

# Measuring brain lesion progression with a supervised tissue classification system

Evangelia I. Zacharaki<sup>1</sup>, Stathis Kanterakis<sup>1</sup>, R. Nick Bryan<sup>1</sup>, Christos Davatzikos<sup>1</sup>

<sup>1</sup> Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA  
{Eva.Zacharaki, Efstathios.Kanterakis, R.Nick.Bryan, Christos.Davatzikos}@uphs.upenn.edu

**Abstract.** Brain lesions, especially White Matter Lesions (WMLs), are associated with cardiac and vascular disease, but also with normal aging. Quantitative analysis of WML in large clinical trials is becoming more and more important. In this paper, we present a computer-assisted WML segmentation method, based on local features extracted from conventional multi-parametric Magnetic Resonance Imaging (MRI) sequences. A framework for preprocessing the temporal data by jointly equalizing histograms reduces the spatial and temporal variance of data, thereby improving the longitudinal stability of such measurements and hence the estimate of lesion progression. A Support Vector Machine (SVM) classifier trained on expert-defined WML's is applied for lesion segmentation on each scan using the AdaBoost algorithm. Validation on a population of 23 patients from 3 different imaging sites with follow-up studies and WMLs of varying sizes, shapes and locations tests the robustness and accuracy of the proposed segmentation method, compared to the manual segmentation results from an experienced neuroradiologist. The results show that our CAD-system achieves consistent lesion segmentation in the 4D data facilitating the disease monitoring.

**Key words:** white matter lesions, classification, segmentation, MRI, SVM, CAD, lesion progression

## 1 Introduction

Population studies have shown that brain lesions, especially WMLs, are associated with several diseases, such as arterial fibrillation, arterioscleroses, impaired cognition and others [1,2]. The increased interest in brain lesion research may improve diagnosis and prognosis possibilities for patients with cardiovascular symptoms. Since brain lesion patterns are very heterogeneous, ranging from punctuate lesions in the deep white matter to large confluent periventricular lesions, the scoring of such lesions is complicated. For longitudinal studies aiming to capture relatively small changes in brain lesion patterns, accurate information of lesion volume and location is essential. It is known that expert-based delineation of brain lesions is difficult to reproduce across raters, or even within the same rater, and that combination of readings from independent reader may be necessary in a longitudinal study.

The use of an automated segmentation method that detects brain lesions with a

high sensitivity and specificity could be advantageous. Most of the methods in the literature have been developed for the detection of Multiple Sclerosis (MS) lesions by combining multi-parametric MR images, i.e. images obtained via different MR protocols. The advantage of integrating information from multiple sequences is that it can reduce the uncertainty and increase the accuracy of the segmentation. They usually apply supervised voxel-wise classification in which the desired segmentation is known (expert manual delineation) and used as a training set to build the segmentation model [3,4,5,6]. However, relatively less attention has been given to brain lesion segmentation in elderly individuals, and AD or diabetic patients. Since MS lesions present different characteristics from lesions in elderly and/or diabetic individuals, those methods are not directly applicable to our studies, albeit they have formed the foundation for our development. Because of the decreased contrast between white matter (WM) and gray matter (GM) in MRI in elderly, techniques that require the segmentation of WM and GM for the extraction of the WMLs perform moderately well when applied to geriatric patients, especially when they were originally designed and trained to extract lesions in MS patients. Mohamed *et al.* [7] presented a method for differentiating WMLs using a supervised classification method with relatively good sensitivity but somewhat limited specificity to lesions.

Moreover, only a few methods have combined space and time into the lesion characterization process [8,9]. These approaches focused primarily on quantifying the temporal variations of MS lesions, important in differentiating active from chronic lesions. In contrast to the complicated MRI dynamics of lesions in MS, the monitoring of WMLs does not require spatiotemporal modeling, since the effects in WMLs are irreversible.

In this paper, we present a computer-assisted WML segmentation approach that has been designed to process longitudinal MR scans of elderly diabetes patients [10]. Our method uses a combination of image analysis and pattern classification using SVM. Image intensities from multiple MR acquisition protocols, after co-registration, are used to form a voxel-wise feature vector that helps to discriminate lesion from various normal tissue image profiles during segmentation. In general, there are three steps in our approach. First, we jointly preprocess baseline and follow-up data. The preprocessing step includes co-registration of different MR modalities of the same patient, skull-stripping, intensity normalization, as well as inhomogeneity correction. Second, a set of training samples is manually delineated by an expert reader on the baseline images, and then used to build a classification model via SVM [11] and the AdaBoost algorithm [12]; this step is applied only once, during training. Third, the SVM model is used to perform the voxel-wise segmentation of the longitudinal data. The methodology is validated against the expert human readings. In the current study we present results obtained using training samples defined by a single expert. However, we have also performed experiments using samples defined by two different experts, which have shown that the lesion load detected by the proposed method showed high correlation across experts.

## 2 Methods

MRI's used herein were obtained from individuals with diabetes, with an inter-scan period of approximately 3 years. All 42 participants' exams consisted of transaxial T1-w, T2-w, PD and FLAIR scans with 23 of the participants having currently a follow-up study. All scans except T1-w were performed with a 3 mm slice thickness, no slice gap, a  $240 \times 240$  mm FOV and a  $256 \times 256$  scan matrix. T1-w scans were performed with a 1.5mm slice thickness, same slice gap, FOV and scan matrix. The data are preprocessed and features are extracted in order to train an SVM classifier, as explained next.

### 2.1 Data preprocessing and feature extraction

*Preprocessing:* The multiple images acquired from the same individual are co-registered, in order to compensate for possible motion between scans. Using as similarity metric the correlation ratio and normalized mutual information, affine registration [13] implemented in FSL [14] is employed for co-registration of multi-modality images. The FLAIR image of each subject is used as reference space, to which all other sequences are transformed. Then each sequence of the follow-up data is co-registered with the corresponding aligned baseline sequence. The intra-modal registration can generally achieve better accuracy. After co-registration, a deformable model based skull-stripping algorithm called BET [15], implemented in FSL [14], is used to generate an initial brain tissue mask from the co-registered T1-w image, and then this brain tissue mask is used to extract the brain region from all other modality images. Finally, for each image volume, inhomogeneities are corrected by N3 [16].

A fundamentally important step in our segmentation algorithm, as well as in most supervised classification methods, is the standardization of the image histograms. To this end, a linear transformation (translation and scaling) is calculated, that minimizes the  $L_2$ -norm of the histogram difference between transformed image and template image, and reverse. The histograms are first smoothed and the bin representing the background is excluded from the least-squares error minimization. In order to achieve high temporal stability in the histogram normalization, we constrain the baseline and the follow-up histograms to be normalized to the template used in the study in exactly the same way. Follow-up histograms are normalized to their respective - standardized to the template - baselines, a problem that is relatively easy to solve, since baseline and follow-up images belong to the same individual. Thus, histograms between images of the same subject get aligned consistently, and the temporal variance is reduced. We will refer to this approach of histogram equalization using (intra-subject) 4D data as *TVR* (temporal variance reduction), as opposed to the standard 3D approach aligning baseline and follow-up images independently to the template histogram. This second approach aims at reducing the inter-subject variance, and will be refer to as *IVR*. Since the inter-subject variability (between subject and template) is much larger than the intra-subject temporal variability, a global histogram matching based on *IVR* fails, and tends to produce inconsistent 4D WML segmentation. We will show the proposed *TVR* approach is more appropriate for measuring temporal WML

change.

*Feature extraction:* In general, the amount of intensity overlap between WMLs and normal tissue varies greatly across different modalities. In T1-w images, WMLs have intensities similar to GM, and in T2-w and PD images, WMLs look very similar to CSF. Although the FLAIR image has the least intensity overlap between WMLs and normal tissues, it has been suggested in the literature that FLAIR is less sensitive in the posterior fossa [17], may lead to “overestimation” of lesion load, and has a higher inter-vendor variability [18]. Furthermore, FLAIR may present hyperintensity artifacts [19] that might lead to false positives, thereby rendering it difficult to use only the FLAIR images to segment WMLs. Therefore, it is important to integrate information from different modalities, in order to minimize the ambiguity in identifying WMLs from using only a single modality image. A feature vector is computed for each non-background voxel in a 3D reference space for each subject. In order to make the features robust to noise, each sequence is smoothed by a Gaussian filter with a very small kernel (0.5mm).

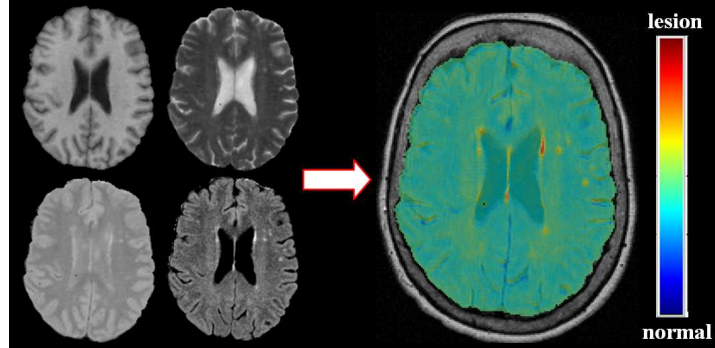
## **2.2 Lesion segmentation (training and testing)**

SVM has been shown to be a powerful technique for learning from data and in particular, for solving binary classification problems [11]. In a binary classification task like the one in our study (normal tissue/lesion tissue), the aim is to find an optimal separating hyperplane between the two data sets. In our application we use linear SVM.

*Training SVM via AdaBoost:* A nonlinear pattern classifier is constructed from the entire training set, i.e. by using all lesion voxels of all training scans as examples of imaging profiles to be recognized in new scans, along with a large number of normal tissue voxels. Because the number of normal tissue voxels is far higher than the number of lesion voxels, it is essential to select only a representative set of normal tissue voxels comparable to the number of lesion voxels. This selection is not random, but it is rather guided by the classification results themselves, using the AdaBoost algorithm [12]. This approach is based on a sequence of classifiers that rely increasingly on misclassified voxels, since those are presumably the voxels on which the classifier must focus. During this adaptive boosting procedure, each sample receives a weight that determines its probability of being selected in a training set for the next iteration. If a training sample is accurately classified, then its likelihood of being used again in subsequent iterations is reduced; conversely, if a training sample is inaccurately classified, then its likelihood of being used again is increased.

*Segmentation (testing):* In the testing stage, T1-w, T2-w and PD images of a new (not in the training set) subject are firstly preprocessed (co-registered etc.) by the procedure described before, and then the pseudo-likelihood of each voxel being WML is measured by the generated SVM classifier. The output of SVM is a scalar measure of abnormality (as shown in Fig.1), which is further binarized by an optimal threshold to produce the WML segmentation.

Subsequently, two post-processing steps are applied in order to remove remaining false positives. Extra-axial hyperintense regions, like fat in the orbits, can not always



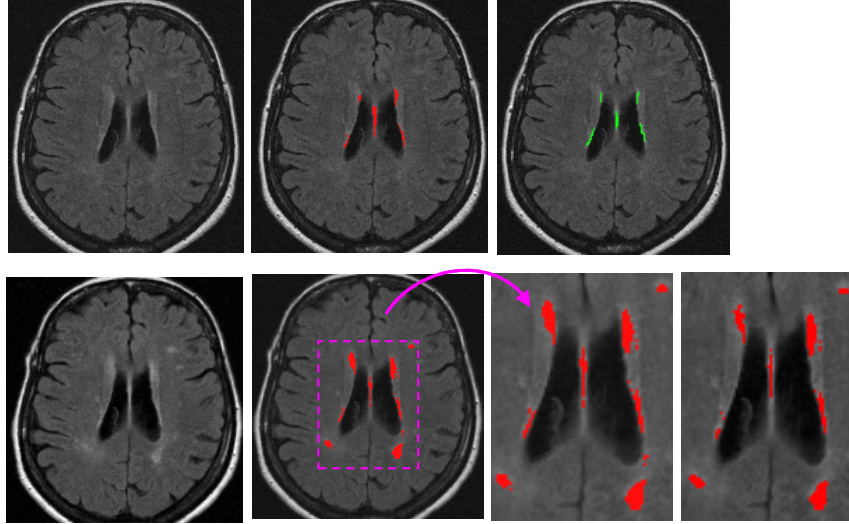
**Fig. 1.** Integration of multi-parametric MRI for creation of voxel-wise abnormality maps based on SVM classifier. From left to right and top to bottom T1-w, T2-w, PD, FLAIR image and generated abnormality map (shown in color scale). The color scale is only relative, i.e. appropriate (non-linear) scaling is required, such as fitting a sigmoid function using Platt's method [20], to represent actual (pseudo)probabilities.

be completely removed by the skull-stripping algorithm used in preprocessing stage. Imaging profiles belonging to these regions are more similar to WMLs than that of normal tissue and therefore they are eliminated from the segmentation mask after SVM classification. This is done by morphological operations combined by adaptive thresholding in skull-stripped FLAIR image. Finally, remaining false positives are further reduced by applying spatial constraints using an unsupervised clustering technique. Specifically, the automated segmentation tool FAST [21] was applied to segment the FLAIR image into 6 classes. The 6<sup>th</sup> class includes hyperintense regions in FLAIR, such as bright GM regions and lesions, and was used as a mask to constraint the segmentation.

It should be noted that all steps in the WML segmentation procedure are automated and the same parameters are used for all subjects. Only one parameter has been shown to be important and vary across subjects, which is the threshold for binarizing the abnormality map generated by the SVM classifier. This threshold is optimized for each subject in the training set by maximizing the *Jaccard score* ( $Jac$ ) [12]. The average threshold maximizing the *Jaccard score* is then used as the default value for segmenting new data. The *Jaccard score* is defined as  $Jac = TP / (TP + FP + FN)$ , where  $TP$ ,  $FP$  and  $FN$  stand for true positive, false positive, and false negative, respectively.

### 3 Results

The dataset consisted of 23 subjects having baseline and follow-up data and 19 subjects having only baseline data. We used one subset to train a linear SVM classifier for lesion segmentation, a second subset to optimize the threshold (based on the *Jaccard score*) for generating a binary lesion mask from the abnormality map and the remaining subjects ( $N=16$ ) to assess the 4D the segmentation method. In the future we plan to perform leave-one-out cross validation for training and testing in order to



**Fig. 2.** Evaluation of 4D WML segmentation. The 1<sup>st</sup> row shows the baseline and the 2<sup>nd</sup> row the follow-up images. The automatic WML segmentation (red) in baseline is very similar to the expert-defined WML (green). The zoomed figures in 2<sup>nd</sup> row show the result of two histogram equalization approaches: the proposed *TVR* (left) versus the standard *IVR* (right).

exploit more effectively the dataset.

Fig. 2 shows the progression of white matter lesions in an elderly subject. The first row illustrates the baseline image and the corresponding lesion segmentation (red for our method and green for the expert-defined WML). It can be noticed the 3D WML segmentation has high sensitivity and specificity. The second row shows the follow-up image and the segmentation produced by our method by either aligning the histogram of the follow-up image to the intensity-normalized baseline image (proposed approach, shown on the left) or, as usually performed, by aligning the histogram of the follow-up image independently to the template image (usual approach, shown on the right). The proposed approach has higher sensitivity in WML segmentation.

Since expert-defined ground truth is not available for the follow-up studies, the segmentation in the follow-up cannot be directly assessed. However, we calculate the lesion volume in baseline and follow-up studies and assess the rate of change by the gradient of the time function. Since only disease progression (increasing lesion load) is expected, the temporal consistency of the multi-parametric segmentation scheme can be easily evaluated. The lesion load for baseline and follow-up data is shown in Table 1. The results show that the proposed *TVR* approach for jointly normalizing the histograms of baseline and follow-up images, gives increasing lesion volume for all subjects. On the contrary, the alignment of histograms of each baseline and follow-up images independently to the template applying a global transformation, does not optimally estimate the MRI dynamics for each subject and therefore false reduction in lesion load is observed. An example is shown in Fig. 3 (bottom right) where the follow-up images are processed independently of the baseline using the *IVR* approach. It can be seen that one of the lesions has not been detected in the follow-up study.

## 4 Discussion and Conclusions

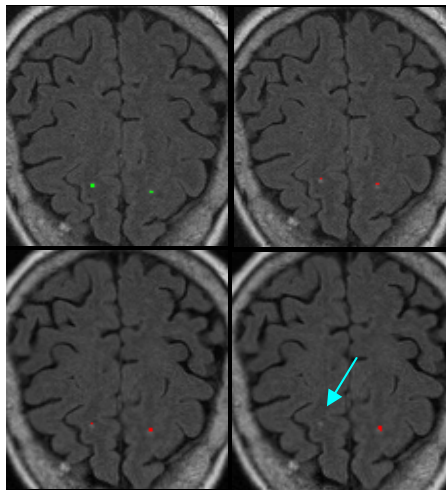
The objective, quantitative, and reproducible evaluation of WMLs has been a challenge in many neuroimaging studies. Although qualitative readings have been employed by many studies, the relatively limited sensitivity and inter-rater agreement is an obstacle, particularly in longitudinal studies or in studies seeking to detect subtle effects. This paper presents a CAD-system for monitoring the progression of WMLs that is based on the integration of different MR acquisition protocols and training of a nonlinear pattern classification algorithm. Our experiments show that the proposed system segments WML accurately even when the load is small. More important, by jointly normalizing the histogram of baseline and follow-up images, lesion load increase was consistently observed for all subjects.

We are currently in the process of applying this method to data from different centers in multi-site studies seeking to quantify vascular disease.

**Acknowledgments.** The authors would like to thank the committee of ACCORD-MIND project, which is funded by the NIA through an intra-agency agreement with NIH/NIH [Y3-HC-3065], for providing the datasets, valuable comments and giving us permissions to publish this paper. This work was also supported, in part, by R01-14973.

**Table 1.** Evaluation of WML progression in 16 participants

Volume in baseline (mm <sup>3</sup> )	Volume increase in follow-up (%)	
	TVR	IVR
1622	91.22	38.05
1371	11.73	-4.81
654	375.00	340.32
2716	63.69	55.15
427	142.59	116.05
1113	7.58	-11.14
3533	15.52	27.16
2078	170.43	132.74
272	20.39	-18.45
3987	42.79	27.84
1192	90.27	27.21
1553	144.99	53.65
1830	42.80	18.88
1972	78.74	28.61
308	44.44	-56.41
198	8.00	5.33



**Fig. 3.** Example illustrating misleading disease regression when using the IVR approach. 1<sup>st</sup> row: expert-defined WML (green) and WML segmentation in baseline. 2<sup>nd</sup> row: WML segmentation in follow-up with TVR approach (left) and IVR approach (right). The TVR approach is more consistent, whereas the IVR approach misses a lesion in the follow-up that was there in the baseline.

## References

1. de Groot, J.C., et al.: Cerebral white matter lesions and cognitive function: the Rotterdam Scan Study. *Annals of Neurology* 47(2), 145-151 (2000)
2. Longstreth, W.T., Jr., et al.: Clinical correlates of white matter findings on cranial magnetic resonance imaging of 3301 elderly people. The Cardiovascular Health Study. *Stroke*. 27(8), 1274-82 (1996)
3. Zijdenbos, A.P., Forghani, R., Evans, A.C.: Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *IEEE Trans. on Medical Imaging* 21(10), 1280-1291 (2002)
4. Wei, X., et al.: Quantitative Analysis of MRI Signal Abnormalities of Brain White Matter With High Reproducibility and Accuracy. *Journal of Magnetic Resonance Imaging* 15(2), 203-209 (2002)
5. Anbeek, P., et al.: Probabilistic segmentation of white matter lesions in MR imaging. *NeuroImage* 21(3), 1037-1044 (2004)
6. Admiraal-Behloul, F., et al.: Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. *NeuroImage* 28(3), 607-617 (2005)
7. Mohamed, F.B., et al.: Increased differentiation of intracranial white matter lesions by multispectral 3D-tissue segmentation: preliminary results. *Magnetic Resonance Imaging* 19(2), 207-18 (2001)
8. Gerig, G., et al.: Exploring the discrimination power of the time domain for segmentation and characterization of lesions in serial MR data. In: *MICCAI*, Cambridge, MA (1999).
9. Meier, D.S., Guttman, C.R.G.: MRI time series modeling of MS lesion development. *NeuroImage* 32(2), 531-537 (2006)
10. Williamson, J., et al.: The Action to Control Cardiovascular Risk in Diabetes Memory in Diabetes Study (ACCORD-MIND): Rationale, Design, and Methods. *American Journal of Cardiology* (2007)
11. Vapnik, V.N.: *Statistical Learning Theory*, New York, Wiley. 736 (1998)
12. Tan, P.-N., Steinbach, M., Kumar, V.: *Introduction to Data Mining* (2005)
13. Jenkinson, M., Smith, S.M.: A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* 5(2),143-156 (2001)
14. Smith, S.M., et al.: Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23(S1), 208-219 (2004)
15. Smith, S.M.: BET: Brain Extraction Tool. FMRIB technical report TR00SMS26
16. Sled, J., Zijdenbos, A., Evans, A.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. on Medical Imaging* 17(1), 87-97 (1998)
17. Gawne-Cain, M.L., et al.: MRI lesion volume measurement in multiple sclerosis and its correlation with disability: a comparison of fast fluid attenuated inversion recovery (fFLAIR) and spin echo sequences. *J. Neurol. Neurosurg. Psychiatry* 64(February), 197-203 (1998)
18. Rovaris, M., et al.: Relevance of Hypointense Lesions on Fast Fluid-Attenuated Inversion Recovery MR Images as a Marker of Disease Severity in Cases of Multiple Sclerosis. *Am. J. Neuroradiology* 20(5), 813 - 820 (1999)
19. Bakshi, R., et al.: Intraventricular CSF Pulsation Artifact on Fast Fluid-Attenuated Inversion-Recovery MR Images: Analysis of 100 Consecutive Normal Studies. *Am. J. Neuroradiology* 21(3), 503-508 (2000)
20. Platt, J.: Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In: Schuurmans D. (ed.) *Advances in Large Margin Classifiers*. MIT Press, Cambridge, MA (2000)
21. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Trans. on Medical Imaging*, 20(1), 45-57 (2001)