



Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches

Xavier Lladó^{a,*}, Arnau Oliver^a, Mariano Cabezas^a, Jordi Freixenet^a, Joan C. Vilanova^b, Ana Quiles^c, Laia Valls^c, Lluís Ramió-Torrentà^d, Àlex Rovira^e

^a Dept. of Computer Architecture and Technology, University of Girona, Spain

^b Girona Magnetic Resonance Center, Girona, Spain

^c Dept. of Radiology, Dr. Josep Trueta University Hospital, Girona, Spain

^d Multiple Sclerosis and Neuroimmunology Unit, Dr. Josep Trueta University Hospital, Institut d'Investigació Biomèdica de Girona, Girona, Spain

^e Magnetic Resonance Unit, Dept of Radiology, Vall d'Hebron University Hospital, Barcelona, Spain

ARTICLE INFO

Article history:

Received 28 July 2010

Received in revised form 14 July 2011

Accepted 8 October 2011

Available online 18 October 2011

Keywords:

Brain MRI

Multiple sclerosis

Review

Automated lesion segmentation

ABSTRACT

Automatic segmentation of multiple sclerosis (MS) lesions in brain MRI has been widely investigated in recent years with the goal of helping MS diagnosis and patient follow-up. However, the performance of most of the algorithms still falls far below expert expectations. In this paper, we review the main approaches to automated MS lesion segmentation. The main features of the segmentation algorithms are analysed and the most recent important techniques are classified into different strategies according to their main principle, pointing out their strengths and weaknesses and suggesting new research directions. A qualitative and quantitative comparison of the results of the approaches analysed is also presented. Finally, possible future approaches to MS lesion segmentation are discussed.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Multiple sclerosis (MS) is a chronic, persistent inflammatory-demyelinating and degenerative disease of the central nervous system (CNS), characterised pathologically by areas of inflammation, demyelination, axonal loss, and gliosis scattered throughout the CNS, often causing motor, sensorial, vision, coordination, deambulation, and cognitive impairment [28]. The two main clinical phenomena of prototypic MS are relapses and progression. Relapses are considered to be the clinical expression of acute focal or multifocal inflammatory demyelination, disseminated within the CNS. The remission of symptoms early in the disease is likely to be the result of remyelination, resolution of inflammation, and compensatory mechanisms such as the redistribution of axolemmal sodium channels and cortical plasticity. These recovery mechanisms are less effective after recurrent attacks.

Multiple sclerosis is the most frequent, non-traumatic, neurological disease capable of causing disability in young adults. It is relatively common in Europe, the United States, Canada, New Zealand and parts of Australia, but rare in Asia and the tropics and subtropics of all continents. Within regions having a temperate climate, the incidence and prevalence of MS increase with latitude – both north and south of the equator. Multiple sclerosis is between two and three-times more common in women than in men, but men have a tendency for later disease onset with a poorer prognosis. The incidence of MS is low in childhood, increases rapidly after the age of 18, reaches a peak between 25 and 35 and then slowly declines, becoming rare at 50 and older. It is estimated that there are between 1.3 and 2.5 million cases of MS in the world, with some 350,000 cases

* Corresponding author.

E-mail address: llado@eia.udg.edu (X. Lladó).

in Western Europe [35]. According to the latest epidemiological studies, the prevalence and incidence of MS has been increasing worldwide.

MS does not usually shorten life significantly, but there is a substantial impact on personal, social and work activities in patients and their families and its socioeconomic importance is second only to trauma in young adults. The etiology of MS is still unknown, but it appears most likely to be the result of interplay between as yet unidentified environmental factors and susceptible genes. Along with the demyelinating episodes, there may be damage to the exposed axons, leading to transection of the axons and retrograde neuronal degeneration. This process can be irreversible and is responsible for the accrual of disability that occurs as the disease progresses.

Conventional Magnetic Resonance Imaging (MRI) techniques [50], such as T2-weighted (T2-w) and gadolinium-enhanced T1-weighted (T1-w) sequences, are highly sensitive in detecting MS plaques and can provide quantitative assessment of inflammatory activity and lesion load. MRI-derived metrics have become the most important paraclinical tool for diagnosing MS, for understanding the natural history of the disease and for monitoring the efficacy of experimental treatments. Both acute and chronic MS plaques appear as focal high-signal intensity areas on T2-w sequences, reflecting their increased tissue water content. The increase in the signal indicates edema, inflammation, demyelination, reactive gliosis and/or axonal loss in proportions that differ from lesion to lesion. They are typically discrete and focal at the early stages of the disease, but become confluent as the disease progresses. The total T2 lesion volume of the brain increases by approximately 5–10% each year in the relapsing forms of MS [65]. Gadolinium-enhanced T1-w imaging is highly sensitive in detecting inflammatory activity. This technique detects disease activity 5 to 10 times more frequently than clinical evaluation of relapses, suggesting that most of the enhancing lesions (EL) are clinically silent. Longitudinal and cross-sectional MR studies have shown that the formation of new MS plaques is often associated with contrast enhancement, mainly in the acute and relapsing stages of the disease.

Approximately 10–20% of T2 hyperintense lesions (HL) are also visible on T1-w images as areas of low signal intensity compared with normal-appearing white matter (WM). These so-called T1 black holes (BH) have a different pathological substrate that depends, in part, on the lesion age. The hypointensity is present in up to 80% of recently formed lesions and probably represents marked edema, with or without myelin destruction or axonal loss. In most cases, the acute (or wet) black holes become isointense within a few months as inflammatory activity abates, edema resolves and reparative mechanisms like remyelination become active. Less than 40% evolve into persisting or chronic black holes [13]. Chronic black holes correlate pathologically with the most severe demyelination and axonal loss, indicating areas of irreversible tissue damage. T1-w sequences have a higher specificity than T2-w sequences for detecting lesions with irreversible tissue damage and may serve as surrogate markers of disability progression in clinical trials. Fig. 1 shows examples of MRI scans of a patient with MS lesions.

Atrophy of the brain and spinal cord is an important part of MS pathology and a clinically relevant component progression of the disease [17]. CNS atrophy, which involves both grey matter (GM) and white matter (WM), is a progressive phenomenon that worsens with increasing disease duration, and progresses at a rate of between 0.6% and 1.2% of brain loss per year. Quantitative measures of whole-brain atrophy can be acquired by automated or semi-automated methods that display this progressive loss of brain tissue bulk in vivo in a sensitive and reproducible manner.

In clinical trials as well as in every-day clinical practice, MRI scans are visually assessed for qualitative analysis and manually marked if a quantitative analysis is required. Quantitative analysis has become invaluable in the assessment of disease progression [75,76] and the evaluation of therapies over the last 25 years [45,23]. In fact, MRI metrics have become common primary endpoints in phase II immunomodulatory drug therapy trials [91]. Moreover, magnetisation transfer MRI, diffusion tensor MRI (DTI), proton MR spectroscopy, and functional MRI are nowadays also contributing to elucidate the mechanisms that underlie injury, repair, and functional adaptation in patients with MS [37,64].

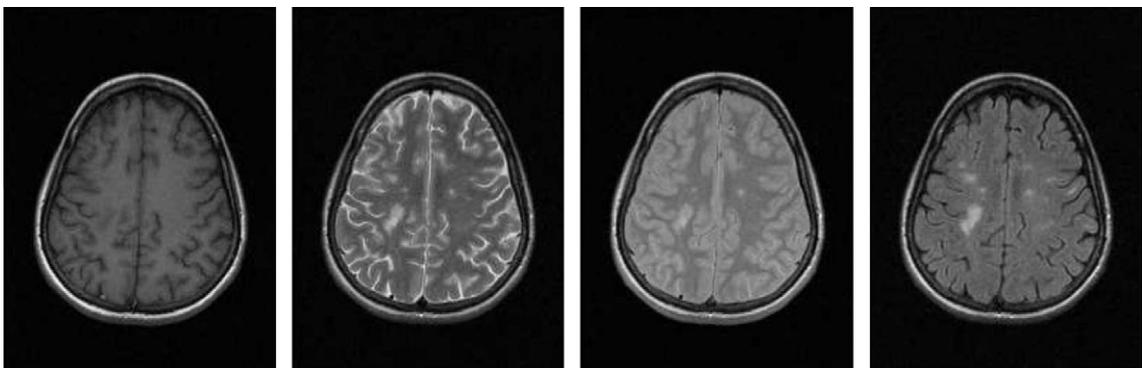


Fig. 1. MR images of a brain slice with MS lesions: (a) T1-w, (b) T2-w, (c) PD-w, and (d) FLAIR, respectively. Note that soft tissues are more distinguishable in the T1-w image, while lesions are usually better appreciated in the FLAIR one.

In quantitative analyses of focal lesions, in both cross-sectional and longitudinal studies, manual or semi-automated segmentations have been used to compute the total number of lesions and the total lesion volume. The manual delineation of MS lesions, however, is both challenging and time-consuming because of the large number of MRI slices required to compose three-dimensional information for each patient. Moreover, it is prone to intra-observer variability (the same study analysed by the same neuroradiologist at different times) and inter-observer variability (the same study analysed by different neuroradiologists), and requires strategies such as the STAPLE algorithm [107] to fuse different segmentation results into one. Another procedure used in clinical practice is scan-rescan reproducibility [68], which image processing practitioners often overlook when assessing the true reproducibility of the whole measurement process. Scan-rescan reproducibility consists of performing multiple scans of the same subject within a period of time and then assessing the reproducibility of both human observers and automated segmentations.

The development of fully automated MS segmentation methods, which can segment large amounts of MRI data and do not suffer from intra- and inter-observer variability, has become an active research field [95,8]. Unfortunately, the results of these fully automated methods show less agreement with manually segmented scans than those obtained with segmentations by independent observers. Moreover, when evaluating MS segmentation methods, there are still no *in vivo* methods to obtain reliable ground truth data mainly because of the large intra- and inter-observer variability. It should be noted that disagreements between segmentations may have important influences on certain evaluation measures due to the small volume of the lesions.

In this paper, the state-of-the-art of strategies for automated MS lesion segmentation is reviewed with the aim of pointing out their strengths and weaknesses and suggest new research directions. Moreover, recent significant works in this field are described and the different techniques are classified according to the strategy used. In particular, a first classification between supervised and unsupervised segmentation strategies is established, dividing the supervised group into atlas-based approaches and those based on training by means of features extracted from manual segmentations. Moreover, the unsupervised group is further divided into those techniques that use tissue segmentation to obtain the lesions and those that use only the lesion properties for the segmentation. In addition to describing and classifying these approaches, a description of the algorithms used to segment the lesions as well as the features and the type of MR images used is also provided.

Various reviews of brain MRI segmentation have been presented in the past. For instance, Bezdek et al. [19] analysed 90 papers on MRI segmentation using pattern recognition techniques. The authors suggested dividing the algorithms into two categories: supervised methods (such as Bayes classifiers with labelled maximum likelihood estimators, the *k*-nearest neighbour rule (*k*NN), and artificial neural networks (ANN)) and unsupervised methods (i.e. Bayes classifiers with unlabelled maximum likelihood estimators or the fuzzy C-means (FCM) algorithms). In addition Clarke et al. [25] reviewed not only methods for MRI segmentation, but also general pre-processing algorithms, validation methods and registration between different MR images. However, these reviews were only related to soft brain tissue segmentation. More recently, Souplet et al. [93] presented a review of semi-automated and automated MS lesion segmentation approaches, analysing MS lesions, pre-processing steps and segmentation approaches. However, to the best of our knowledge, this paper is the first attempt to review the most relevant works in automated MS lesion segmentation that provides an evaluation of the experimental results.

The inability to compare evaluation results due to the use of different data sets and different evaluation measures has been a major obstacle to reviewing MS lesion segmentation methods. Ideally, methods should be applied to a common database and compared to a ground truth. This is however very difficult due to the lack of common public databases of real images along with their ground truth and the fact that only few methods are publicly available. Our contribution is though close to this idea. Here we will compare quantitatively the segmentation approaches accordingly to their reported results in the literature. Furthermore, the recent MS Lesion Segmentation Challenge [95] provided a common framework for MS lesion segmentation algorithms, allowing comparisons to be made between different approaches. In the results section of this paper, the most typical measures used for evaluating MS lesion segmentation results are described. Moreover, the works analysed are compared in a qualitative and quantitative way.

The rest of this paper is organised as follows. Section 2 reviews the pre-processing steps needed when automatically segmenting MS lesions. Section 3 shows the classification of the segmentation approaches and also reviews the features and algorithms used. In Section 4, the different image databases and measures used to evaluate the results are presented and used to compare the performances of the works analysed. Discussions are given in Section 5, whilst the paper finishes with some conclusions and suggestions for future work.

2. Pre-processing

The segmentation of MR brain images is difficult because of variable imaging parameters, overlapping intensities, noise, partial voluming, gradients, motion, echoes, blurred edges, normal anatomical variations and susceptibility artifacts [81]. Therefore, before applying any approach to MS lesion segmentation, there are generally two pre-processing steps that are carried out: first, the removal of those image artefacts and second, the removal of non-brain tissue, such as the skull, from the image. Other optional pre-processing steps may be applied such as the equalisation of soft brain tissues or registration between different MR images. In this section, these steps are briefly described.

It is well known that the capture process itself corrupts MR images with various artefacts (such as flow and motion artefacts, susceptibility artefacts, partial volume effects, the point-spread function of different pulse sequences, etc.). This can lead to inaccurate segmentation. However, from the image processing viewpoint, it is common to simplify all these problems [89], defining the intensities, Y , of each voxel as:

$$Y = \alpha X + \beta \quad (1)$$

where X is the real intensity, α is a multiplicative smooth bias field that causes intensity inhomogeneities due to the sensitivity of the reception coil and β represents additive noise. While it is often assumed that β follows a Gaussian distribution, there have been different approaches for the correction of the bias field α . This is an important issue since pixels belonging to different tissues may be assigned with the same grey-level value when varying this term. Two different studies have recently reviewed different ways to overcome this preprocessing step problem [52,105]. Both studies classify these methods into various groups: segmentation-based, filtering-based, surface fitting-based, histogram-based and other specific techniques. However, as pointed out by Hou [52], none of the methods has been shown to be superior to the others and exclusively applicable.

Skull stripping is another important pre-processing step since fat, skull, skin and other non-brain tissues may cause misclassifications in some approaches due to the intensity similarities with brain structures [29]. This is also the case of brain sub-cortical structures, where advanced analysis is not readily applicable since the intensity characteristics that overlap between different structures may reduce the reliability of automated segmentation methods. Two recent works have analysed and compared the state-of-the-art methods to extract brain regions from MR images. The first study, by Boesen et al. [20], compared four systems: Statistical Parametric Mapping (SPM2) [9], the Brain Extraction Tool (BET) [90], the Brain Surface Extractor (BSE) [83], and their own Minneapolis Consensus Strip (McStrip) [73]. They validate these systems with three data sets of T1-w images. The second study, by Hartley et al. [49], compared only the accuracy of BET and BSE against 296 PD-w images. In both studies, manual segmentations were used as the “gold standard”. Boesen et al. [20] concluded that the McStrip - which is a hybrid algorithm incorporating intensity thresholding, nonlinear warping, and edge detection - consistently outperformed SPM2, BET, and BSE. Furthermore, other studies [54,109] also suggest to perform this region removal task before applying the inhomogeneity correction. In this way, the correction would be carried out only on those voxels belonging to the internal brain tissues.

As already mentioned, other pre-processing steps can be applied to improve the classification of tissues and the subsequent lesion segmentation. Examples of such pre-processing pipeline approaches can be found in studies by Zijdenbos et al. [113] and Hou and Huang [53]. The steps in the pipeline approach include registration between different MR images [15], intensity normalisation [82], and the transformation to a standard space to facilitate atlas-based segmentation. For example, the Talairach coordinate system is used to describe the location of brain structures that are independent from individual differences in the size and overall shape of the brain [98]. Some approaches also use intra-sequence and inter-sequence pre-processing registration steps. Intra-sequence registration enables the correction of the misregistration between the acquisition steps, since standard multi-slice acquisition sequences are acquired in multiple, interleaved passes (this step is only used when applicable to the actual acquisition sequence, such as a multi-slice dual-echo/ PD-w sequence). On the other hand, the inter-sequence registration can help compensate for possible (and likely) patient motion between scans.

Ultimately, as in many image processing applications, the choice of all these pre-processing steps is a trade-off between algorithmic complexity (i.e. the number of tunable parameters), flexibility and processing speed (although nowadays this is not a critical issue since a few hours of processing time for a large series of images is acceptable). Moreover, this selection may also depend on the demands made of the developed pipeline or on user-defined criteria in terms of accuracy, reproducibility and consistency over time or across data sets acquired on different scanners.

3. MS lesion segmentation

In this section, the recent state-of-the-art of automated MS lesion segmentation is reviewed. Firstly, the main image features used as input for the different segmentation algorithms are analysed. Afterwards, a classification of the different strategies is proposed and the most significant works in this field are described. The approaches reviewed are summarised in Table 1, which offers a compact at-a-glance overview of these studies.

3.1. Features used for lesion segmentation

Based on the assumption that different brain tissues have different grey-level intensities, the most common feature used for lesion segmentation is the voxel intensity [11]. Furthermore, the appearance of the tissue and the lesions depends on the MR image used (see Fig. 1, which shows four examples of different MRI scans of a patient with MS lesions). For instance, the white matter (WM) appears as the brightest tissue in T1-w images, the darkest in T2-w images and an intermediate grey-level in the fluid-attenuated inversion recovery (FLAIR) image. In contrast, the cerebro-spinal fluid (CSF) is the brightest tissue in T2-w images, whereas in both T1-w and the FLAIR images it is the darkest tissue. The grey matter (GM) appears as the intermediate grey in both T1-w and T2-w images, while it is slightly brighter than the WM in the FLAIR image. On the other

Table 1

Summary of the approaches to MS lesion segmentation with respect to the sequences, algorithms and lesions. The approaches are classified according to the strategies. The acronyms for the algorithms stand for (in alphabetical order): adaptive mixtures method (AMM), artificial neural networks (ANN), constrained Gaussian mixture models (CGMM), expectation maximisation (EM), fast trimmed likelihood estimator (FAST-TLE), fuzzy C-means (FCM), Fisher linear discriminant (FLD), hidden Markov chains (HMC), k-nearest neighbours (kNN), mean shift (MeS), morphological greyscale reconstruction (MGR), Markov random fields (MRF) and principal components analysis (PCA). The acronyms for the lesions and sequences stand for: attenuation of fluid by fast inversion recovery with magnetisation transfer imaging with variable echoes (AFFIRMATIVE), diffusion tensor imaging (DTI), fractional anisotropy (FA), mean diffusivity (MD), black holes (BH), enhanced lesions (EL), and hyperintense T2 lesions (HL).

		Article	Algorithms	Images	Lesions		
Supervised	Atlas	Van Leemput et al. [62]	EM + MRF	PD, T1, T2	HL		
		Zijdenbos et al. [113]	ANN	PD, T1, T2	HL		
		Wu et al. [109]	kNN	PD, T1c, T2	EL & BH & HL		
		Shiee et al. [86]	FCM	PD, T1, T2, FLAIR	HL		
		Shiee et al. [87]	FCM	T1, T2, FLAIR	HL		
		Bricq et al. [22]	FAST-TLE + HMC	T2, FLAIR	HL		
		Prastawa and Gerig [72]	Region partitioning	T1, T2	HL		
		Kroon et al. [57]	PCA	T1, T2, FLAIR, DTI (FA, MD)	HL		
		Souplet et al. [92]	EM + GMM	T1, T2, FLAIR	HL		
		Tomas and Warfield [100]	Bayes	T1, T2, FLAIR	HL		
		Akselrod-Ballin et al. [3]	FLD + Decision Forest	PD, T1, T2, FLAIR	HL		
		Shiee et al. [88]	FCM	T1, T2, FLAIR	HL		
		Manual segmentations	Kamber [55]	Different classifiers	PD, T1, T2	HL	
	Goldberg et al. [46]		ANN	PD, T2, FLAIR	EL & HL		
	Alfano et al. [4]		Spatial Clustering	PD, T1, T2	HL		
	Anbeek et al. [7]		kNN	PD, T1, T2, FLAIR, IR	HL		
	Anbeek et al. [6]		kNN	PD, T1, T2, FLAIR, IR	HL		
	Sajja et al. [78]		Parzen windows	PD, T2, FLAIR	HL		
	Datta et al. [31]		Parzen windows + MGR	PD, T1, T2, FLAIR	BH		
	Anbeek et al. [5]		kNN	FLAIR	HL		
	Scully et al. [79]		Bayes	T1, T2, FLAIR	HL		
	Morra et al. [67]		AdaBoost	T1, T2, FLAIR, DTI (FA, MD)	HL		
	Subbanna [96]		Simulated annealing + MRF	PD, T1, T2	BH & HL		
	Lecoeur et al. [61]		Graph Cuts	PD, T1, T2	HL		
	Unsupervised		Tissue	Freifeld et al. [38]	EM + CGMM	PD, T1, T2	HL
				Khayati et al. [56]	Bayes + AMM + MRF	FLAIR	HL
		García-Lorenzo et al. [42]		EM + MeS	T1, T2, PD	HL	
García-Lorenzo et al. [44,41]		EM		PD, T1, T2	HL		
García-Lorenzo et al. [43]		EM		T1, T2, FLAIR	HL		
Bedell and Narayana [16]		Threshold Subtraction		T1, T1c, AFFIRMATIVE	EL		
Lesion		Bourdraa et al. [21]	FCM	PD, T2	HL		
		He and Narayana [51]	MGR	T1, T1c, AFFIRMATIVE	EL		
		Datta et al. [30]	MGR	T1, T1c, T2, FLAIR	EL		
		Saha and Bandyopadhyay [77]	FCM	T2	HL		

hand, the lesions can appear as either hyperintense or hypointense signals, depending on the properties of the lesion itself and the type of MR image used. For instance, lesions appear as hyperintense signals in PD-w, T2-w and FLAIR images. Note that for segmenting MS lesions, the FLAIR sequence is better at differentiating between lesions and healthy tissue despite the fact that it introduces bony and flow artefacts into the image, thus complicating the lesion segmentation in the sub-cortical regions. In contrast, in T1-w images, active lesions are displayed with hyperintense signals, while necrotic ones appear as hypointense signals.

Analysing the literature, one may distinguish between single-channel or multi-channel approaches, i.e. approaches that use only one MR image or those that combine several images. Single-channel approaches are mainly used to segment the brain tissues. For instance, T1-w images are widely used for this purpose, since they show the best contrast between the three main brain tissues: WM, GM and CSF. This initial tissue segmentation may then be used to help obtain the final lesion segmentation, and T2-w and PD-w are the classical images for detecting MS lesions. Another example of the single-channel approach is the segmentation of MS lesions using just the FLAIR sequence [56]. The multi-channel approaches, on the other hand, use at least two of the PD-w, T1-w, T2-w, and FLAIR images. One of the benefits of using more than one of the different MR images is that it increases the intensity feature space, producing a better discrimination between brain tissues. Furthermore, more than one kind of image may be required because MS lesions can appear independently in different images [109], depending on their subtype. As shown in Table 1, most of the approaches combine different MR images to perform both the tissue and lesion segmentation.

Additional features are used in some approaches to include spatial information in the algorithms. This is usually done using Markov Random Fields (MRF) to model neighbourhood interactions [63,112]. If the parameters controlling the strength of the spatial interactions are properly selected, smoother structures are obtained. Alternatives to MRF that also include spatial information are the Fuzzy Connectedness (FC) segmentation methods [101] or the inclusion of probabilistic atlas [57].

Moreover, most of the algorithms can be roughly divided into either global or local depending on the information they use (i.e. the feature extraction process). Global methods extract features from the whole image and then use this information to classify each voxel independently. In contrast, local methods use only local information to create, in many cases, an undetermined number of local regions. The challenge that these local methods must overcome is to combine these local regions to build a global and meaningful segmentation.

3.2. Classification of lesion segmentation approaches

In this section, the MS lesion segmentation methods are described according to the classification shown in Table 1. Notice that the table offers an overview of all these works with respect to the strategy, the type of MR images used, the type of lesions detected and the algorithms. The criteria used to select these methods are based on several aspects: (1) representative works for each of the identified strategies; (2) the reported experimental results and the evaluation measures used by the authors; and (3) the data sets used to perform the experiments (synthetic, real cases, and data from the MS Lesion Segmentation Challenge [95], which enables a quantitative comparison of the evaluation results).

The table is divided between supervised and unsupervised segmentation strategies. Supervised approaches are those based on using some kind of a priori information or knowledge to perform the MS lesion segmentation. The group of supervised strategies is further subdivided into two sub-groups of approaches. In the first group all the approaches use atlas information and therefore require the application of a registration process to the analysed image to perform the segmentation. In the second group all the approaches perform an initial training step on features extracted from manually segmented images annotated by neuroradiologists. The methods in this second group employ the image intensities previously segmented by an expert to train a classifier that segments the tissues and lesions of the MR images.

With regard to the unsupervised strategies, where no prior knowledge is used, two different sub-groups can also be identified. A sub-group of methods that segment brain tissue to help lesion segmentation and another sub-group that use only the lesion properties for segmentation. In the first sub-group, there are methods that either segment the tissue first and then the MS lesions, or segment the tissue and the lesions at the same time. In the second sub-group, the methods directly segment the lesions according to their properties, without providing tissue segmentation. The advantage of segmenting the tissue is that neuroradiologists can also evaluate the GM tissue volumetry and monitor the progression of cerebral atrophy.

3.2.1. Supervised strategies based on atlas

Looking at the strategies based on atlas information, it is possible to distinguish between the use of both statistical and topological atlases. A statistical atlas provides the prior probability of each voxel to belong to a particular tissue class. This statistical atlas is built from a set of manual segmentations of the structures of interest, where the boundaries of each structure are used to make a smooth probability map and to account for anatomical variations beyond those present within the training set. Notice that the use of an atlas can be helpful in classifying tissues in the presence of noise or inhomogeneities [32] (an atlas takes spatial information into account), or in order to segment lesions as deviations from normal human brains. On the other hand, a topological atlas is a parcellation of the brain that is edited to encode a specific topology for each structure and group of structures. This topological atlas is usually used to preserve topology and to lower the influence of competing intensity clusters in regions that are spatially disconnected, while the statistical atlas affects the segmentation of adjacent structures that have similar intensity. As shown in the flowchart of Fig. 2(a), notice that in atlas-based segmentation methods, the analysed MR image has to be registered with the atlas before the segmentation is done. Hence, the challenge for these atlas-based approaches is to align the atlas and the images, thereby converting the segmentation problem into a registration problem.

Van Leemput et al. [62] provide an example of an atlas-based approach for MS lesion segmentation.¹ They proposed an intensity-based tissue classification using a stochastic model extracted from an expectation maximisation (EM) algorithm [34], while simultaneously detecting MS lesions as outliers that were not well explained by the model. In this method, a prior classification is derived from a digital brain atlas that contains information about the expected location of WM, GM and CSF. Their approach also corrects for MRI field inhomogeneities, estimates tissue-specific intensity models from the data itself and incorporates contextual information into the classification using an MRF.

The k -nearest neighbour (kNN) [36] is probably the commonest of supervised classification methods. With this algorithm a test sample is classified by the majority class of its closest neighbours. kNN makes strong assumptions about the data, i.e. that there is no correlation among different multivariate channels and that all variances are the same. Based on this classifier, Wu et al. [109] proposed the automatic segmentation of MS lesions into three subtypes: enhancing lesions, black holes and hyperintense lesions. An intensity-based statistical kNN classifier is combined here with atlas segmentation to extract WM masks. Based on the assumption that lesions are only found within WM regions, the authors discard all the lesions outside the masks. Moreover, partial volume problems (i.e. arising from the fact that a voxel may be composed of more than one tissue type) are corrected using morphological operators. Another supervised approach relying on atlas information is pro-

¹ This software is publicly available at <http://www.medicalimagecomputing.com/downloads/ems.php> as an add-on to the well-known SPM package (<http://www.fil.ion.ucl.ac.uk/spm/>).

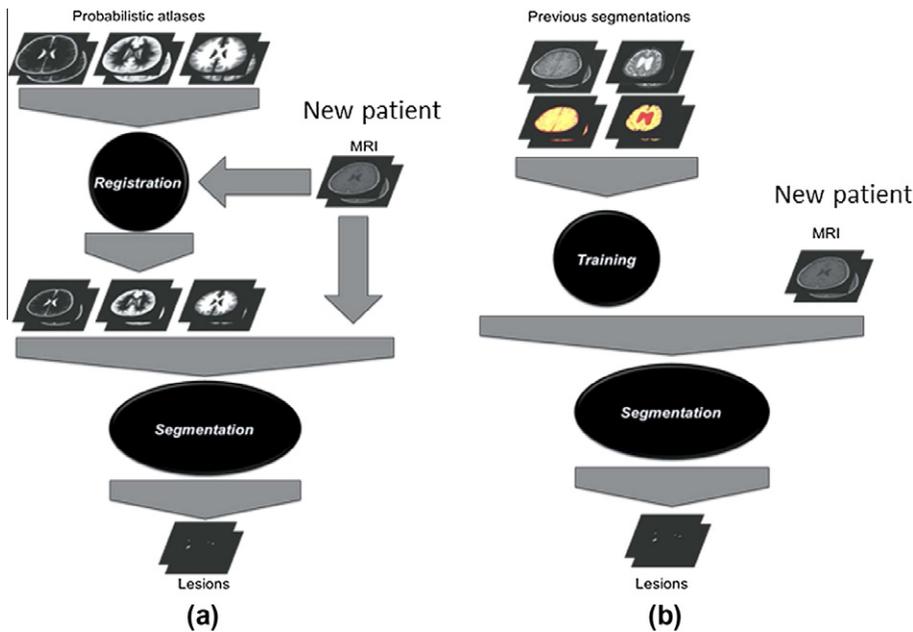


Fig. 2. Flowchart of the supervised strategies: (a) based on atlas information and (b) based on learning from manual segmentation.

vided by Zijdenbos et al. [113]. This method² uses the probability – extracted from the atlas – that tissue is WM, GM or CSF and uses the intensities from T1-w, T2-w and PD-w as inputs to an ANN classifier, which performs the lesion segmentation.

There are more examples of atlas-based approaches. The method proposed by Shiee et al. [86–88] segments brain tissues in an iterative way, interleaving a fuzzy segmentation and defining topologically consistent regions. MS lesions are identified as dark holes inside the WM. The authors use multi-channel images to segment the major structures of the whole brain. Basically, their method is an atlas-based segmentation technique employing a topological atlas and a statistical atlas, together with the well-known fuzzy C-means (FCM) algorithm [18] which performs the classification. As reported by Shiee et al., the advantage of using the topological atlas is that all segmented structures are spatially constrained, thereby allowing subsequent processing to perform cortical reconstruction and cortical unfolding. The automated atlas-based segmentation method presented by Bricq et al. [22] performs tissue classification using an algorithm based on the Trimmed Likelihood Estimation of a mixture model [48], while the lesions are outliers to the model. Neighbourhood information is encoded by the hidden Markov chain (HMC) model and also incorporates the use of a statistical atlas. Prastawa and Gerig [72] proposed a fully automated lesion segmentation method that combines outlier detection and region partitioning, and is based on using an atlas of healthy subjects to detect lesions as outliers. The algorithm is iterative and alternates between estimating the intensity probability density functions (PDF), computing voxelwise spatial probabilities, correcting for intensity inhomogeneities and partitioning the images into spatially coherent regions. Notice that the intensity PDF for the healthy brain tissue is computed using samples obtained from the atlas. The Minimum Covariance Determinant robust estimation scheme is then applied to the intensity samples of healthy tissues to determine inliers and outliers. The inlier samples are used to form the PDF of the brain tissue intensities, while the outliers are assigned to the lesion class, similar to the approach of Van Leemput et al. [62]. Moreover, a watershed algorithm that takes the neighbourhood information into account is used to initialise the segmentation. Therefore, segmented regions are used rather than individual voxels. The authors argued that the use of regions helps to significantly reduce false positives inherently linked to conventional voxel-based classification.

Kroon et al. [57] introduced a method based on a local feature vector for automated lesion segmentation of multi-channel MRI data. Their local feature vector contains neighbourhood voxel intensities, histogram and probabilistic atlas information. The histogram information is added to provide the model with low pass intensity information of a certain region, while the atlas probability allows it to exclude false MS voxel detection in areas where MS is less probable. Principal Component Analysis (PCA), with a log-likelihood ratio, is then used to classify each voxel. Using a different algorithm, Souplet et al. [92] proposed a method designed to detect hyperintense signal areas on a T2-FLAIR sequence.³ They first apply an EM algorithm to perform the segmentation in T1-w and T2-w images. This segmentation, using an atlas, allows them to initialise – for each voxel – the probability of belonging to the different tissue intensities modelled as Gaussian mixture models. From the resulting tissue segmentations, the authors propose to automatically apply a threshold for the T2-FLAIR sequence to detect the most plausible

² This software is publicly available at <http://www.bic.mni.mcgill.ca/ServicesSoftwareAdvancedImageProcessingTools/HomePage>, as part of the INSECT application.

³ This software is publicly available at <http://www-sop.inria.fr/asclepios/software/SepINRIA>.

lesions in the hyperintense signals. In this final step an enhanced FLAIR image is generated to allow a better lesion segmentation.

Following the idea of Prastawa and Gerig [72], Tomas and Warfield [100] used the samples extracted from an atlas to train a classifier. This procedure to extract healthy tissue training samples is based on the work of Scully et al. [79], who extended the intensity feature space to obtain better discrimination between tissue clusters. Subsequently, their method uses an atlas to create a distance map from which training samples were selected. With this training data, a Bayes classifier performed the MS lesion segmentation. This approach also introduced an atlas-based post-processing step to remove false positive lesions. Finally, Akselrod-Ballin et al. [3] proposed a multi-scale approach that combines segmentation with classification to automatically segment MS lesions in multi-channel MR images. Their method uses segmentation to obtain a hierarchical decomposition of the MR images, which produces a rich set of features describing the regions in terms of intensity, shape, location, neighbourhood relations, and anatomical context. The atlas information is applied in this step to obtain statistical features. Afterwards, segmentation is performed using a decision forest along with Fisher linear discriminant analysis to deal with multiple features.

In conclusion, atlas-based approaches can be used to segment both the tissue and the lesions. Moreover, atlases make it possible to treat the lesions as outliers in the tissue, to introduce spatial information into the segmentation process and to reduce the false positive lesion segmentations. As a drawback, these approaches rely on building an atlas, which is not a simple task. In addition, they also introduce the registration problem into the MS lesion segmentation. Note that this registration step is even more difficult when dealing with cases with severe atrophy, large numbers of lesions, etc.

3.2.2. Supervised strategies based on learning from manual segmentation

The second group of supervised approaches uses manually-segmented images annotated by neuroradiologists to segment the MS lesions. Fig. 2(b) shows the flowchart followed by these algorithms. Note that unlike atlas-based approaches these approaches do not need any registration process between the analysed images and the atlas. In contrast, these methods use mainly the image intensities of different MR images to train a classifier for the segmentation purpose. As reported by several authors, the use of prior knowledge to guide the segmentation of MS lesions improves the robustness of the algorithms, thus reducing the volume of false positive lesions compared to purely data-driven segmentations. Note that some of the approaches classified in this category may be similar to strategies described in the previous section. However, the majority of the strategies included here rely on a training process performed using features extracted from manually-segmented MR images. Furthermore, some of these methods include the use of registration algorithms that focus on the intra-sequence and inter-sequence pre-processing registration steps. Table 1 shows that a large number of proposals have followed this strategy, most of them being multi-channel approaches. Furthermore, different classifiers or a combination of them, for example, ANN, kNN, AdaBoost, Bayesian classifiers or decision trees, have been used to perform the segmentation.

The first example in Table 1 is the proposal by Kamber et al. [55], where the inputs for the training step are not voxel intensities but rather their probabilities of belonging to WM, GM, or CSF tissue categories. Using this prior information as a pattern, their approach trains and tests a set of different classifiers to segment MS lesions. Goldberg et al. [46] use local thresholding to select the brightest regions of the image. Afterwards, the lesions are segmented by looking for closed contours and using different morphological properties such as area, perimeter and shape. For segmentation, an ANN is trained and used to classify the regions. Alfano et al. [4] also used previous segmentations of normal tissues to extract features used to train a spatial clustering algorithm. The authors stated in their experiments that their approach was also suitable for monitoring changes in the disease over time.

Anbeek et al. applied the kNN classifier in two studies. The aim of the first was exclusively to detect MS lesions [7] and spatial features were included in the classifier to achieve better lesion segmentations; in the second study, the aim was to model all brain tissues [6] and in this case, the kNN classifier was used to classify tissues and lesions simultaneously. It is important to mention that the authors tested their multi-channel approach using information from T1-w, inversion recovery (IR), PD-w, T2-w and FLAIR images, concluding that the incorporation of the T1-w, PD-w or T2-w did not significantly improve the segmentation results of the different brain tissue types.

From a different viewpoint, Sajja et al. [78] proposed segmenting CSF and lesions using a Parzen windows classifier and then segmenting WM and GM using a parametric method. The assumption behind this approach is that GM and WM, but not lesions and CSF, follow a Gaussian distribution. Therefore, the authors first classify CSF and hyperintense T2-w lesions using a Parzen classifier and then the remaining brain parenchyma – excluding CSF and lesions – is classified into GM and WM using the PD-w and T2-w images and an MRF together with an EM algorithm. This method also exploits contextual information by using a fuzzy-connectedness to minimise the false negative lesion classifications. In another work, the authors also proposed segmenting black holes (considered as regional minima) in MS using a similar strategy [31]. After applying the previous segmentation method, black hole segmentation is achieved by using grey-scale morphological reconstruction (MGR) in T1-w images.

The work of Scully et al. [79] introduced a new parametric method to the field of MS lesion segmentation. This method uses a vector image joint histogram, built over a training set, as an explicit model of the feature vectors indicating lesion. This model is then used to generate samples to train a naive Bayesian classifier which proceeds to classify the vector image composed of the T1-w, T2-w and FLAIR images. Using a different strategy, Morra et al. [67] proposed a framework to automatically segment sub-cortical structures in brain MR images. Their method uses an AdaBoost algorithm to learn a unified

appearance and context model which is then used to perform the lesion segmentation. Their feature pool includes intensity, position and neighbourhood features.

Subbanna et al. [96,97] presented a fully automated framework for identifying MS lesions in multi-channel MR images. Manual segmented images are used to extract intensity histograms of both tissue and lesions. From the histograms, multi-variate Gaussian distributions are estimated and used in the MRF classification step, which incorporate local spatial variations and neighbourhood information. Finally, Lecoeur et al. [61] also presented an optimised supervised lesion segmentation method using multi-channel MR images. Their proposal creates an optimised spectral gradient colour space from single-channel images. Based on this transformation, they then apply a Graph Cuts segmentation. As argued by the authors, the Graph Cuts algorithm provides an optimal solution for the joint use of regional and border information in a way similar to how MRFs work.

To summarise, the use of manually annotated data allows expert knowledge to be incorporated into MS segmentation approaches. Moreover, as shown in Table 1, having initial segmentations allows a large variety of classifiers (ANN, kNN, EM, Bayes, etc.) to be applied. As in the atlas-based approaches, selecting a good initial MRI training set is an important step. Another issue with many lesion segmentation algorithms, particularly those that employ training data to model lesion intensity profiles, is that they are dependent on a specific acquisition sequence. These approaches must be modified or re-trained to process data acquired using alternative pulse sequences.

3.2.3. Unsupervised strategies segmenting tissue

There are several works in which the main principle consists of applying an unsupervised algorithm to segment the tissue and the lesions, as it is schematically shown in the flowchart of Fig. 3(a). These approaches usually detect lesions as outliers on each tissue rather than adding a new class to the classification problem. For instance, Freifeld et al. [38] first initialise their algorithm based on a pre-segmentation using k -means and its subsequent decomposition into a mixture of many spatially-oriented Gaussians per tissue (constrained GMM) in order to capture the spatial layout [47]. The intensity is considered as a global parameter and is constrained to be the same value for a set of related Gaussians per tissue. In order to detect the lesions, a set of rules that distinguishes between normal tissue regions and lesions is defined. Following initialisation, voxel-wise GMM parameters are learned via an EM algorithm. Finally, an active contour algorithm is used to delineate lesion boundaries.

García-Lorenzo et al. [42] combined a modified version of the EM-based method (mEM) to the trimmed likelihood estimator with the Mean Shift algorithm [27] to segment MS lesions. A local segmentation approach (Mean Shift) is used to generate local regions in the images, while an EM variant is employed to classify the regions obtained into healthy tissue or lesions. In another work by García-Lorenzo et al. [44,43], the authors presented a modified version of the Spatio Temporal Robust Expectation Maximisation (STREM) [2] to perform the MS lesion segmentation. Their approach is based on three main processes: robust estimation of healthy tissues using the mEM introduced above, refinement of outlier detection and application of lesion rules. This work has been recently extended with a more extensive validation and explained in more detail [41]. The authors included in another work the possibility to semiautomatically improve the segmentation by using an

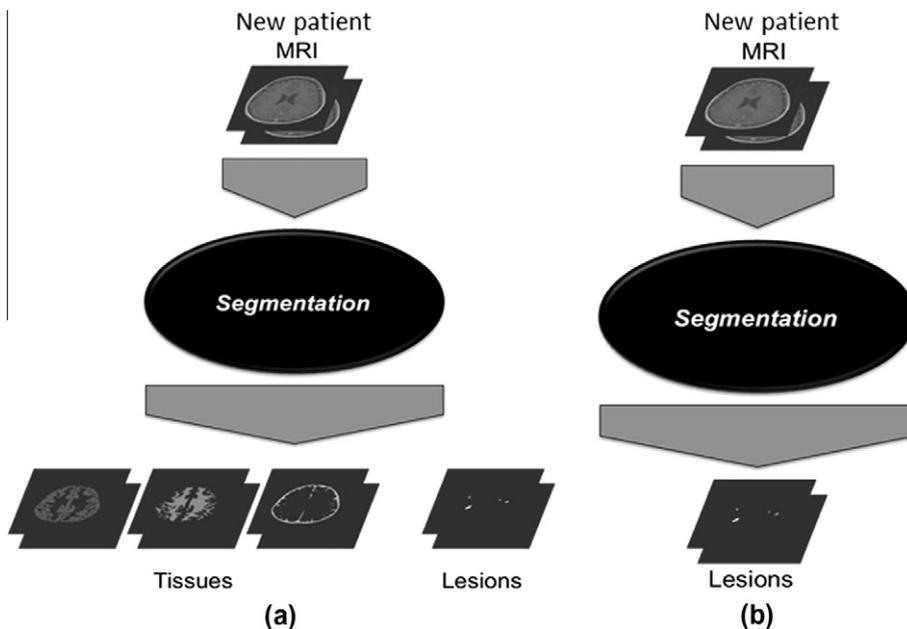


Fig. 3. Flowchart of the unsupervised strategies: (a) based on tissue properties and (b) based on lesions properties.

interactive Graph Cuts approach [40]. Khayati et al. [56] combined an adaptive mixtures method (AMM), MRF and a Bayesian classifier to simultaneously classify the three main brain tissues and the MS lesions using only FLAIR images. In particular, they first propose to segment the brain into four classes: WM, GM, CSF and “others”. Afterwards, inside the “others” class, lesions are dealt with as outliers not correctly explained by the model.

In conclusion, this set of algorithms are based on classifying the outliers coming from a previous tissue segmentation in order to provide the lesion segmentation. Notice that the results of the lesion segmentation depend highly on the quality of the tissue segmentation. Furthermore, not all the tissue segmentation methods take abnormal cases (with severe lesions, atrophy, etc.) into account.

3.2.4. Unsupervised strategies segmenting only lesions

This last group of unsupervised approaches is based on using only the lesion properties to perform the MS lesion segmentation, avoiding hence the tissue segmentation step (Fig. 3(b) summarises the flowchart of this strategy). For instance, Bedell and Narayana [16] present an automated segmentation and quantification of contrast-enhanced lesions based on performing threshold subtraction to eliminate enhancing structures such as choroid plexus. The authors reported that all MS lesions larger than 5 mm³ were successfully identified and the automated analysis produced no false positive or false negative lesions above this volume in 13 different patients. The method used by Boudraa et al. [21] performs a FCM algorithm two times to detect lesions. The first FCM algorithm has the goal of obtaining two clusters: one that groups together CSF and lesions, and another that groups together WM and GM. Afterwards, a second FCM algorithm is applied to distinguish between lesions and CSF. A final post-processing step based on anatomical knowledge is performed to remove extra segmented structures. This strategy assumes that, in the first step, all the lesions have been grouped together within a specific cluster and then, in the second step, this cluster is resegmented to distinguish between healthy tissue and lesions, taking spatial information into account. Saha et al. [77] automatically determine the number of clusters by introducing genetics into the algorithm. Membership values of points to different clusters are computed based on a point symmetry based distance rather than using the Euclidean distance. The chromosomes encode the centers of a number of clusters, whose value may vary.

He and Narayana [51] also proposed a method for the automated identification and segmentation of contrast-enhanced MS lesions in brain MR images. This method relies on an adaptive local segmentation derived from morphological grey-scale reconstruction operations to identify both lesion and non-lesion enhancements. Similarly, Datta et al. [30] developed a method for the identification and quantification of gadolinium (Gd) enhancements. This is also a multi-channel MRI approach and aims to identify enhancements using morphological operations. These enhancing lesions are further segmented based on Fuzzy Connectedness. Their experimental results show that accurate segmented Gd enhancements are obtained.

In regard to this strategy, one should noted that unlike the methods presented in the previous section, these approaches do not rely on an initial WM, GM and CSF tissue segmentation. Moreover, they are useful for segmenting special lesions such as black holes and enhancing lesions. As a drawback, however, they have to specifically define the properties used in each image, which is not an easy task. It should be noted that some artefacts may share the same lesion properties.

3.3. Summary of the strategies

Four strategies used to deal with the automated MS lesion segmentation have been presented. In what follows a brief summary of these strategies, summarising the main advantages and drawbacks is provided (see Table 2). Moreover, a more detailed description of a representative algorithm of each strategy is presented.

The approaches have been classified between supervised and unsupervised algorithms. The inherent advantage of supervised algorithms is that they can automatically learn the characteristics of both normal tissue and lesions. However, their main problem is that they rely on having a good training set, which may be difficult to obtain. Two supervised strategies have been identified according to the procedure the annotations are introduced into the algorithms: with or without using

Table 2

Advantages and drawbacks of the different reviewed MS lesion segmentation strategies.

	Description	Strengths	Weaknesses
Supervised based on atlas	The algorithms use atlas information	Includes local information Probabilistic framework	Needs registration Needs previous segmentations to create an atlas
Supervised based on training	The algorithms classify using manual training	No registration needed A wide list of classifiers can be used	Previous segmentations may contain errors Systems depend on the acquisition sequence
Unsupervised based on tissue	The lesions are found as tissues outliers	Tissue guides the lesion segmentation	Depends on the quality of the tissue segmentation
Unsupervised based on lesion	The lesions are found using empirical rules	Special sublesions can be segmented	Specific rules for each image and lesion type Artifacts may share lesion properties

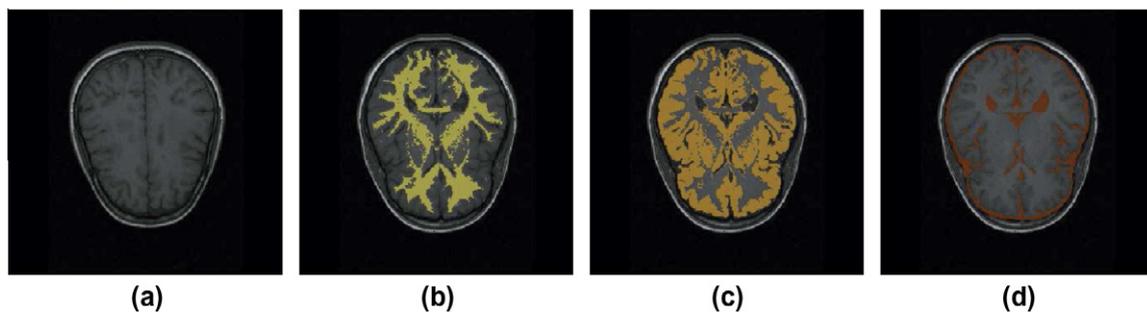


Fig. 4. A brain MRI slice and the obtained masks of (b) white matter, (c) grey matter, and (d) cerebro-spinal fluid.

a registration step. The advantage of atlas-based approaches is that spatial information is inherently used, although registration is also a challenging task. On the other hand, training-based approaches allow to use real characteristics of the tissues and the lesions, but spatial information has to be imposed in a further step since it is not included in the training process. We have also seen in Section 3.2 that there is a group of techniques which are unsupervised and therefore do not depend on a training step, being more generalised algorithms. This group of unsupervised techniques has been subdivided into two different strategies according to the use of the tissue information. The advantage of using tissue information is that may help in localising the lesions. However, the correct segmentation of the tissue is critical in these approaches. On the other hand, defining rules according to the lesion features allows to identify special lesions, although the rules may change with the used modality and scanning machine.

An example of supervised MS lesion segmentation by using atlas registration is the method of Souplet et al. [94], which used the EM algorithm to maximise the log-likelihood between the MRI data (T1-w, T2-w and FLAIR images) and a Gaussian model of 10 classes: WM, GM, CSF, 6 GM/CSF partial volume classes and one outlier class representing mostly the vessels. In this approach, the probability of each voxel to belong to the different classes is first initialised thanks to the a priori registration with the atlas. After the initialisation process, the two steps of the EM algorithm are iterated. In the maximisation step, the parameters of each class (mean and covariance) are computed from the voxel intensities and their probabilities of belonging to the different classes. The parameters of the partial volume classes are computed as a proportion of the pure tissue parameters, while the parameters of the outlier class are obtained as a fraction of the CSF parameters. Therefore, in the expectation step, the probability of each voxel to belong to the different classes is updated using the Gaussian function with the class parameters and the atlas values as prior probabilities. Finally, when the algorithm converges, a bi-dimensional Gaussian distribution is estimated for each class including GM, WM, CSF and partial volumes as well as all the voxel probabilities related to those distributions. Fig. 4 shows an example of the obtained WM, GM, and CSF masks using this algorithm. After the EM segmentation, the obtained GM mask is used to compute the mean and standard deviation values of this region in the FLAIR image which in turn are used to determine a threshold that allows the lesion segmentation. The FLAIR image contrast is first enhanced using morphological operations and then thresholded for obtaining an initial lesion mask. In order to reduce false positives, this mask is further refined using the WM mask with the cavities filled (representing a “healthy WM” mask) and the CSF and WM masks coming from the EM. Therefore, the lesion voxels are defined as those that are present in the “healthy WM” but not present in the real WM nor CSF masks.

The approach of Subbanna et al. [96] is a clear example of a supervised framework for identifying MS lesions in multi-channel MRI using manually segmented images. These manual segmentations are used to extract intensity histograms of both tissue and lesions using the intensities from PD-w, T1-w and T2-w images. A multivariate Gaussian distribution is then fitted to those histograms. In the case of lesions, two different Gaussians are estimated: black holes (visible on both T1-w and T2-w images) and other lesions (only visible in T2-w images). These intensity distributions are then used in a MRF optimisation scheme, where the energy of a given segmentation is optimised using a simulated annealing approach. This energy, is defined by three different terms: a data term which is computed using the voxel intensities and the intensity distributions previously computed; a gradient term which is computed from the intensity distributions and the neighbouring voxel intensities; and finally a weighting term which is defined as the number of voxels that belong to a given class.

García-Lorenzo et al. [41] proposed an unsupervised method to classify each voxel of the brain as belonging to one of the four classes: MS lesions, WM, GM and CSF, using only the intensities from PD-w, T1-w and T2-w images. Firstly, the image intensities are modelled with a GMM, where each Gaussian represents one of the normal appearing brain tissues (GM, WM and CSF). In order to obtain the maximum likelihood estimation of the distribution parameters for each tissue (mean, covariance matrix and mixing proportions), the EM algorithm is used. In this approach the authors proposed two contributions to improve the initialisation step of the EM and the sensibility to outliers when estimating the distribution parameters. First of all, instead of using an atlas as initialisation, several random initialisations of the distribution parameters are used, iterating the algorithm 50 steps using only the T1-w image. Afterwards, the classification with the highest likelihood is used to create a histogram for each tissue on each image to obtain its mean and variance. The covariance matrix for each tissue is then defined as a diagonal matrix with the variances of each image. In order to increase the robustness to outliers, the trimmed likelihood estimator is used when computing the tissue parameters. Instead of using all of the voxels, a fraction of them with the

lowest probability is rejected. In other words, the likelihood is only computed with the voxels most likely to belong to the model. Once the tissues have been segmented, the outliers with a Mahalanobis distance higher than a certain threshold are considered as possible lesions. This initial estimate is subsequently refined using heuristic rules. For instance, lesions should be hyperintense in PD-w and T2-w images, lesions smaller than a certain area (i.e. 9 mm^3) should be discarded, and lesions should be contiguous to WM and not to the brain border.

With a different strategy, Datta et al. [30] proposed an unsupervised method which does not rely on a tissue segmentation for identification and quantification of gadolinium enhancements using FLAIR images together with pre- and postcontrast T1-w images. The enhanced lesions on the postcontrast images are identified as regional maxima using an iterative elementary geodesic morphological reconstruction. A regional maximum on an image is defined as the group of connected voxels with signal intensities greater than the immediate neighbours. The original postcontrast T1-w image is eroded with a three-dimensional structuring element to obtain a T1'-w image. Afterwards, elementary geodesic dilation consisting on first dilating the image with the structuring element, followed by the point-wise minimum with the T1-w image is applied to the T1'-w image. This process is iterated until no change in the morphologically reconstructed image is found. Afterwards, regional maxima are obtained subtracting this image to the original postcontrast T1-w image. These hyperintense areas also include contrast passing within the vascular system and certain structures that lack a blood-brain barrier. In order to reduce these false positives lesions are also segmented in FLAIR using the method proposed by Sajja et al. [78]. Those postcontrast maxima that are not correlated with a FLAIR lesion are discarded. Moreover, the remaining false positives may be also discarded using a change ratio map obtained from the subtraction of pre- and postcontrast T1-w images. Finally, the lesion boundaries are refined using a fuzzy connectivity approach proposed by Udupa et al. [102].

4. Results

MS lesion segmentation approaches are usually evaluated using different quantitative measures and both synthetic and real MRI volumes. The most common data sets used in the works analysed and the typical measures computed for the evaluation are described in this section. Finally, a comparison and discussion of the results presented by the different approaches, highlighting the most interesting aspects, is also presented.

4.1. MS databases

As mentioned in the introduction, one of the main difficulties when comparing MS lesion segmentation approaches is the lack of a common database (with ground truth data) for training and testing the algorithms. Fortunately, this is becoming less of an issue with the appearance of new databases designed to meet this particular objective.

Synthetic databases are useful as an initial framework for testing the various approaches to segmentation. BrainWeb [26] is becoming a standard synthetic database for evaluating both tissue classification and MS lesion segmentation algorithms. BrainWeb contains simulated T1-w, T2-w and PD-w brain MRI data based on two anatomical models: (1) a normal brain and (2) a brain containing MS lesions. Moreover, this database provides different models according to parameters such as slice thicknesses, noise levels and levels of non-uniformity intensity. These simulated data sets are available in three orthogonal views – axial, sagittal, and coronal – although the majority of algorithms use only the axial view.

Even though synthetic data sets are useful for an initial evaluation, they are not as challenging as real data sets. In most cases, algorithms that are correctly tuned for synthetic data may not be successfully applied to a real in vivo acquired volume. Therefore, segmentation algorithms have to be tuned and tested with MRI volumes of real patients. A review of the work of recent years has highlighted a noticeable lack of a public MS MRI databases supplied by hospitals. However, the recent MS Lesion Segmentation Challenge [95] has provided a common framework for evaluating these algorithms, and for making comparisons among them. The MRI data for this competition was acquired separately by the Children's Hospital Boston (CHB) and the University of North Carolina (UNC). In total, 53 brain MRI volumes that were randomised into three different groups: 20 training MRI volumes (including ground truth – see Fig. 5 for examples of manual lesion segmentations); 25 testing volumes (without ground truth) that were downloadable before the contest for training/validating the different algorithms; and eight testing volumes used for the real contest from which the different participants extracted their on-site results.

4.2. Evaluation measures

The results are evaluated in different ways in the reviewed papers. However, all the measures are based on comparing the result of automated segmentations with the ground truth, which is usually annotated by an expert. In order to avoid intra- and inter-observer variability, segmentations from more than one expert should be used, since this provides a more consistent ground truth. Strategies such as the STAPLE algorithm [107] allows annotations from different experts to be fused. This is an important issue specially when considering the small volume of each lesion which may produce significant differences in the evaluation measures. Fig. 5 shows an example of large discrepancy between two different expert annotations of the same patient (inter-rater variability).

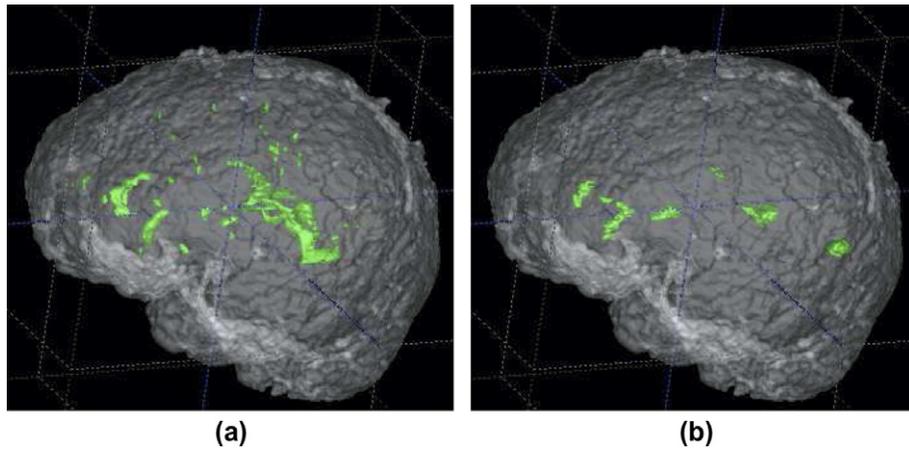


Fig. 5. Generated 3D volume with MS lesions segmented by two different experts showing a large inter-rater variability. Note here the importance of using more than one manual annotation when evaluating the automatic algorithms (example extracted from the MICCAI challenge).

Table 3

Common measures used in the evaluation of MS lesion segmentation methods.

Name	Computation	
Hard measure	Accuracy	$\frac{ TN + TP }{ TN + TP + FP + FN }$
	Percentage agreement	
	Dice similarity coefficient (DSC)	$\frac{2 \times TP }{2 \times TP + FP + FN }$
	Error rate	$\frac{ FP + FN }{ FP + FN + TP + TN }$
	Sensitivity	$\frac{ TP }{ TP + FN }$
	Overlap fraction	
	Percentage of correct estimation	
	True positive fraction/rate	
	Specificity	$\frac{ TN }{ TN + FP }$
	True negative fraction/rate	
	False positive fraction/rate	$1 - \text{Specificity}$
	Under estimation fraction (UEF)	$\frac{ FN }{ TN + FN }$
	False negative volume fraction	
	Over estimation fraction (OEF)	$\frac{ FP }{ TN + FP }$
	Extra fraction	
	Overlap objects fraction	$\frac{N_{obj}(TP)}{N_{obj}(Ref)}$
Probabilistic	Probabilistic similarity index	$\frac{2 \times \sum P_{x,gs-1}}{\sum 1_{x,gs-1} + \sum P_x}$
	Probabilistic overlap fraction	$\frac{\sum P_{x,gs-1}}{\sum 1_{x,gs-1}}$
	Probabilistic extra fraction	$\frac{\sum P_{x,gs-0}}{\sum 1_{x,gs-1}}$

Automated segmentations and ground truth can be compared by either comparing each voxel in each lesion (voxel-to-voxel), or using the whole detected lesion (lesion-to-lesion). Note that in both cases, voxels and lesions can be classified as a true positive (TP), false positive (FP), true negative (TN) or false negative (FN). Obviously, the objective is to obtain the maximum number of TPs and TNs, and at the same time reduce the number of FPs and FNs. However, in practice, one should find the best trade-off between these values, since increasing the number of TPs usually increases the number of FPs, while reducing the number of FNs also reduces the number of TNs. In fact, there is permanent debate about the best option: should we to obtain more TPs or reduce FNs? One could argue that it is preferable to reduce the FNs at the expense of increasing FPs. However, increasing the number of FPs also leads to reduced confidence among neuroradiologists in computerised tools.

Table 3 summarises the most common measures used to evaluate the MS lesion segmentation algorithms. A distinction between two main groups has been done: those that evaluate a “hard” result (i.e. each voxel is assigned to only one tissue type) and those that provide a probabilistic result (i.e. each voxel has a membership value for belonging to the different tissue types). However, one should notice that all these measures are highly related. For instance, the probabilistic similarity index (PSI) is equivalent to the Dice similarity coefficient (DSC) if the final segmentation is binary. Note also that different researchers used the same evaluation measure but under different names.

Table 4

Summary of the results presented in the articles analysed. For the simulated BrainWeb database, the noise and bias field of the tested volume ($n \times b \times f$ means noise x and bias field y) is included. For the real data the number of slices (s) or volumes (v) and the origin of the database are shown. The two results for Van Leemput et al. are obtained by using the segmentation by two different experts as a ground-truth, while Wu et al. distinguish the results between T2 lesions (T2) and black holes (BH). DSC stands for Dice similarity coefficient and the $m \sim n$ means that there were between m and n slices.

Article	Synthetic	Real (database)	Measures	Results
Kamber et al. [55]	–	12 × 56 s (Montreal Neurological Inst.)	Error rate	0.02–0.04
Goldberg et al. [46]	–	14 × 10 s (Sheba Medical Center)	Sensitivity Specificity	0.87 0.96
Bedell and Narayana [16]	–	13 v (Univ. of Texas Medical School at Houston)	Qualitative	–
Bourdraa et al. [21]	–	10 × 22 s (Hôpital d'Antiquaille)	Sensitivity	0.65
Alfano et al. [4]	–	84 × 16 s (Univ. Federico II)	Sensitivity	0.81
Van Leemput et al. [62]	–	50 v (BIOMORPH project)	DSC	0.47 & 0.58
Zijdenbos et al. [113]	–	29 v (Montreal Neurological Inst.)	DSC	0.60
He and Narayana [51]	–	5 v (Univ. of Texas Medical School at Houston)	Kappa	0.9
Anbeek et al. [7]	–	20 × 38 s (Univ. Medical Center Utrecht)	DSC Sensitivity OEF	0.80 0.79 0.19
Anbeek et al. [6]	–	10 × 5 s (Univ. Medical Center Utrecht)	DSC Sensitivity Specificity	0.808 0.815 0.999
Datta et al. [31]	–	14 v (Sanjay Gandhi Post-Graduate Inst. of Medical Sciences)	DSC Sensitivity OEF UEF	0.73 ± 0.11 0.72 ± 0.13 0.27 ± 0.21 0.28 ± 0.13
Sajja et al. [78]	–	23 v (Sanjay Gandhi Post-Graduate Inst. of Medical Sciences)	DSC Sensitivity OEF UEF	0.78 ± 0.12 0.88 ± 0.13 0.38 ± 0.27 0.11 ± 0.13
Wu et al. [109]	–	6 × 2 v (Slotervaart Hospital)	Sensitivity Specificity	0.70 (T2) 0.62 (BH) 0.98 (T2) 0.99 (BH)
Datta et al. [30]	–	22 v (Sanjay Gandhi Post-Graduate Inst. of Medical Sciences)	DSC Sensitivity OEF UEF	0.76 ± 0.18 0.74 ± 0.22 0.25 ± 0.62 0.26 ± 0.22
Khayati et al. [56]	–	20 × 12 ~ 20 s (Koorosh Diagnostics and Medical Imaging Center)	DSC Sensitivity OEF	0.7504 0.7402 0.2303
Garcia-Lorenzo et al. [42]	–	7 v (McConnel Brain Imaging Centre)	DSC	0.55 ± 0.05
Garcia-Lorenzo et al. [44]	–	3 v (Magnetic Resonance and Image Analysis Research Centre)	DSC	0.56
Tomas and Warfield [100]	–	9 v (Boston Children's Hospital)	Sensitivity FPR	0.65 ± 0.29 0.71 ± 0.21
Subbanna et al. [96]	–	10 v (Montreal Neurological Inst.)	DSC FPR FNR	0.71 0.00 0.10
Akselrod-Ballin et al. [3]	–	25 + 16 v × 24 s (Scientific Institute Ospedale San Raffaele)	DSC Sensitivity Specificity Accuracy	0.53 ± 0.1 0.55 ± 0.13 0.98 ± 0.01 0.97 ± 0.01
Shiee et al. [88]	– Brainweb (n3bf0) Brainweb (n9bf0)	10 v (National Multiple Sclerosis Society) – –	DSC DSC DSC	0.63 0.77 0.73
Garcia-Lorenzo et al. [42]	Brainweb (n3bf0) Brainweb (n3bf20) Brainweb (n3bf40)	– – –	DSC	0.87 0.85 0.63
Shiee et al. [86]	Brainweb (n3bf0) Brainweb (n9bf0)	– –	DSC	0.720 0.591
Shiee et al. [88]	Brainweb (n3bf0)	–	DSC	0.812

Other measures used that are not related to the ones described in the table are those based on distance measures (in voxels or in millimetres). The aim of these measures is to evaluate how far the boundaries of an obtained lesion segmentation are from those of the real one. Although in general, these measures are not common in most of the analysed works, they were used in the MS Lesion Segmentation Challenge [95].

In conclusion, although DSC has become something of a standard measure for evaluating MS lesion segmentation methods, none of the measures are perfect for this purpose. In fact, as stated by Cárdenes et al. [24], various different measures (i.e. measures based on intensity values, distances and connectivity) should be combined to obtain a more objective and reliable assessment.

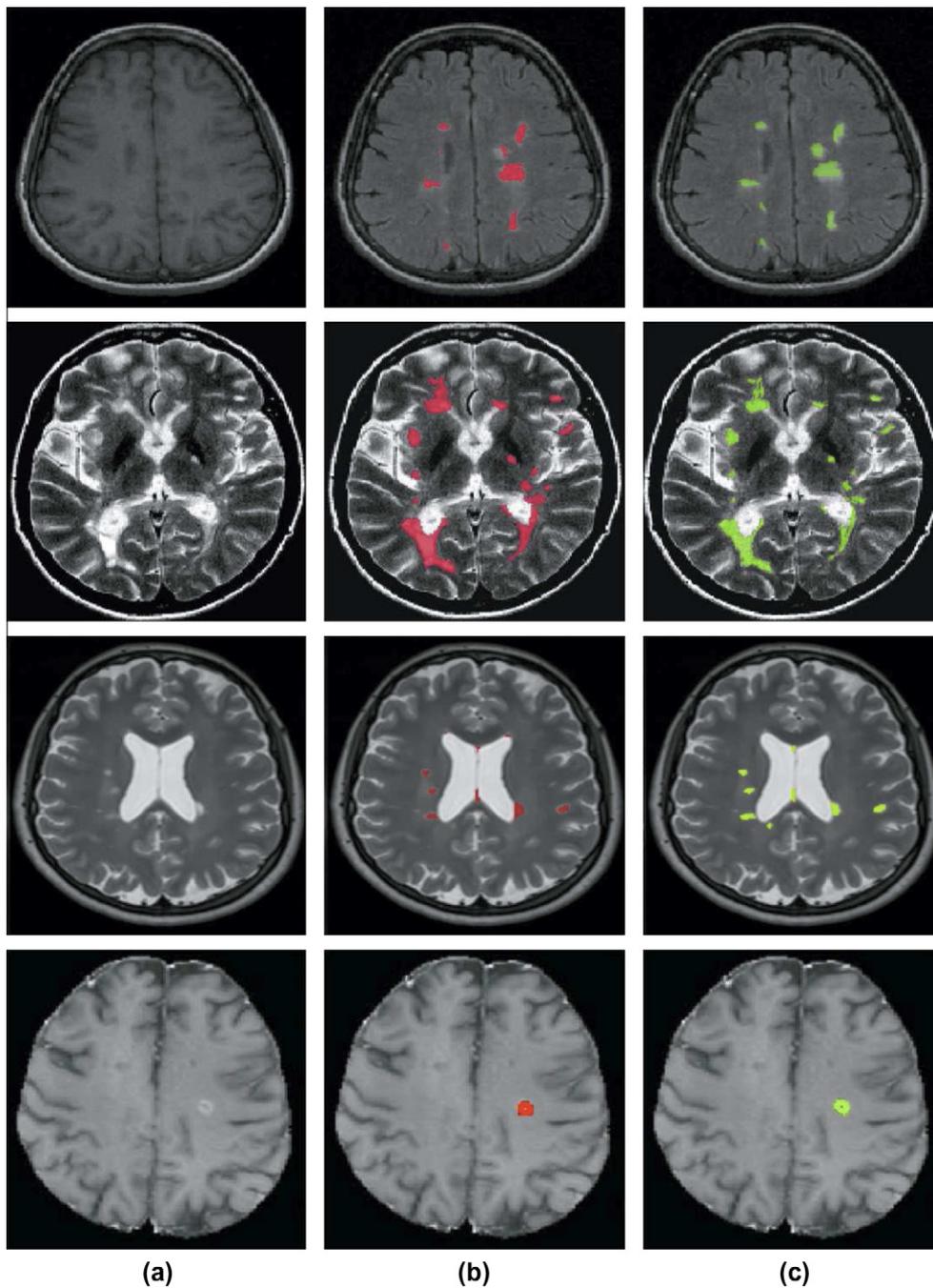


Fig. 6. MS lesion segmentation examples. Each row shows the result of a representative algorithm of each reviewed strategy. From top to bottom: supervised approach based on atlas [94], supervised approach based on training [96], unsupervised approach based on tissue [41], and unsupervised approach based on lesion [30]. (a) Shows the original images, (b) shows the results of the automatic MS lesion segmentation algorithms, and (c) the corresponding ground-truth annotations.

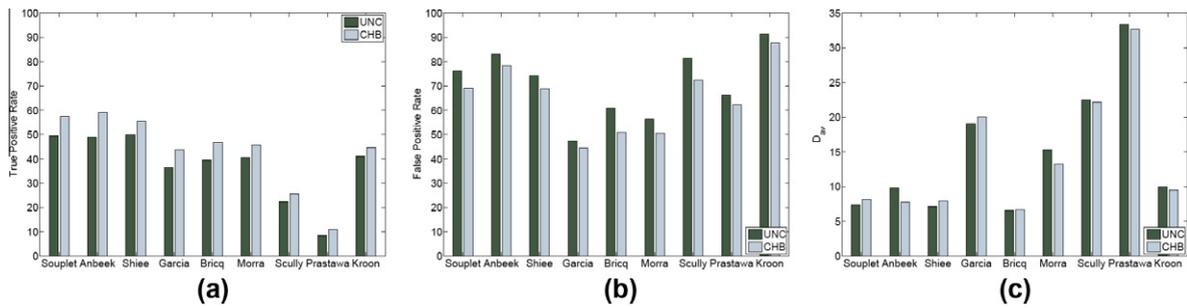


Fig. 7. Results extracted from the works presented in the 2008 MS Challenge. Each plot details the results when using ground truth provided by either the UNC or the CHB experts. (a) Shows the true positive rate per lesion, (b) the false positive rate per lesion and (c) the average distance between obtained and ground truth segmentations. The results are ordered according to the final ranking in the on-site testing competition.

4.3. Analysis of the results

This section provides a qualitative comparison of the results obtained by the approaches analysed. Table 4 summarises the data, the evaluation measures and the results obtained. As already mentioned, a quantitative evaluation among these approaches is a difficult task due to the variability in the data sets and in the measures used.

Table 4 clearly shows that the most commonly-used evaluation measures are the DSC and sensitivity (also known as the Overlap Fraction). The results obtained using real data range from 0.47 to 0.808 for the Dice coefficient. The highest Dice coefficients reported involving a large amount of data were obtained by Sajja et al. [78], who used 23 volumes from which they obtained a mean Dice coefficient of 0.78. The work presented by Van Leemput et al. [62] provides an example of how the variability of the ground truth affects the results obtained. Using the same automatically segmented data and comparing them with two different expert annotations, the DSC varies more than 10% (from 0.47 to 0.58). These results illustrate the usefulness of algorithms that combine the ground truth.

Several works also use sensitivity and specificity as evaluation measures (see Table 4). Observe that the specificity reported by the different methods is always close to 1. This is because this measure evaluates the ratio between the number of voxels correctly classified as healthy divided by the number of voxels automatically classified as healthy. Therefore, considering that lesions are small spots within the whole volume, the specificity value always tends to be close to 1.

The last four methods shown in Table 4 used the simulated BrainWeb phantom (a synthetic model) to provide quantitative results. As already mentioned, the BrainWeb helps to compare the performance of different approaches more effectively. For example, each of these four methods included simulations with 3% added Gaussian noise and with zero bias field. The results show that the approach of García-Lorenzo et al. [42] clearly outperforms the others, obtaining a DSC of 0.87. The methods of Freifeld et al. [38] and the two by Shiee et al. [86,88] obtained DSC values of 0.77, 0.72 and 0.81 respectively under the same conditions. Furthermore, García-Lorenzo et al. [42] also report the difference between using simulated data and using real MRI data when evaluating algorithms. Notice that when evaluating the same approach but using seven real volumes from the McConnel Brain Imaging Centre, the DSC drastically dropped to 0.55.

Qualitative examples showing the result of automatic MS lesion segmentation algorithms are illustrated in Fig. 6. This figure provides visual examples from the representative methods of each strategy described in Section 3.3. The first row shows an example of the supervised approach based on atlas proposed by Souplet et al. [94], the second row corresponds to the supervised approach based on training from Subanna et al. [96], the third row illustrates the results from the unsupervised segmentation approach based on tissue of García-Lorenzo et al. [41], while the last row shows the result of the unsupervised approach based on lesion properties from Datta et al. [30]. Note that this last algorithm was specifically designed to detect gadolinium-enhanced lesions in post-contrast T1 images.

Finally, the quantitative results published during the 2008 MS Challenge⁴ [95] are presented in this section. The reported results are summarised in Fig. 7, which consists of three plots showing the results obtained when using three different evaluation measures: (a) the true positive rate (per lesion), (b) the false positive rate (per lesion) and (c) the average symmetric surface distance, which measures how far away the correctly segmented lesions are from the ground truth. The results in the plots have been ordered according to the final ranking in the on-site testing competition (from left to right). Notice that these plots illustrate slight differences in the results when using the ground truth from the UNC experts or from the CHB centre. Slightly better performances were obtained when using the annotations from the CHB expert during the training. As suggested by Styner et al. [95], this may be due to the fact that the UNC ground truth was obtained from two experts, while the CHB ground truth was obtained from only one expert. In fact, the training and testing processes for the UNC can be performed using ground truth from different experts while this is not possible at the CHB centre. In general, Fig. 7(a) shows us, that there is room for improvement regarding the true positive rates. The best performances were obtained by Souplet et al. [92], Anbeek et al. [5] and Shiee et al. [87], each of whom had true positives rates of around 60%. There is also room for improvement regarding the false positive

⁴ All this information can be found in: <http://grand-challenge2008.bigr.nl/> and <http://www.midajournal.org/browse/publication/638>.

rates (see Fig. 7(b)), the best results being obtained by García-Lorenzo et al. [43], Bricq et al. [22] and Morra et al. [67], each with false positive rates of around 50%. This is to say, from each pair of marked lesions, only one was correctly placed. However, as stated in [78], this problem can be successfully overcome by applying an automated post-processing step for false positive reduction. Finally, Fig. 7(c) shows the average symmetric surface distance. Using this measure based on the distance from the ground truth to the lesion surface contour, the best results were obtained by Souplet et al. [92], Anbeek et al. [5], Shiee et al. [87], Bricq et al. [22] and Kroon et al. [57], with results of between 5 and 10 mm. Considering that the nominal diameter of an MS lesion is about 7 mm [106], the results obtained, although promising, are still far from the requirements of a perfect automated volumetric tool.

In addition to the results obtained from the manual segmentations, the MS Challenge organisers computed a composite segmentation using the well-known STAPLE algorithm [107]. Specifically, the input for STAPLE included all the manual segmentations, as well as the segmentations provided by the workshop participants. Hence, it represented a composite of two human experts and nine automated segmentation methods. With respect to this evaluation experiment, the best sensitivities (which were from the works of Anbeek et al., Morra et al. and Kroon et al.) were around 0.5 while the specificities were close to 1, showing the ability of these algorithms to identify lesions correctly. These experiments using the STAPLE segmentation illustrate the improvement in reducing the false positive rate. In summary, the quantitative evaluation performed in the MS segmentation Challenge 2008 [95] has revealed both the challenges facing the participants as well as the need to develop new approaches to MS lesion segmentation.

5. Discussion

As seen in previous sections, the most widely-used feature in all segmentation methods is voxel intensity, which is commonly employed with a multi-channel approach. In addition, features based on modelling the voxel neighbourhood are also used in some approaches to introduce (local) spatial information to the algorithms. Regarding the modalities, T1-w images are widely used for the tissue segmentation and also for the black holes and enhanced lesion segmentation. T2-w and PD-w images are typically used for detecting MS lesions. However, the major drawback of these images is the similarity in the intensities of lesions and CSF. Due to this similarity, the discrimination between ventricles and lesions may be difficult, especially when they are connected. Some approaches perform another segmentation step to solve this problem. FLAIR images also provide good discrimination between lesions and healthy tissue but, as some authors have pointed out, they have problems when dealing with sub-cortical structures.

In regard to the segmentation algorithms shown in Table 1, note that the vast majority belong to the clustering family [39]. However, clustering algorithms do not naturally deal with the spatial information needed to obtain proper segmentations. This can be introduced using Markov Random Fields or Fuzzy Connectedness. The most widely-used clustering approaches are based on the FCM algorithm [18] and the EM algorithm [34]. The FCM is usually used to group the different tissues of the brain into three (WM, GM, CSF) or five (WM, GM, WM + GM, CSF, GM + CSF) different classes [11]. The aim of using five classes is to take partial volume effects into account, allowing a voxel to be composed of more than one tissue type. The EM algorithm also allows different models to be used for different tissues, which is indeed very useful since WM and GM can be assumed to follow Gaussian distributions while CSF does not follow any known distribution [78]. Despite the fact that EM methods are usually more accurate in the presence of noise than those of the C-Means family, they share with them the possibility of converging to a local maximum (or minimum). In a way similar to the EM, the AMM algorithm also allows distributions to be estimated. As argued by Khayati et al. [56] the advantages of the AMM over the EM are the non-requirement of initial knowledge for the number of terms and the initial estimate for the parameters. Moreover, as all the data are not used simultaneously to update the parameter estimation, less computational time is required. Although the clustering approaches introduced above are unsupervised algorithms, there are also supervised approaches which use these clustering algorithms to perform the classification. See for instance the works of Shiee et al. [86,87] and of Souplet et al. [92].

While iterative approaches are inherently unsupervised approaches (even though they may use a priori information), supervised approaches have also been proposed for the purpose of MS lesion segmentation. Most methods rely on pattern recognition techniques to detect voxels that are either outliers to the tissue models or are similar to the lesion model (derived from a training set). In general, these methods tend to be more robust if a good training set is provided. However, collecting enough representative samples for a fully automated and accurate system may be difficult due to differences between scanning machines, patient anatomy or expert reading. A large variety of supervised classifiers such as kNN, ANN or Bayes are used for the MS lesion segmentation problem.

Atlas registration is another way of tackling MS lesion segmentation in brain MR images. These approaches are useful since the lesion intensities often overlap with the intensities of other structures in the brain. Hence, atlases might provide valuable contextual information to eliminate possible false positives. As a drawback, the registration process is a computationally intensive procedure. Moreover, the results obtained by these methods are affected by the physiological variability of each subject and may lead to erroneous results in the case of diseased brains. This is because atlases are based on normal brains and lesions are very variable in size, shape and location, making it difficult to construct atlases from diseased brains (although this has been analysed in [57]).

An analogous problem to atlas registration is the temporal registration of MRI volumes. This procedure is necessary to track and compare the atrophy of the WM and GM, as well as to monitor the evolution of MS lesions. However, this

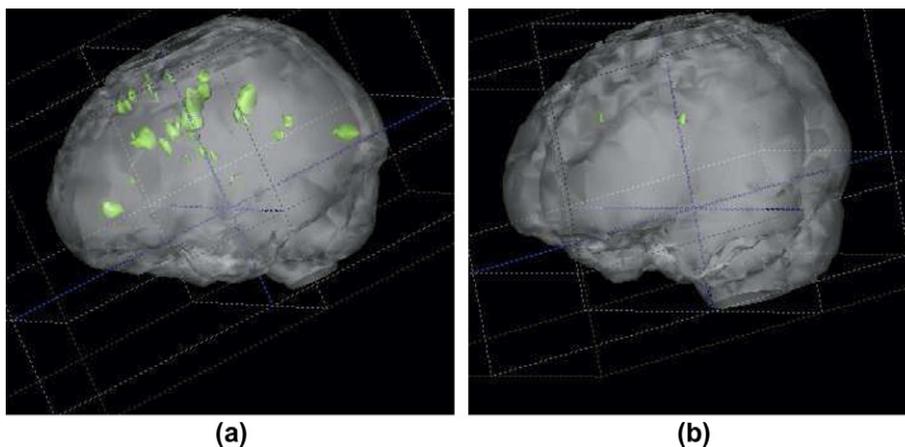


Fig. 8. Generated 3D lesion volume for two different patients (a and b) with different lesion load.

registration can also be used to detect lesions, which are defined as those regions with apparent local volume variation [99,74]. Note that a registration step will be also needed for including information coming from other modality images [14].

Regarding the evaluation measures, notice that the Dice coefficient and sensitivity are the two most commonly-used measures for lesion segmentation. Both measures take the number of true positives and false negatives into account, while the main difference is that the DSC also considers the number of false positives. On the other hand, specificity is usually high because this measure takes the number of true negatives into account and since lesions are small spots in the image, specificity has values close to 1 (the accuracy measure behaves in the same way as well). We think that the best way to show the number of false positives is using the over estimation fraction (OEF) or using both the sensitivity and the Dice coefficient.

Analysing the results, note that those methods that accomplished DSC values above 0.7 (which is considered a good segmentation result) use prior knowledge to help the detection and delimitation of the lesions. This prior knowledge is mostly defined by a set of manual segmentations that can aid to model a tissue intensity distribution or constrain it using either a classification scheme [7,6,96] or a registration based [31,56] approach. However, this information may also be encoded as a set of pre-defined rules [31] such as that enhancing lesions should appear brighter in T1-w images with gadolinium enhancement than in T1-w images without enhancement.

We have also pointed out the difficulty of making quantitatively comparisons of results that are obtained using different databases, mainly because there is not a standardised evaluation measure nor a common public database. Note that even when using the same evaluation measures, cases with significant different lesion load like the ones shown in Fig. 8, can bias the results. Observe that it is likely to have a higher false positive rate in (b) than in (a) due to the small lesion volume of that patient. In respect to this difficulty of performing evaluations, the MS Challenge has helped to create a data set that can become a standard for evaluating the automated segmentation of MS lesions. In this competition, supervised strategies obtained the best results with the best ranked method being an atlas-based strategy presented by Souplet et al. [92]. This method uses a tissue based strategy aided by a probabilistic atlas that provides prior knowledge during the tissue segmentation step. As pointed out in Section 3.3, the advantage of the supervised algorithms is that they can automatically learn the characteristics of both normal tissue and lesions, while the algorithms which follow an atlas-based approach incorporate also spatial information to perform the segmentation task (as is done by Souplet et al. [92]). Nowadays, the website for the challenge is still available and there is the possibility to submit new results in order to evaluate the performance of new techniques.

5.1. Similar diseases

From a general point of view, the problem of segmenting MS lesions can be closely related to the detection of other diseases in brain MR images. Damaged brains may contain focal tissue lesions that represent lesions produced by the loss and inflammation of tissue, like multiple sclerosis and stroke or large space-occupying lesions like tumours.

As stated in Section 3, most MS lesion segmentation methods are supervised approaches relying on the use of prior knowledge to segment brain tissues and consider then lesions as outliers. Similar strategies have been used to detect other diseases. For instance, Shen et al. [84,85] proposed to automatically detect stroke lesions by comparing voxel-to-voxel the inconsistency between the result of applying an unsupervised tissue segmentation of the patient scans and the probability priors obtained within an atlas. On the other hand, Seghier et al. [80] extended the atlas-based tissue segmentation approach of Ashburner and Friston [10] to detect also stroke lesions. Admiraal-Behloul et al. [1] and Lao et al. [60] detect the WM signal abnormalities that frequently occur in vascular disease using a supervised training-based approach.

Space-occupying brain lesions induce large deformations and lack of clear detail of anatomy because of infiltration and edema [69]. The problem in this specific disease is different to the one tackled in this review. Effectively, most of the methods assume a semiautomatic segmentation of the tumour to facilitate the registration of the damaged brain to an atlas to find how the different tissues and structures of the brain are affected by the presence of the lesion [59,33,12,70,111,110]. In contrast, Moon et al. [66] extended the tissue segmentation approach of Van Leemput et al. [104] for automatically detect brain tumours. Hence, in a similar way they were dealing with the segmentation of MS lesions [103], the authors used the same EM approach but extending the number of classes with the tumour class and introducing the prior spatial probabilities of the tumour location. In a similar way, Prastawa et al. [71] proposed to detect tumour and edema by extending their MS lesion segmentation approach [72], which determines the diseased regions as outliers to the model.

6. Future trends

In this paper several strategies to perform the automated MS lesion segmentation have been reviewed. For instance, atlas-based segmentation is becoming a standard paradigm for exploiting spatial prior knowledge in MR brain image segmentation. Atlases provide helpful information about anatomy and its variability. Several works already showed that atlas selection, or a group of atlases, is crucial for improving the segmentation results. Thus, research interests are currently on the construction of temporal (also called 4D) atlases with the goal of capturing both inter-subject and longitudinal anatomical variability [58]. We think that novel atlas will aim at encode the anatomical differences locally in order to overcome the limitation of most current atlas where all voxels in all atlas are considered equally important.

The development of more powerful computers might also allow the introduction of region-based methods to deal with brain MRI segmentation. These methods, which inherently take spatial information into account, are useful for obtaining more compact segmentations than those provided by clustering algorithms. Moreover, this review has shown that few algorithms have used edge information to perform or refine the segmentation. Proposals of new algorithms based on the integration of both strategies might be useful to improve the MS lesion segmentation. Furthermore, using a model combining prior information (intensity and shape information) and a matching procedure based on an atlas approach appears to be a good option that would integrate and include the advantages of both strategies.

Regarding the imaging modalities, the analysis of the results has shown that FLAIR provides a good discrimination between lesions and healthy tissue and is used in numerous approaches to perform the automated lesion segmentation. Recent reports have also stated that 3D FLAIR imaging reduces the artifacts and provides an excellent signal-to-noise ratio compared with 2D FLAIR images. Notice that 3D FLAIR images provide 3D volume data with isotropic information and minimize the partial volume effect between small lesions and surrounding tissue. Therefore, the use of 3D FLAIR imaging may improve the estimates of the WM and GM as well as the MS lesions.

While there are many articles focusing on the lesion segmentation problem, most of them do not incorporate an automated method for interpreting the lesion change (or evolution). This is still an open issue in the research community. General problems associated to these lesion detection and change detection techniques are the lesion shape, which is usually ambiguous and has ill-defined boundaries, and the lesion position, since the lesion can appear or disappear arbitrarily and may shrink and enlarge overtime. Moreover, the effect of a lesion does not always appear as an intensity change on the tissue where it locates (the so-called *tissue transformation*), but can also influence the appearance of surrounding tissues (known as the *mass effect*) [99]. Thus, observing the lesion evolution without change in intensity but with displacement on the surrounding tissues (deformation) is more difficult. In real cases, both tissue transformation (changes in intensity) and tissue deformation generally occur. Hence, the mass effect of the lesion should also be taken into account in order to define a precise lesion evolution. Furthermore, detecting real image changes is a hard work due to noise and residual artifacts in the MR images, and also due to the fact that images of a patient at different times are not always directly comparable due to the patient movement. Hence, in many cases, a robust image registration algorithm must be used.

The most common approach for the change detection in serial imaging is the visual inspection performed by experts. The processed data such as already detected lesions are presented to the radiologists in order to render a decision with respect to the lesion load change. Although this is a very subjective method, some improvements can be used in order to reduce the misinterpretations made by the expert. For instance, statistical change detection techniques may be used in order to reduce false positives in the subtracted images. In a different way, approaches may use already segmented lesions in order to quantify the lesion evolution by means of their volume changes. This volumetric quantification approaches have already been proven to be useful for detecting positive and negative disease activity [108]. Note that this quantification process can be either done by subtracting individual lesion volumes or by subtracting total lesion volume between the time-series images. However, notice that when computing the total MS lesion volume of a patient it is possible that some lesions enlarge while others shrink at the same time. Therefore, this quantification process may not detect a change in lesion volume even if there are growing and shrinking lesions. Therefore, comparing lesion volumes individually may be a more precise way of doing the quantification. Furthermore, when using volumetric measures one should note that the process relies on the results of a previous segmentation method which may not provide the desired result introducing therefore errors in the quantification.

7. Conclusions

This paper has reviewed the automated approaches for MS lesion segmentation, classifying them according to the strategy used. In addition, the results obtained by these approaches have been summarised and compared, reviewing also the most common data sets and evaluation measures used in this field. We observed that the automated segmentation of different MS lesion types in MRI is a challenging task due to heterogeneous intensity values among the different MR images (enhancing lesions, black holes and hyperintense lesions). Despite recent progress, there is not yet a specific automated lesion segmentation approach robust enough to emerge as a standard for clinical practice. The main reasons for this are the unsatisfactory results they produce, the high computational demand required, or their insufficient generalisation capability. Moreover, most of these approaches are unable to locate MS lesions inside the sub-cortical structures of the brain.

Finally, we would like to point out that there is room for improvement in the automated segmentation of MS lesions in MR images. Advances in new algorithms, as well as new developments in MRI acquisition protocols will without doubt help neuroradiologists to improve the diagnose and follow-up of MS patients, both in clinical studies investigating future MS therapies and in every-day clinical practice. This may save time for the neuroradiologists and provide less subjective measures which will provide better comparisons and analysis of MS disease evolution.

Acknowledgements

We would like to thank Dr. D. García-Lorenzo, Prof. C. Barillot, Prof. P. A. Narayana, and Dr. N.K. Subbanna, for providing the images of their results (Fig. 6). We would like to thank also the reviewers for their critical evaluation of the manuscript. This work has been supported by the Instituto de Salud Carlos III Grant PI09/91018, Grant VALTEC09-1-0025 from the Generalitat de Catalunya, and Grant CEM-Cat 2011 from the Fundació Esclerosi Múltiple. M. Cabezas holds a FI Grant 2011FI-B1 00167.

References

- [1] F. Admiraal-Behloul, D.M.J. van den Heuvel, H. Olofsen, M.J.P. van Osch, J. van der Grond, M.A. van Buchem, J.H.C. Reiber, Fully automatic segmentation of white matter hyperintensities in MR images of the elderly, *NeuroImage* 28 (3) (2005) 607–615.
- [2] L. Aït-Ali, S. Prima, P. Hellier, B. Carsin, G. Edan, C. Barillot, STREM: a robust multidimensional parametric method to segment MS lesions in MRI, in: *Int. Conf. Med. Image Comput. Comput. Assist. Interv.* 2005, pp. 409–416.
- [3] A. Akselrod-Ballin, M. Galun, J. Gomori, M. Filippi, P. Valsasina, R. Basri, A. Brandt, Automatic segmentation and classification of multiple sclerosis in multichannel MRI, *IEEE Trans. Biomed. Eng.* 56 (10) (2009) 2461–2469.
- [4] B. Alfano, A. Brunetti, M. Larobina, M. Quarantelli, E. Tedeschi, A. Ciarmiello, E.M. Covelli, M. Salvatore, Automated segmentation and measurement of global white matter lesion volume in patients with multiple sclerosis, *J. Magn. Reson. Imag.* 12 (6) (2000) 799–807.
- [5] P. Anbeek, K. Vincken, M. Viergever, Automated MS-lesion segmentation by k-nearest neighbor classification, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008, pp. 1–8.
- [6] P. Anbeek, K.L. Vincken, G.S. van Bochove, M.J. van Osch, J. van der Grond, Probabilistic segmentation of brain tissue in MR imaging, *NeuroImage* 27 (4) (2005) 795–804.
- [7] P. Anbeek, K.L. Vincken, M.J. van Osch, R.H. Bisschops, J. van der Grond, Probabilistic segmentation of white matter lesions in MR imaging, *NeuroImage* 21 (3) (2004) 1037–1044.
- [8] Asclepios-SepINRIA. <<http://www.sop.inria.fr/asclepios/software/SepINRIA/>>, 2008 (accessed 04.03.10).
- [9] J. Ashburner, K.J. Friston, Voxel-based morphometry: the methods, *NeuroImage* 11 (6) (2000) 805–821.
- [10] J. Ashburner, K.J. Friston, Unified segmentation, *NeuroImage* 26 (3) (2005) 839–851.
- [11] M. Bach-Cuadra, L. Cammoun, T. Butz, O. Cuisenaire, J.P. Thiran, Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images, *IEEE Trans. Med. Imag.* 24 (12) (2005) 1548–1565.
- [12] M. Bach-Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J.G. Villemure, J.P. Thiran, Atlas-based segmentation of pathological MR brain images using a model of lesion growth, *IEEE Trans. Med. Imag.* 23 (10) (2004) 1301–1314.
- [13] F. Bagnato, N. Jeffries, N. Richert, R. Stone, J. Ohayon, H. McFarland, J. Frank, Evolution of T1 black holes in patients with multiple sclerosis imaged monthly for 4 yr, *Brain* 126 (Pt 8) (2003) 1782–1789.
- [14] S. Banerjee, D.P. Mukherjee, D.D. Majumdar, Fuzzy c-means approach to tissue classification in multimodal medical imaging, *Inf. Sci.* 115 (1–4) (1999) 261–279.
- [15] A. Bardera, M. Feixas, I. Boada, M. Sbert, Image registration by compression, *Inf. Sci.* 180 (7) (2010) 1121–1133.
- [16] B.J. Bedell, P.A. Narayana, Automatic segmentation of gadolinium-enhanced multiple sclerosis lesions, *Magn. Reson. Med.* 39 (6) (1998) 935–940.
- [17] R. Bermel, R. Bakshi, The measurement and clinical relevance of brain atrophy in multiple sclerosis, *Lancet Neurol.* 5 (2) (2006) 158–170.
- [18] J.C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
- [19] J.C. Bezdek, L.O. Hall, L.P. Clarke, Review of MR image segmentation techniques using pattern recognition, *Med. Phys.* 20 (4) (1993) 1033–1048.
- [20] K. Boesen, K. Rehm, K. Schaper, S. Stoltzner, R. Woods, E. Luders, D. Rottenberg, Quantitative comparison of four brain extraction algorithms, *NeuroImage* 22 (3) (2004) 1255–1261.
- [21] A. Boudraa, S.M.R. Dehak, Y. Zhu, C. Pachai, Y. Bao, J. Grimaud, Automated segmentation of multiple sclerosis lesions in multispectral MR imaging using fuzzy clustering, *Comput. Biol. Med.* 30 (1) (2000) 23–40.
- [22] S. Bricq, C. Collet, J. Armspach, Ms lesion segmentation based on hidden Markov chains, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008, pp. 1–9.
- [23] G. Calcagno, A. Staiano, G. Fortunato, V. Brescia-Morra, E. Salvatore, R. Liguori, S. Capone, A. Filla, G. Longo, L. Sacchetti, A multilayer perceptron neural network-based approach for the identification of responsiveness to interferon therapy in multiple sclerosis patients, *Inf. Sci.* 180 (21) (2010) 4153–4163.
- [24] R. Cárdenes, R. de Luis-García, M. Bach-Cuadra, A multidimensional segmentation evaluation for medical image data, *Comput. Methods Prog. Biomed.* 96 (2) (2009) 108–124.
- [25] L.P. Clarke, R.P. Velthuisen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, R.W. Thatcher, M.L. Silbiger, MRI segmentation: methods and applications, *Magn. Reson. Imag.* 13 (3) (1995) 343–368.
- [26] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, Design and construction of a realistic digital brain phantom, *IEEE Trans. Med. Imag.* 17 (3) (1998) 463–468. <<http://www.bic.mni.mcgill.ca/brainweb>>.
- [27] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Trans. Pattern Anal. Machine Intel.* 24 (5) (2002) 603–619.
- [28] A. Compston, A. Coles, Multiple sclerosis, *Lancet* 359 (9313) (2006) 1221–1231.

- [29] S. Datta, P.A. Narayana, Automated brain extraction from T2-weighted magnetic resonance images, *J. Magn. Reson.* 33 (4) (2011) 822–829.
- [30] S. Datta, B.R. Sajja, R. He, R.K. Gupta, J.S. Wolinsky, P.A. Narayana, Segmentation of gadolinium-enhanced lesions on MRI in multiple sclerosis, *J. Magn. Reson. Imag.* 25 (5) (2007) 932–937.
- [31] S. Datta, B.R. Sajja, R. He, J.S. Wolinsky, R.K. Gupta, P.A. Narayana, Segmentation and quantification of black holes in multiple sclerosis, *NeuroImage* 29 (2) (2006) 467–474.
- [32] S. Datta, G. Tao, R. He, J. Wolinsky, P.A. Narayana, Improved cerebellar tissue classification on magnetic resonance images of brain, *J. Magn. Reson.* 29 (5) (2009) 1035–1042.
- [33] B.M. Dawant, S.L. Hartmann, S. Pan, S. Gadamsetty, Brain atlas deformation in the presence of small and large space-occupying tumors, *Comput. Aided Surg.* 7 (1) (2002) 1–10.
- [34] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum-likelihood from incomplete data via EM algorithm, *J. R. Statist. Soc. B* 39 (1) (1977) 1–38.
- [35] T. Dua, P. Rompani, Atlas: Multiple Sclerosis Resources in the World 2008, World Health Organization, 2008.
- [36] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, 2nd ed., John Wiley & Sons, New York, 2001.
- [37] M. Filippi, F. Agosta, Imaging biomarkers in multiple sclerosis, *J. Magn. Reson. Imag.* 31 (4) (2010) 770–788.
- [38] O. Freifeld, H. Greenspan, J. Goldberger, Lesion detection in noisy MR brain images using constrained GMM and active contours, in: *IEEE Int. Symp. Biomed. Imaging*, 2007, pp. 596–599.
- [39] K.S. Fu, J.K. Mui, A survey on image segmentation, *Pattern Recog.* 13 (1981) 3–16.
- [40] D. García-Lorenzo, J. Lecoer, L.D. Arnold, D.L. Collins, C. Barillot, Multiple sclerosis lesion segmentation using an automatic multimodal graph cuts, in: *12th International Conference on Medical Image Computing and Computer Assisted Intervention*, Lecture Notes in Computer Science, vol. 5762, SpringerLink, 2009, pp. 584–591.
- [41] D. García-Lorenzo, S. Prima, D. Arnold, L. Collins, C. Barillot, Trimmed-likelihood estimation for focal lesions and tissue segmentation in multi-sequence MRI for multiple sclerosis, *IEEE Trans. Med. Imag.* 30 (8) (2011) 1455–1467.
- [42] D. García-Lorenzo, S. Prima, D.L. Collins, D.L. Arnold, S.P. Morrissey, C. Barillot, Combining robust expectation maximization and mean shift algorithms for multiple sclerosis brain segmentation, in: *Work. Med. Image Anal. Mult. Scler.* 2008, pp. 82–91.
- [43] D. García-Lorenzo, S. Prima, S. Morrissey, C. Barillot, A robust expectation-maximization algorithm for multiple sclerosis lesion segmentation, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008, pp. 1–9.
- [44] D. García-Lorenzo, S. Prima, L. Parkes, J.C. Ferré, S.P. Morrissey, C. Barillot, The impact of processing workflow in performance of automatic white matter lesion segmentation in multiple sclerosis, in: *Work. Med. Image Anal. Mult. Scler.*, 2008, pp. 104–112.
- [45] Y. Ge, Multiple sclerosis: The role of MR imaging, *Am. J. Neuroradiol.* 27 (6) (2006) 1165–1176.
- [46] D. Goldberg-Zimring, A. Achiron, S. Miron, M. Faibel, H. Azhari, Automated detection and characterization of multiple sclerosis lesions in brain MR images, *Magn. Reson. Imag.* 16 (3) (1998) 311–318.
- [47] H. Greenspan, A. Ruf, J. Goldberger, Constrained Gaussian mixture model framework for automatic segmentation of MR brain images, *IEEE Trans. Med. Imag.* 25 (9) (2006) 1233–1245.
- [48] A.S. Hadi, A. Luceño, Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms, *Comput. Stat. Data Anal.* 25 (3) (1997) 251–272.
- [49] S.W. Hartley, A.I. Scher, E.S.C. Korf, L.R. White, L.J. Launer, Analysis and validation of automated skull stripping tools: A validation study based on 296 MR images from the Honolulu Asia aging study, *NeuroImage* 30 (4) (2006) 1179–1186.
- [50] R.H. Hashemi, W.G. Bradley Jr., C.J. Lisanti, MRI: the basics, third ed., Lippincott Williams & Wilkins, Philadelphia, PA, 2010.
- [51] R. He, P.A. Narayana, Automatic delineation of Gd enhancements on magnetic resonance images in multiple sclerosis, *Med. Phys.* 29 (7) (2002) 1536–1546.
- [52] Z. Hou, A review on MR image intensity inhomogeneity correction, *Int. J. Biomed. Imag.* 2006 (1) (2006) 1–11.
- [53] Z. Hou, S. Huang, Characterization of a sequential pipeline approach to automatic tissue segmentation from brain MR images, *Int. J. Comput. Assist. Radiol. Surg.* 2 (5) (2008) 305–316.
- [54] B. Johnston, M.S. Atkins, B. Mackiewicz, M. Anderson, Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI, *IEEE Trans. Med. Imag.* 15 (2) (1996) 154–169.
- [55] M. Kamber, R. Shinghal, D.L. Collins, G.S. Francis, A.C. Evans, Model-based 3-D segmentation of multiple sclerosis in magnetic resonance brain images, *IEEE Trans. Med. Imag.* 4 (3) (1995) 442–453.
- [56] R. Khayati, M. Vafadust, F. Towhidkhal, S.M. Nabavi, 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 (3) (2008) 379–390.
- [57] D. Kroon, E. van Oort, K. Slump, Multiple sclerosis detection in multispectral magnetic resonance images with principal components analysis, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008, pp. 1–14.
- [58] M. Kuklisova-Murgasova, P. Aljabar, L. Srinivasan, S.J. Counsell, V. Doria, A. Serag, I.S. Gousias, J.P. Boardman, M.A. Rutherford, A.D. Edwards, J.V. Hajnal, D. Rueckert, A dynamic 4D probabilistic atlas of the developing brain, *NeuroImage* 54 (4) (2011) 2750–2763.
- [59] S. Kyriakou, C. Davatzikos, Nonlinear elastic registration of brain images with tumor pathology using a biomechanical model, *IEEE Trans. Med. Imag.* 18 (7) (1999) 580–592.
- [60] Z. Lao, D. Shen, D. Liu, E.R. Melhem, L.J. Launer, R.N. Bryan, C. Davatzikos, Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machines, *Acad. Radiol.* 15 (3) (2008) 300–313.
- [61] J. Lecoer, J.C. Ferré, C. Barillot, Optimized supervised segmentation of MS lesions from multispectral MRIs, in: *Work. Med. Image Anal. Mult. Scler.* 2009, pp. 5–14.
- [62] K.V. Leemput, F. Maes, D. Vandermeulen, A. Colchester, P. Suetens, Automated segmentation of multiple sclerosis lesions by model outlier detection, *IEEE Trans. Med. Imag.* 20 (8) (2001) 677–688.
- [63] K.V. Leemput, F. Maes, D. Vandermeulen, P. Suetens, Automated model-based tissue classification of MR images of the brain, *IEEE Trans. Med. Imag.* 18 (10) (1999) 897–908.
- [64] M. Martin-Fernandez, E. Muñoz-Moreno, L. Cammoun, J.-P. Thiran, C.-F. Westin, C. Alberola-López, Sequential anisotropic multichannel Wiener filtering with Rician bias correction applied to 3D regularization of DWI data, *Med. Image Anal.* 13 (1) (2009) 19–35.
- [65] H. McFarland, L. Stone, P. Calabresi, H. Maloni, C. Bash, J. Frank, MRI studies of multiple sclerosis: implications for the natural history of the disease and for monitoring effectiveness of experimental therapies, *Mult. Scler.* 2 (4) (1996) 198–205.
- [66] N. Moon, E. Bullitt, K. van Leemput, G. Gerig, Automatic brain and tumor segmentation, in: *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2002, pp. 372–379.
- [67] J. Morra, Z. Tu, A. Toga, P. Thompson, Automatic segmentation of MS lesions using a contextual model for the MICCAI grand challenge, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008, pp. 1–7.
- [68] K. Nakamura, E. Fisher, Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients, *NeuroImage* 44 (3) (2009) 769–776.
- [69] J.P.W. Pluim, J.B.A. Maintz, M.A. Viergever, Mutual-information-based registration of medical images: a survey, *IEEE Trans. Med. Imag.* 22 (8) (2003) 986–1004.
- [70] C. Pollo, M. Bach-Cuadra, O. Cuisenaire, J.-G. Villemure, J.P. Thiran, Segmentation of brain structures in presence of a space-occupying lesion, *NeuroImage* 24 (4) (2005) 990–996.
- [71] M. Prastawa, E. Bullitt, S. Ho, G. Gerig, A brain tumor segmentation framework based on outlier detection, *Med. Image Anal.* 8 (3) (2004) 275–283.
- [72] M. Prastawa, G. Gerig, Automatic MS lesion segmentation by outlier detection and information theoretic region partitioning, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008, pp. 1–8.

- [73] K. Rehm, K. Schaper, J. Anderson, R. Woods, S. Stoltzner, D. Rottenberg, Putting our heads together: a consensus approach to brain/non-brain segmentation in T1-weighted MR volumes, *NeuroImage* 22 (3) (2004) 1262–1270.
- [74] D. Rey, G. Subsol, H. Delingette, N. Ayache, Automatic detection and segmentation of evolving processes in 3D medical images: application to multiple sclerosis, *Med. Image Anal.* 6 (2) (2002) 163–179.
- [75] A. Rovira, A. León, MR in the diagnosis and monitoring of multiple sclerosis: an overview, *Eur. J. Radiol.* 67 (3) (2008) 409–414.
- [76] A. Rovira, J. Swanton, M. Tintor, E. Hueriga, F. Barkhof, M. Filippi, J.L. Frederiksen, A. Langkilde, K. Miszkziel, C. Polman, M. Rovaris, J. Sastre-Garriga, D. Miller, X. Montalban, A single, early magnetic resonance imaging study in the diagnosis of multiple sclerosis, *Arch. Neurol.* 66 (5) (2009) 587–592.
- [77] S. Saha, S. Bandyopadhyay, A new point symmetry based fuzzy genetic clustering technique for automatic evolution of clusters, *Inf. Sci.* 179 (9) (2009) 3230–3246.
- [78] B.R. Sajja, S. Datta, R. He, M. Mehta, R.K. Gupta, J.S. Wolinsky, P.A. Narayana, Unified approach for multiple sclerosis lesion segmentation on brain MRI, *Ann. Biomed. Eng.* 34 (1) (2006) 142–151.
- [79] M. Scully, V. Magnotta, C. Gasparovic, P. Pelligrino, D. Feis, H. Bockholt, 3D segmentation in the clinic: a grand challenge II at MICCAI 2008 – MS lesion segmentation, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge, 2008*, pp. 1–9.
- [80] M.L. Seghier, A.J. Ramackhansingha, A.P.L. Criniona, C.J. Pricea, Lesion identification using unified segmentation-normalisation models and fuzzy clustering, *NeuroImage* 41 (4) (2008) 1253–1266.
- [81] D.D. Sha, J.P. Sutton, Towards automated enhancement, segmentation and classification of digital brain images using networks of networks, *Inf. Sci.* 138 (1–4) (2001) 45–77.
- [82] M. Shah, Y. Xiao, N. Subbanna, S. Francis, D.L. Arnold, D.L. Collins, T. Arbel, Evaluating intensity normalization on MRIs of human brain with multiple sclerosis, *Med. Image Anal.* 15 (2) (2011) 267–282.
- [83] D.W. Shattuck, S.R. Sandor-Leahy, K.A. Schaper, D.A. Rottenberg, R.M. Leahy, Magnetic resonance image tissue classification using a partial volume model, *NeuroImage* 13 (5) (2001) 856–876.
- [84] S. Shen, A.J. Szameitat, A. Sterr, Detection of infract lesions from single MRI modality using inconsistency between voxel intensity and spatial location – A 3D automatic approach, *IEEE Trans. Inform. Technol. Biomed.* 12 (4) (2008) 532–540.
- [85] S. Shen, A.J. Szameitat, A. Sterr, An improved lesion detection approach based on similarity measurement between fuzzy intensity segmentation and spatial probability maps, *Magn. Reson. Imag.* 28 (2) (2010) 245–254.
- [86] N. Shiee, P. Bazin, J.L. Cuzzocreo, D.S. Reich, P.A. Calabresi, D.L. Pham, Topologically constrained segmentation of brain images with multiple sclerosis lesions, in: *Work. Med. Image Anal. Mult. Scler, 2008*, pp. 71–81.
- [87] N. Shiee, P. Bazin, D.L. Pham, Multiple sclerosis lesion segmentation using statistical and topological atlases, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge, 2008*, pp. 1–10.
- [88] N. Shiee, P.-L. Bazin, A. Ozturk, D.S. Reich, P.A. Calabresi, D.L. Pham, A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions, *NeuroImage* 49 (2) (2010) 1524–1535.
- [89] J. Sled, A. Zijdenbos, A. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, *IEEE Trans. Med. Imag.* 17 (1) (1998) 87–97.
- [90] S.M. Smith, Fast robust automated brain extraction, *Hum. Brain Map.* 17 (3) (2002) 143–155.
- [91] M.P. Sormani, L. Bonzano, L. Roccatagliata, G.R. Cutter, G.L. Mancardi, P. Bruzzi, Magnetic resonance imaging as a potential surrogate for relapse in multiple sclerosis: a meta-analytic approach, *Ann. Neurol.* 65 (3) (2009) 270–277.
- [92] J. Souplet, C. Lebrun, N. Ayache, G. Malandain, An automatic segmentation of T2-FLAIR multiple sclerosis lesions, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge, 2008*, pp. 1–11.
- [93] J. Souplet, C. Lebrun, S. Chanalet, N. Ayache, G. Malandain, Revue des approches de segmentation des lésions de sclérose en plaques dans les séquences conventionnelles IRM, *Rev. Neurol.* 165 (1) (2009) 7–14.
- [94] J.C. Souplet, C. Lebrun, N. Ayache, G. Malandain, An automatic segmentation of T2-FLAIR multiple sclerosis lesions, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge, 2008*, pp. 1–11.
- [95] M. Styner, J. Lee, B. Chin, M. Chin, O. Commowick, H. Tran, V. Jewells, S. Warfield, Editorial: 3D segmentation in the clinic: a grand challenge II: MS lesion segmentation, in: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge, 2008*, pp. 1–8.
- [96] N. Subbanna, M. Shah, S.J. Francis, S. Narayanan, D.L. Collins, D.L. Arnold, T. Arbel, MS lesion segmentation using Markov Random Fields, in: *Work. Med. Image Anal. Mult. Scler., 2009*, pp. 15–26.
- [97] N.K. Subbanna, S.J. Francis, D. Precup, D.L. Collins, D.L. Arnold, T. Arbel, Adapted MRF segmentation of MS lesions using local contextual information, in: *Med. Image Underst. Anal.* (2011) 1–5.
- [98] J. Talairach, P. Tournoux, Co-planar stereotaxic atlas of the human brain, Mark Rayport, Trans. Thieme., Stuttgart, 1988.
- [99] J.P. Thirion, G. Calmon, Deformation analysis to detect and quantify active lesions in three-dimensional medical image sequences, *IEEE Trans. Med. Imag.* 18 (5) (1999) 429–441.
- [100] X. Tomas, S.K. Warfield, Fully-automatic generation of training points for automatic multiple sclerosis segmentation, in: *Work. Med. Image Anal. Mult. Scler., 2009*, pp. 5–14.
- [101] J.K. Udupa, L. Nyul, U. Ge, R.I. Grossman, Multiprotocol MR image segmentation in multiple sclerosis: Experience with over 1000 studies, *Acad. Radiol.* 8 (11) (2001) 1116–1126.
- [102] J.K. Udupa, L. Wei, S. Samarasekera, Y. Miki, M.A. van Buchem, R.I. Grossman, Multiple sclerosis lesion quantification using fuzzy-connectedness principles, *IEEE Trans. Med. Imag.* 16 (5) (1997) 598–609.
- [103] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, P. Suetens, Automated segmentation of multiple sclerosis lesions by model outlier detection, *IEEE Trans. Med. Imag.* 20 (8) (2001) 677–688.
- [104] K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens, Automated model-based tissue classification of MR images of the brain, *IEEE Trans. Med. Imag.* 18 (10) (1999) 897–908.
- [105] U. Vovk, F. Pernus, B. Likar, A review of methods for correction of intensity inhomogeneity in MRI, *IEEE Trans. Med. Imag.* 26 (3) (2007) 405–421.
- [106] L. Wang, H. Lai, A. Thompson, D. Miller, Survey of the distribution of lesion size in multiple sclerosis: implication for the measurement of total lesion load, *J. Neurol. Neurosurg. Psych.* 63 (4) (1997) 452–455.
- [107] S.K. Warfield, K.H. Zou, W.M. Wells, Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation, *IEEE Trans. Med. Imag.* 23 (7) (2004) 903–921.
- [108] H.L. Weiner, C.R.G. Guttmann, S.J. Khoury, E.J. Orav, M.J. Hohol, R. Kikinis, F.A. Jolesz, Serial magnetic resonance imaging in multiple sclerosis: correlation with attacks, disability, and disease stage, *J. Neuroimmunol.* 104 (2000) 164–173.
- [109] Y. Wu, S.K. Warfield, L.L. Tan, W. Wells, D. Meier, R. van Schijndel, F. Barkhof, C. Guttmann, Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI, *NeuroImage* 32 (3) (2006) 1205–1215.
- [110] E.I. Zacharaki, C.S. Hoge, D. Shen, G. Biros, C. Davatzikos, Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth, *NeuroImage* 46 (3) (2009) 762–774.
- [111] E.I. Zacharaki, D. Shen, S.K. Lee, C. Davatzikos, ORBIT: a multiresolution framework for deformable registration of brain tumor images, *IEEE Trans. Med. Imag.* 27 (8) (2008) 1003–1017.
- [112] Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm, *IEEE Trans. Med. Imag.* 20 (1) (2001) 45–57.
- [113] A.P. Zijdenbos, R. Forghani, A.C. Evans, Automatic “pipeline” analysis of 3-D MRI data for clinical trials: Application to multiple sclerosis, *IEEE Trans. Med. Imag.* 21 (10) (2002) 1280–1291.