Shi, M. Rosen, J. E Tomaszewski, and M. D Feldman, "Comparing classification performance of feature ensembles: Detecting prostate cancer from high resolution MRI," in *Computer Vision Methods in Medical Image Analysis (In conjunction with ECCV)*, 2006, vol. 4241, pp. 25–36.
- [4] D. L Langer, T. H van der Kwast, A. J Evans, J. Trachtenberg, B. C Wilson, and M. A Haider, "Prostate cancer detection with multiparametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI," *Journal of Magnetic Resonance Imaging*, vol. 30, no. 2, pp. 327–334, 2009.
- [5] S. Viswanath, B. N Bloch, M. Rosen, J. Chappelow, R. Toth, N. Rofsky, R. Lenkinski, E. Genega, A. Kalyanpur, and A. Madabhushi, "Integrating structural and functional imaging for computer assisted detection of prostate cancer on multi-protocol," in *SPIE Medical Imaging*, Miami, Florida, 2009, vol. 7260, pp. 72603I–72603I–12.
- [6] S. Ozer, D. L Langer, X. Liu, M. A Haider, T. H van der Kwast, A. J Evans, M. N Wernick Y. Yang, and I. S Yetik, "Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI," *Medical Physics*, vol. 37, no. 4, pp. 1873, 2010.
- [7] Y. Artan, M. A Haider, D. L Langer, T. H van der Kwast, A. J Evans, Y. Yang, M. N Wernick, J. Trachtenberg, and I. S Yetik, "Prostate cancer localization with multispectral MRI using Cost-Sensitive support vector machines and conditional random fields," *IEEE Trans. on Image Processing*, vol. 19, no. 9, pp. 2444–2455, Sept. 2010.
- [8] R. Lopes, A. Ayache, N. Makni, P. Puech, A. Villers, S. Mordon, and N. Betrouni, "Prostate cancer characterization on mr images using fractal features," *Medical Physics*, vol. 38, no. 1, pp. 83–95, 2011.
- [9] P.C Vos, J.O Barentsz, N. Karssemeijer, and H.J Huisman, "Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis," *Physics in Medicine and Biology*, vol. 57, pp. 1527–1542, 2012.
- [10] E. Niaf, O. Rouvière, F. Mège-Lechevallier, F. Bratan, and C. Lartizien, "Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI," *Physics in Medicine and Biology*, vol. 57, no. 12, pp. 3833–3851, 2012.
- [11] E. Niaf, C. Lartizien, F. Bratan, L. Roche, M. Rabilloud, F. Mège-Lechevallier, and O. Rouvière, "Prostate Focal Peripheral Zone Lesions : Characterization at Multiparametric MR Imaging - Influence of a Computer-aided Diagnosis System," *Radiology*, vol. 271, no. 3, pp. 761–769, 2014.
- [12] Y. Tarabalka, M. Fauvel, J. Chanussot, and J.A. Benediktsson, "SVM-and MRF-based method for accurate classification of hyperspectral images.," *IEEE Trans. on Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 736–740, 2010.
- [13] A. Madabhushi, M. D Feldman, D. N Metaxas, J. Tomaszewski, and D. Chute, "Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI," *IEEE Trans. on Medical Imaging*, vol. 24, no. 12, pp. 1611–1625, 2005.
- [14] R. Flamary and A. Rakotomamonjy, "Support vector machine with spatial regularization for pixel classification," in *International Workshop on Advances in Regularization, Optimization, Kernel Methods and Support Vector Machines : theory and applications (ROKS)*, 2013.
- [15] V. N Vapnik, *Statistical Learning Theory*, Wiley-Interscience, 1 edition, Sept. 1998.
- [16] O. Chapelle, "Training a support vector machine in the primal," *Neural Computing*, vol. 19, no. 5, pp. 1155–1178, 2007.
- [17] G. Camps-Valls, T. Bandos Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3044–3054, 2007.
- [18] M. Dundar, J. Theiler, and S. Perkins, "Incorporating spatial contiguity into the design of a support vector machine classifier," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2006, pp. 364–367.
- [19] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [20] J. Platt, "Probabilistic outputs for support vector machines," *Advances in Large Margin Classifiers*, pp. 61–74, 1999.