Efficient Lesion Segmentation using Support Vector Machines

Jean-Baptiste Fiot, Laurent D. Cohen, Parnesh Raniga, Jurgen Fripp

Efficient Lesion Segmentation using Support Vector Machines

Jean-Baptiste Fiot [1,2] Laurent D. Cohen [1] Parnesh Raniga [2] Jurgen Fripp [2]
[1] *CEREMADE, UMR 7534 CNRS Université Paris Dauphine, France*
[2] *CSIRO Preventative Health National Research Flagship ICTC, The Australian e-Health Research Centre - BioMedIA, Royal Brisbane and Women's Hospital, Herston, QLD, Australia*

Support Vector Machines (SVM) are a machine learning technique that has been used for segmentation and classification of medical images, including segmentation of white matter hyper-intensities (WMH). Current approaches using SVM for WMH segmentation extract features from the brain and classify these followed by complex post-processing steps to remove false positives. The method presented in this paper combines the use of domain knowledge, advanced pre-processing (based on tissue segmentation and atlas propagation) and SVM classification to obtain efficient and accurate WMH segmentation. Features generated from up to four MR modalities (T1-w, T2-w, PD and FLAIR), differing neighbourhood sizes and the use of multi-scale features were compared. We found that although using all 4 modalities gave the best overall classification (average Dice scores of 0.54 ± 0.12, 0.72 ± 0.06 and 0.82 ± 0.06 respectively for small, moderate and severe lesion loads, using 3x3x3 neighbourhood intensity features); this was not significantly different (p = 0.50) from using just T1-w and FLAIR sequences (Dice scores of 0.52 ± 0.13, 0.71 ± 0.08 and 0.81 ± 0.07 for the same lesion loads and feature type). Furthermore, there was a negligible difference between using 5x5x5 and 3x3x3 features (p = 0.93). Finally, we show that careful consideration of features and preprocessing techniques leads to more efficient classification which outperforms the one based on all features with post-processing, and also saves storage space and computation time.

Keywords: Lesion, Segmentation, Classification, Support Vector Machines, Brain Imaging

1 INTRODUCTION

WMH appear brightly on T2-weighted (T2-w) and fluid attenuated inversion recovery (FLAIR) MRI modalities. They are a possible risk factor for Alzheimer's Disease (AD), with progression associated with vascular factors and cognitive decline (Lao et al. 2008). To quantify these changes in large scale population studies, it is desirable to have fully automatic and accurate segmentation methods to avoid time-consuming, costly and non-reproducible manual segmentations. However, WMH segmentation using a single modality is challenging because their signal intensity range overlaps with that of normal tissue: in T1-weighted (T1-w) images, WMH have intensities similar to grey matter (GM), and in T2-w and PD images, WMH look similar to cerebro-spinal fluid (CSF). The FLAIR images have been shown to be most sensitive to WMH (Anbeek et al. 2004), but can also present hyper-intensity artifacts that can lead to false positives. To improve the WMH segmentation performance, additional discriminative information is extracted from multiple MR modalities.

The most successful lesion segmentation methods in the literature have been developed for the detection of multiple sclerosis lesions, with a recent grand challenge comparing the performance of various techniques (Styner et al. 2008). Lesion segmentation algorithms can be categorised into unsupervised clustering or (semi-)supervised voxel-wise classification. Unsupervised methods suffer from the issue of model selection. Supervised methods such as neural networks (Dyrby et al. 2008) and k-NN (Anbeek et al. 2004) have been proposed. Neural networks are efficient but setting their parameters is difficult. The k-NN method performs relatively well, but is computationally expensive.

We present an SVM based segmentation scheme inspired by the work in (Lao et al. 2008; Zacharaki et al. 2008). Lao et al. applied four steps: pre-processing (co-registration, skull-stripping, intensity normalisation and inhomogeneity correction), SVM training

with Adaboost, segmentation and elimination of false positives. Our implementation utilises a similar but more advanced pre-processing pipeline and a simpler training procedure. As one of the primary causes of errors in other approaches is false positive cortical regions, we incorporate advanced pre-processing including patient specific tissue segmentation and atlas based population tissue priors to minimize the false positive regions that are usually found with naive classifier. As a result of this the advanced post processing required by other techniques (Lao et al. 2008) are not necessary. We also evaluated the relative value of each MRI acquisition protocol for segmentation. This scheme is quantitatively validated on a significantly larger dataset with healthy aging, mild cognitive impairment and AD subjects.

2 CLASSIFICATION AND SUPPORT VECTOR MACHINE THEORY

Lesion segmentation can be formulated as a binary classification problem. SVM (Schölkopf and Smola 2001) solves it in a supervised way: given l labelled features $(x_i, y_i) \in X \times \{-1, 1\}$, it builds a function $f\colon X \to \mathbb{R}$ such that $y(.) = sign(f(.))$ is an optimal labeling function. The function f is computed via the optimization problem:

$$f^* = \underset{f \in \mathcal{H}_K}{\operatorname{argmin}} \frac{1}{l} \sum_{i=1}^{l} V(f(x_i), y_i) + \gamma \|f\|_K^2 \qquad (1)$$

where $K : X \times X \to \mathbb{R}$ is a Mercer Kernel, $\mathcal{H}_K$ its associated Reproducing Kernel Hilbert Space of functions $X \to \mathbb{R}$ and its corresponding norm $\| \ \|_K$, and V is the hinge loss defined as $V(f(x), y) = \max\{0, 1 - y \times f(x)\}$. The loss function V controls the labeling performance, and the second term controls the smoothness of the solution.

The optimization problem is convex because of the convexity of the hinge loss function. However as the objective function is not differentiable, the problem is reformulated with additional slack variables $\xi_1, \ldots, \xi_l \in \mathbb{R}$:

$$f^* = \underset{\substack{f \in \mathcal{H}_K \\ \xi_1, \ldots, \xi_n \in \mathbb{R}}}{\operatorname{argmin}} \frac{1}{l} \sum_{i=1}^{l} \xi_i + \gamma \|f\|_K^2 \qquad (2)$$

$$\text{subject to: } \xi_i \geq V(f(x_i), y_i) \ \forall i \in \{1, \ldots, l\}$$

The Riesz representation theorem states that the solution of (1) exists in $\mathcal{H}_K$, and can be written:

$$f^*(.) = \sum_{i=1}^{l} \alpha_i K(., x_i) \text{ with } \alpha_i \in \mathbb{R} \qquad (3)$$

By plugging the expansion of f from (3) in (2), the optimisation problem becomes a finite dimension optimisation problem. Let the matrix K be defined as $K_{i,j} = K(x_i, x_j)$. The optimisation problem is now:

$$\min_{\substack{\alpha_1, \ldots, \alpha_l \in \mathbb{R} \\ \xi_1, \ldots, \xi_l \in \mathbb{R}}} \frac{1}{l} \sum_{i=1}^{l} \xi_i + \gamma \alpha^T K \alpha \text{ subject to:} \qquad (4)$$

$$\begin{cases} \xi_i - 1 + y_i \sum_{j=1}^{l} \alpha_j K(x_i, x_j) \geq 0 & \forall i \in \{1, \ldots, l\} \\ \xi_i \geq 0 & \forall i \in \{1, \ldots, l\} \end{cases}$$

Let $\mu, \nu \in \mathbb{R}^l$ be the Lagrangian multipliers. The Lagrangian of this problem is:

$$L(\alpha, \xi, \nu, \mu) = \frac{1}{l} \sum_{i=1}^{l} \xi_i + \gamma \alpha^T K \alpha$$

$$- \sum_{i=1}^{l} \mu_i \left(\xi_i - 1 + y_i \sum_{j=1}^{l} \alpha_j K(x_i, x_j) \right) - \sum_{i=1}^{l} \nu_i \xi_i \qquad (5)$$

Solving $\nabla_\alpha L = 0$ leads to $\alpha_i^*(\mu, \nu) = \frac{y_i \mu_i}{2\gamma} \ \forall i \in \{1, \ldots, l\}$. Solving $\nabla_\xi L = 0$ leads to $\mu_i + \nu_i = \frac{1}{l}$. The Lagrance dual function is:

$$q(\mu, \nu) = \inf_{\alpha, \xi \in \mathbb{R}^l} L(\alpha, \xi, \nu, \mu) = \qquad (6)$$

$$\begin{cases} \sum_{i=1}^{l} \mu_i - \frac{1}{4\gamma} \sum_{i,j=1}^{l} y_i y_j \mu_i \mu_j K(x_i, x_j) \text{ if } \mu_i + \nu_i = \frac{1}{l} \\ -\infty \text{ otherwise} \end{cases}$$

The dual problem consists in maximising $q(\mu, \nu)$ subject to $\mu \geq 0, \nu \geq 0$, and is equivalent to:

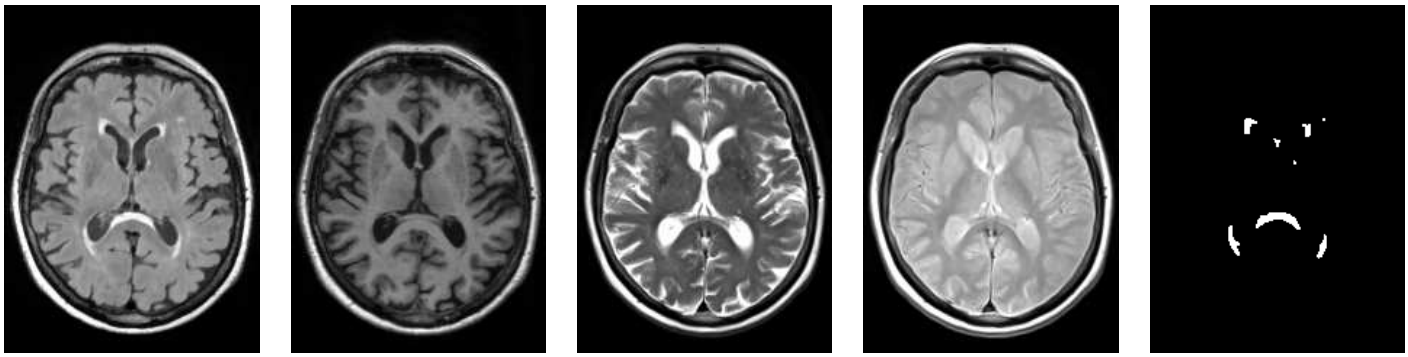
$$\max_{0 \leq \mu \leq \frac{1}{l}} \sum_{i=1}^{l} \mu_i - \frac{1}{4\gamma} \sum_{i,j=1}^{l} y_i y_j \mu_i \mu_j K(x_i, x_j) \qquad (7)$$

Therefore the problem that α must solve is:

$$\max_{\alpha_1, \ldots, \alpha_l \in \mathbb{R}} 2 \sum_{i=1}^{l} \alpha_i y_i - \sum_{i,j=1}^{l} \alpha_i \alpha_j K(x_i, x_j) \qquad (8)$$

$$= \max_{\alpha_1, \ldots, \alpha_l \in \mathbb{R}} 2\alpha^T y - \alpha^T K \alpha$$

The training vectors with $\alpha_i \neq 0$ are called the support vectors. The optimization maximizes the margin, which is the distance between the decision boundary and the support vectors.



(a) FLAIR (b) T1-w (c) T2-w (d) PD (e) Manual segmentation

Figure 1: Axial slices from one subject illustrating the different MR modalities and manual segmentation. Lesions can be seen in the FLAIR and T2-w as a bright signal.

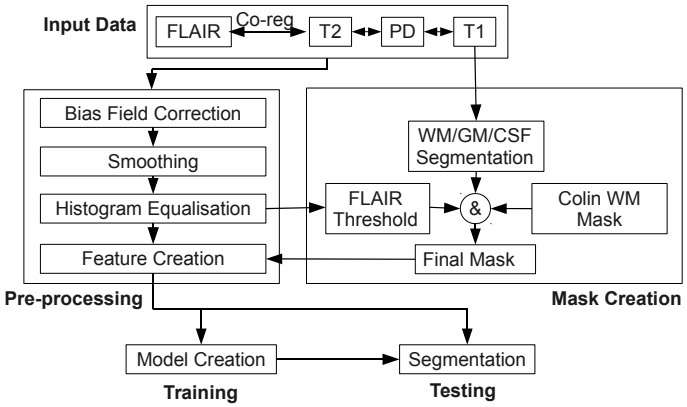


Figure 2: Summary of the WMH segmentation pipeline

3 MATERIALS AND METHODS

3.1 Data

The dataset comes from the AIBL study (Ellis et al. 2009), where T1-w (160x240x256 image, spacing 1.2x1x1mm in the sagittal, coronal and axial direction, TR = 2300ms, TE = 2.98ms, flip angle = 9°), FLAIR (176x240x256, 0.90x0.98x0.98 mm, TR = 6000ms, TE = 421ms, flip angle = 120°, TI = 2100ms), T2-w (228x256x48, 0.94x0.94x3, TR = 3000ms, TE = 101ms, flip angle = 150°) and PD (228x256x48 0.94x0.94x3, TR = 3000ms, TE = 11ms, flip angle = 150°) were acquired for 125 subjects. Lesions were manually segmented by PR, reviewed by a neuro-radiologist and used as ground truth in the classification.

3.2 Proposed Algorithm

The proposed algorithm, summarised in Fig. 2, consists of the following steps:

Pre-processing: images were rigidly co-registered (Ourselin et al. 2001), bias-field corrected (Salvado et al. 2006), smoothed using anisotropic diffusion and histogram equalised to a reference subject. T1-w images were segmented into WM, GM, CSF using an Expectation-Maximisation approach with priors (Acosta et al. 2009). For each modality, features were extracted within the mask defined below, and scaled to [0, 1]. Multi-modality features were created by concatenation of single modality features. Neighbourhood intensities features (3x3x3 and 5x5x5 sizes) and pyramidal features (with 4 levels, taking one voxel per level, Gaussian kernel convolutions of $\sigma = \{0.5, 1, 1.5\}$) were tried.

Mask creation: a global threshold on FLAIR images provides a high sensitivity, but poor specificity, which means it can be used to define areas of interest. To further reduce the areas of interest, we define the region W as the intersection of the dilated Colin WM mask (registered rigidly (Ourselin et al. 2001) then non-rigidly (Rueckert et al. 1999)) and the WM mask (defined from the segmentation of the previous step). Using the mean μ_W and standard deviation σ_W of the FLAIR intensities on W, an intensity threshold of $\mu_W + 2\,\sigma_W$ on W is used to define the mask M:

$$\begin{cases} \forall x \in W, M(x) = 1 \text{ if } FLAIR(x) > \mu_W + 2\,\sigma_W \\ M = 0 \text{ everywhere else} \end{cases}$$

Training: a subset of 10 000 features, with half belonging to the lesion class, the other half belonging to the non-lesion class, randomly selected and equally distributed among the training samples was used to generate the classifiers. A Matlab implementation solving SVM in its primal formulation was used (Melacci 2009). The chosen kernel was the (gaussian) radial basis function. The width of the kernel and the regularisation weight were selected via a 10-fold cross validation.

Testing (Segmentation): the images in the test set were fully segmented within the mask created. Pixels outside this region were set to the non-lesion class. As post-processing, all the connected components of lesions with less than 10 voxels were removed.

3.3 Validation

The dataset was randomly split equally into training and test sets. A classifier was built using the training set, and then used to segment the test set. Then training set and test set were swapped, another classifier was built, and the rest of the segmentations were computed. Results were then merged. Model performances were compared using the Dice score (Dice 1945) DSC = $2\frac{\lambda(S\cap GT)}{\lambda(S)+\lambda(GT)}$ (with S the computed segmentation, GT the ground truth and λ counting the number of voxels in a volume), the number of true/false positive/negative (TP, FP, TN, FN) voxels, the specificity ($\frac{TN}{TN+FP}$) and the sensitivity ($\frac{TP}{TP+FN}$) computed on the full image. Higher is better for DSC, TP, TN, sensitivity and specificity. Lower is better for FP and FN. Statistical significance was analysed via the p-values of paired t-tests (Ott and Longnecker 2008). We performed experiments to test the influence of the combination of modalities, the influence of the feature type and the influence of using the mask in pre-processing instead of in the post-processing.

4 RESULTS

Figure 3 shows DSC, FP and TP for various combinations of modalities (using 3x3x3 neighourhood features). TN, FN, sensitivity and specificity are similar, so corresponding graphs are not displayed. As the overall lesion load impacts the segmentation performance, as previously report in (Anbeek et al. 2004), results are displayed for low (<3mL), moderate (3-10mL) and severe (>10mL) lesion loads. When using one modality, FLAIR gives the best performance. Combining several modalities generates less FP and more TP. Table 1 indicates that the T1-w + FLAIR combination is statistically better than FLAIR on low and moderate lesion load, but T2-w + FLAIR is not. T1-w + T2-w + FLAIR combination is statistically better than FLAIR on the overall dataset. The model with the 4 modalities performs the best (Fig. 3), but not significantly better than T1-w + FLAIR (p=0.50, see Table 1).

Figure 4 shows the performances of different feature types (using the 4 modalities). With neighbourhood intensity features, a 5x5x5 size slightly increases the DSC compared to 3x3x3, but the difference is not statistically significant (p=0.93). Pyramidal features with 4 dimensions do not perform as well as neighbourhood intensity features, but the DSC difference is not statistically significant (p=0.21 when compared with 3x3x3 features, p=0.18 with 5x5x5).

As illustrated in Fig. 5, using the mask in the pre-processing instead of post-processing decreases FP dramatically and leads to a much better DSC. (Comparison using FLAIR and 3x3x3 neighbourhood). The computation time in the prediction step being linear in the number of features to label, computing predictions for a significantly lower number of features (only within the mask) saves computation time. On average on the 125 patients, this represents a 41 times computation speed-up.

5 DISCUSSION

We have presented a machine learning scheme applied to the WMH segmentation problem. Our approach is inspired by the previous work on SVM but has a number of differences. It combines the use of tissue segmentation, atlas propagation techniques and SVM classification to get efficient and accurate segmentation results.

This work also quantifies the relative performance variations with regard to different modalities or feature types. Regarding the modalities, our results confirm that using all of the four modalities adds discriminative information and improves the segmentation results, as reported in (Lao et al. 2008). However, our quantitative results show that using only FLAIR and T1-w can give similar performance at a lower cost. One reason could be the lower axial resolution of our T2-w and PD images. Regarding the features types, there is a trade off between the complexity, storage place and computation time versus the performance.

As other important contribution of this work, the mask we define and use in the pre-processing has several positive impacts. First, it improves the classifier performance as the training features are selected in regions of interest, which leads to better classifiers. Second, computation time and storage space required are significantly lower (41 times lower on our dataset) as features and predictions are computed in a restricted area. Finally, using our mask in the pre-processing makes most of the complex post-processing steps required in current state-of-art methods redundant.

6 ACKNOWLEDGEMENTS

Data used in this article was obtained from the AIBL study funded by the CSIRO, `www.aibl.csiro.au`.

REFERENCES

Acosta, O., P. Bourgeat, M. A. Zuluaga, J. Fripp, O. Salvado, and S. Ourselin (2009). Automated voxel-based 3D cortical thickness measurement in a combined Lagrangian-Eulerian PDE approach using partial volume maps. *Medical Image Analysis 13*(5), 730 – 743.

Anbeek, P., K. L. Vincken, M. J. P. van Osch, R. H. C. Bisschops, and J. van der Grond (2004). Probabilistic segmentation of white matter lesions in MR imaging. *NeuroImage 21*(3), 1037 – 1044.

Dice, L. R. (1945, July). Measures of the amount of ecologic association between species. *Ecol-*

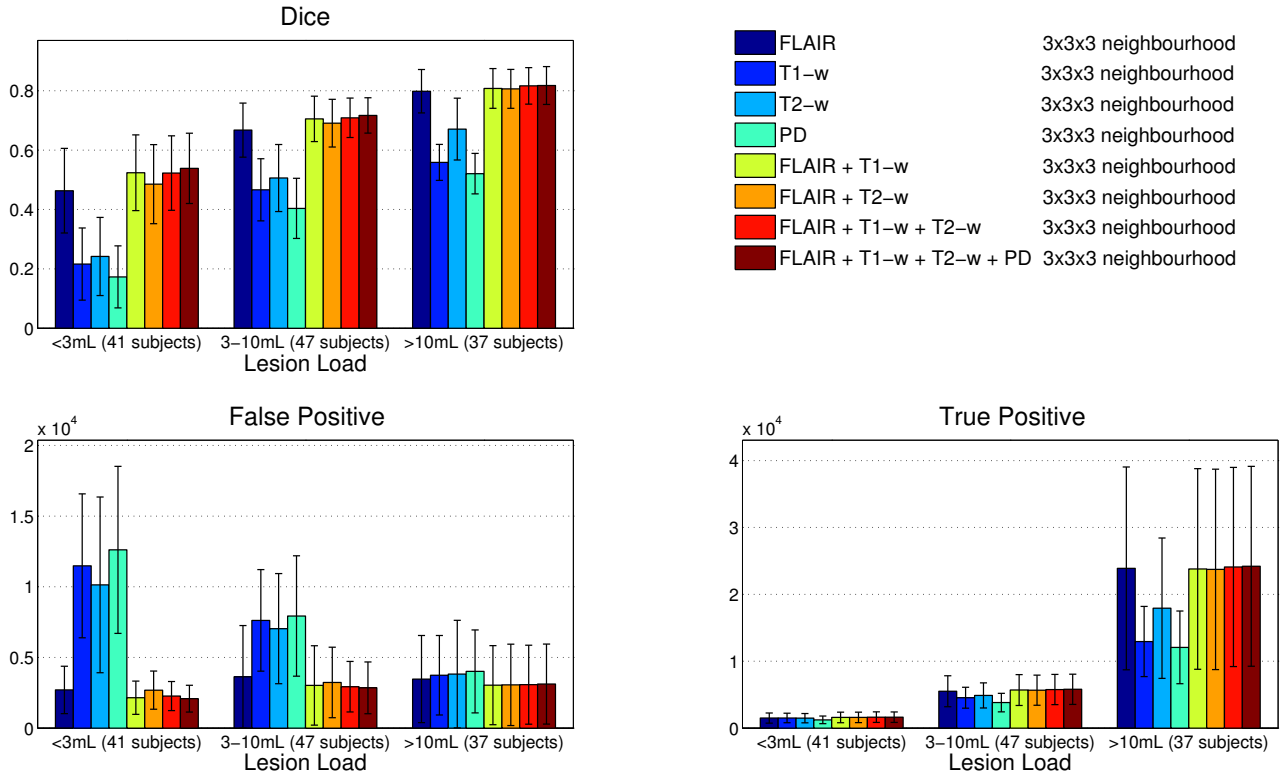


Figure 3: Segmentation performance with different modality combinations (using the 3x3x3 neighbourhood intensity feature type).

Table 1: p-values of paired t-tests using 3x3x3 features. Statistically significant differences ($p < \alpha = 0.05$) in bold green.

Modalities		p-values of t-tests for lesion load in mL (number of subjects)			
Model 1	Model 2	$<$ 3 (35)	3-10 (47)	$>$10 (43)	Any (125)
FLAIR	FLAIR, T2-w	0.47	0.19	0.62	0.38
FLAIR	FLAIR, T1-w	**0.047**	**0.032**	0.57	0.070
FLAIR	FLAIR, T1-w, T2-w	**0.048**	**0.014**	0.26	**0.047**
FLAIR	FLAIR, T1-w, T2-w, PD	**0.011**	**0.002**	0.23	**0.014**
FLAIR, T1-w	FLAIR, T1-w, T2-w, PD	0.59	0.41	0.51	0.50

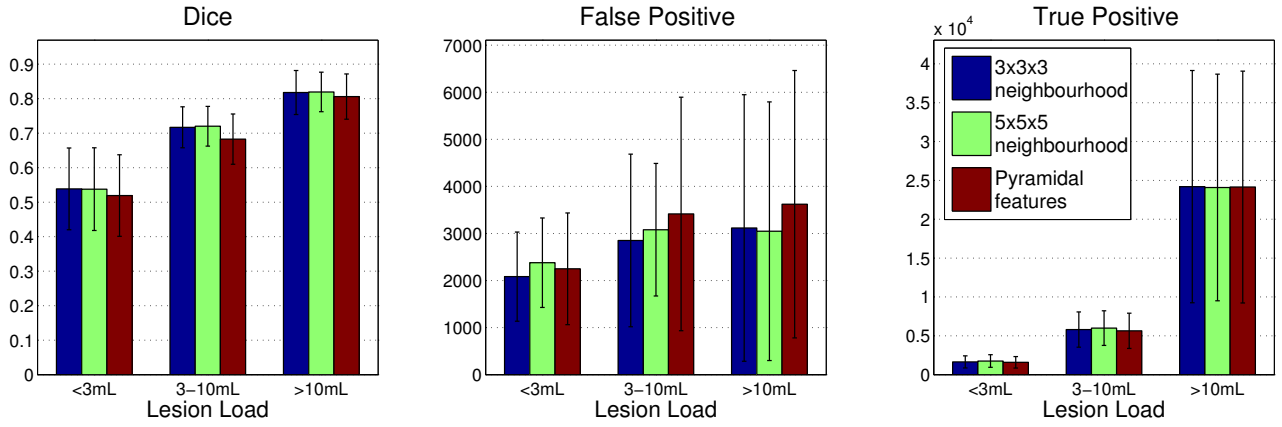


Figure 4: Segmentation performance with different feature types (using the 4 modalities).

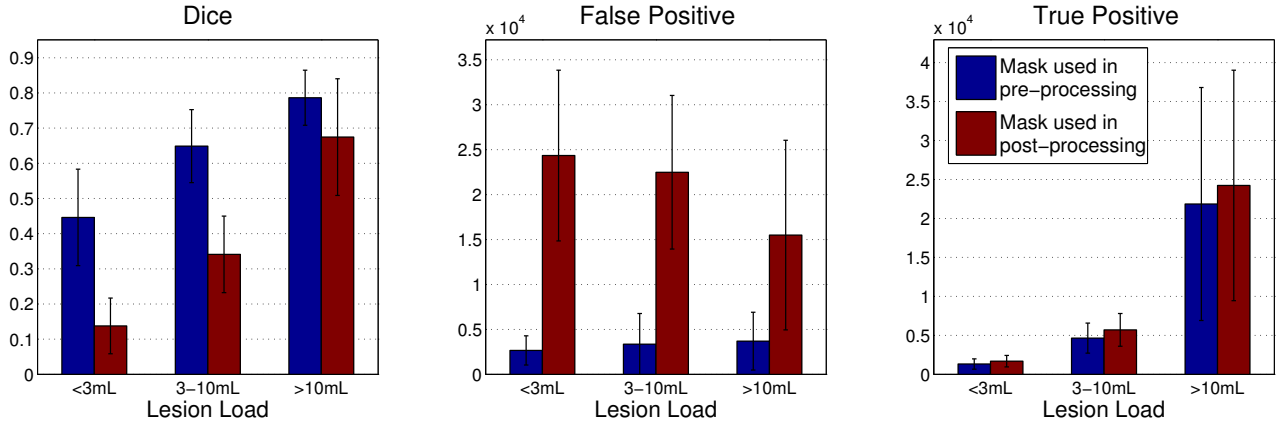


Figure 5: Using our mask M in the pre-processing gives better results than using it only as a post-processing step.

ogy 26(3), 297–302.

Dyrby, T. B., E. Rostrup, W. F. Baaré, E. C. van Straaten, F. Barkhof, H. Vrenken, S. Ropele, R. Schmidt, T. Erkinjuntti, L.-O. Wahlund, L. Pantoni, D. Inzitari, O. B. Paulson, L. K. Hansen, and G. Waldemar (2008). Segmentation of age-related white matter changes in a clinical multi-center study. *NeuroImage 41*(2), 335 – 345.

Ellis, K. A., A. I. Bush, D. Darby, D. De Fazio, J. Foster, P. Hudson, N. T. Lautenschlager, N. Lenzo, R. N. Martins, P. Maruff, C. Masters, A. Milner, K. Pike, C. Rowe, G. Savage, C. Szoeke, K. Taddei, V. Villemagne, M. Woodward, and D. Ames (2009). The Australian imaging, biomarkers and lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatrics 21*(4), 672–87.

Lao, Z., D. Shen, D. Liu, A. F. Jawad, E. R. Melhem, L. J. Launer, R. N. Bryan, and C. Davatzikos (2008). Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. *Academic Radiology 15*(3), 300 – 313.

Melacci, S. (2009, September). Manifold regularization: Laplacian SVM. `http://www.dii.unisi.it/~melacci/lapsvmp/index.html`.

Ott, R. L. and M. T. Longnecker (2008, December). *An Introduction to Statistical Methods and Data Analysis* (6 ed.). Duxbury Press.

Ourselin, S., A. Roche, G. Subsol, X. Pennec, and N. Ayache (2001). Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing 19*(1-2), 25 – 31.

Rueckert, D., L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes (1999, August). Nonrigid registration using free-form deformations: application to breast MR images. *IEEE transactions on medical imaging 18*(8), 712–721.

Salvado, O., C. Hillenbrand, S. Zhang, and D. Wilson (2006). Method to correct intensity inhomogeneity in MR images for atherosclerosis characterization. *Medical Imaging, IEEE Transactions on 25*, 539–552.

Schölkopf, B. and A. J. Smola (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning)*. MIT Press.

Styner, M., J. Lee, B. Chin, M. Chin, O. Commowick, H. Tran, S. Markovic-Plese, V. Jewells, and S. Warfield (2008, sep). 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. In *MIDAS Journal, Special Issue on 2008 MICCAI Workshop - MS Lesion Segmentation*, pp. 1–5.

Zacharaki, E. I., S. Kanterakis, R. N. Bryan, and C. Davatzikos (2008). Measuring brain lesion progression with a supervised tissue classification system. In *Proceedings of MICCAI 2008*, pp. 620–627.