# Segmenting Brain Tumors with Conditional Random Fields and Support Vector Machines

Chi-Hoon Lee[1], Mark Schmidt[1], Albert Murtha[2], Aalo Bistritz[3], Jöerg Sander[1], and Russell Greiner[1]

[1] Department of Computing Science, University of Alberta
[2] Division of Radiation Oncology
[3] Division of Diagnostic Imaging, Department of Oncology, Cross Cancer Institute, Edmonton AB, Canada

**Abstract.** Markov Random Fields (MRFs) are a popular and well-motivated model for many medical image processing tasks such as segmentation. Discriminative Random Fields (DRFs), a discriminative alternative to the traditionally generative MRFs, allow tractable computation with less restrictive simplifying assumptions, and achieve better performance in many tasks. In this paper, we investigate the tumor segmentation performance of a recent variant of DRF models that takes advantage of the powerful Support Vector Machine (SVM) classification method. Combined with a powerful Magnetic Resonance (MR) preprocessing pipeline and a set of 'alignment-based' features, we evaluate the use of SVMs, MRFs, and two types of DRFs as classifiers for three segmentation tasks related to radiation therapy target planning for brain tumors, two of which do not rely on 'contrast agent' enhancement. Our results indicate that the SVM-based DRFs offer a significant advantage over the other approaches.

## 1 Introduction

Support Vector Machines (SVMs) are a popular tool for classification tasks due to their appealing generalization properties; this has led several groups to propose using SVMs for brain tumor segmentation [1,2,3]. However, SVMs assume that data (here, individual voxels) is independently and identically distributed (iid), which is not appropriate for tasks such as segmenting medical images. In particular, SVMs can not consider dependencies in the labels of adjacent pixels/voxels. Markov Random Fields (MRFs), a popular classification technique that models such dependencies, have been used in many medical image segmentation tasks [4,5,6], and have also been used in systems for brain tumor segmentation [5,6,7]. However, *generative* MRFs often do not have the discriminative power of *discriminative* techniques such as SVMs. Conditional Random Fields (CRFs [8]) and their multi-dimensional extension, Discriminative Random Fields (DRFs), are *discriminative* alternatives to MRFs, which have outperformed MRFs for several tasks [9,10]. In the remainder of this section, we

Y. Liu, T. Jiang, and C. Zhang (Eds.): CVBIA 2005, LNCS 3765, pp. 469–478, 2005.

review MRFs (Sect. 1.1), CRFs and DRFs (Section 1.2), and SVMs (Section 1.3). Section 2 then describes our recently proposed Support Vector Random Field (SVRF) model, which combines the advantages of both SVMs and CRFs [11]. Section 3 presents an evaluation of these techniques within a system for brain tumor segmentation that uses an extensive MR preprocessing pipeline and a set of multi-scale image-based and 'alignment-based' features.

### 1.1 Markov Random Fields (MRFs)

Markov Random Fields (MRFs) are widely used in medical image processing applications [4,5,6]. They are ideal for many tasks, and are particularly relevant to segmentation tasks as they allow the classification of one element to depend on the labels of neighboring elements of the observation (image, volume, or sequence). By contrast, traditional classification techniques assume the data is iid, and therefore do not model dependencies in the labels of neighboring elements. MRFs typically use a generative approach, modeling the joint probability of the features of the set of voxels $\mathbf{x} = \{x_1, \ldots, x_n\}$ and their corresponding labels $\mathbf{y}$: $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})\, p(\mathbf{y})$. However, these systems often make simplifying assumptions to make the calculation of the joint probability tractable. This usually involves assuming that the likelihoods have a simple factorized form, such as $p(\mathbf{x}|\mathbf{y}) = \prod_i p(x_i|y_i)$, which involves restrictive independence assumptions, and does not allow the modeling of complex dependencies between the features and the labels. For the MRF in our experiments, we used a Gaussian assumption to factor $p(\mathbf{x}|\mathbf{y})$ (as opposed to a non-parametric alternative such as Parzen Windowing [4]), and used the Hammersley-Clifford method [12] to factor $p(\boldsymbol{y})$, producing the following model for the posterior, given a set of labeled training data $S = \{\langle x_i, y_i \rangle\}_i$.

$$p(\mathbf{y}|\mathbf{x}) \quad = \quad \frac{1}{Z} \exp \left[ \sum_{i \in S} \log(p(x_i|y_i)) \; + \; \sum_{c \in C} V_c(\mathbf{y}_c) \right] \tag{1}$$

where $C$ is a set of cliques in the neighborhood (here defined as the set of 8 planar neighbors), $V_c(\mathbf{y})$ is a clique potential function of labels for the clique $c \in C$, and $Z$ normalizes over all possible labelings. The Gaussian assumption allows us to use Maximum Likelihood (ML) parameter estimation.

### 1.2 Conditional and Discriminative Random Fields (CRF, DRF)

Conditional Random Fields (CRFs) are a *discriminative* alternative to the traditionally *generative* MRFs [8]. Rather than modelling the joint likelihood of the features and labels $p(\mathbf{x}, \mathbf{y})$, *discriminative* models directly model the posterior probability of the labels given the features $p(\mathbf{y}|\mathbf{x})$. This subtle difference alleviates the need to model the distribution over the observations. This is important in medical imaging applications, since anatomic structures can have complex shapes that are not easy to model and may not be appropriately modelled by a factorized form of $p(\mathbf{x}|\mathbf{y})$. Since CRFs directly model the posterior, they can relax many of the major simplifying assumptions often made in MRFs. This allows

the (tractable) modelling of complex dependencies (a) *between the features of an element and its label*, (b) *between the labels of adjacent elements*, and (c) *between the labels of adjacent elements and their features*, or even other features of the observation.

Discriminative Random Fields (DRFs) are a multi-dimensional extension of 1-dimensional CRFs for lattice-structured data [9]. This extension, combined with the popularity of MRFs in medical imaging applications and the major advantages that CRFs can have in certain situations over MRFs, suggests that DRFs could have a major impact on a number of medical imaging tasks. In our experiments, we used the following DRF model:

$$p(\mathbf{y}|\mathbf{x}) \quad = \quad \frac{1}{Z} \exp\left( \sum_{i \in S} A_i(y_i, \mathbf{x}) \; + \; \sum_{i \in S} \sum_{j \in N_i} I_{ij}(y_i, y_j, \mathbf{x}) \right) \tag{2}$$

where $A_i$ is the 'Association' (Observation-Matching) potential for modelling dependencies between the $i$-th class label $y_i$ and the set of all observations $\mathbf{x}$. The DRF method uses a Generalized Linear Model (GLM) based on Logistic Regression for this potential [9]. The 'Interaction' (Local-Consistency) potential for modelling dependencies between the labels of neighboring elements, $I_i$, is also a GLM. Non-linear models for both potentials can be induced through a change of basis. Simultaneously determining the optimal parameters of the Association potential and the Interaction potential can be done numerically as a convex optimization problem. The performance of the GLM in DRFs compared to the probability distribution in the first term of MRFs will depend on the application. However, note the important difference between the clique potentials in MRFs and the Interaction potential in DRFs. MRFs indiscriminately smooth over neighboring cliques while DRFs consider the features when taking into account interactions in the labels $\sum_{i \in S} \sum_{j \in N_i} I_{ij}(y_i, y_j, \mathbf{x})$. This is a subtle but important point, since it means a DRF can learn how to optimally use image (and image gradient) information when modeling label dependencies.

DRFs are a powerful method for modeling dependencies in spatial data. There are, however, several problems associated with this method: it is hard to find a good initial labeling during inference, and due to the simultaneous learning of parameters, it tends to overestimate the Interaction potential parameters which can degrade edges during inference (unless regularization is used very carefully). Furthermore, the GLM may not estimate appropriate parameters in data with a high-dimensional feature space or where features may be correlated (as with textural features or multi-modality data) [13]. Because of these factors, in some tasks DRFs will not be advantageous compared to models such as Support Vector Machines.

### 1.3 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are a popular tool for classification of data that is independent and identically distributed. SVMs are less sensitive to class imbalance than GLMs, and due to the properties of error bounds, SVMs tend

to outperform GLMs, especially in cases where the classes overlap (often the case in medical imaging applications) [14]. SVMs try to maximize the margin between classes (here using the simple linear feature space $x_i \cdot x_j$), by finding the optimal $\alpha_i$ values in the following Quadratic Programming problem (represented in dual Lagrangian form where $C$ is a constant that bounds the misclassification error) [14]:

$$\begin{gathered} \max \sum_{i=1}^{N} \alpha_i \; - \; \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \, \alpha_j \, y_i \, y_j \, (x_i \cdot x_j) \\ \text{subject to} \\ 0 \leq \alpha_i \leq C \quad \text{and} \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \end{gathered} \tag{3}$$

Unlabelled instances are classified using the learned parameters $\alpha_i$ and bias $b$, by taking the sign of the following decision function [14]:

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i (x \cdot x_i) + b$$

## 2 Support Vector Random Fields (SVRFs)

An SVM is an iid classifier, which does not consider interactions in the labels of adjacent data points. Conversely, DRFs and MRFs consider these interactions, but do not have the same appealing generalization properties as SVMs. This section will review our Support Vector Random Field (SVRF) model, an extension of SVMs that uses a DRF framework to model interactions in the labels of adjacent data points [11]:

$$p(\mathbf{y}|\mathbf{x}) \quad = \quad \frac{1}{Z} \exp \left\{ \sum_{i \in S} \log(O(y_i, \Upsilon_i(\mathbf{x}))) \; + \; \sum_{i \in S} \sum_{j \in N_i} V(y_i, y_j, \mathbf{x}) \right\} \tag{4}$$

where $\Upsilon_i(\mathbf{x})$ computes features from the observations $\mathbf{x}$ for location $i$, $O(y_i, \Upsilon_i(\mathbf{x}))$ is an SVM-based Observation-Matching potential, and $V(y_i, y_j, \mathbf{x})$ is the Local-Consistency potential over a pair-wise neighborhood structure, where $N_i$ are the 8 neighbors around location $i$.

### 2.1 Observation-Matching

The Observation-Matching function maps from the observations (features) to class labels. We would like to use SVMs for this potential. However, the decision function in SVMs produces a distance value, not a posterior probability suitable for the DRFs' framework. To convert the output of the decision function to a posterior probability, we used a modified version of the method in [15]. This efficient method minimizes the risk of overfitting and is formulated as follows:

$$O(y_i = 1, \Upsilon_i(\mathbf{x})) \quad = \quad \frac{1}{1 \; + \; \exp(A \times f(\Upsilon_i(\mathbf{x})) \; + \; B)} \tag{5}$$

The parameters $A$ and $B$ are estimated from training data represented as pairs $\langle f(\Upsilon_i(\mathbf{x})), t_i \rangle$, where $f(\Upsilon_i(\mathbf{x}))$ is the real-valued SVM response (here, distance to the separator), and $t_i$ denotes a related probability that $y_i = 1$, represented as the relaxed probabilities: $t_i = \frac{N_+ + 1}{N_+ + 2}$ if $y_i = 1$, and $t_i = \frac{1}{N_- + 2}$ if $y_i = -1$, where $N_+$ and $N_-$ are the number of positive and negative class instances. Using these training instances, we can solve the following optimization problem to estimate parameters $A$ and $B$:

$$\min - \sum_{i=1}^{l} [t_i \log O(t_i, \Upsilon_i(\mathbf{x})) \ + \ (1 - t_i) \log(1 - O(t_i, \Upsilon_i(\mathbf{x})))] \tag{6}$$

Platt [15] used a Levenberg-Marquardt approach that tried to set $B$ to guarantee that the Hessian approximation was invertible. However, dealing with the constant directly can cause problems, especially for unconstrained optimization problems [13]. Hence, we employed Newton's method with backtracking line search for simple and robust estimation. To avoid overflows and underflows of *exp* and *log*, we reformulated (6) as

$$\min \sum_{i=1}^{l} [t_i(A \times f(\Upsilon_i(\mathbf{x})) \ + \ B) \ + \ \log(1 + \exp(-A \times f(\Upsilon_i(\mathbf{x})) \ - \ B))] \tag{7}$$

### 2.2 Local-Consistency

We use a DRF model for Local-Consistency, since we do not want to make the (traditional MRF) assumption that the label interactions are independent of the features. We adopted the following pairwise Local-Consistency potential:

$$V(y_i, y_j, \mathbf{x}) \quad = \quad y_i y_j \, (\nu \cdot \Phi_{ij}(\mathbf{x})) \tag{8}$$

where $\nu$ is the vector of Local-Consistency parameters to be learned, while $\Phi_{ij}(\mathbf{x})$ calculates features for sites $i$ and $j$. DRFs use a $\Phi_{ij}$ that penalizes for high absolute differences in the features. As we are additionally interested in encouraging label continuity, we used the following function that encourages continuity while discouraging discontinuity: ($\max(\Upsilon(\mathbf{x}))$ denotes the vector of max values of the features):

$$\Phi_{ij}(\mathbf{x}) \quad = \quad \frac{\max(\Upsilon(\mathbf{x})) \ - \ | \, \Upsilon_i(\mathbf{x}) - \Upsilon_j(\mathbf{x}) \, |}{\max(\Upsilon(\mathbf{x}))} \tag{9}$$

Observe that this function is large when neighboring elements have very similar features, and small when there is a wide gap between their values.

### 2.3 Learning: Parameter Estimation

SVRFs use a sequential learning approach to parameter estimation. This involves first solving the SVM Quadratic Programming problem (3). The resulting decision function is then converted to a posterior probability using the training data and estimated relaxed probabilities. The Local-Consistency parameters are then

estimated from the $m$ training pixels from each of the $K$ training images using pseudolikelihood [12]:

$$\widehat{\nu} \quad = \quad \arg\max_{\nu} \prod_{k=1}^{K} \prod_{i=1}^{m} p(y_i^k | y_{N_i}^k, \mathbf{x}^k, \nu) \tag{10}$$

We ensure that the log-likelihood is convex by assuming a Gaussian prior over $\nu$: that is, $p(\nu|\tau)$ is a Gaussain distribution with 0 means and $\tau^2 I$ variance (see [9]). Thus, the local-consistency parameters are estimated using its log likelihood:

$$\widehat{\nu} = \arg\max_{\nu} \sum_{k=1}^{K} \sum_{i=1}^{m} \left\{ O_i^n + \sum_{j \in N_i} V(y_i^k, y_j^k, \mathbf{x}^k) - \log(z_i^k) \right\} \; - \; \frac{1}{2\tau} \nu^T \nu \tag{11}$$

where $z_i^k$ is a partition function for each site $i$ in image $k$, and $\tau$ is a regularizing constant that ensures the Hessian is not singular. Keeping the Observation-Matching ($O_i^k = O(y_i, \Upsilon_i(\mathbf{x}))$) constant, the optimal Local-Consistency parameters can be found by gradient descent.

We close by noting that the $M^3N$ [10] framework resembles SVRFs, as it also incorporates label dependencies and uses a max-margin approach. However, the $M^3N$ approach uses a margin that magnifies the difference between the target labels and the best runner-up, while we use the 'traditional' 2-class SVM approach of maximizing the distance from the classes to a separating hyperplane. An efficient approach for training and inference in a special case of $M^3N$s was presented in [16]. However, the simultaneous learning and the inference strategy used still make computations with this model expensive compared to SVRFs.

## 3 Brain Tumor Segmentation

Segmenting brain tumors is an important medical imaging problem, currently done manually by expert radiation oncologists for radiation therapy target planning. Markov Random Fields [5,6,7] and SVMs [1,2,3,17] have been used in systems to perform this task. We have recently evaluated DRFs and SVRFs for the relatively easy case of segmenting "enhancing tumor areas" [11]. We extend this by providing improved results for this easy case (due to using better preprocessing and features), and results for two much harder segmentation cases. This section will present (i) our experimental data and design, (ii) a summary of the MR preprocessing pipeline and the multi-scale image-based and 'alignment-based' features that afford a significant improvement over those previous results and allow us to address more challenging tasks, and (iii) experimental results comparing SVMs, MRFs, DRFs, and SVRFs within this context for three different segmentation tasks.

Our experimental data set consisted of T1, T1c (T1 after injecting contrast agent), and T2 images (each 258 by 258 pixels) from 7 patients (Fig. 1), each

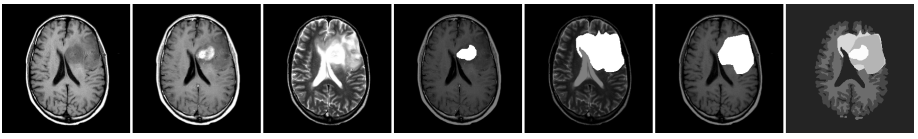

**Fig. 1.** Left to right: T1 image, T1 image with contrast agent, T2 image, enhancing area label, edema label, gross tumor label, full brain segmentation

having either a grade 2 astrocytoma, an anaplastic astrocytoma, or a glioblastoma multiforme. The data was preprocessed with an extensive MR preprocessing pipeline (described in [3], and making use of [18,19]) to reduce the effects of noise, inter-slice intensity variations, and intensity inhomogeneity. In addition, this pipeline robustly aligns the different modalities with each other, and with a template image in a standard coordinate system (allowing the use of alignment-based features, mentioned below).

We used the most effective feature set from the comparative study in [17]. This multi-scale feature set contains traditional image-based features in addition to three types of 'alignment-based' features: spatial probabilities for the 3 normal tissue types (white matter, gray matter and cerebrospinal fluid), spatial expected intensity maps, and a characterization of left-to-right symmetry (all measured at multiple scales). As with many of the related works on brain tumor segmentation (such as [1,2,6,20]), we employed a patient-specific training scenario, where training data for the classifier is obtained from the patient to be segmented. In order to be fair, all classifiers received the same training and testing pixels, and the testing pixels came from a different area of the volume than the training pixels — here, distant MR slices (this prevents the Random Field models from achieving high scores due to over-fitting.)

In our experiment, we applied 6 classifiers — a Maximum Likelihood classifier (degenerate MRF), a Logistic Regression model (degenerate DRF), an SVM (degenerate SVRF), an MRF, a DRF, and an SVRF — to 13 different volumes, based on various time points from 7 patients.

For each of the Random Field methods, we initialized inference with the corresponding degenerate classifier (ie. Maximum Likelihood, Logistic Regression, or SVM), and used the computationally efficient Iterated Conditional Modes (ICM) algorithm to find a locally optimal label configuration [12].

The 6 classifiers were evaluated over the 13 time points for the following 3 tasks, where the ground truth was defined by an expert radiologist. The first task was the relatively easy task of segmenting the 'enhancing' tumor area — ie. the region that appears hyper-intense after injecting the contrast agent (and including the non-enhancing or necrotic areas contained within the enhancing contour). The second task was the segmentation of the entire edema area associated with the tumor, which is significantly more challenging due to the high degree of similarity between the intensities of edema areas and normal cerebrospinal fluid. The final task was segmenting the Gross Tumor area as defined by the radiologist. This can be a subset of the edema but a superset of the

**Table 1.** Jaccard Percentage Scores for Enhancing tumor, Edema areas, and Gross Tumor areas (high scores in bold). ML denotes Maximum likelihood and LR denotes Logistic regression.

| | Enhancing Tumor Area | | | | | | Edema Area | | | | | | Gross Tumor Area | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study | ML | MRF | LR | DRF | SVM | SVRF | ML | MRF | LR | DRF | SVM | SVRF | ML | MRF | LR | DRF | SVM | SVRF |
| 1-1 | 23.1 | 24.6 | 44.4 | 46.1 | 49.7 | **52.8** | 21.9 | 21.6 | 35.7 | 36.7 | 57.0 | **58.2** | 19.3 | 19.5 | 39.4 | 40.9 | **40.7** | 40.5 |
| 2-1 | 0 | 0 | 61.3 | 61.5 | 86.4 | **87.7** | 33.3 | 34.2 | 59.2 | 61.4 | 88.4 | **89.2** | 35.4 | 35.7 | 65.1 | 66.1 | **78.2** | 76.9 |
| 3-1 | 69.2 | 69.7 | 61.8 | 61.8 | 82.0 | **84.8** | 34.4 | 34.4 | 75.5 | 77.1 | 80.7 | **82.2** | 44.5 | 46.1 | 72.9 | 73.4 | 77.9 | **78.7** |
| 3-2 | 40.1 | 40.3 | 84.8 | 84.6 | 84.7 | **87.8** | 47.6 | 48.1 | 73.6 | 74.1 | 79.3 | **83.1** | 51.2 | 51.3 | 76.3 | 76.2 | 78.1 | **80.8** |
| 4-1 | 26.9 | 27.3 | 49.1 | 50.4 | 77.8 | **81.7** | 28.3 | 29.1 | 38.6 | 41.2 | 53.0 | **55.4** | 37.4 | 38.7 | 39.4 | 40.1 | **41.4** | 41.2 |
| 4-2 | 58.9 | 59.7 | 68.3 | 70.2 | 75.7 | **77.9** | 43.2 | 46.8 | 45.3 | 46.7 | 53.7 | **57.7** | 38.0 | 40.2 | 39.7 | 39.4 | 62.1 | **64.9** |
| 4-3 | 49.2 | 50.7 | 71.3 | 71.6 | 87.2 | **88.1** | 35.4 | 35.4 | 69.9 | **70.7** | 68.2 | 69.1 | 66.0 | 68.5 | 73.3 | **73.5** | 64.4 | 64.5 |
| 4-4 | 65.6 | 68.2 | 87.5 | 87.1 | 86.0 | **89.1** | 44.1 | 43.7 | 78.6 | 79.0 | 76.7 | **79.3** | 46.7 | 45.8 | 83.8 | 83.5 | 86.0 | **89.0** |
| 5-1 | 67.0 | 67.5 | 52.2 | 51.4 | 81.8 | **84.3** | 47.8 | 48.6 | 63.6 | 65.7 | 73.8 | **76.9** | 50.1 | 50.9 | 65.3 | 68.3 | 82.8 | **84.8** |
| 6-1 | 37.4 | 37.6 | 76.4 | 76.2 | 78.2 | **80.4** | 40.3 | 40.1 | 79.3 | 79.7 | 81.2 | **83.7** | 46.6 | 47.6 | 79.6 | 79.4 | 87.6 | **88.2** |
| 7-1 | 63.2 | 63.0 | 75.5 | 76.7 | 80.0 | **81.4** | 74.9 | 77.7 | 91.2 | 92.4 | 93.8 | **94.9** | 66.4 | 66.3 | 71.9 | 73.2 | **74.6** | 74.1 |
| 7-2 | 37.7 | 39.3 | 75.9 | 75.8 | 85.5 | **87.3** | 39.2 | 40.4 | 80.9 | 82.7 | 82.1 | **82.8** | 49.6 | 52.4 | 68.3 | 67.9 | 72.7 | **72.9** |
| 7-3 | 45.3 | 45.6 | 81.8 | 81.5 | 87.7 | **89.6** | 54.1 | 53.9 | 79.3 | 80.7 | 84.6 | **86.5** | 43.4 | 43.7 | 73.5 | 72.7 | 81.6 | **83.2** |
| Ave: | 44.9 | 45.7 | 68.6 | 68.8 | 80.2 | **82.5** | 41.9 | 42.6 | 67.0 | 68.3 | 74.8 | **76.9** | 5.7 | 46.7 | 65.3 | 65.7 | 71.4 | **72.3** |

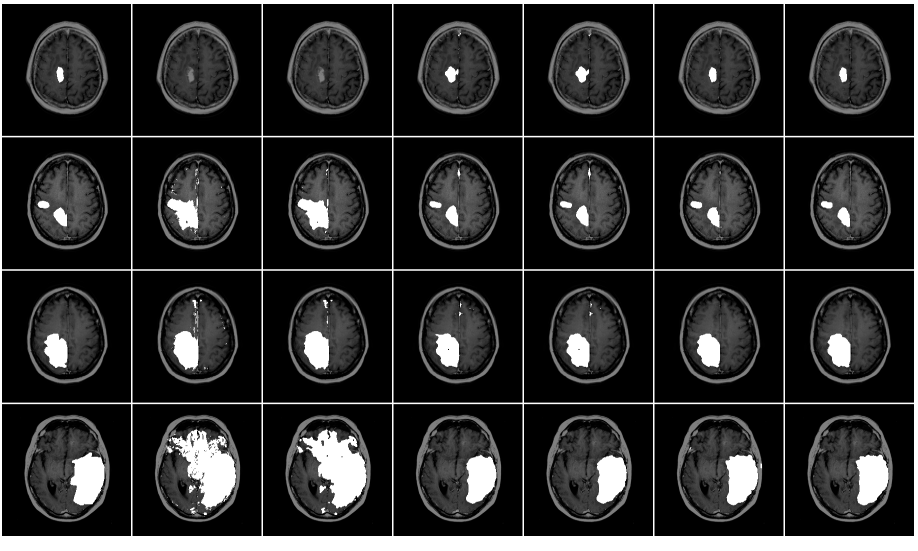

**Fig. 2.** Classification results for 4 different test slices, where each row shows a different test slice. Top to bottom: 2-1 Enhancing tumor, 7-2 Enhancing tumor, 7-1 Edema, and 4-4 Gross Tumor. Left column to right: Expert Segmentation, ML, MRF, LR, DRF, SVM, SVRF.

enhancing area, and is inherently a very challenging task, even for human experts, given the modalities examined. We used the Jaccard similarity measure to assess the classifications in terms of true positives (tp), false positives (fp), and false negatives (fn): $J = \frac{tp}{tp+fp+fn}$.

Table 1 presents the classification results for the three tasks (example test slice results are shown in Fig 2). For each of the three tasks, SVRFs showed the best performance on average, while SVMs were the second most effective method. The differences in the average scores between all methods across the

three tasks were significant at the $p < 0.05$ level based on a paired example $t$-test. Note that SVRFs were the best in all 13 enhancing tumor cases, 12 of the 13 edema cases, and in the challenging Gross Tumor cases, SVRFs were best 8 times, SVMs best 4 times, and DRFs 1 time. The results from the second patient "2-1", produced an interesting observation: significant overlap between Gaussians in the high dimensional feature space leads ML and subsequently MRFs to misclassify *all* areas as non-tumors. This example shows that inappropriate modeling of $p(\mathbf{x}|\mathbf{y})$ can generates poor performance (see the first row of Fig 2). Although the segmentation tasks for edema and gross tumor areas are very hard, the discriminative approaches, and in particular SVRFs, still produce segmentations that are highly similar to the manual segmentations on average for all 3 tasks.

## 4 Conclusion

We are currently focusing on methods to allow inter-patient testing scenarios with SVRFs. This necessitates intensity standardization methods as in [17,5], and developing more computationally efficient parameter estimation models. Note that the SVRF results could be improved through the use of non-linear kernels (as in [1,2]), and through more effective inference methods. We are also interested in exploring applications of CRFs in other medical imaging tasks.

This work introduces SVRFs, a method that combines the random field relaxation properties of DRFs (to associate labels of neighboring voxels) with the discriminative properties of SVMs. We then presented experimental results on 3 challenging tasks related to brain tumor segmentation, and found that SVRFs offer a significant performance advantage over 5 other plausible classifiers, including both SVMs and other random field models.

## Acknowledgments

R. Greiner is supported by NSERC and the Alberta Ingenuity Centre for Machine Learning (AICML). C.-H. Lee is supported by NSERC, AICML, and iCORE. Our thanks to the Cross Cancer Institute for providing the brain tumor data, Dale Schuurmans for helpful discussions on optimization and parameter estimation, Marianne Morris for help in data acquisition, and other members of the BTGP team.

## References

1. Garcia, C., Moreno, J.: Kernel based method for segmentation and modeling of magnetic resonance images. LNCS **3315** (2004) 636–645
2. Zhang, J., Ma, K., Er, M., Chong, V.: Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine. International Workshop on Advanced Image Technology (2004) 207–211
3. Schmidt, M.: Automatic brain tumor segmentation. Master's thesis, University of Alberta (2005)

4. Held, K., Kops, E., Krause, B., Wells, W., Kikinis, R., Mullter-Gartner, H.: Markov random field segmentation of brain mr images. IEEE TMI **16** (1997) 878–886
5. Gering, D.: Recognizing Deviations from Normalcy for Brain Tumor Segmentation. PhD thesis, MIT (2003)
6. Chen, T., Metaxas, D.N.: Gibbs prior models, marching cubes, and deformable models: A hybrid framework for 3d medical image segmentation. In: MICCAI. (2003) 703–710
7. Capelle, A., Colot, O., Fernandez-Maloigne, C.: Evidential segmentation scheme of multi-echo MR images for the detection of brain tumors using neighborhood information. Information Fusion **5** (2004) 203–216
8. Lafferty, J., Pereira, F., McCallum, A.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. ICML (2001)
9. Kumar, S., Hebert, M.: Discriminative fields for modeling spatial dependencies in natural images. NIPS (2003)
10. Taskar, B., Guestrin, C., Koller, D.: Max margin markov networks. NIPS (2003)
11. Lee, C., Schmidt, M., Greiner, R.: Support vector random fields for spatial classification. PKDD (2005)
12. Li, S.Z.: Markov Random Field Modeling in Image Analysis. Springer-Verlag, Tokyo (2001)
13. R.Fletcher: Practical Methods of Optimization. John Wiley & Sons (1987)
14. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge, UK (2004)
15. Platt, J.: Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. MIT Press, Cambridge, MA (2000)
16. Anguelov, D., Taskar, B., Chatalbashev, V., Koller, D., Gupta, D., Heitz, G., Ng, A.: Discriminative learning of markov random fields for segmentation of 3d scan data. CVPR (2005)
17. Schmidt, M.: Segmenting brain tumors using alignment-based features. Technical report, University of Alberta (2005)
18. McAuliffe, M., Lalonde, F., McGarry, D., Gandler, W., Csaky, K., Trus, B.: Medical image processing, analysis and visualization in clinical research. IEEE CBMS (2001) 381–386
19. Statistical Parametric Mapping: http://www.fil.ion.bpmf.ac.uk/spm/ (Online)
20. Kaus, M., Warfield, S., Nabavi, A., Black, P., Jolesz, F., Kikinis, R.: Automated segmentation of MR images of brain tumors. Radiology **218** (2001) 586–591