NeuroImage 59 (2012) 3774–3783

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis

Paul Schmidt [a,b,1], Christian Gaser [c,d,1], Milan Arsic [a], Dorothea Buck [a], Annette Förschler [e], Achim Berthele [a], Muna Hoshi [a], Rüdiger Ilg [a], Volker J. Schmid [b], Claus Zimmer [e], Bernhard Hemmer [a], Mark Mühlau [a,*]

[a] Department of Neurology, Technische Universität München, Munich, Germany
[b] Department of Statistics, Ludwig-Maximilians-University, Munich, Germany
[c] Department of Psychiatry, Friedrich-Schiller-University, Jena, Germany
[d] Department of Neurology, Friedrich-Schiller-University, Jena, Germany
[e] Department of Neuroradiology, Technische Universität München, Munich, Germany

ARTICLE INFO

Article history:
Received 21 April 2011
Revised 28 October 2011
Accepted 9 November 2011
Available online 18 November 2011

Keywords:
Lesion segmentation
FLAIR
Multiple Sclerosis
Voxel-based morphometry

ABSTRACT

In Multiple Sclerosis (MS), detection of T2-hyperintense white matter (WM) lesions on magnetic resonance imaging (MRI) has become a crucial criterion for diagnosis and predicting prognosis in early disease. Automated lesion detection is not only desirable with regard to time and cost effectiveness but also constitutes a prerequisite to minimize user bias. Here, we developed and evaluated an algorithm for automated lesion detection requiring a three-dimensional (3D) gradient echo (GRE) T1-weighted and a FLAIR image at 3 Tesla (T). Our tool determines the three tissue classes of gray matter (GM) and WM as well as cerebrospinal fluid (CSF) from the T1-weighted image, and, then, the FLAIR intensity distribution of each tissue class in order to detect outliers, which are interpreted as lesion beliefs. Next, a conservative lesion belief is expanded toward a liberal lesion belief. To this end, neighboring voxels are analyzed and assigned to lesions under certain conditions. This is done iteratively until no further voxels are assigned to lesions. Herein, the likelihood of belonging to WM or GM is weighed against the likelihood of belonging to lesions. We evaluated our algorithm in 53 MS patients with different lesion volumes, in 10 patients with posterior fossa lesions, and 18 control subjects that were all scanned at the same 3T scanner (Achieva, Philips, Netherlands). We found good agreement with lesions determined by manual tracing (R2 values of over 0.93 independent of FLAIR slice thickness up to 6 mm). These results require validation with data from other protocols based on a conventional FLAIR sequence and a 3D GRE T1-weighted sequence. Yet, we believe that our tool allows fast and reliable segmentation of FLAIR-hyperintense lesions, which might simplify the quantification of lesions in basic research and even clinical trials.

Introduction

Multiple Sclerosis (MS) is an inflammatory demyelinating disease of the central nervous system that affects over 2.5 million people worldwide and is one of the leading causes of serious neurologic disability in young adults (Confavreux and Vukusic, 2008; Weiner, 2009). The disease is characterized by unpredictable episodes of clinical relapses and remissions followed by continuous progression of disability over time (secondary progressive MS) in most instances (Compston and Coles, 2008; Noseworthy et al., 2000). Demyelinating lesions (plaques) within cerebral white matter (WM) are the hallmark of MS and its detection by T2-weighted magnetic resonance imaging (MRI) has become a crucial diagnostic criterion (Polman et al., 2011). Moreover, T2-hyperintense lesion volume has been demonstrated to correlate with severity of symptoms, progression of disability and gray-matter (GM) atrophy (Bendfeldt et al., 2010; Chard et al., 2002; Fisher et al., 2008; Fisniku et al., 2008). Accordingly, T2-hyperintense lesion volume has been of interest in basic research and has been determined in most pivotal trials on disease-modifying drugs since the late nineties (Ebers, 1998; Jacobs et al., 2000; Kappos, 1998).

Automatization of T2-hyperintense lesion volumetry is desirable with regard to time and cost effectiveness but also constitutes a prerequisite to minimize user bias. Up to now, a number of algorithms have been proposed (Table 1) but no gold standard has been established. Therefore, in the vast majority of clinical trials, lesions were traced manually slice by slice — sometimes with the help of semi-automated tools for contour detection.

Here, we aimed at the development and validation of an automated algorithm for segmentation of T2-hyperintense WM lesions in MS based on a T2-weighted fluid-attenuated (FLAIR) and a three-

* Corresponding author at: Department of Neurology, Klinikum rechts der Isar, Technische Universität München, Ismaningerstr. 22, D-81675 Munich, Germany. Fax: +49 89 4140 4867.
E-mail address: muehlau@lrz.tum.de (M. Mühlau).
[1] These authors contributed equally to this work.

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.11.032

Table 1
Studies on automated lesion segmentation in MS.

References in chronological order	Method	Sequences	Validation technique	Number of patients/ controls	
(Van Leemput et al., 2001)	EM alg. with MRF and OD	T1, T2, PD	Manual segmentation	23	/
(Ferrari et al., 2003)	SVM	T1, T2, FLAIR	Manual segmentation	18	/
(Anbeek et al., 2004)	KNN	T1, T2, PD, FLAIR	Manual segmentation	19	/
(Ait-Ali et al., 2005)	Robust EM, trimmed likelihood estimator and OD	T1, T2, PD	Simulated data (BrainWeb)	/	/
(Li et al., 2005)	EM alg. with MRF	T1, T2, FLAIR	Visual inspection	6	2
(Wu et al., 2006)	Supervised KNN classifier	T1, T2, PD	Manual segmentation	6	/
(Herskovits et al., 2008)	Classification statistics based on training data	T1, T2, FLAIR	Manual segmentation	1	
(Khayati et al., 2008)	Adaptive mixture model	FLAIR	Manual segmentation	20	/
(Wels et al., 2008)	Probabilistic boosting tree	T1, T2, FLAIR	Manual segmentation	6	/
(Freifeld et al., 2009)	Constrained Gaussian mixture model and OD	T1, T2, PD or T1, T2, PD, FLAIR	Mainly simulated data (BrainWeb) manual segmentation	?	/
(Garcia-Lorenzo et al., 2009)	Automatic multimodal graph cuts	T1, T2, PD	Simulated (BrainWeb) and manual segmentation	10	/
(Akselrod-Ballin et al., 2009)	Decision forest classifier based on training data	T1, T2, PD FLAIR	Manual segmentation	25 16	/
(Geremia et al., 2010)	Spatial decision forests	T1, T2, FLAIR	Manual segmentation	20	/

Note. In two online libraries (http://apps.isiknowledge.com, http://www.ncbi.nlm.nih.gov/), we searched for the combination of the following terms: Multiple Sclerosis, MRI, automated (or automatic), lesion. The retrieved articles including their references were then studied. alg., algorithm; EM, expectation maximization; KNN, K-nearest neighbor; MRF, Markov random field; OD, outlier detection; SVM, support vector machine.

dimensional (3D) gradient echo (GRE) T1-weighted sequence derived from a 3 Tesla (T) scanner. Such protocols have increasingly been used in clinical practice, since evidence suggests superiority of 3 T scanners over those with lower field strength (Wattjes and Barkhof, 2009; Wattjes et al., 2006a) and superiority of FLAIR sequences over conventional T2-weighted sequences (Bakshi et al., 2001; Filippi et al., 1996; Stevenson et al., 1997; Wattjes et al., 2006b; Woo et al., 2006).

Materials and methods

Subjects

The study was performed in accord with the Helsinki Declaration of 1975 and approved by the local ethics committee. Beforehand, written informed consent was obtained from the control subjects to participate and from the patients to subject their MRI scans, acquired in routine clinical practice, to scientific studies.

We obtained MRI scans from 18 control subjects (age in years: 23–58; median, 30; 32.6 ± 10.6), and 52 MS patients (age in years: 23–64; median, 41; 32.6 ± 11.4; EDSS: 0–4; median, 2). Five of the 52 MS patients were also included in another group of 10 MS patients with posterior fossa lesions according to their medical records (age in years: 24–64; median, 36; 37.9 ± 13.1; EDSS: 0–4; median, 2).

Magnetic resonance imaging

All brain images were acquired on the same 3 T scanner (Achieva, Philips, Netherlands). We used a 3D GRE T1-weighted sequence (orientation, 170 contiguous sagittal 1 mm slices; field of view, 240×240 mm; voxel size, $1.0 \times 1.0 \times 1.0$ mm; repetition time (TR), 9 ms; echo time (TE), 4 ms), and a 3D FLAIR sequence (orientation, 144 contiguous axial 1.5 mm slices; field of view, 230×185 mm; voxel size, $1.0 \times 1.0 \times 1.5$ mm; TR,10^4 ms; TE, 140 ms; TI, 2750 ms). For additional analyses (see Evaluation section), FLAIR images were also resliced to 3 and 6 mm in axial and sagittal orientations.

Lesion segmentation algorithm

For better understanding, we give a conceptual overview of the three major steps of our algorithm in this paragraph (Fig. 1). First (for details see Preprocessing with already available software section), preprocessing is performed with the standard software of SPM8 and its extension VBM8. To surpass smoothing of the individual images by warping, the algorithm operates in the space of the original T1-weighted image, i.e. in native space. Each voxel of the individual native T1-weigted image is assigned to one of the three tissue classes of GM, WM, or CSF. The FLAIR image is bias-corrected for MR field inhomogeneity and coregistered to the T1-weighted image. Since a-priori probability of each voxel of belonging to WM, the tissue class containing the lesions, is utilized later, the SPM tissue probability map of WM (TPM_{WM}) is warped into native space. Second (for details see Lesion belief maps and initialization section), FLAIR intensity distributions are calculated for each of the three tissue classes to detect FLAIR-hyperintense outliers which are further weighed according to their spatial probability of being WM. This results in lesion belief maps (B_{WM}, B_{CSF}, B_{GM}). Now, the three lesion belief maps are summed up (B). The binary version of the GM lesion belief map is used as a seed, the initial lesion map (L_{init}). Third (for details see Lesion growing section), the lesion growth model expands the L_{init}, a conservative assumption for lesions, toward the lesion belief map (B), a liberal assumption for lesions. To this end, neighboring voxels are analyzed and assigned to lesions under certain conditions. This is done iteratively until no further voxels are assigned to lesions. Herein, the likelihood of belonging to WM or GM is weighed against the likelihood of belonging to lesions. Two model parameters had to be set (for details see Determination of the initial threshold and of the final threshold section). The algorithm was programmed in MATLAB (www.mathworks.de/products/matlab/). We provide a pseudo-code description of our algorithm in Appendix A1 and performance parameters in Appendix A2. In the following subsections, we describe our algorithm in detail. Alternatively, the reader may continue with the Evaluation section.

Preprocessing with already available software

We use SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) and its VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm). At option, VBM8 can provide images of the same modality which are bias-corrected for MR field inhomogeneity either in native space or normalized to MNI space; further, partial volume estimate (PVE) images (Tohka et al., 2004) can be generated which are either in native space or normalized to MNI space. In this protocol, images are corrected for bias-field inhomogeneity, registered using linear (12-parameter affine) and nonlinear transformations, and tissue-classified into GM, WM, and CSF within the same generative model (Ashburner and Friston, 2005). This segmentation procedure is further refined by accounting for partial volume effects (Tohka et al., 2004), by applying adaptive maximum a posteriori estimations (Rajapakse et al., 1997), and by applying hidden Markov random field (MRF) model (Rajapakse et al., 1997) as proposed recently (Lüders et al., 2009).

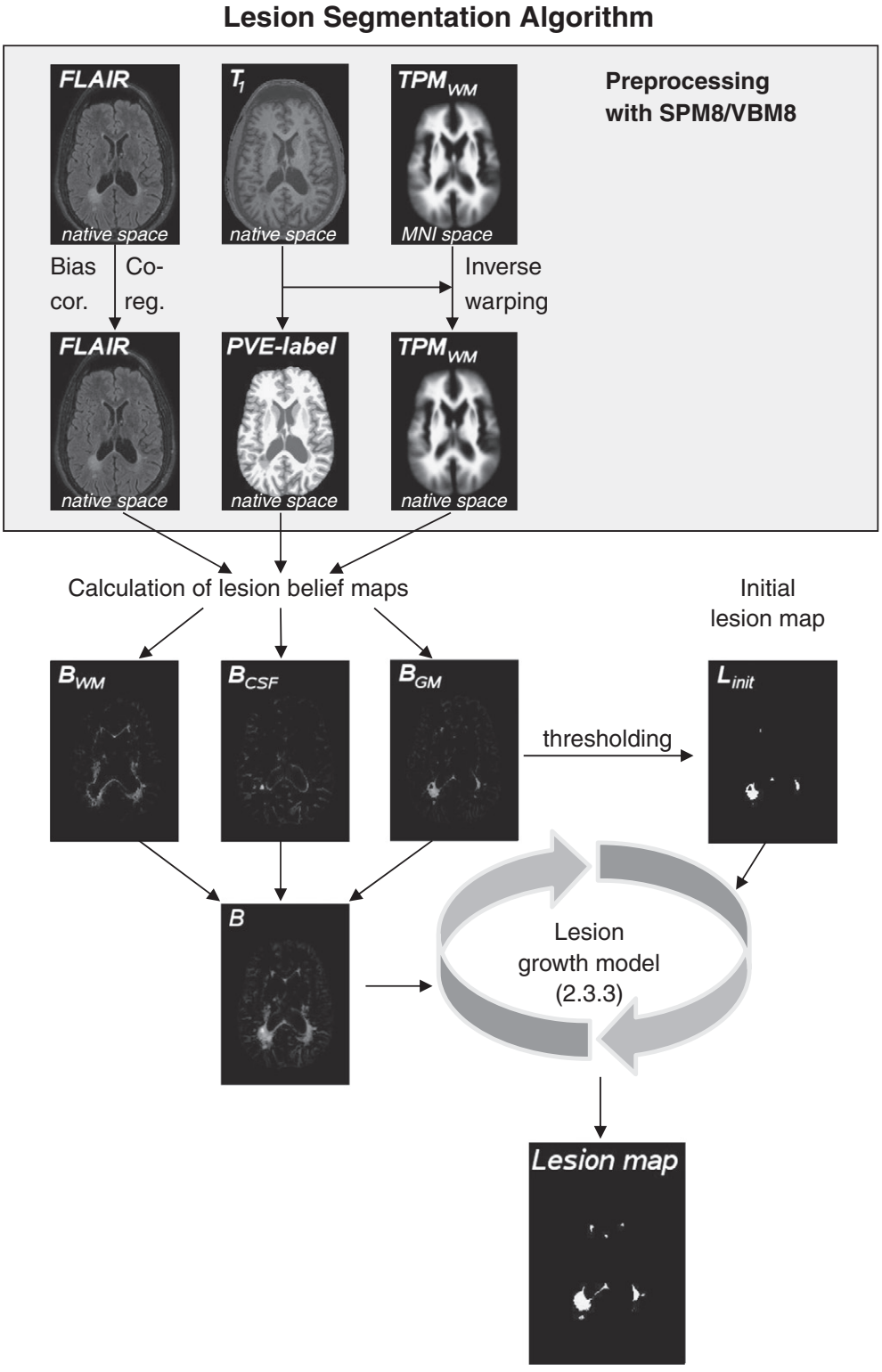

Fig. 1. The flow diagram of the lesion segmentation algorithm is shown. Preprocessing with the standard software of SPM8 and VBM8 is illustrated in the gray box. At first, the individual native T1 image is used to generate a partial volume estimate (PVE) label. To this end, some normalization is necessary. To surpass smoothing of the individual images by warping, the algorithm operates in native space exclusively. Thus, preprocessing includes the coregistration of FLAIR images to T1 images, PVE label estimation but output in native space, as well as inverse warping of the white matter (WM) tissue probability map (TPM_{WM}) to native space by the use of the inverse deformation matrix derived from PVE label estimation. Next, FLAIR intensity distribution is calculated for each of the three tissue classes to detect FLAIR-hyperintense outliers which are further weighed according to their spatial probability of being WM resulting in belief maps (B_{WM}, B_{CSF}, B_{GM}). Now, the three lesion belief maps are summed up (B). The binary version (threshold $\kappa = 0.3$) of the GM lesion belief map is used as initial lesion map (L_{init}). Finally, the lesion growth model expands the L_{init}, a conservative assumption for lesions, toward the lesion belief map (B), a liberal assumption for lesions (see text for details).

For tissue classification, the T1-weighted image is used to estimate a PVE image in which a number in the range between 1 and 3 is assigned to each voxel. In accordance with image intensity, the integers (1, 2, 3) stand for CSF, GM and WM, respectively. Values between those integers indicate the partial volume effect. Of note, voxel values are estimated primarily on the basis of intensity values. Yet a-

priori (i.e. spatial) information on tissue-classes is used for scalp editing so that some normalization is necessary although our algorithm only uses the PVE image in native space. For normalization, we choose the option of low-dimensional warping since visual inspection of the PVE images, normalized this way, did not yield a single case in which MS lesions disturbed normalization. Next, the FLAIR image is bias-corrected by the same routine of VBM8 and coregistered to the native T1-weighted image by the standard routine of SPM8. Since a-priori probability of each voxel belonging to WM, the tissue class containing the lesions, is utilized later, the SPM tissue probability map of WM (TPM_{WM}) is warped into native space by the use of the inverse deformation matrix derived from PVE label estimation. We will refer to this image as native TPM_{WM}.

Lesion belief maps and initialization

By the use of the PVE image, we estimate the distributions of FLAIR intensity for each tissue class. Of note, we expect lesion voxels to behave as hyperintense outliers from these distributions. Let x_i denote the estimated PVE label of voxel i, we then assign a discrete label z_i to each voxel as follows:

$$z_i = \begin{cases} \mathrm{CSF}, & \text{if } x_i < 1.5 \\ \mathrm{GM}, & \text{if } 1.5 \le x_i < 2.5 \\ \mathrm{WM} & \text{if } x_i \ge 2.5. \end{cases} \quad (1)$$

Next, FLAIR images are scaled via voxel-wise division by the mean FLAIR intensity of the GM class ($z_i = \mathrm{GM}$). Let $y = (y_1, \ldots, y_n)$ denote the scaled FLAIR intensities. Further, we denote the means of y along the three tissue classes by $\bar{y}_k$, $k \in \{\mathrm{CSF.GM, WM}\}$. The lesion belief value for voxel i represents the amount of hyperintensity in terms of the distance from the class mean $\bar{y}_k$ weighted by the estimated PVE label and by the a-priori (i.e. spatial) probability for WM:

$$b_{k,i} = (y_i - \bar{y}_k)^+ \cdot x_i \cdot Pr(z_i = \mathrm{WM})$$

Here $(y_i - \bar{y}_k)^+$ is $y_i - \bar{y}_k$ if $y_i > \bar{y}_k$ and zero otherwise and $Pr(z_i = \mathrm{WM})$ is the probability that voxel i belongs to WM according to the native TPM_{WM}. In this way, we obtain lesion belief maps $B_k = \{b_{k1}, \ldots, b_{kn}\}$ for all tissue classes k. Voxel values of B_{GM} increase with (1) a high a-priori (i.e. spatial) probability for WM, (2) medium intensity at T1, and (3) hyperintensity at FLAIR. Hence, increasing B_{GM} values support the assumption that the respective voxel belongs to a WM lesion. Similar interpretations can be made for B_{CSF}. Voxel values of B_{CSF} increase with a high a-priori probability for WM, hypointensity at T1, and hyperintensity at FLAIR. Hence, increasing B_{CSF} values support the assumption that the respective voxel belongs to a WM lesion ("black holes"). Likewise, voxel values of B_{WM} increase with a high a-priori probability for WM, hyperintensity at T1, and hyperintensity at FLAIR. Hence, increasing B_{WM} values support the assumption that the respective voxel belongs to a WM lesion ("dirty WM"). Besides the lesion belief maps for the three tissue classes, we compute a total lesion belief map $B = \{b_1, \ldots, b_n\}$ by summing up the three maps: $b_i = b_{\mathrm{CSF},i} + b_{\mathrm{GM},i} + b_{\mathrm{WM},i}$.The lesion belief map B can be interpreted as a liberal assumption of lesion voxels.

The proposed lesion growth algorithm requires initialization, i.e. seed regions from where the lesions are expanded. Since extensive preliminary experiments and analyses did not yield lesions without any part assigned to GM according to the PVE label, we choose B_{GM} for initialization of lesions after application of the threshold κ so that we obtain the initial lesion map $L_{init} = \{l_{init,1}, \ldots, l_{init,n}\}$ by

$$l_{init,i} = 1 \Leftrightarrow b_{\mathrm{GM},i} > \kappa.$$

This map can be interpreted as a conservative assumption of lesion voxels. Since the choice of the threshold κ is potentially critical, we investigate the impact κ on the final segmentation in the Evaluation section.

Lesion growing

Now, the lesion growth model expands the L_{init}, a conservative assumption for lesions, toward the lesion belief map (B), a liberal assumption for lesions. Each voxel in the neighborhood of the initialized lesions is labeled *lesion (Les)* or *other*. This latter class consists of the three main tissue classes CSF, GM and WM. Thus, the discrete label z_i for voxel i can be either *Les* or *other*. We approximate the distribution of *Les* by a gamma distribution with shape and scale parameters α and β, respectively, and the distribution of *other* by a mixture of three Gaussians:

$$p_{other}(y_i|\theta) = \sum_k \pi_k \cdot \phi\left(y_i|\mu_k, \sigma_k^2\right).$$

In this notation, ϕ stands for the probability density function of the Gaussian (normal) distribution with mean μ_k and variance σ_k^2 and π_k is the proportion of the kth class. The vector θ collects the parameters μ_k and σ_k^2 for all tissue classes. Since the classification of voxel i in respect to the three tissue classes CSF, GM, and WM is known from Eq. (1), the unknown parameters in θ can be estimated using the maximum likelihood estimators:

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i:z_i=k} y_i \quad \hat{\sigma}_k^2 = \frac{1}{n_k - 1} \sum_{i:z_i=k} (y_i - \hat{\mu}_k)^2.$$

The mixture proportions are estimated by $\hat{\pi}_k = n_k/n$ where n is the total amount of brain voxels that belong to either CSF or GM or WM. The parameters of the gamma distribution are also estimated by maximum likelihood as it is implemented in the MATLAB function gamfit. Next, we describe the way in which the initialized lesions are allowed to grow toward the final lesion map.

We use an iterated growth algorithm. In each iteration, voxels that share a common border with a lesion voxel are considered to be a lesion. Instead of accepting or rejecting the proposed value ($z_i = Les$) the algorithm assigns the following value to the ith voxel:

$$\pi_i^{Les} = Pr(z_i = Les) = \min\left(1, \frac{p_{Les}\left(y_i|\hat{\alpha}^{(t-1)}, \hat{\beta}^{(t-1)}\right) \cdot b_i}{p_{other}\left(y_i|\hat{\theta}^{(t-1)}\right)}\right) \quad (2)$$

Here, i stands for indices of voxels with at least one neighbor j with $\pi_j^{Les} > 0$. The value b_i in Eq. (2) ensures that a lesion could only grow along those voxels which have a positive lesion belief value. With other words, lesions are only allowed to grow within our liberal lesion assumption. After each iteration, the parameters in θ and α and β are re-estimated. For the estimation of α and β only those voxels with $\pi_i^{Les} \ge 0.5$ are considered. Likewise, the updating of $\hat{\theta}$ is based on voxels with $\pi_i^{Les} < 0.5$. For initialized lesion voxels, we set $\pi_i^{Les} = 1$. The algorithm stops when no more voxels with $\pi_i^{Les} > 0.01$ are observed. This results in a lesion probability map. Next, we expand our algorithm by incorporation of the information of neighboring voxels.

Assuming that a voxel, which is completely surrounded by lesion voxels, is more likely to be part of a lesion than of the other classes, we utilize a MRF. We will briefly summarize the important aspects of MRFs that have been extensively reviewed elsewhere (Winkler, 2003). In a random field, the labels $z = (z_1, \ldots, z_n)$ are viewed as a realization of a collection of random variables $Z = (Z_1, \ldots, Z_n)$ in which each random variable has the domain L, or $L = \{Les, other\}$ in this case. Z is a MRF if the following two conditions are met

1. $p(\mathbf{z}) > 0$ for all possible realizations z,
2. $p(z_i|\mathbf{z}, z_i) = p(z_i|\mathbf{z}_{N_i})$

While the first assumption is required for technical reasons, the second states that the probability of z_i, given all other labels, depends only on voxels that are in the neighborhood N_i of z_i. Here, z_{N_i} is the set of voxels that are in the neighborhood N_i, $z_{N_i} = \{z_{i'} | i' \in N_i\}$. According to the Hammersley–Clifford-Theorem (Hammersley and Clifford, 1971), a MRF is equivalent to a Gibbs Random Field. Thus, its distribution follows a Gibbs distribution and the conditional distribution of z_i for a given neighborhood can be written as

$$p\left(z_i | z_{N_i}\right) = \frac{\exp\left\{-U\left(z_i | z_{N_i}\right)\right\}}{\sum_{k \in L} \exp\left\{-U\left(k | z_{N_i}\right)\right\}}. \quad (3)$$

Here, $U(k|z_{N_i})$ is the so-called energy function that is primarily responsible for the resulting segmentation. One simple choice for this energy function leads to the Ising model that has been extensively studied in the context of brain image segmentation by others (Zhang et al., 2001; Woolrich et al., 2005):

$$U\left(k | z_{N_i}\right) = \varphi \sum_{j \in N_i} I\left(z_j \neq z_i\right) \quad (4)$$

in which $I(a)$ is the indicator function that is 1 if statement a is true and 0 otherwise. This energy function favors the segmentation of voxel i as a lesion if more voxels in N_i are already marked as a lesion. Instead of discrete labels the proposed algorithm produces values $\pi_1^{Les}, \ldots, \pi_n^{Les}$ in the interval [0,1]. Thus, integration of energy function (4) in our algorithm would require another threshold to produce a binary lesion map. To address this problem, we modify energy function (4) by using the probabilities of the neighboring voxels instead of the discrete labels:

$$U\left(\pi_i^{Les} | \pi_{N_i}^{Les}\right) = \varphi \sum_{j \in N_i} \left(1 - \pi_j^{Les}\right). \quad (5)$$

We follow the choice of others (Khayati et al., 2008; Li et al., 2005; Zhang et al., 2001) and choose $\varphi = 1$. To include the MRF in the lesion growth model, we use Eq. (3) as additional information and expand Eq. (2) to

$$\pi_i^{Les} = \min\left(1, \frac{p_{Les}\left(y_i | \hat{\alpha}^{(t-1)}, \hat{\beta}^{(t-1)}\right) \cdot b_i \cdot \exp\left(-\sum_{j \in N_i}\left(1 - \pi_j^{Les}\right)\right)}{p_{other}\left(y_i | \hat{\theta}^{(t-1)}\right) \cdot \exp\left(-\sum_{j \in N_i} \pi_j^{Les}\right)}\right).$$

Determination of the initial threshold and of the final threshold

The initial threshold κ is a cutoff that produces the initial lesion map, L_{init}, from the GM lesion belief map, B_{GM} (see Preprocessing with already available software section). In detail, values near 0 imply many voxels (all voxels of GM with a TPM_{WM} value over 0 and with a FLAIR intensity value above GM average) whereas values around 1 imply a very conservative initialization. We tested the images of all patients with values of κ ranging from 0.05 to 0.95 with an increment of 0.05 (Fig. 2A). Applying values below 0.1 led to identification cortical hyperintensities which are inherent to FLAIR images whereas lesions were missed at values above 0.8. In conclusion, the effect of different κ values seemed to be limited in the range from 0.1 to 0.8. Yet, we analyzed the influence of κ on the agreement with manual segmentation (see Evaluation section) as measured with the Dice coefficient (DC). As shown in Fig. 2A, there is a plateau of DC values for κ values between 0.25 and 0.4. Eventually, we choose the value of 0.3 for κ, as it goes along with the greatest mean, the greatest minimum and the smallest bandwidth of DCs.

To enable comparison with manual segmentation (see Evaluation section), the lesion probability maps must be transformed into binary maps. To this end, we chose the threshold of 1.00. The distribution of all voxels with lesion probability greater than 0 across all subjects (Fig. 2B) strongly suggested this threshold since there was a sharp increase in frequency of voxel values in the range from >0.95 to ≤1.00 compared to voxel values in the range from >0.90 to ≤0.95; further, of the latter class, 99.65% of the voxel values were 1.00.

Evaluation

Since no gold standard for segmentation of T2-hyperintense lesions exists, we compared our algorithm with a semi-automatic manual tracing pipeline, which is based on commercially available software (Amira 5.3.3, Visage Imaging, Inc.) and which has been applied for basic research studies (Bendfeldt et al., 2010) and clinical trials (Li et al., 2006). At first, the manual segmentation was independently performed by two investigators, who were blinded to the study group. Then, a difference image of the two binary lesion maps was generated for each subject and both experts together decided which differences were assigned to lesions or not.

We performed a correlation and regression analysis to compare the volumetric agreement between automated and manual segmentation. For better estimation of intercept, slope, and R2, we included the data of our control subjects. Since 3D acquisition of FLAIR sequences has not been used commonly, we repeated estimation of R2 after reslicing of the FLAIR images to 3 and 6 mm slice thickness in sagittal and axial orientations, respectively.

To determine agreement between automatic and manual segmentation, we used standard validation techniques (Anbeek et al., 2004; Ashburner and Friston, 2005). We extracted the true positives (*TP*) and true negatives (*TN*) as well as the false positives (*FP*) and false negatives (*FN*). Then, we calculated the similarity measures of sensitivity (SE), SE = TP/(TP + FN), specificity (SP), SP = TN/(TN + FP), and accuracy (AC), AC = (TN + TP)/(TN + TP + FP + FN). Furthermore, we calculated the Dice coefficient (DC), which equally weighed the number of false negatives and false positives without accounting for true negatives (Dice, 1945; Zijdenbos et al., 1994):

$$DC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}.$$

All of these similarity measures have values between 0 and 1 with higher values indicating better quality.

Moreover, we analyzed a group of 10 MS patients with posterior fossa lesions and 18 control subjects.

Results

Based on T1-weighted and FLAIR images, T2-hyperintense WM lesions were segmented. Correlation analysis of lesion volumes (52 MS patients and 18 control subjects) derived from automated segmentation with those derived from manual tracing yielded excellent results with R2 values greater than 0.93 irrespective of orientation and slice thickness of the FLAIR sequence (Fig. 3, 3 mm slice thickness not shown). The slope of the regression line of 0.948 did not differ significantly from 1.0 (95% confidence interval, 0.892 to 1.004) and the intercept of −0.154 not from 0 (95% confidence interval, −1.04 to 0.732). Moreover, a high degree of agreement between manual tracing and automated segmentation was demonstrated with regard to sensitivity, specificity, accuracy, and DC (Table 2). To further evaluate differences between both methods with respect to size and location of the lesions, we determined the DC in the 52 MS patients and related them to the lesion volume indicating decreasing DCs with decreasing lesion volume (Fig. 4). However, in the patient group, 71% (n = 37) showed an excellent DC of greater than 0.7 (see Fig. 5 for 2 examples). Of the remaining 29% (n = 15), 10% (n = 5) had a DC below 0.6. Of the latter, the

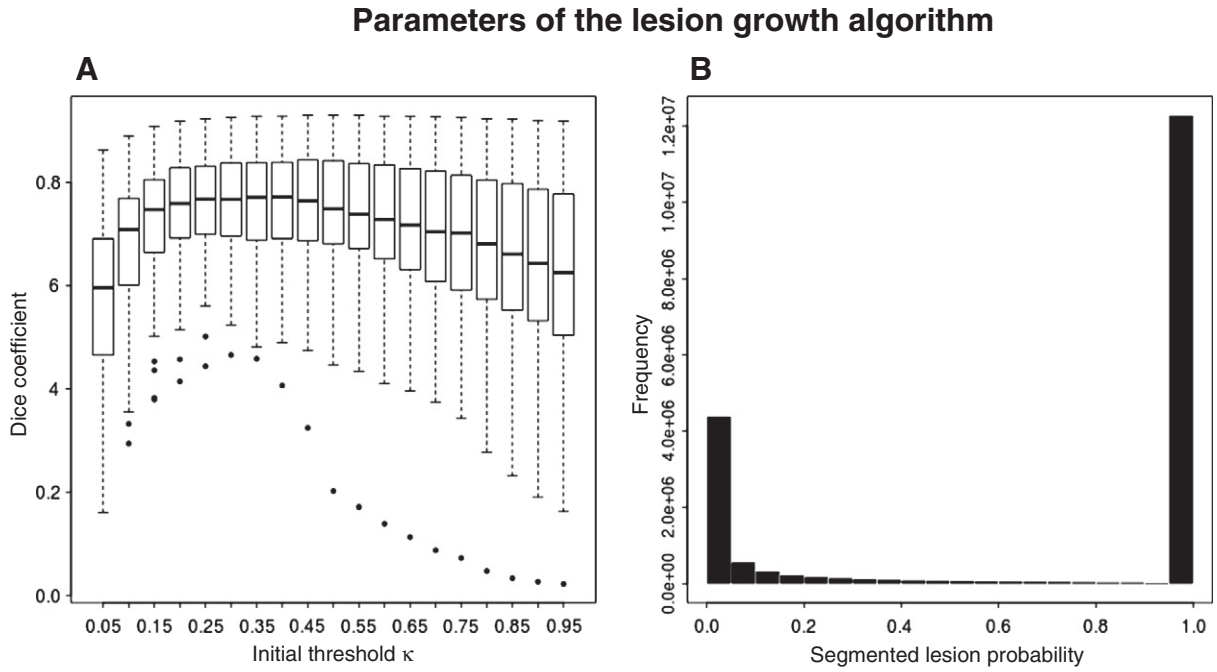

Fig. 2. Parameters of the lesion growth algorithm are shown. In Panel A, scatter plots of Dice coefficients from all patients over different initial thresholds κ are shown. The κ value of 0.3 goes along with Dice coefficients that show the greatest mean, greatest minimum, and smallest bandwidth. Panel B shows a histogram of the lesion probabilities of all voxels greater zero across all subjects. Note that 99.65% of the voxels in the range from >0.95 to ≤1.00 as represented by the right bar were 1.00 so that we chose the threshold of 1.00.

lowest DC observed was 0.46 (Fig. 6, Panel A) and the highest lesion volume 13.44 ml (Fig. 6, Panel B). Analysis of posterior fossa lesions from 10 MS patients yielded that our algorithm detected 11 of 13 posterior fossa lesions and 85% of the lesion volume (overall sensitivity, 0.85; overall DC, 0.94) but no false positive lesions. In the control group, hyperintense foci volume ranged from 0.0 to 1.53 ml (0.25-Quantil, 0.029 ml; median, 0.058 ml; 0.75-quantil, 0.176 ml) and was almost exclusively limited to anterior and posterior periventricular capping, as well as septal hyperintensity as illustrated by the images of the control subject with the highest hyperintense foci volume (Fig. 6, Panel C).

Discussion

We developed and evaluated an algorithm for automated segmentation of T2-hyperintense lesions in MS. We will review the strategy of our algorithm, assess the results of its evaluation, and, finally, speculate on its potential opportunities.

Our algorithm (Fig. 1) requires high-resolution T1-weighted images, which have been regarded most suitable for VBM (Ashburner and Friston, 2000) and which have become broadly available not only in neuroimaging research but also in clinical practice. It also requires FLAIR images, which have increasingly been used in MRI

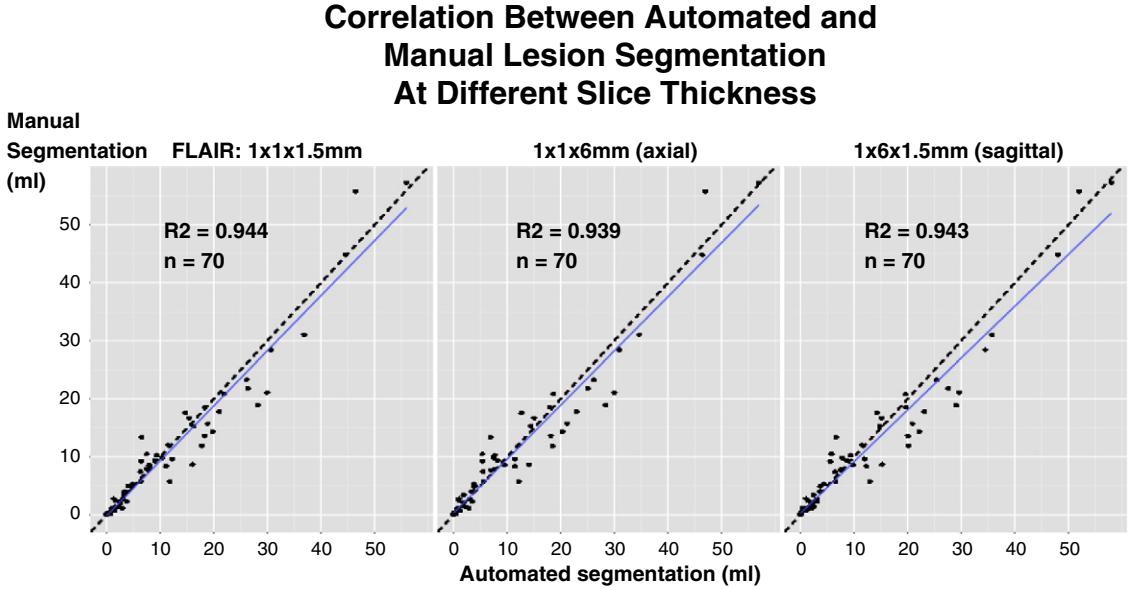

Fig. 3. Manually traced lesion volumes are plotted over automatically segmented lesion volumes at different orientations and slice thickness of the FLAIR sequence. Data from 52 MS patients and 18 control subjects are included. The respective R2 values are given.

Table 2
Statistics of similarity between automated segmentation and manual tracing derived from 53 MS patients.

Lesion volume	Sensitivity TP/(TP + FN)	Specificity TN/(TN + FP)	Accuracy (TN + TP)/(TN + TP + FN + FP)	Dice coefficient 2TP/(2TP + FP + FN)
(ml)	(min. mean max.)	(min. mean max.)	(min. mean max.)	(min. mean max.)
<5	0.4289 0.7332 0.9673	0.9997 0.9999 1.0000	0.9995 0.9998 1.0000	0.4658 0.6665 0.8025
5–10	0.3889 0.7592 0.9497	0.9994 0.9998 1.0000	0.9990 0.9995 0.9999	0.5243 0.7594 0.8910
10–15	0.7359 0.8870 0.9655	0.9991 0.9996 0.9999	0.9986 0.9993 0.9997	0.6738 0.8157 0.8727
>15	0.9012 0.9494 0.9841	0.9990 0.9994 0.9997	0.9990 0.9993 0.9997	0.7838 0.8498 0.9253
Total	0.3889 0.8033 0.9841	0.9990 0.9997 1.0000	0.9986 0.9995 1.0000	0.4658 0.7531 0.9253

Note. FN, false negative; FP, false positive; min., minimum; max., maximum; TN, true negative; TP, true positive.

protocols for MS patients (Filippi et al., 1996; Stevenson et al., 1997; Wattjes et al., 2006b). Further, T1-weighted images were based on a GRE sequence as commonly applied at 3 T. In preliminary experiments, we failed to establish a robust segmentation of MS lesions from a single sequence since the algorithms applied attributed a considerable number of voxels to the lesion compartment in any case which resulted in tremendous false positive misclassification of voxels to the lesion class especially in patients with low lesion volume and also in control subjects. Inspired by van Leemput et al. (Van Leemput et al., 2001), we surpassed lesion misclassification by incorporating two sequences in our algorithm. At first, we assigned all voxels to one of the three tissue classes of GM, WM, and CSF by the use of PVE labels derived from T1-weighted images, then, estimated the distribution of FLAIR intensities for each tissue class separately, and, finally, detected FLAIR hyperintense outliers within each tissue class. This way, the number of voxels correctly assigned to lesions can vary from zero to large values. To account for variable intensity within FLAIR images with regard to both normal tissue and lesions, we created an iterative algorithm that expanded the lesion belief from a conservative assumption toward a liberal assumption by voxel-wise weighed the likelihood of belonging to gray or white matter against the likelihood of belonging to lesions. Further, a hidden MRF segmentation model as well as a priori knowledge on WM location was incorporated. It may seem surprising that our algorithm starts with the binary lesion belief map derived from the GM tissue class since this implies the assumption of, at least, some T1-hypointense voxels within every lesion. In contrast, evidence suggests that not all T2-hyperintense MS lesions are T1-hypointense (Bagnato et al., 2003; Sahraian et al., 2010). However, respective studies on these T1-hypointense lesions, also called black (or dark) holes, have almost exclusively been performed at field strengths of up to 1.5 T with turbo spin echo sequences whereas we applied a GRE sequence at 3 T and did not observe lesions without a T1-hypointense part. Of note, the initial lesion estimate is allowed to expand toward voxels, which are not T1-hypointense so that a lesion must only display a T1-hypointense part rather than complete T1-hypointensity. Based on our experience of possibly tremendous misclassification particularly in patients with low lesion volume, we also evaluated our algorithm with real data including control subjects, and patients with lesion volumes ranging from low to high.

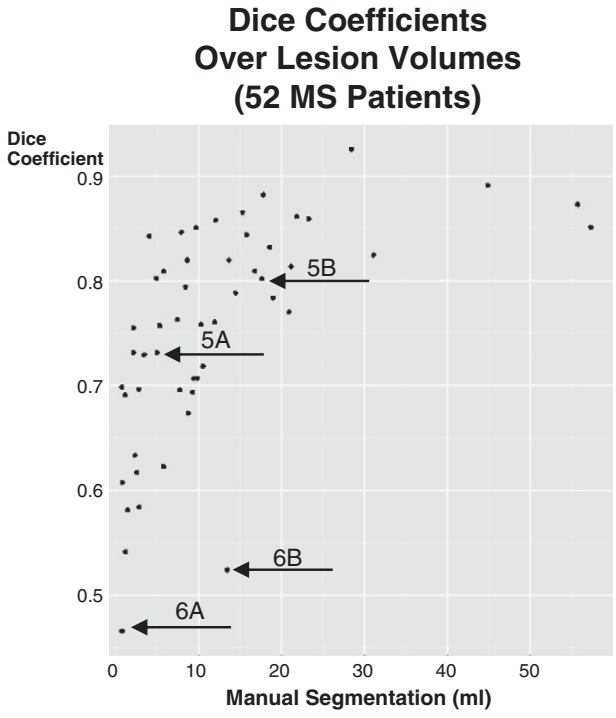

Fig. 4. Dice coefficients (DCs) of 52 patients are plotted over lesion volumes derived from manual tracing. The points indicated by numbered arrows correspond to Panels of Figs. 5 and 6.

Intriguingly, evaluation of an algorithm on MS lesion segmentation is hampered by the lack of a commonly accepted gold standard so that we compared the results derived from our algorithm with those derived from manual tracing by the use of a contour detection tool as suggested by others (Bendfeldt et al., 2010; Li et al., 2006). Mere correlation analyses of both methods showed excellent results (Fig. 3). The high R2 value indicates that both results share more than 93% of their variability. Furthermore, neither the slope differed significantly from 1 nor the intercept from zero suggesting that the lesion extent determined by our algorithm largely resembles what an experienced examiner assigns to be a lesion. Yet mere correlation measures of global lesion volumes are insufficient for evaluation since they measure association of overall volume but not spatial agreement (Bartko, 1991). Therefore, we calculated the standard validation parameters of sensitivity, specificity, accuracy, and Dice coefficient (Table 2). Of note, the validation parameters as determined here are based on volumes, i.e. 3D data, so that they may seem to understate the quality of our algorithm. For example, since lesion borders are often fuzzy, it is well conceivable that the radius of a small lesion is measured to be 3 mm by the algorithm under investigation and 4 mm by the algorithm taken as the gold standard resulting in a sensitivity of 0.42 ($3^3/4^3$). This value nearly equals the lowest sensitivity measured in our patients (0.43); hence, sensitivity of our algorithm can be regarded to be very good. On the other hand, the measures of specificity and accuracy are excellent with regard to the mere numbers but still of little value since these parameters are strongly influenced by the number of true negatives, which is inherently very large as is the volume of the whole brain compared to the lesion volume. Only the DC equally weighed the number of false positives and false negatives without accounting for the absolute number of true negatives so that this similarity measure seems most suitable to evaluate the overall quality of our lesion segmentation algorithm. Nevertheless, the DC has two limitations. 1) In our study, DC determination becomes more critical with decreasing lesion volume which was also reported by others (Anbeek et al., 2004; Wu et al., 2006). This is well conceivable assuming that lesion

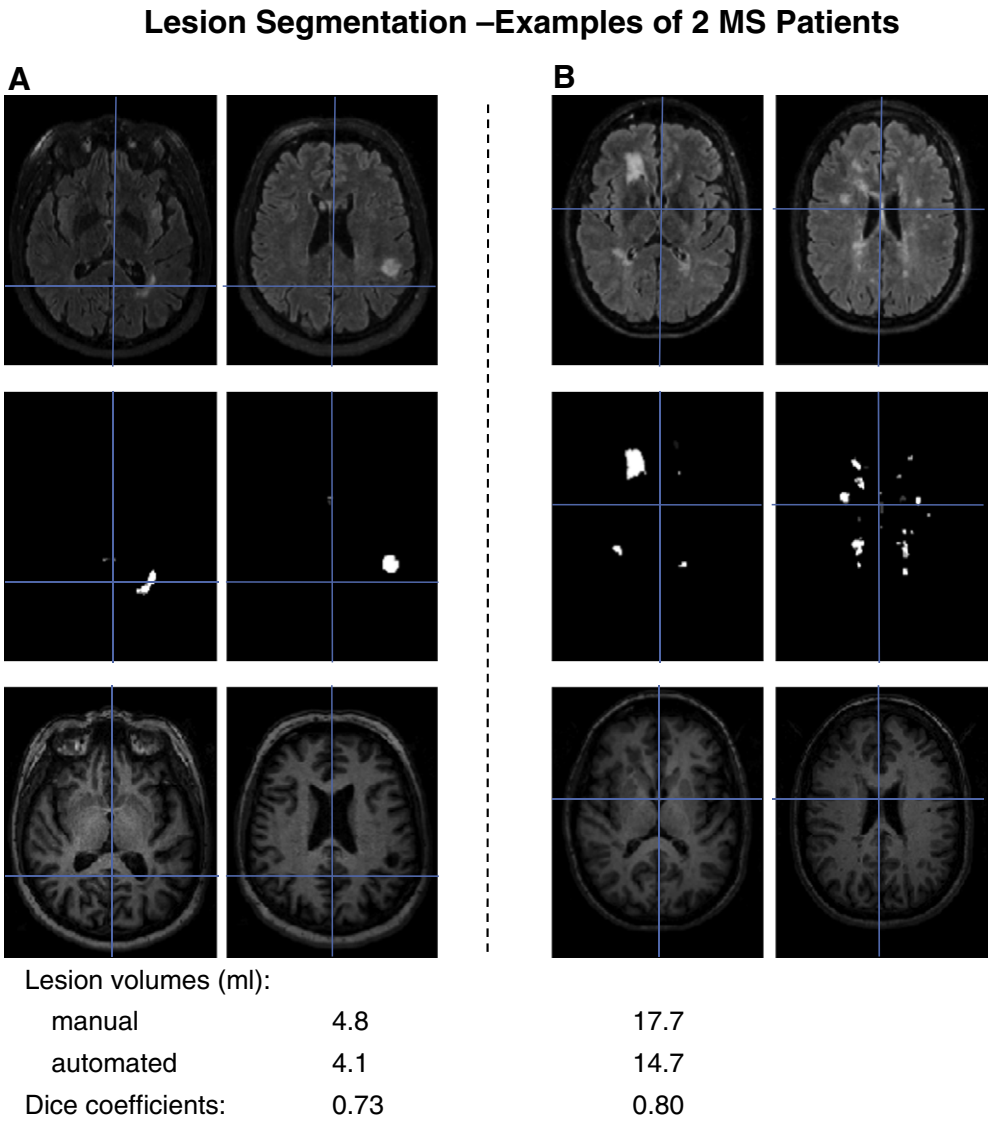

Fig. 5. Exemplary lesion segmentation is displayed from two MS patients with Dice coefficients over 0.7 (Panels A and B). Two axial slices are shown for each patient (upper row, FLAIR images; middle row, lesion maps; lower row, T1-weighted images). Panel letters correspond to arrows in Fig. 4. Lesion volumes and Dice coefficients are given at the bottom.

borders are determined with equal absolute errors irrespective of lesion size. This is likely to apply for our algorithm and, possibly even more so, for manual tracing (Woo et al., 2006) as also illustrated by our outlier patients (arrows in Fig. 4, Panels A and B of Fig. 6) in whom low DC values result from both false positives and false negatives. Further, in the extreme case, in which no lesion exists and no lesion is detected, as was the case in 1 of our 18 control subjects, the DC is not even defined (division by zero), although the result is perfect. 2) Apart from the fact that higher DC values imply better agreement, no commonly accepted rules on the interpretation of the DC exist. Some authors regard DC values over 0.7 as "excellent" (Anbeek et al., 2004; Bartko, 1991) others regard DC values over 0.4 as "moderate", over 0.6 as "substantial", and over 0.8 as "almost perfect" (Landis and Koch, 1977) while others emphasize that "conventional interpretative guidelines" may be misleading as DC values "obtained from samples with different base rates may not be comparable" (Uebersax, 1987). In addition, we evaluated our algorithm separately with regard to lesions in the posterior fossa because two studies found FLAIR imaging to be less sensitive here (Filippi et al., 1996; Stevenson et al., 1997). Although this finding lacked significance in a later study at 3 T (Wattjes et al., 2006b), we speculated that difficulties in detecting posterior fossa lesions could come more into play when applying an automated tool which was the case to some degree as 2 of 13 lesions were missed. Yet our patients with posterior fossa lesions were selected according to their medical records which were also based on conventional T2-weighted images. Our examiners may have detected posterior fossa lesions more easily at FLAIR images knowing that lesions are likely to exist. Hence, sensitivity values for posterior fossa lesions possibly reflect lower sensitivity of FLAIR compared to conventional T2-weighted sequences to some degree. Moreover, we applied our algorithm in 18 control subjects. Here, the hyperintensities identified were almost exclusively limited to anterior and posterior periventricular capping, as well as septal hyperintensity. These results are well in accordance with a study on normal findings on FLAIR images at 3 T (Neema et al., 2009). In conclusion, our degree of agreement with manual tracing is remarkably good given that most DC values exceeded 0.7, that the DC values below 0.6 were observed almost exclusively in patients with low lesion volume, that most posterior fossa lesions were detected, and that no misclassification occurred in control subjects.

Finally, we repeated our evaluation analyses with the same images after reslicing of the FLAIR images to larger slice thickness at different orientations, since 3D FLAIR sequences have not been broadly established. This yielded highly comparable R2 values. These results

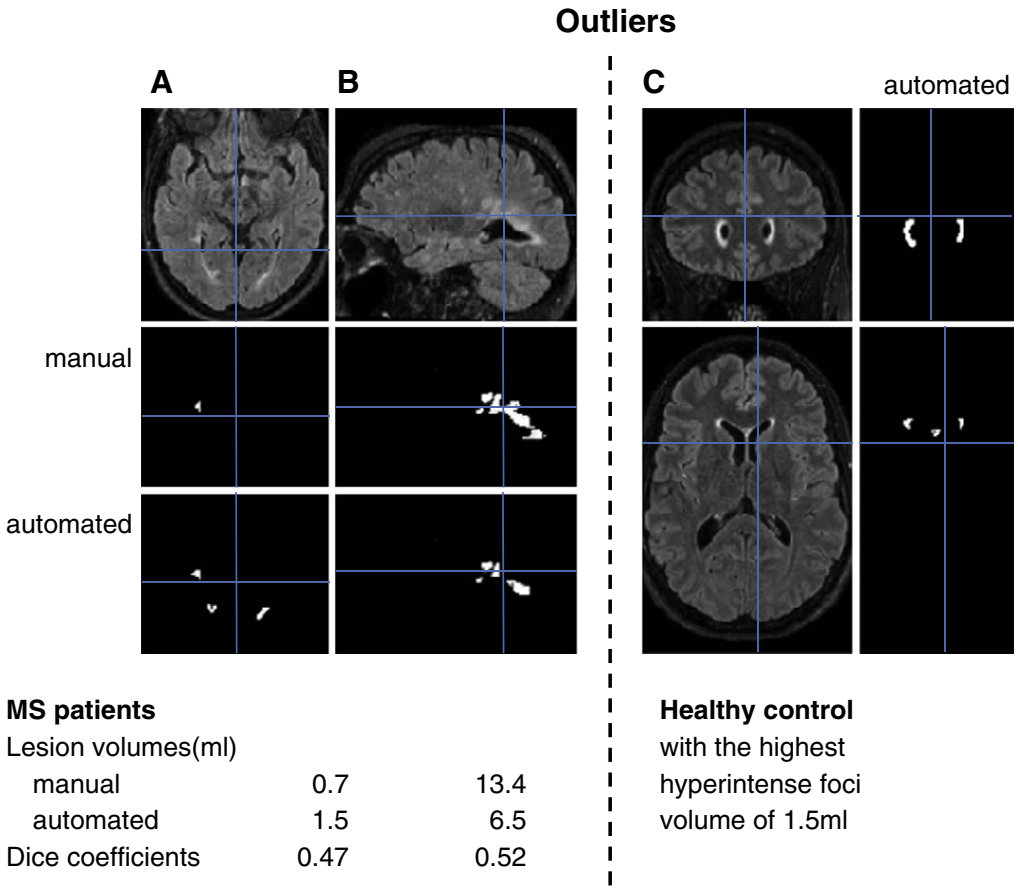

Fig. 6. Outliers of lesion segmentation are shown. Panel letters correspond to arrows in Fig. 4. Panel A shows the patient with the lowest Dice coefficient (DC), which results from posterior periventricular lesions detected by automated segmentation but not by manual tracing. Panel B shows the patient with the highest lesion volume among those with a DC below 0.6; the lesion volume results from many lesions, of which the manually segmented lesions appear to be larger. Panel C shows the control subject with the highest hyperintense foci volume which is restricted to anterior periventricular capping and septal hyperintensity. Lesion volumes and Dice coefficients are given at the bottom.

suggest that our algorithm may also operate well on data derived from protocols of other 3 T scanners based on a 3D GRE T1-weighted and a conventional FLAIR sequence. In this way, our tool may help to eventually take advantage of modern MRI protocols for MS patients in basic research and even clinical trials. However, these potential opportunities require validation with data from other protocols based on a conventional FLAIR sequence and a 3D GRE T1-weighted sequence at 3 T. Currently, we are programming an SPM toolbox of our algorithm, including the opportunity to adapt κ, which will be freely available to the scientific community.

In summary, we have developed a promising tool for automated detection of T2-hyperintense lesions in MS based on a modern 3 T MRI protocol including a 3D GRE T1-weighted and a FLAIR sequence.

Disclosure statement

The authors declare that there are neither actual nor potential conflicts of interest.

Acknowledgments

C.G. is supported by the German BMBF grant 01EV0709. M.A. was supported by Merck Serono. This work was in part supported by a grant from the German Ministry for Education and Research (BMBF, "German Competence Network Multiple Sclerosis" (KKNMS), Control-MS, 01GI0917).

Appendix A1. Pseudo-code description of the lesion growth algorithm

$L = L_{\text{init}}$

DO WHILE stopping criterion not satisfied

FOR all voxels i with $\pi_i^{\text{Les}} = 0$ that have at least one voxel in their neighborhood N_i with $\pi_j^{\text{Les}} > 0$, $j \in \text{N}_i$

$$\pi_i^{\text{Les}} = \min\left(1, \frac{p_{Les}\left(y_i|\hat{\alpha}^{(t-1)}, \hat{\beta}^{(t-1)}\right) \cdot b_i \cdot \exp\left(-\sum_{j\in N_i}\left(1-\pi_j^{Les}\right)\right)}{p_{other}\left(y_i|\hat{\theta}^{(t-1)}\right) \cdot \exp\left(-\sum_{j\in N_i}\pi_j^{Les}\right)}\right)$$

END FOR

UPDATE $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$

UPDATE stopping criterion

END DO

with

B: lesion belief map with values $b_i, i = 1,\ldots, n$ (Lesion belief maps and initialization section)

y_i: normalized FLAIR intensity of voxel i (Lesion belief maps and initialization section)

p_{Les}: gamma probability density function for the lesion class with shape and scale parameters alpha and beta, respectively (Lesion growing section)

p_{other}: mixture of three Gaussians for the other tissue classes (CSF, GM and WM) with parameter vector $\theta = \{\mu_{CSF}, \mu_{GM}, \mu_{WM}, \sigma^2_{CSF}, \sigma^2_{GM}, \sigma^2_{WM}\}$ (Lesion growing section)

L_{init}: initialized lesion map (Lesion belief maps and initialization section)

L: lesion probability map with values $\pi_i^{Les}, i = 1, \ldots, n$

N_i: first order neighborhood of voxel i

Stopping criterion: maximal number of iterations or greatest new lesion probability <0.01.

Appendix A2. Performance parameters

The algorithm performs well on a computer with a 3.2 GHz processor and 8 GB RAM. Preprocessing by SPM8 and VBM8 takes about 10 min, the lesion growth algorithm another 2–3 min depending on the number of iterations. In our analysis, the median number of iterations was 16, the maximum 50. We assume that performance is similar on computers with less memory capacity and processing power. However, we recommend the use of at least 2 GB RAM.

References

Ait-Ali, L.S., Prima, S., Hellier, P., Carsin, B., Edan, G., Barillot, C., 2005. STREM: a robust multidimensional parametric method to segment MS lesions in MRI. In: Duncan, J.S., Gerig, G. (Eds.), Medical Image Computing and Computer-Assisted Intervention — Miccai 2005, pp. 409–416. Pt 1.

Akselrod-Ballin, A., Galun, M., Gomori, J.M., Filippi, M., Valsasina, P., Basri, R., Brandt, A., 2009. Automatic segmentation and classification of multiple sclerosis in multichannel MRI. IEEE Trans Biomed Eng 56, 2461–2469.

Anbeek, P., Vincken, K.L., van Osch, M.J., Bisschops, R.H., van der Grond, J., 2004. Probabilistic segmentation of white matter lesions in MR imaging. Neuroimage 21, 1037–1044.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—the methods. Neuroimage 11, 805–821.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.

Bagnato, F., Jeffries, N., Richert, N.D., Stone, R.D., Ohayon, J.M., McFarland, H.F., Frank, J.A., 2003. Evolution of T1 black holes in patients with multiple sclerosis imaged monthly for 4 years. Brain 126, 1782–1789.

Bakshi, R., Ariyaratana, S., Benedict, R.H., Jacobs, L., 2001. Fluid-attenuated inversion recovery magnetic resonance imaging detects cortical and juxtacortical multiple sclerosis lesions. Arch. Neurol. 58, 742–748.

Bartko, J.J., 1991. Measurement and reliability: statistical thinking considerations. Schizophr. Bull. 17, 483–489.

Bendfeldt, K., Blumhagen, J.O., Egger, H., Loetscher, P., Denier, N., Kuster, P., Traud, S., Mueller-Lenke, N., Naegelin, Y., Gass, A., Hirsch, J., Kappos, L., Nichols, T.E., Radue, E.W., Borgwardt, S.J., 2010. Spatiotemporal distribution pattern of white matter lesion volumes and their association with regional grey matter volume reductions in relapsing–remitting multiple sclerosis. Hum. Brain Mapp. 31, 1542–1555.

Chard, D.T., Griffin, C.M., Parker, G.J., Kapoor, R., Thompson, A.J., Miller, D.H., 2002. Brain atrophy in clinically early relapsing–remitting multiple sclerosis. Brain 125, 327–337.

Compston, A., Coles, A., 2008. Multiple sclerosis. Lancet 372, 1502–1517.

Confavreux, C., Vukusic, S., 2008. The clinical epidemiology of multiple sclerosis. Neuroimaging Clin. N. Am. 18, 589–622 ix-x.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

Ebers, G.C., 1998. Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group. Lancet 352, 1498–1504.

Ferrari, R.J., Wei, X., Zhang, Y., Scott, J.N., Mitchell, J.R., 2003. Segmentation of multiple sclerosis lesions using support vector machines. Proceedings of SPIE 5032, 16–26.

Filippi, M., Yousry, T., Baratti, C., Horsfield, M.A., Mammi, S., Becker, C., Voltz, R., Spuler, S., Campi, A., Reiser, M.F., Comi, G., 1996. Quantitative assessment of MRI lesion load in multiple sclerosis, a comparison of conventional spin-echo with fast fluid-attenuated inversion recovery. Brain 119 (Pt 4), 1349–1355.

Fisher, E., Lee, J.C., Nakamura, K., Rudick, R.A., 2008. Gray matter atrophy in multiple sclerosis: a longitudinal study. Ann. Neurol. 64, 255–265.

Fisniku, L.K., Brex, P.A., Altmann, D.R., Miszkiel, K.A., Benton, C.E., Lanyon, R., Thompson, A.J., Miller, D.H., 2008. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. Brain 131, 808–817.

Freifeld, O., Greenspan, H., Goldberger, J., 2009. Multiple sclerosis lesion detection using constrained GMM and curve evolution. Int J Biomed Imaging 2009, 715124.

Garcia-Lorenzo, D., Lecoeur, J., Arnold, D.L., Collins, D.L., Barillot, C., 2009. Multiple sclerosis lesion segmentation using an automatic multimodal graph cuts. Med Image Comput Comput Assist Interv 12, 584–591.

Geremia, E., Menze, B., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N., 2010. Spatial decision forests for MS lesion segmentation in multi-channel MR images. Med Image Comput Comput Assist Interv 13, 111–118.

Hammersley, J.M., Clifford, P., 1971. Markov fields on finite graphs and lattices.

Herskovits, E.H., Bryan, R.N., Yang, F., 2008. Automated Bayesian segmentation of microvascular white-matter lesions in the ACCORD-MIND study. Adv. Med. Sci. 53, 182–190.

Jacobs, L.D., Beck, R.W., Simon, J.H., Kinkel, R.P., Brownscheidle, C.M., Murray, T.J., Simonian, N.A., Slasor, P.J., Sandrock, A.W., 2000. Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis, CHAMPS Study Group. N. Engl. J. Med. 343, 898–904.

Kappos, L., 1998. Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. European Study Group on interferon beta-1b in secondary progressive MS. Lancet 352, 1491–1497.

Khayati, R., Vafadust, M., Towhidkhah, F., Nabavi, M., 2008. Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model. Comput. Biol. Med. 38, 379–390.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

Li, D.K., Held, U., Petkau, J., Daumer, M., Barkhof, F., Fazekas, F., Frank, J.A., Kappos, L., Miller, D.H., Simon, J.H., Wolinsky, J.S., Filippi, M., 2006. MRI T2 lesion burden in multiple sclerosis: a plateauing relationship with clinical disability. Neurology 66, 1384–1389.

Li, L., Wei, X., Li, X., Rizvi, S., Liang, Z., 2005. Mixture segmentation of multispectral MR brain images for multiple sclerosis. Journal of Systemics, Cybernetics and Informatics 3, 65–68.

Lüders, E., Gaser, C., Narr, K.L., Toga, A.W., 2009. Why sex matters: brain size independent differences in gray matter distributions between men and women. J. Neurosci. 29, 14265–14270.

Neema, M., Guss, Z.D., Stankiewicz, J.M., Arora, A., Healy, B.C., Bakshi, R., 2009. Normal findings on brain fluid-attenuated inversion recovery MR images at 3 T. AJNR Am. J. Neuroradiol. 30, 911–916.

Noseworthy, J.H., Lucchinetti, C., Rodriguez, M., Weinshenker, B.G., 2000. Multiple sclerosis. N. Engl. J. Med. 343, 938–952.

Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F.D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A.J., Waubant, E., Weinshenker, B., Wolinsky, J.S., 2011. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann. Neurol. 69, 292–302.

Rajapakse, J.C., Giedd, J.N., Rapoport, J.L., 1997. Statistical approach to segmentation of single-channel cerebral MR images. IEEE Trans. Med. Imaging 16, 176–186.

Sahraian, M.A., Radue, E.W., Haller, S., Kappos, L., 2010. Black holes in multiple sclerosis: definition, evolution, and clinical correlations. Acta Neurol. Scand. 122, 1–8.

Stevenson, V.L., Gawne-Cain, M.L., Barker, G.J., Thompson, A.J., Miller, D.H., 1997. Imaging of the spinal cord and brain in multiple sclerosis: a comparative study between fast FLAIR and fast spin echo. J. Neurol. 244, 119–124.

Tohka, J., Zijdenbos, A., Evans, A., 2004. Fast and robust parameter estimation for statistical partial volume models in brain MRI. Neuroimage 23, 84–97.

Uebersax, J.S., 1987. Diversity of decision-making models and the measurement of interrater agreement. Psychol. Bull. 101, 140–146.

Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P., 2001. Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Trans. Med. Imaging 20, 677–688.

Wattjes, M.P., Barkhof, F., 2009. High field MRI in the diagnosis of multiple sclerosis: high field-high yield? Neuroradiology 51, 279–292.

Wattjes, M.P., Lutterbey, G.G., Harzheim, M., Gieseke, J., Traber, F., Klotz, L., Klockgether, T., Schild, H.H., 2006a. Higher sensitivity in the detection of inflammatory brain lesions in patients with clinically isolated syndromes suggestive of multiple sclerosis using high field MRI: an intraindividual comparison of 1.5 T with 3.0 T. Eur. Radiol. 16, 2067–2073.

Wattjes, M.P., Lutterbey, G.G., Harzheim, M., Gieseke, J., Traber, F., Klotz, L., Klockgether, T., Schild, H.H., 2006b. Imaging of inflammatory lesions at 3.0 Tesla in patients with clinically isolated syndromes suggestive of multiple sclerosis: a comparison of fluid-attenuated inversion recovery with T2 turbo spin-echo. Eur. Radiol. 16, 1494–1500.

Weiner, H.L., 2009. The challenge of multiple sclerosis: how do we cure a chronic heterogeneous disease? Ann. Neurol. 65, 239–248.

Wels, M., Huber, M., Hornegger, J., 2008. Fully automated segmentation of multiple sclerosis lesions in multispectral MRI. Pattern Recognition and Image Analysis 18, 347–350.

Winkler, G., 2003. Image Analysis, Random Fields and Markov Chain Monte Carlo Methods. A Mathematical Introduction. Springer-Verlag, 2nd Edition.

Woo, J.H., Henry, L.P., Krejza, J., Melhem, E.R., 2006. Detection of simulated multiple sclerosis lesions on T2-weighted and FLAIR images of the brain: observer performance. Radiology 241, 206–212.

Woolrich, M.W., Behrens, T.E.J., Beckmann, C.F., Smith, S.M., 2005. Mixture models with adaptive spatial regularization for segmentation with an application to FMRI data. IEEE Trans. Med. Imaging 24 (1), 1–11.

Wu, Y., Warfield, S.K., Tan, I.L., Wells, W.M., Meier, D.S., van Schijndel, R.A., Barkhof, F., Guttmann, C.R., 2006. Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. Neuroimage 32, 1205–1215.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57.

Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C., 1994. Morphometric analysis of white-matter lesions in MR-images — method and validation. IEEE Trans. Med. Imaging 13, 716–724.