Chapter 3

Adapted from:

**Fully automatic segmentation of white matter hyperintensities in MR images of the elderly**

F Admiraal-Behloul, PhD
DMJ van den Heuvel, MSc
H Olofsen, MSc
MJP van Osch, PhD

J van der Grond, PhD
MA van Buchem, MD PhD
JHC Reiber, PhD

*Institutional affiliations:* From the departments of ¹Radiology division of Image Processing, ²Radiology, Leiden University Medical Center, The Netherlands.

*NeuroImage 2005 (in press)*
Abstract

Quantitative image analysis is increasingly important in large clinical trials. Several methods are available for performing white matter hyperintensity (WMH) volume quantification. These methods vary in the amount of human interaction involved. In this paper, we describe a fully automatic segmentation that was used to quantify WMHs in a large clinical trial with elderly subjects. Our segmentation method combines information from 3 different MR images: proton density (PD), T2-weighted and fluid attenuated inversion recovery (FLAIR) images. Our method uses an established artificial intelligent technique (fuzzy inference system) and does not require extensive computations. The reproducibility of the WMH segmentation was evaluated in 9 patients who underwent scan-rescan with repositioning. Here, an intraclass correlation coefficient (ICC) of 0.91 was obtained. The effect of differences in image resolution (i.e. FLAIR 3mm versus FLAIR 6mm) on WMH volume measurements was tested in 44 patients. Volumetric agreement between FLAIR 3mm and FLAIR 6mm was excellent (ICC = 0.99). The accuracy of the segmentation was evaluated in 100 patients for whom manual delineation of WMHs was available; the obtained ICC was 0.98 and the similarity index was 0.75. The software requires less than 2 minutes for the entire delineation process of one individual and demonstrated very high volumetric and spatial agreement with expert delineation.
Introduction

Cerebral white matter hyperintensities (WMHs) on magnetic resonance images (MRI) of the elderly have attracted the interest of many researchers for more than a decade now. They were initially reported in people with dementia and were suggested to be part of a dementing process\(^1\)\(^2\). Later, these lesions were also revealed to be frequent in healthy old adults\(^3\). Besides their correlation with advanced age\(^4\), WMHs were suggested to be associated with various diseases such as transient ischemic attacks and ischemic stroke\(^5\)-\(^7\), Alzheimer’s disease\(^8\)-\(^10\), late onset depression\(^11\), impairment of gait and balance\(^12\) and migraine\(^13\).

MRI is highly sensitive to the lesions affecting the cerebral white matter. Damaged white matter usually has a prolonged T2 relaxation due to increased tissue water content and degradation of the macromolecular structure of myelin. Therefore, white matter lesions are well depicted with conventional PD and T2 weighted spin echo (SE) or fast SE sequences, but are even more conspicuous on fluid attenuated inversion recovery (FLAIR) images.

To assess these signal abnormalities of the white matter, several visual rating scales have been proposed\(^14\)-\(^19\). The scales are defined to serve various purposes and have different properties. Mäntylä et al\(^19\) compared 13 different scales and concluded that the heterogeneous properties of the different scales resulted in inconsistencies in previous studies and called for standardization.

Efforts are continuously made to compare, tune, standardize or even invent new visual rating scales\(^5\),\(^7\),\(^14\),\(^15\),\(^18\),\(^20\). However, visual rating scales will probably remain inferior to quantitative methods. Melhem et al\(^21\) reported that trained reviewers can reliably detect a 15% or more increase in diameter for any original lesion size and 10% or more for lesions with a diameter larger than 6mm. Smaller changes were undetectable for the 3 trained reviewers who participated in the experiment. This work demonstrates clearly the limitation of the human eye in detecting small changes in lesion size. However, in drug trials such small changes might play a decisive role.

Quantitative image analysis and its advantages have been reported in the literature for more than a decade. Several methods have been proposed for performing quantitative white matter lesion load measurements\(^22\)-\(^27\). These methods vary from fully manual (manual outline techniques) to fully automatic (no user interaction). The fully manual methods are labor intensive and time consuming. They require well-trained reviewers and remain moderately accurate\(^28\). These methods are therefore quite impractical in large-scale studies. Many semi-automatic algorithms have been later introduced to reduce the processing time and increase the intra- and interrater variability\(^29\)-\(^33\). Nowadays, these techniques are considered as the best alternative to manual contouring or visual scoring in large scale studies. However, even if they are less time consuming and more reproducible\(^22\),\(^25\),\(^33\),\(^34\) they still remain labor intensive for trials on large numbers of subjects. The need for
more automated methods (ultimately fully automatic) is growing as large-scale longitudinal trials in degenerative brain disease are becoming more and more important.

There are only few fully automatic WMH quantification techniques and most of them are designed for detection of WMHs in multiple sclerosis (MS)\textsuperscript{24, 26, 35, 36}. Due to the decreased contrast between white and gray matter in MR images of the elderly, techniques that require the segmentation of white and gray matter for the extraction of the WMHs, perform moderately when applied to geriatric patients; especially, when they were originally designed and trained to extract lesions in MS patients.

We present a fully automatic lesion quantification method that has been designed to process MR scans of the elderly and used in a large clinical study: the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER) a randomized, double blind, placebo-controlled trial\textsuperscript{37}. Our method is based on simple algorithmic techniques that do not require extensive computations. To evaluate our automatic segmentation approach, we investigated i) the effect of scan-rescan with repositioning in 9 patients, ii) the effect of image resolution (slice thickness) in 44 patients and iii) the accuracy of the delineation in 100 patients for which manual delineation was available.

**Materials and Methods**

**Material** Inclusion and exclusion criteria for the PROSPER study have been described in detail elsewhere\textsuperscript{37}. MRI was performed on a clinical MR-system operating at 1.5 Tesla field strength (Philips Medical Systems, Best, The Netherlands) and had the following characteristics: Dual fast spin echo images (TE 27 ms, TR 3000 ms, 48 contiguous 3 mm slices with no gap, matrix 256x256, FOV 220). FLAIR: TE 100 ms, TR 8000, 48 contiguous 3mm slices with no gap, matrix 256x256, FOV 220) were obtained from all subjects. To test the scan-rescan effects on the quantitative analysis, 9 patients were scanned and rescanned few minutes later, with repositioning. In 44 patients, an extra FLAIR with a lower resolution has been acquired (TE 120 ms, TR 8000, 22 contiguous 6mm slices with no gap, matrix 256x256, FOV 220). To evaluate the accuracy of the segmentation we randomly selected 100 patients with no infarcts. Then, two trained reviewers, using an inhouse developed interactive software tool, generated manual delineation of the WMHs on the T2-weighted images with visual inspection of the corresponding FLAIR images. The intra- and interrater reliability for this expert manual delineation was very high. That is, trained reviewers delineated the WMHs twice on ten MR scans. Intraclass Correlations Coefficients (ICC) were all above 0.9 (table 1).
Table 1. Intra- and interrater reliability for the manual delineation of WMHs.

<table>
<thead>
<tr>
<th></th>
<th>WMH volume</th>
<th>Number of WMH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrarater 1 (n=10)</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Intrarater 2 (n=5)</td>
<td>1</td>
<td>0.99</td>
</tr>
<tr>
<td>Interrater</td>
<td>0.99</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Numbers are Intraclass Correlation Coefficients. WMH; white matter hyperintensities.

Segmentation method We developed a two-level image segmentation technique: an adaptive level that is robust to differences in image intensity ranges and image contrast, and a reasoning level that mimics expert reasoning and that remains unchanged when applied to images acquired on different MR scanners (or different software releases on the same scanner). The reasoning level is implemented using a well-known artificial intelligence technique: Fuzzy Inference system (FIS)38, 39. In a FIS, the number of linguistic variables, the corresponding linguistic values and the set of rules are preferably defined with the help of the expert. In this work, the experts were able to explain how they would classify a voxel as WMH or cerebrospinal fluid (CSF). Our FIS uses 3 linguistic variables to classify a voxel: t2_intensity, flair_intensity and voxel_position; t2_intensity and flair_intensity are described by three linguistic values: BRIGHT, MEDIUM-BRIGHT and DARK. voxel_position takes 2 linguistic values: WM (for white matter) and IC (for Intracranial). The adaptive level maps the exact intensity values to linguistic values such as bright, dark, etc. The reasoning level operates using these linguistic values in the fuzzy if-then-rules to derive a label to every voxel.

The general workflow of our segmentation algorithm is presented in figure 1. In our approach the different images are combined in a cascaded way and are used only when they are considered to be critical for the segmentation. As showed in figure 1, our algorithm consists of 6 main steps:

1. Image registration: First both sequences were automatically cropped to the intersecting area based on physical slice position extracted from the DICOM header. To correct for possible patient movement (mainly pitch and yaw), the reformatted FLAIR image was automatically co-registered to the PD using rigidbody (6 parameters) registration provided by the AIR library40, 41. We refer to the co-registered FLAIR image by CR-FLAIR. Note that we cropped the images before applying the automatic registration to speed up the process. It is well known that when there is missing data in one of the scans (i.e the scans do not cover the same area of the head) the automatic registration would require more time and if stopped after a maximum number of iterations it would have a higher risk of failure.
2. **Template mapping**: Spatial information on brain structures were integrated in the form of a 3D brain tissue probability model (template or atlas). For each voxel the prior probabilities of each main brain tissue class are stated in this atlas. In this work, we used the prior maps of IC, WM, gray matter (GM) and CSF, provided by the Montreal Neurological Institute (MNI). The images are in the Talaraich space. In order to map these priors to our image we used a 12-parameter affined transformation with the standard deviation of ratio images for cost function. The success of our segmentation method is highly dependent on the outcome of this registration. A minimum convergence rate of the registration procedure has been pragmatically defined based on the value of the cost function. If the registration step did not reach this minimum value after a limited number of iterations, the scan is automatically flagged as “possible failure”. This flag is later used for the quality control process. The co-registered IC and WM templates were used as membership functions for the IC and WM position linguistic-values used by our fuzzy inference system.

3. **Intensity to linguistic values mapping (fuzzification)**: The Fuzzy C-mean (FCM) algorithm is applied on every MR image (PD, T2 and the CR_FLAIR). The intensity values of the PD image are clustered into foreground (bright) and background (dark). The T2-weighted-image intensities are automatically clustered in 3 clusters: bright, medium-bright, and dark. The CR_FLAIR image-intensities are also automatically clustered in 3 clusters: bright, medium-bright, and dark. The co-registered templates are used to initialize the cluster centers and the membership matrix, and guide the first iterations of the FCM algorithm.
4. **Brain striping**: Using the outcome of the fuzzy clustering step for the PD image, a binary image involving only foreground voxels is created. Mathematical morphology filters are then applied to delete (or reduce as much as possible) the intracranial/skin "connections", very common in the elderly as the skull does not always appear as dark as in young subjects. A region-growing algorithm automatically seeded in the brain and constrained to remain within the mapped IC- prior map is finally applied. For the IC segmentation, only the PD image is considered. The T2-weighted and FLAIR images do not carry any extra information. They actually present very often low signal intensities (close or similar to the image background) in the basal ganglia area which can cause holes in the segmented IC cavity.

5. **CSF and lesion detection**: Using the fuzzy membership functions defined at the adaptive level, two fuzzy inference rules are applied to segment the CSF and the WMH respectively. 
   
   If \( \text{voxel\_position} \) is IC and \( t2\_intensity \) is BRIGHT and \( \text{FLAIR\_intensity} \) is DARK then \( \text{segmented\_voxel} \) is CSF,
   
   If \( \text{voxel\_position} \) is WM and \( t2\_intensity \) is BRIGHT and \( \text{FLAIR\_intensity} \) is BRIGHT then \( \text{segmented\_voxel} \) is WMH.

6. **User (application) preferences**: In this step we can integrate user preferences and exclude known false positives. A user can define manually any brain region on the template image (Talairach space) that is not relevant to the disease in question. This region is mapped together with the brain-tissue spatial distribution images. Any lesions detected in the mapped area will be automatically excluded. In this work, hyperintensities between the lateral ventricles were automatically ignored. Furthermore, lesions can be ignored based on their size (lesions smaller than 6 voxels were automatically excluded).

**Results**

The algorithm was implemented in C++ and was integrated to our inhouse software package SNIPER (Software for Neuro Image Processing in Experimental Research. Table 2 shows the results of the volumetric agreement in scan- rescan data and FLAIR 6mm versus FLAIR 3mm. All ICC values were above 0.9 suggesting an excellent agreement.

The accuracy of the method has been evaluated on 100 patients. In large-scale study it is often required to correct for IC volume. The agreement between the manual delineation and the automatic measurement of the IC volume was excellent (ICC = 0.98). Figure 2a shows the outcome of the regression analysis for the IC segmentation. The regression line (continuous line) versus the equality line (dotted line) are close to be identical. Figure 2b shows the Bland & Altman plot where the differences of the two measurements (Auto-GS) are plotted against the average of the two measurements. The plot confirms the absence of a
systematic bias and a quite narrow 95% limit of agreement [-81.88 mL, +79.16 mL] with an average difference of -2.72 mL.

Reliability of the WMH volume segmentation (i.e. agreement between the ‘gold standard’ manual delineation and the fully automatic measurement) was investigated in 3 different categories of WMH severity: large (total WMH volume >18 mL), medium (total WMH volume > 4 mL - 18 mL) and small (total WMH volume < 4 mL). In our sample, 40% of the subjects presented small, 35% presented medium and only 25% presented a large WMH lesion load. Typical examples of automatic segmentation for patients with small, medium and large total WMH volumes are shown (see appendix, figure 1). The agreement between the manual delineation (i.e. gold standard) and the fully automatic volume measurement was excellent. The ICC were all above 0.9 (table 3).

The regression analysis (figure 3a) shows a strong correlation between the automatic and the manual delineations of WMHs. The Bland and Altman plot (figure 3b) shows no systematic bias but a slight underestimation of the average WMH volume (average difference is -1.08 mL). The error tends to increase as the magnitude of lesion load measurements increases; the mean difference is approximately proportional to the magnitude of the lesion load, and this results in a relatively wide limit of agreement interval when looking at small lesion loads (95% limit of agreement [-5.8 mL, +3.6 mL]).

\[
\text{SI}_{(GS, Auto)} = \frac{2 \times (\text{GS} \cap \text{Auto})}{\text{GS} + \text{Auto}},
\]

In order to evaluate the spatial agreement of WMHs between the ‘gold standard manual delineation’ and the automatic segmentation we used the similarity index (SI) defined as follow:

Where GS is used for gold standard and Auto for the fully automatic segmentation. \(\cap\) refers to the volume of the overlap between Auto and GS. The SI takes values between 0 and 1; 0 for no overlap at all (total disagreement) and 1 for a perfect alignment of the two (total agreement). An SI value of 0.7 or higher indicates a very good to excellent agreement\(^4\). We calculated the SI for every single patient and generated the mean and standard deviation for each sub-group as well as for the whole data set (table 4). The values reported in table 4 suggest an excellent spatial agreement. The agreement was lower for the small lesion load category since small error (only few misclassified voxels) can still be relatively large if the total lesion area is small. Figure 4 shows a plot of the calculated SI values per patient. The largest spatial disagreement for WMHs occurred in patients with a total WMH volume < 2 mL. Only few misclassified voxels can have a large impact on the SI when the total area is small.
Table 2. Scan – rescan reliability and volumetric agreement between FLAIR 6mm and FLAIR 6mm and FLAIR 3mm for the fully automatic segmentation of IC and WMH volumes.

<table>
<thead>
<tr>
<th></th>
<th>IC volume</th>
<th>WMH volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scan – Rescan (n=9)</td>
<td>0.95</td>
<td>0.91</td>
</tr>
<tr>
<td>FLAIR 6mm - FLAIR 3mm (n=44)</td>
<td>1</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Numbers are Intraclass Correlation Coefficients. FLAIR; fluid attenuated inversion recovery. WMH; white matter hyperintensities. IC; Intracranial.

Table 3. Reliability of the WMH volume segmentation; agreement between manual delineation ('gold standard') and the fully automatic measurements of WMHs for 3 different categories according to WMH severity.

<table>
<thead>
<tr>
<th>WMH volume</th>
<th>low (n=40)</th>
<th>medium (n=35)</th>
<th>high (n=25)</th>
<th>all (n=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>0.91</td>
<td>0.93</td>
<td>0.94</td>
<td>0.98</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.91</td>
<td>0.95</td>
<td>0.97</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Numbers are intraclass correlation coefficients (ICC) and Spearman correlations. All p-values are <0.0001. WMH; white matter hyperintensities.

Table 4. Similarity Index measurements (i.e. spatial agreement) between manual delineation ('gold standard') and fully automatic measurements of IC volume and WMH.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Similarity index (SI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC volume</td>
<td>100</td>
<td>0.97 (0.01)</td>
</tr>
<tr>
<td>Small WMH</td>
<td>40</td>
<td>0.70 (0.09)</td>
</tr>
<tr>
<td>Medium WMH</td>
<td>35</td>
<td>0.75 (0.07)</td>
</tr>
<tr>
<td>Large WMH</td>
<td>25</td>
<td>0.82 (0.05)</td>
</tr>
<tr>
<td>Total WMH</td>
<td>100</td>
<td>0.75 (0.09)</td>
</tr>
</tbody>
</table>

Data are presented as mean (SD). WMH; white matter hyperintensities. IC; intracranial.
Figure 2a, b. Automatic (AUTO) versus manual delineation (i.e. gold standard; GS) segmentation of intracranial volume (ICV). a) Linear regression line (continuous) and the equality line (dotted) are presented in the figure. b) Bland & Altman plot; the difference of the two measurements (AUTO - GS) are plotted against the average of the two measurements.
Figure 3a, b. Automatic (AUTO) versus manual delineation (i.e. gold standard; GS) segmentation of white matter hyperintensity volume (WMH). a) Linear regression line (continuous) and the equality line (dotted) are presented in the figure. b) Bland & Altman plot; the difference of the two measurements (AUTO - GS) are plotted against the average of the two measurements.
Figure 4. Similarity index (SI) for all subjects (n=100); spatial agreement between manual delineation of WMH versus fully automatic measurement of WMH
Discussion

Why using PD, T2-weighted and FLAIR images for automatic assessment of WMHs?

Feature selection is a critical step while developing an automatic segmentation algorithm. Using more features will not always improve the system. Sometimes the performance of the learning algorithm can decrease because of adding unnecessary extra features. The optimal set of features should be identified. The main advantages of integrating information from multiple images are to reduce the uncertainty and increase the accuracy of the segmentation. These advantages are related to the notions of redundancy and complementarity. PD, T2 and FLAIR images provide redundant information concerning WMH: they are hyperintense in all three scans. However, the extent of the lesions may not look the same on the different images. It has been suggested in the literature that FLAIR is less sensitive in the posterior fossa, may have an “overestimation” of lesion load, and has a higher inter-vendor variability. Furthermore, FLAIR may present hyperintense artifacts that might lead to false positives. The combination of FLAIR with a redundant source (T2) will increase the certainty of the lesion delineation and reduce false positives. Lacunar infarcts and Virchow Robin spaces have increased signal intensity on T2 and PD images but low signal intensity on FLAIR images. This makes FLAIR a complementary source of information for a better characterization of WMH. One might argue the use of the PD images. Indeed, in our approach it used only for the IC segmentation. The IC cavity can also be extracted in the T2-weighted image. However, since T2-weighted images very often present with hypointense basal ganglia in the elderly that might lead the intensity mapping part of the segmentation to create holes in the segmented IC. Extra rules and morphological operations will be required to close these holes without adding the skull to the segmented IC. In the clinical routine, the T2-weighted image is very often acquired in a dual sequence, the use of the PD preferred as the IC segmentation algorithm is more straightforward and faster (less rules and less operations). Our segmentation algorithm is adaptive to differences in signal intensity ranges, template based, fast and has been thoroughly evaluated.

Adaptive segmentation

One can categorize most brain segmentation algorithms in two main categories: supervised pixel classification and unsupervised clustering. In supervised pixel classification, a set of images in which the desired segmentation is known (expert manual delineation) is used as training material to build and tune the segmentation algorithm. To be truly effective, supervised training algorithms require a representative sample of examples that cover most of the cases (ideally all) in order to perform well in practice. They would also require image intensity standardization to be applied on new data sets otherwise the segmentation might fail dramatically. In unsupervised segmentation, generally clustering techniques are used. A clustering algorithm clusters data points into a given number of clusters according to similarity measures. Points that look most alike will be...
grouped in the same cluster/class. The only information required by such algo-
rithms is the number of clusters to extract. This type of methods is less depend-
ent (if at all) on image intensity standardization and is more adaptive to different
protocols or images of different scanners. Therefore, we chose to implement an
unsupervised classifier.

Our algorithm is organized in two levels: an adaptive level and a reasoning level.
The reasoning takes place at a linguistic level that remains unchanged across
images acquired on different scanners or using different sequence parameters.
The adaptive level, based on fuzzy clustering, is used to correct for differences in
image intensity ranges. It maps the absolute values to the common linguistic
space, in which the reasoning level operates.

Template based segmentation
In literature, the use of spatial information varies from none\textsuperscript{31, 32} to a brain atlas
where the spatial distribution of white WM, GM and CSF is given in the form of
probability maps (probabilistic atlases)\textsuperscript{34, 36-39}. It is commonly accepted that the
integration of the spatial information is crucial for the success of a fully automat-
ic segmentation algorithm. In the study of Anbeek et al\textsuperscript{27}, cartesian and polar
coordinates of the voxels have been considered in a K-nearest neighbour (KNN)
approach. Although it proved to be better than using only signal intensities, it is
still limiting as white matter lesions can appear anywhere in the white matter.
Since it is very unlikely that the chosen samples of the training set will cover the
whole white matter, the method is prone to false negatives. The best solution,
so far, is to use brain templates, where prior distribution of the GM, WM and CSF
are known.

The success of template based segmentation algorithms depends on the outcome
of the template-image registration. Most of the reported template-based brain
image segmentation methods use the MNI templates, which have been generat-
ed from MR images of young healthy volunteers. Brain atrophy and abnormalities
in white matter are prevalent changes in elderly people. Therefore, MR images of
elderly subjects are not congruent with the MNI brain templates. A “naive” use of
the MNI CSF, GM and WM prior distribution map to segment the brain in the eld-
ery, might lead to an underestimation of CSF, and an overestimation of grey mat-
ter. Therefore, in our method we use these priors only to initialize and guide the
few first iterations of the FCM algorithm. The white matter template is further
used to define a coarse area in which WMH are likely to appear. The exact delin-
eation of WM is not necessary, as the final decision on lesion segmentation
remains dependent on the signal intensities of the FLAIR and T2-weighted
images.

Fast segmentation
If including a “weak” feature in a multi-channel classifier does not decrease its
performance, it definitely will increase its computational time. Our segmentation
approach combines PD, T2 and FLAIR images as well as the prior spatial informa-
tion (templates) of the IC, WM, GM and CSF. Unlike all previously published multi-spectral segmentations, the different voxel features are combined in a cascaded way and are used only when they are considered to be critical for the classifier. This way, our system operates in low dimensions and therefore, requires less computational time.

Our segmentation algorithm does not require an accurate registration of the template to the patient image. A coarse mapping is enough to guide the segmentation. By stopping the multi-resolution registration at half image resolution, we could obtain an acceptable registration in only few seconds.

Validation
To be used in clinical settings or clinical trials, a segmentation algorithm must be thoroughly validated. One of the main problems in validating fully automatic approaches is the lack of a gold standard (the absolute truth is unknown). Some researchers use phantoms or synthetic images where the structures to be detected are simulated and fully controlled (size, position, number, growth, appearance etc.)\(^6\). However, most phantoms (or synthetic images) are often simplified models of the process or the structure to be detected and do not exhibit the wide range of anatomy and acquisition variability found in real data. Therefore, the only gold standard remains the expert delineation.

Most of the proposed fully automatic techniques have been validated on small numbers of patients (10 to 30) for whom expert manual delineation is obtained. Unless the evaluation data set is representative and/or the algorithm is proven to be robust to variations, the good performances reported in the literature could be the result of an over tuning (tweaking) of the parameters to perform best for a particular set of MR scans.

In this study the reproducibility of the automatic segmentation has been evaluated in 9 patients who have been scanned and rescanned with repositioning. We also showed that the method is robust to differences in FLAIR image slice thickness in 44 patients. The accuracy of the segmentation has been evaluated on 100 patients selected randomly from a data collected in a large clinical trial for whom expert delineation of the WMH and the IC, on all slices, was available. Three categories of WMHs severity (Small, Medium and Large) have been evaluated. The quality of our segmentation method is illustrated by high volumetric and spatial agreements with the manual delineation.

Limitations
Our approach has some known limitations. Due to the required FLAIR-DUAL scan registration, the FLAIR image is resliced (reformatted). This requires image interpolation, which has the tendency of blurring out and darkening small hyperintensities of only few voxels. Our method might miss these small lesions. Therefore, a more sophisticated interpolation (edge preserving) would be required.
As for any template based segmentation algorithm, template registration is one of the most crucial steps. Our registration step is fine-tuned to perform reasonably well and very fast (only few seconds). We assume the axial slice orientation does not deviate significantly from the AC-PC (anterior commissure, posterior commissure) line and that the yaw (z direction) correction would not be too large (-10 to 10 degrees). Any significant deviation form these assumptions may lead to a misregistration and thus lead to a poor segmentation. In most failure cases, a different initialisation of the registration algorithm leads to a successful image registration and segmentation.

We also assume that the scans cover the entire brain. If it is not the case (missing data) the registration may fail when it tries to map a complete brain (template) on an incomplete data set without spatial information on the missing parts. If the different scans have been consistently scanned with the same missing part, cropping the template to the consistently imaged area leads to successful registration and segmentation. If not, more sophisticated and time-consuming registration would be required.

Our segmentation does not integrate an in-plane inhomogeneity correction, as we did not experience the need for it in WMH and CSF segmentation. If white matter and grey matter segmentation are required than this step might be necessary. For interslice inhomogeneity, the fuzzy clustering in the adaptive level is applied on a slice by slice basis and guided by the prior spatial distributions (templates). The effect of severe inplane inhomogeneity has not been evaluated but will certainly affect the performances of our system.

**Conclusion**

In this paper we presented a fully automatic WMH segmentation algorithm. Our system has been carefully designed to be fast, accurate and reliable. The system has been originally designed and applied to automatically quantify white matter lesion load in elderly subjects who participated in the large scale study PROSPER. It is also applicable to the quantification of white matter lesion in other diseases such as MS, Migraine, NPSLE (neuropsychiatry systemic lupus erythematosus) or CADASIL (*Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts en Leukoencephalopathy*). Very good to excellent performances have been demonstrated in 100 patients, with variable lesion load. Our system has also some known limitations; therefore quality control by visual inspection remains mandatory.
Reference List


