Efficient brain lesion segmentation using multi-modality tissue-based feature selection and 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

SUMMARY

Support Vector Machines (SVM) are a machine learning technique that have 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 advanced pre-processing, tissue-based feature selection and SVM classification to obtain efficient and accurate WMH segmentation. Features from 125 patients, 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 four 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); 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). Furthermore, there was a negligible difference between using $5 \times 5 \times 5$ and $3 \times 3 \times 3$ features ($p = 0.93$). Finally, we show that careful consideration of features and pre-processing techniques not only saves storage space and computation time but also leads to more efficient classification which outperforms the one based on all features with post-processing. Copyright © 2012 John Wiley & Sons, Ltd.

1. INTRODUCTION

White matter hyper-intensities (WMH) are regions in the brain white matter (WM) that appear with bright signal on T2-weighted (T2-w) and fluid attenuated inversion recovery (FLAIR) MRI modalities. They are a possible risk factor for Alzheimer's Disease (AD) and vascular dementia, with progression associated with vascular factors and cognitive decline [1]. 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 proton-density (PD) images, WMH look similar to cerebro-spinal fluid (CSF). FLAIR images have been shown to be most sensitive to WMH [2], 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 [3]. Lesion segmentation algorithms can be categorised into unsupervised clustering or (semi-)supervised voxel-wise classification [4]. Unsupervised methods suffer from the issue of model selection. Supervised methods such as neural networks [5], k-NN [2], Naive Bayes classifier [6] and Parzen windows [7, 8] have been proposed. Neural networks can be efficient but designing an appropriate network architecture and setting the parameters are difficult.

We present an SVM based segmentation scheme whose preliminary results were presented as a conference paper in [9], and inspired by the work in [1, 10]. 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 use information from multiple modalities to define a mask of potential WMH. This mask is built from patient specific tissue segmentation and atlas based population tissue priors. It leads to three main advantages compared to existing techniques. First, such careful feature selection enables to have a more accurate model without the use of boosting. Second it limits the areas where the classification is performed on the training set, which means a faster overall brain classification. Third, it reduces the false positive regions that are usually found with naive classifiers, so the advanced post processing required by other techniques [1] are not necessary. We also evaluated the relative value of each MRI acquisition protocol for WM lesion segmentation. This scheme is quantitatively validated on a significantly larger dataset including healthy aging, mild cognitive impairment and AD subjects. These results were compared with other supervised classification algorithms such as kNN, Naive Bayes, Parzen windows and decision tree.

2. CLASSIFICATION AND SUPPORT VECTOR MACHINE THEORY

Lesion segmentation can be formulated as a binary classification problem. The SVM technique [11] solves it in a supervised way: given l labelled features $(x_i, y_i) \in X \times \{-1, 1\}$, it builds a function $f : X \to \mathbb{R}$ such that $y(.) = sign(f(.))$ is an optimal labeling function. The function f is a solution of the optimization problem:

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

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 $\xi_1, \ldots, \xi_l \in \mathbb{R}$:

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

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:

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

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:

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

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

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

$$L(\alpha, \xi, \nu, \mu) = \frac{1}{l} \sum_{i=1}^{l} \xi_i + \gamma \alpha^T K \alpha - \sum_{i=1}^{l} \mu_i \left(\xi_i - 1 + y_i \sum_{j=1}^{l} \alpha_j K(x_i, x_j) \right) - \sum_{i=1}^{l} \nu_i \xi_i \tag{5}$$

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:

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

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

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

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

Therefore the problem that α must solve is:

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

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.

3. MATERIALS AND METHODS

3.1. Data

The dataset used in this paper comes from the AIBL study [12], where T1-w ($160 \times 240 \times 256$ image, spacing $1.2 \times 1 \times 1$ mm in the sagittal, coronal and axial directions, TR = 2300 ms, TE = 2.98 ms, flip angle = 9°), FLAIR ($176 \times 240 \times 256$, $0.90 \times 0.98 \times 0.98$ mm, TR = 6000 ms, TE = 421 ms, flip angle = 120°, TI = 2100 ms), T2-w ($228 \times 256 \times 48$, $0.94 \times 0.94 \times 3$, TR = 3000 ms, TE = 101 ms, flip angle = 150°) and PD ($228 \times 256 \times 48$, $0.94 \times 0.94 \times 3$, TR = 3000 ms, TE = 11 ms, flip angle = 150°) were acquired for 125 subjects. WM lesions were manually segmented by one of the authors (PR), reviewed by a neuro-radiologist and used as ground truth in the classification (Fig. 1).

3.2. Proposed algorithm

The proposed algorithm uses the standard supervised classification design for segmentation: given images and corresponding segmentations, the goal is to build a classifier to segment new images (Fig. 2). To obtain good performance, adequate preprocessing, mask and feature type have to be defined. This application-specific part is followed by a machine learning process.

As summarised in Fig. 3, the proposed algorithm consists of the following steps:

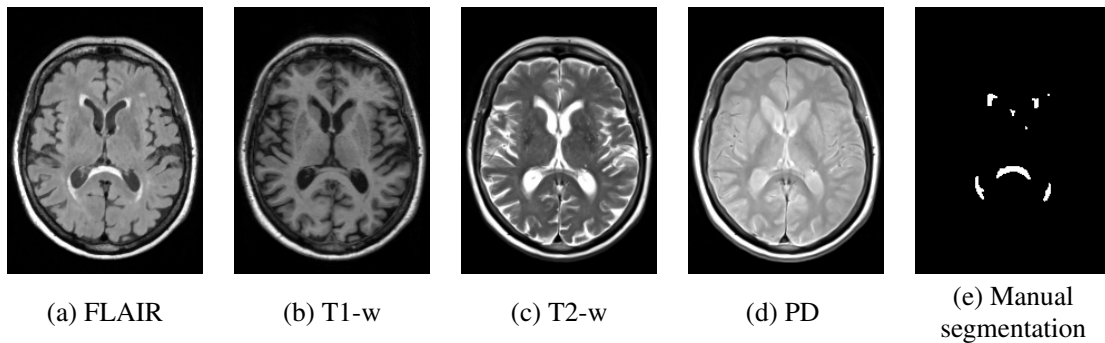(a) FLAIR   (b) T1-w   (c) T2-w   (d) PD   (e) Manual segmentation

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.
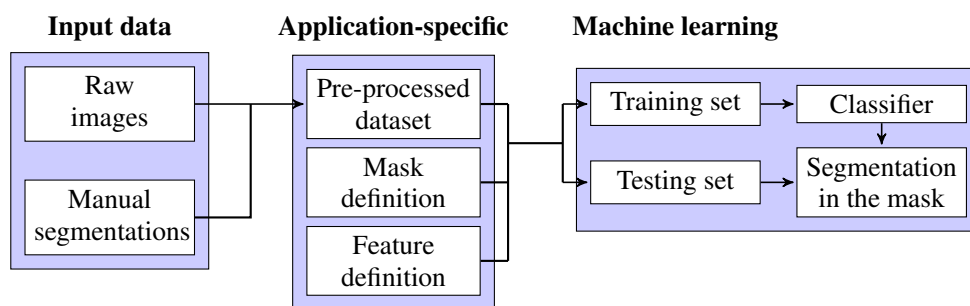

Figure 2. Supervised classication algorithms for segmentation aim at building a classifier from images and corresponding segmentations. To obtain good performance, adequate pre-processing, mask and feature definitions have to be used. This application-specific part is followed by a machine learning process where a classifier is built from training examples, to segment new (testing) images.
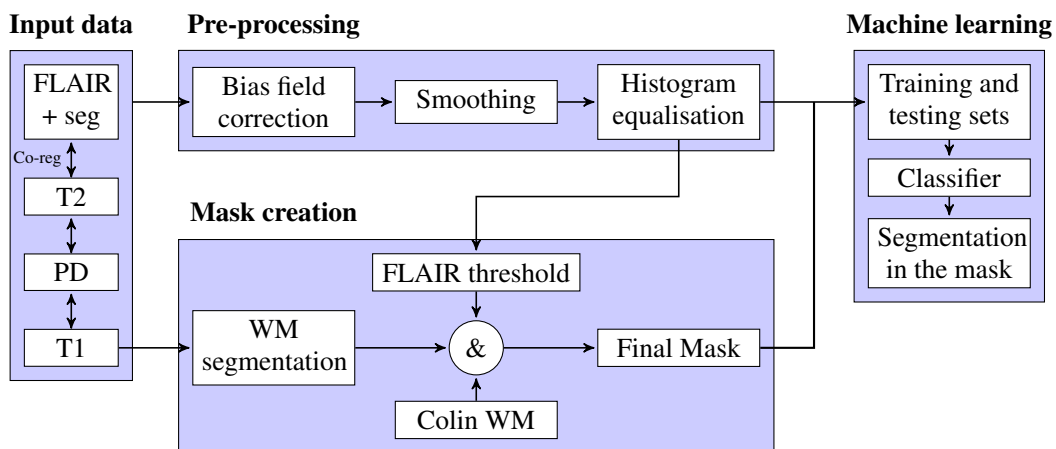

Figure 3. The proposed WMH segmentation pipeline is composed of 3 mains steps: pre-processing, mask creation and machine learning.

Preprocessing: Images were rigidly co-registered [13], bias-field corrected [14], 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 [15]. 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
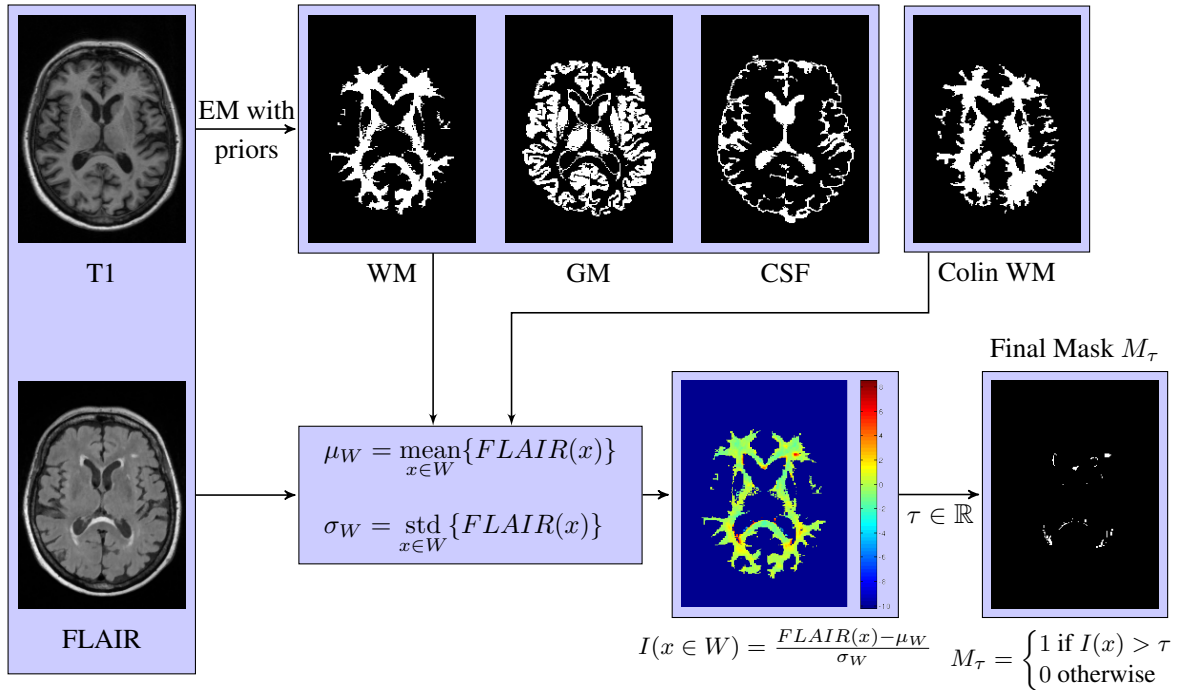
Figure 4. The mask creation uses both FLAIR and T1 modalities to combine intensity-based and tissue-based properties. First, an Expectation-Maximisation (EM) technique on the T1 is used to generate WM/GM/CSF segmentation. On the intersection W of the patient WM and the registered Colin WM, a normalized scalar map is computed from the FLAIR intensities. A final threshold on this map provides the mask M_τ.

features ($3 \times 3 \times 3$ and $5 \times 5 \times 5$ sizes) and pyramidal features (with 4 levels, taking one voxel per level, Gaussian kernel convolutions of $\sigma = \{0.5, 1, 1.5\}$) were examined.

Mask creation: A global threshold on FLAIR images provides a high sensitivity, but poor specificity, which means it can only be used to define areas of interest. To further refine the areas of interest, we define the region W as the intersection of the dilated Colin WM mask [16] (which was registered rigidly [13] then non-rigidly [17] to the subject) and the WM mask (from the tissue segmentation in patient space). Using the mean μ_W and standard deviation σ_W of the FLAIR intensities on W, an intensity threshold of $\mu_W + \tau\,\sigma_W$ on W ($\tau = 2$ was the numerical value used) is used to define the mask M_τ:

$$M_\tau(x) = \begin{cases} 1 \text{ if } FLAIR(x) > \mu_W + \tau\,\sigma_W \text{ and } x \in W \\ 0 \text{ otherwise} \end{cases}$$

Machine learning: 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 [18, 19]. 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. Then each image in the test set was segmented within the patient mask created. Pixels outside this region were set to the non-lesion class. As post-processing, all the connected components segmented as lesion with less than 10 voxels were removed.

Table I. p-values of paired t-tests using $3 \times 3 \times 3$ 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

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 [20] 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 images. 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 [21]. 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. The performance of the SVM classifier was compared to the performance of other supervized classification algorithms. As the overall lesion load impacts the segmentation performance, as previously reported in [2], results are displayed for low (<3mL), moderate (3-10mL) and severe (>10mL) lesion loads. Finally, the impact of the parameter in the mask creation was evaluated by performance bounds of the segmentation performance.

4. RESULTS

4.1. Performance with regard to modality combinations

The segmentation performance was evaluated for various combinations of modalities (using $3 \times 3 \times 3$ neighourhood features). Figure 5 shows DSC, FP and TP values for four single-modality and four multi-modality features. TN, FN, sensitivity and specificity values are similar for the various combinations, so corresponding graphs are not displayed. When using one modality, FLAIR gives the best performance. However, combining several modalities reduces FP and increases TP. Table I indicates that on low and moderate lesion loads the T1-w + FLAIR combination performs statistically better than FLAIR (T2-w + FLAIR does not). On the overall dataset, the T1-w + T2-w + FLAIR combination performs statistically better than FLAIR. The model with the four modalities performs the best (Fig. 5), but not significantly better than T1-w + FLAIR (p=0.50).

4.2. Performance with regard to features type

Using the four modalities, the segmentation performance was evaluated for various feature types (Fig. 6). With neighbourhood intensity features, a $5 \times 5 \times 5$ size slightly increased the DSC compared to $3 \times 3 \times 3$, but the difference was not statistically significant ($p = 0.93$). Pyramidal features with 4 dimensions did not perform as well as neighbourhood intensity features, but the DSC difference was not statistically significant (p=0.21 when compared with $3 \times 3 \times 3$ features, p=0.18 with $5 \times 5 \times 5$ features).
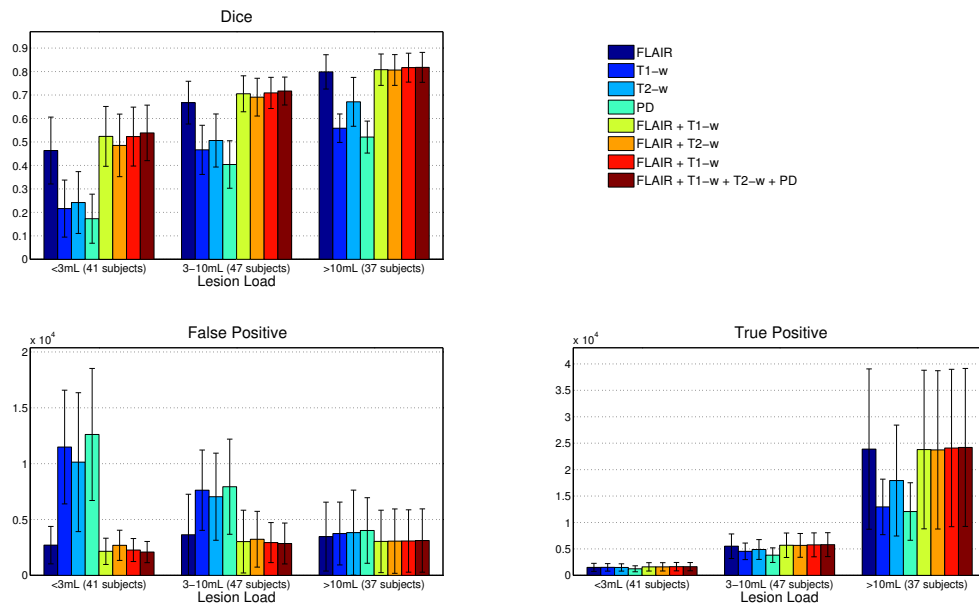
Figure 5. Segmentation performance with different modality combinations (using the $3 \times 3 \times 3$ neighbourhood intensity feature type).
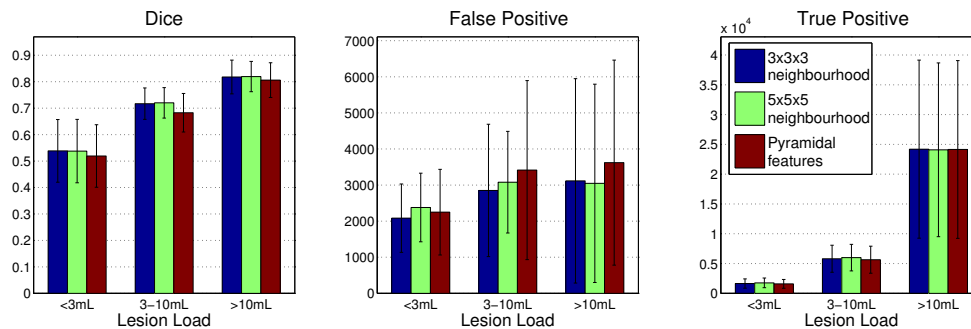

Figure 6. Segmentation performance with different feature types (using the 4 modalities).

4.3. Performance with regard to the mask use

Using the FLAIR modality and $3 \times 3 \times 3$ neighbourhood feature type, the impact of the mask on the segmentation performance has been evaluated. The use of the mask in the pre-processing instead of post-processing significantly decreased FP and led to a much better DSC (Fig. 7). 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) reduced the computation time (41 times computation speed-up on our dataset with $\tau = 2$).

4.4. Performance comparison with other supervised classification algorithms

Using the FLAIR + T1 combination and $3 \times 3 \times 3$ feature type, the performance of the SVM classifier was compared with several other supervised classifiers: kNN (with k=100 as in [2]), Naives Bayes, Parzen window and decision tree.
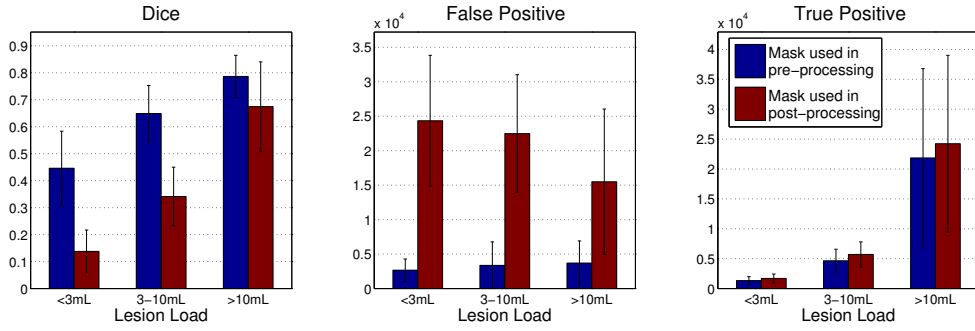
Figure 7. Using our mask M_τ in the pre-processing gives better
results than using it only as a post-processing step.

The kNN classifier classifies a features x to the class with the highest cardinality among the k nearest neighbours of x in the training set.

$$\hat{y}(x) = \underset{y}{\text{argmax}} \; Card\{x_i \in kNN(x) \text{ and } y_i = y\} \tag{8}$$

The Naive Bayes method computes the posterior probability of a feature x belonging to each class y, and classifies according the largest posterior probability (see Eq. (9)). To compute the parameters of the probability density of feature x given class y, the features are assumed conditionally independent given the class.

$$\hat{y}(x) = \underset{y}{\text{argmax}} \, P(y|x) = \underset{y}{\text{argmax}} \, \frac{P(x|y)P(y)}{P(x)} = \underset{y}{\text{argmax}} \, P(x|y)P(y) \tag{9}$$

In the case of the Parzen window, the prior probability is estimated as

$$P(x|y) = \frac{1}{Card\{i|y_i = y\}} \sum_{i,y_i=y} K(x, x_i) \tag{10}$$

where K is for example the Gaussian kernel. The Parzen window classifier is then found using the Bayes rule:

$$\hat{y}(x) = \underset{y}{\text{argmax}} \, P(y|x) = \underset{y}{\text{argmax}} \, \frac{\sum_{i,y_i=y} K(x, x_i)}{\sum_i K(x, x_i)} \tag{11}$$

On this dataset, using combined $3 \times 3 \times 3$ features from FLAIR and T1 modalities, SVM obtained the best DSC results, followed by Parzen windows, kNN, decision tree and Naive Bayes (Fig. 8). In terms of FP, kNN, Parzen window and SVM provide the best results. These tree classifiers also provide the best specifities. However, kNN and Parzen window classifier have a quite-low sensitivity, whereas SVM is most sensitive (followed by classification tree).

4.5. Performance bounds from mask parameter setting

The influence of the τ parameter in the mask creation on the segmentation performance was evaluated. Three performance bounds were computed, which are independent of the modality combination and feature type.

First, the upper bound of the final FP, which is equal to the number of FP for the "lesion-everywhere" classifier (worst-case scenario in terms of FP). Figure 9a shows the FP upper bound for different values of τ. The higher τ is, the smaller is the mask M_τ (which is the area used for the machine learning), and therefore the lower is the FP upper bound. Second, the lower bound of the final FN, which is obtained when using the optimal classifier within the mask (best case scenario in terms of FN). Figure 9b shows the FN lower bound for different values of τ. The higher τ is, the
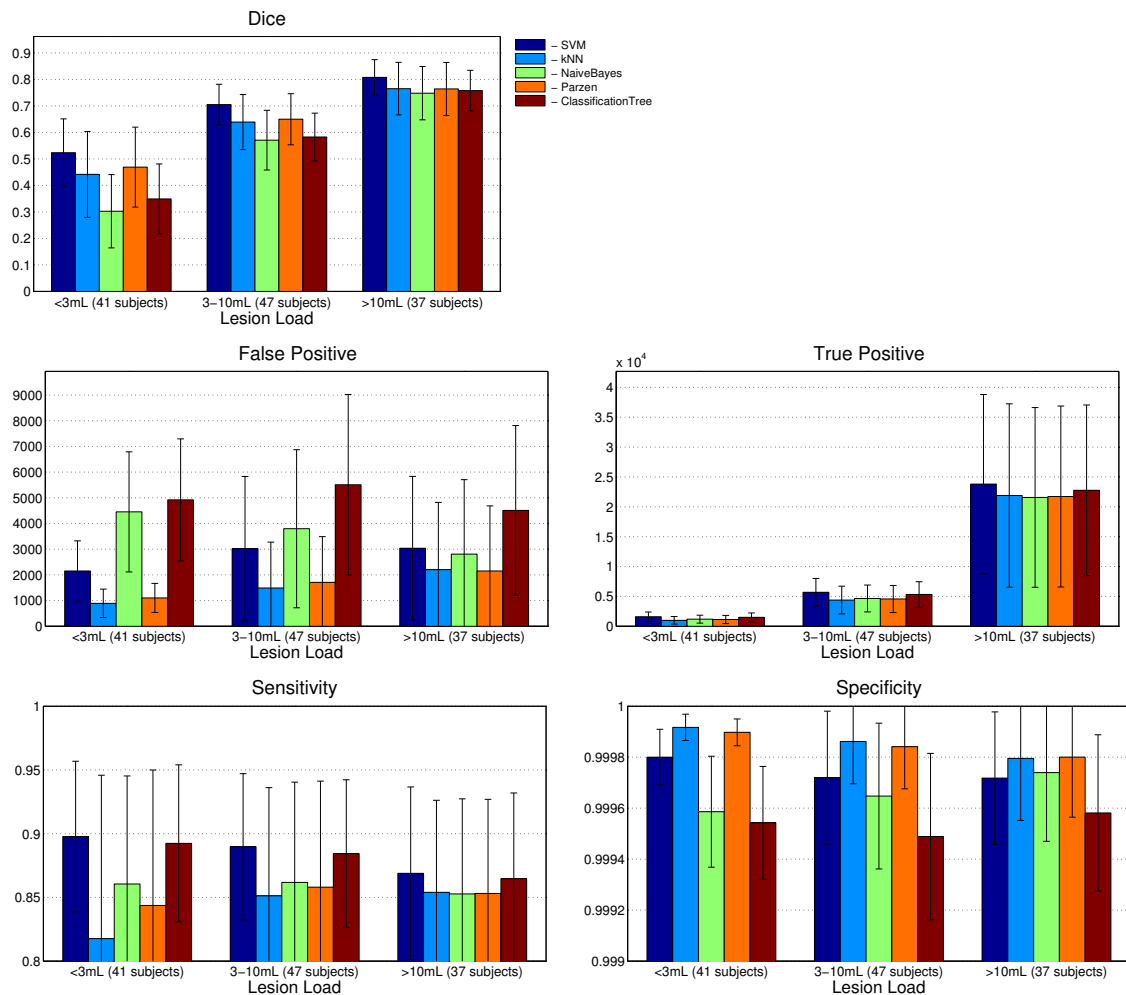
Figure 8. Segmentation performance with different algorithms (using the $3 \times 3 \times 3$ neighbourhood intensity feature type).

smaller is the mask M_τ, the more lesion features will be left out, and therefore the higher is the FN lower bound. Third, the upper bound of the final Dice score, which is similarly obtained using the optimal classifier within the mask. The higher τ is, the lower is the Dice upper bound (Fig. 9c).

These graphs give insight on the impact of the τ parameter. If τ is too high, the final segmentation performance will be low, no matter how good is the classifier segmenting inside M_τ. If τ is too low, the algorithm has a high risk of FP, which is a known drawback in WM lesion segmentation. Setting τ therefore involves a trade-off between the use of tissue-information to reduce the risk of FP and a high upper bound performance. The value $\tau = 2$, which was selected for all experiments, is in the range of acceptable values decreasing the FP upper bound without decreasing too much the Dice upper bound.

5. CONCLUSION

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
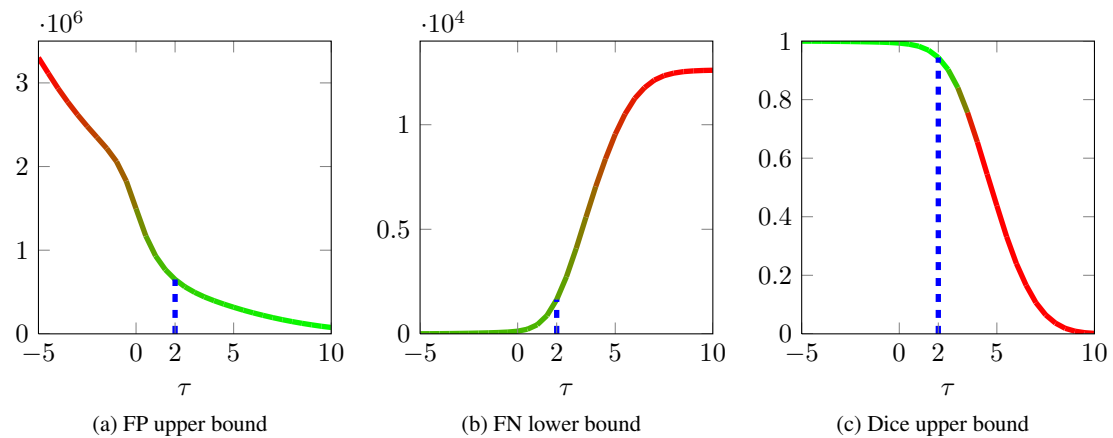
Figure 9. Performance bounds due to the threshold τ in the mask creation. Increasing τ decreases the upper bound of FP (a), increases the lower bound of FN (b), and decreases the lower bound of the Dice score (c). The value $\tau = 2$ was the value selected for all experiments.

accurate segmentation results. Using our pipeline and our dataseet, SVM has a higher classification performance than other supervised algorithms such as kNN, Naive Bayes, Parzen window and decision tree.

In this work we also quantified 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 [1]. 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. We have given insight on the trade-off related to the threshold parameter selection in the mask creation. Increasing the threshold decreases the upper bound of FP (and therefore the potential risk of final FP). However, a low FP upper bound comes at a price as increasing the threshold increases the FN lower bound and decreases the Dice upper bound, which means a treshold too high would cause poor final performance no matter how good is the classifier. Second, computation time and storage space required are significantly lower (41 times lower on our dataset with the chosen threshold) 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.

into the early detection and causation of AD. For a list of members of the AIBL Research Group, see www.aibl.csiro.au/partners.html.

REFERENCES

1. Lao Z, Shen D, Liu D, Jawad AF, Melhem ER, Launer LJ, Bryan RN, Davatzikos C. Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. *Academic Radiology* 2008; **15**(3):300 – 313, doi:DOI:10.1016/j.acra.2007.10.012. URL http://www.sciencedirect.com/science/article/B75BK-4RV7SX5-7/2/99b40e635d683ccf28ff68bc0a4277da.

2. Anbeek P, Vincken KL, van Osch MJP, Bisschops RHC, van der Grond J. Probabilistic segmentation of white matter lesions in MR imaging. *NeuroImage* 2004; **21**(3):1037 – 1044, doi:DOI:10.1016/j.neuroimage.2003.10.012. URL http://www.sciencedirect.com/science/article/B6WNP-4BDY5W0-4/2/b162dcf916098b5d0d4270f2955405b7.

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

4. Llad'o X, Oliver A, Cabezas M, Freixenet J, Vilanova JC, Quiles A, Valls L, Ramió-Torrentà L, Àlex Rovira. Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. *Information Sciences* 2012; **186**(1):164 – 185, doi:10.1016/j.ins.2011.10.011. URL http://www.sciencedirect.com/science/article/pii/S0020025511005548.

5. Dyrby TB, Rostrup E, Baaré WF, van Straaten EC, Barkhof F, Vrenken H, Ropele S, Schmidt R, Erkinjuntti T, Wahlund LO, *et al.*. Segmentation of age-related white matter changes in a clinical multi-center study. *NeuroImage* 2008; **41**(2):335 – 345, doi:DOI:10.1016/j.neuroimage.2008.02.024. URL http://www.sciencedirect.com/science/article/B6WNP-4RY6WPP-2/2/e4414cdc04616feca32799b21c05e7a8.

6. Scully M, Anderson B, Lane T, Gasparovic C, Magnotta V, Sibbitt W, Roldan C, Kikinis R, Bockholt HJ. An automated method for segmenting white matter lesions through multi-level morphometric feature classification with application to lupus. *Frontiers in Human Neuroscience* 2010; **4**(0), doi:10.3389/fnhum.2010.00027. URL http://www.frontiersin.org/Journal/Abstract.aspx?s=537&name=humanneuroscience&ART_DOI=10.3389/fnhum.2010.00027.

7. Sajja B, Datta S, He R, Mehta M, Gupta R, Wolinsky J, Narayana P. Unified approach for multiple sclerosis lesion segmentation on brain mri. *Annals of Biomedical Engineering* 2006; **34**:142–151. URL http://dx.doi.org/10.1007/s10439-005-9009-0, 10.1007/s10439-005-9009-0.

8. Datta S, Sajja BR, He R, Wolinsky JS, Gupta RK, Narayana PA. Segmentation and quantification of black holes in multiple sclerosis. *NeuroImage* 2006; **29**(2):467 – 474, doi:10.1016/j.neuroimage.2005.07.042. URL http://www.sciencedirect.com/science/article/pii/S1053811905005562.

9. Fiot JB, Cohen LD, Raniga P, Fripp J. Efficient Lesion Segmentation using Support Vector Machines. *Computational Vision and Medical Image Processing: VipIMAGE 2011*, Olhão, Portugal, 2011; ISBN 9780415683 951. URL http://hal.archives-ouvertes.fr/hal-00662344.

10. Zacharaki EI, Kanterakis S, Bryan RN, Davatzikos C. Measuring brain lesion progression with a supervised tissue classification system. *Proceedings of MICCAI 2008*, 2008; 620–627, doi:http://dx.doi.org/10.1007/978-3-540-85988-8_74.

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

12. Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, Lautenschlager NT, Lenzo N, Martins RN, Maruff P, *et al.*. 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* 2009; **21**(4):672–87. URL http://www.biomedsearch.com/nih/Australian-Imaging-Biomarkers-Lifestyle-AIBL/19470201.html.

13. Ourselin S, Roche A, Subsol G, Pennec X, Ayache N. Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing* 2001; **19**(1-2):25 – 31, doi:DOI:10.1016/S0262-8856(00)00052-4. URL http://www.sciencedirect.com/science/article/B6V09-41ST5DM-8/2/53745696d1923da242063714f4abd4ec.

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

15. Acosta O, Bourgeat P, Zuluaga MA, Fripp J, Salvado O, Ourselin S. Automated voxel-based 3D cortical thickness measurement in a combined Lagrangian-Eulerian PDE approach using partial volume maps. *Medical Image Analysis* 2009; **13**(5):730 – 743, doi:DOI:10.1016/j.media.2009.07.003. URL http://www.sciencedirect.com/science/article/B6W6Y-4WR66MY-1/2/4898cb5fe75cdad82a3e2345765671c1.

16. Collins D, Zijdenbos A, Kollokian V, Sled J, Kabani N, Holmes C, Evans A. Design and construction of a realistic digital brain phantom. *Medical Imaging, IEEE Transactions on* june 1998; **17**(3):463 –468, doi:10.1109/42.712135.

17. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE transactions on medical imaging* August 1999; **18**(8):712–721, doi:10.1109/42.796284. URL http://dx.doi.org/10.1109/42.796284.

18. Melacci S, Belkin M. Laplacian Support Vector Machines Trained in the Primal. *Journal of Machine Learning Research* March 2011; **12**:1149–1184.

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

20. Dice LR. Measures of the amount of ecologic association between species. *Ecology* July 1945; **26**(3):297–302.

21. Ott RL, Longnecker MT. *An Introduction to Statistical Methods and Data Analysis*. 6 edn., Duxbury Press, 2008. URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0495017582.