Neuroradiology (2012) 54:787–807
DOI 10.1007/s00234-011-0992-6

CONTINUING EDUCATION

Automated detection of multiple sclerosis lesions in serial brain MRI

Xavier Lladó · Onur Ganiler · Arnau Oliver · Robert Martí · Jordi Freixenet · Laia Valls · Joan C. Vilanova · Lluís Ramió-Torrentà · Àlex Rovira

Received: 17 October 2011 / Accepted: 29 November 2011 / Published online: 20 December 2011

Abstract
Introduction Multiple sclerosis (MS) is a serious disease typically occurring in the brain whose diagnosis and efficacy of treatment monitoring are vital. Magnetic resonance imaging (MRI) is frequently used in serial brain imaging due to the rich and detailed information provided.
Methods Time-series analysis of images is widely used for MS diagnosis and patient follow-up. However, conventional manual methods are time-consuming, subjective, and error-prone. Thus, the development of automated techniques for the detection and quantification of MS lesions is a major challenge.
Results This paper presents an up-to-date review of the approaches which deal with the time-series analysis of brain MRI for detecting active MS lesions and quantifying lesion load change. We provide a comprehensive reference source for researchers in which several approaches to change detection and quantification of MS lesions are investigated and classified. We also analyze the results provided by the approaches, discuss open problems, and point out possible future trends.
Conclusion Lesion detection approaches are required for the detection of static lesions and for diagnostic purposes, while either quantification of detected lesions or change detection algorithms are needed to follow up MS patients. However, there is not yet a single approach that can emerge as a standard for the clinical practice, automatically providing an accurate MS lesion evolution quantification. Future trends will focus on combining the lesion detection in single studies with the analysis of the change detection in serial MRI.

Keywords Multiple sclerosis · Serial analysis · MRI · Review

X. Lladó (✉) · O. Ganiler · A. Oliver · R. Martí · J. Freixenet
Computer Vision and Robotics Group, University of Girona,
Edifici P-IV Campus de Montilivi s/n,
17071 Girona, Spain
e-mail: llado@eia.udg.edu

L. Valls
Department of Radiology, Dr. Josep Trueta University Hospital,
Girona, Spain

J. C. Vilanova
Girona Magnetic Resonance Center,
Girona, Spain

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

À. Rovira
Magnetic Resonance Unit, Department of Radiology,
Vall d'Hebron University Hospital,
Barcelona, Spain

Introduction

Multiple sclerosis (MS) is the most frequent, non-traumatic, neurological disease capable of causing disability in young adults. It is a chronic, persistent inflammatory-demyelinating, and degenerative disease of the central nervous system (CNS), characterized pathologically by areas of inflammation, demyelination, axonal loss, and gliosis scattered throughout the CNS, often causing motor, sensorial, vision, coordination, deambulation, and cognitive impairment [11].

Conventional magnetic resonance imaging (MRI) techniques, 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 [62]. Quantitative analyses have become invaluable in the assessment of disease progression [34, 46, 47] and activity [58], and the evaluation of therapies over the last 25 years [7, 18]. Figure 1 shows two scans (T1-w, T2-w, and fluid attenuated inversion recovery (FLAIR) images) of a damaged brain taken with 1 year of difference, together with the manual annotations done by an expert. The last column of the figure illustrates the total 3D lesion load in the basal exploration and the new appearing lesions in the follow-up scan.

While there are many articles focusing on the lesion detection problem, most of them do not incorporate an automated method for interpreting the lesion evolution. The most common approach for the detection of change in serial imaging is visual inspection, which is usually performed manually by experts [41]. 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 [43, 47]. Experts use their high level of anatomical knowledge to identify the lesion evolution. However, the manual detection of change is not only time-consuming; it is also prone to intra-observer and inter-observer variability [51]. Although automated lesion detection techniques reduce this disagreement, an automated change detection method is still necessary to increase the diagnostic precision [14, 15]. Moreover, it is established that automated systems may outperform the human expert. For instance, as reported by Bosc et al. [5], while many small and subtle changes in lesion evolution were missed by the expert, the automated change detection algorithm did not. Therefore, we believe that a comprehensive resource of the literature on automated lesion detection and quantification is important work for researchers who want to improve upon previous work or develop new automated methods for progressive neurological disease analysis.

Change detection techniques may be divided into two categories: methods considering large structural changes and methods for smaller and more localized changes [5]. In accordance with this classification, lesion detection and quantification methods involve algorithms which must consider both small and large localized structural changes (i.e., tumors). General problems associated with these 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 or enlarge over time. In addition, their growth rates are not well characterized, and there can be great similarity between lesions and normal tissues, so they may not always be easily distinguishable. Moreover, the effect of a lesion does not always appear as an intensity change on the tissue where it is located (the so-called *tissue transformation*), but can also influence the appearance of surrounding tissues (known as the *mass effect*) [57]. 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 hard work due to noise and residual artifacts in the MR images and also because the images of a

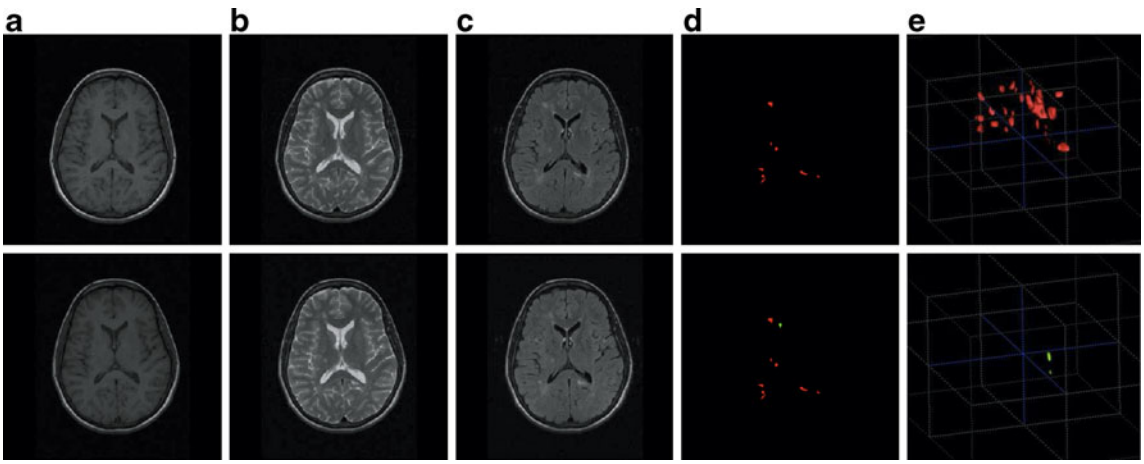

Fig. 1 An example of MS lesion serial analysis. The *upper row* shows a slice of the basal control, while the *lower row* shows the corresponding slice of the following exploration, done 12 months later. **a**, **b**, and **c** show, respectively, T1-w, T2-w, and FLAIR images. **d** Shows the manual lesion annotations of the slices done by an expert radiologist, where the *green annotation* shows a newly appearing lesion. Finally, the *upper* image of **e** represents the 3D lesion load in the basal exploration, while the *lower* image shows the 3D representation of the new lesions in the follow-up exploration

patient at different times are not always directly comparable due to patient movement. In many cases, a robust image registration algorithm must be used [22, 33, 76]. Notice that in this case the quantification accuracy will depend on the alignment accuracy [44]. Therefore, change detection techniques should be accordingly tuned to these facts.

Numerous approaches to lesion detection and quantification have been proposed in the literature [5, 20, 39, 42, 44, 73]. Despite the variety of approaches, none of them provide a fully automatic procedure that includes all the required steps for the diagnosis and treatment follow-up. For instance, some of the works which introduce automated methods for lesion detection (which are typically based on segmentation) do not always provide an automated method for quantifying the lesion evolution [1, 2, 63]. On the other hand, some of the works that focus on the change detection do not always provide an automatic lesion detection method and need user interaction to locate lesions [57] because they are not good enough to segment lesions after change detection [44]. Furthermore, some of the change detection algorithms provide only a resulting image which has to be interpreted visually by experts [38, 55], and a final expert decision is required to assess the lesion evolution [5]. Note also that the change detection algorithms do not usually cover the detection of static lesions, such as black holes. Combining the advantages of different techniques may compensate some of the missing parts of some strategies and may enable the development of less subjective and more automated approaches.

The aim of this article is to point out the capabilities of the developed approaches, providing an up-to-date state-of-the-art review of automated MS lesion detection and quantification methods in serial MRI. Furthermore, we classify the different techniques according to the strategy used, describing also the most representative works in this field. We analyze numerous articles providing a detailed classification of lesion detection and change detection techniques based on the main characteristics of each strategy, pointing out the challenging parts of each method. In addition to introducing and classifying these approaches, we also describe the algorithms used to detect and quantify the lesions as well as the features and the type of MR images used. Furthermore, we compare the results of the analyzed works in terms of accuracy and robustness.

Few articles have reviewed MS lesion detection and quantification methods in brain MRI serial analysis. For instance, Patriarche and Erickson [41] provided a review of the change detection techniques in time-series analysis. However, this review was not particularly focused on the purpose of MS lesion detection. Bosc et al. [5] also provided a simple classification of inter-image comparisons considering lesion evolution. Nevertheless, this work was not a complete review. Recently, Mortazavi et al. [40] have presented a review of MS lesion detection in a single time point, without taking into account the change detection, the lesion evolution, and quantification. Even though some articles have given information about either MS lesion detection or lesion evolution quantification methods [26, 29, 42, 52], none of them proposed a comprehensive review. Furthermore, none of them tried to quantitatively compare the results of the different approaches, as it would be difficult to guess the performance of all these detection and quantifications approaches. Ideally, methods should be applied to a common database and compared with 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. Here we will quantitatively compare the detection approaches accordingly to their reported results in the literature. We will describe the most typical measures used for evaluating MS lesion detection and quantification in time-series MRI, comparing in a qualitative and quantitative way the results of the works analyzed. To the best of our knowledge, our paper is the first attempt to review the most relevant works in the time-series analysis from both MS detection and quantification point of views and which also provides an evaluation of the experimental results.

The rest of this paper is organized as follows. "Classification of MS lesion detection and quantification in serial brain MRI" presents the classification of the lesion detection and quantification approaches, reviewing also the main features and algorithms used. "Lesion detection approaches," "Change detection approaches," and "Quantification approaches" describe the lesion detection, change detection, and quantification approaches, respectively. "Experimental validation" explains how to perform the experimental validation of an approach, explaining the different image databases and common measures used to evaluate the results. A brief analysis and a performance comparison of some key works are given in "Discussion," which also presents possible improvements and further works. The paper finishes with conclusions.

Classification of MS lesion detection and quantification in serial brain MRI

In this section, we propose a classification for categorizing the state-of-the-art of automated serial MS detection and quantification methods in time-series analysis. Afterward, we also analyze the general problems encountered in the segmentation and quantification processes.

Proposed classification

In order to classify the MS lesion detection approaches, we considered the different classifications proposed by Bosc et

al. in 2003 [5] alongside the one proposed by Patriarche and Erickson in 2004 [41]. From this starting point and also from the information collected from the newest works [26, 29, 42, 52], we propose a new classification with the categories and subcategories shown in Fig. 2. In particular, we classify the detection approaches within two primary categories, according to their main principle and characteristics:

- *Lesion detection methods.* We consider lesion detection methods to be the ones that aim to detect both static and dynamic MS lesions on a single-time magnetic resonance (MR) volume of a patient. These segmentation-based methods, which can be supervised or unsupervised algorithms, rely on the intensity homogeneities of the tissues and typically apply data mining techniques (clustering, classification) to distinguish lesions from normal tissues. In time-series analysis, the use of segmentation-based methods mostly involve a subsequently lesion quantification approach that computes the volumetric changes of each segmented lesion between two time points in order to determine the MS lesion evolution.
- *Change detection methods*. These approaches are not based on the analysis of a single time point (one control of a patient) but rely on analyzing the differences between successive MRI controls at both 2D and 3D image levels. From this classification, we further subclassify the main strategies. *The intensity-based methods* consist of analyzing two successive scans by means of subtraction techniques. Among these methods, we further distinguish between *deterministic* approaches, which typically cover the subtraction methods using direct intensity differences between the scans, and *statistical* approaches, which are used for compensating the interpretation problems of point-to-point comparison. The *temporal analysis* approaches are based on detecting active voxels through a time-series analysis of more than two successive scans. Finally, the *deformation-based* approaches aim to obtain a deformation field from a non-rigid registration process between successive controls, which can be directly used to perform the lesion detection and evolution. We have subclassified these approaches according to the way the deformation field is used: *vector displacement field* and *deformation field morphometry*. Note that, depending on the technique used, these approaches may or may not require a subsequent analysis of the quantification.

The approaches reviewed in this work are summarized in Table 1, which offers a compact, at-a-glance overview of these studies. Moreover, the most important features and properties of all the approaches have also been taken into account. Namely, the main characteristics of each analyzed approach: detection strategy and quantification algorithm used, the type of automation (semi-automated or fully automated), whether the method uses a template (an atlas) to improve the accuracy, such as template driven segmentation (TDS) or methods that use healthy control images to compare and correct their results. Finally, we have also included the image types used (T1-w, T2-w, PD-w, FLAIR, etc.) and the lesion types the method can deal with. It should be noted that not all the analyzed works specify always the specific type of lesion. In "Lesion detection approaches" and "Change detection approaches" of this paper, we will analyze in detail all these strategies, describing the primary characteristics of each category.

Lesion quantification

As well as classifying the MS lesion detection approaches, we can categorize the methods according to the quantification of the lesion evolution. Note that this quantification process is essential for radiologists and neurologists to analyze the patient follow-up [43]. We classify the quantification approaches into three main categories: *visual inspection*, *statistical change detection*, and *volumetric approaches*.

The visual inspection is a manual method for determining lesion evolution. The processed images such as registered images or subtracted images are analyzed and interpreted

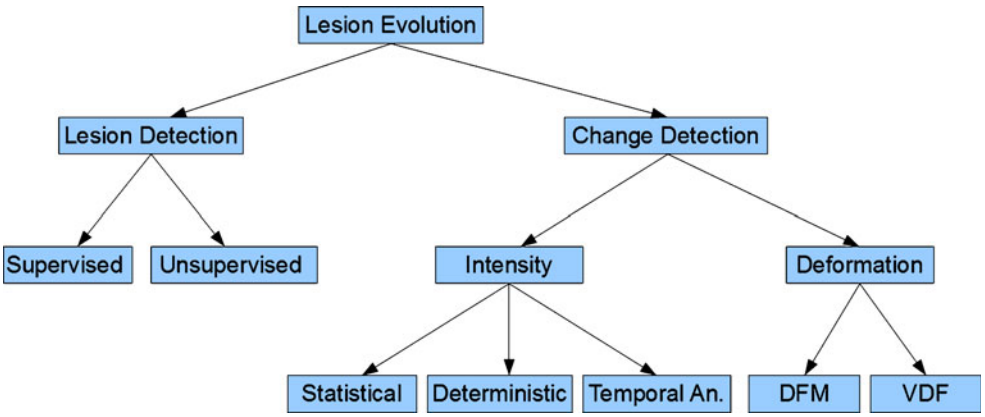

Fig. 2 Proposed classification of MS lesion serial analysis. We clearly distinguish between lesion detection and lesion change detection techniques. The acronyms *VDF* and *DFM* stand for vector displacement field and deformation field morphometry, respectively

Table 1 Classification of the lesion evolution methods

			References	Detection	Quantification	A	Atlas	Sequences	Lesions
Lesion detection	Supervised		[Udupa, 1997]	FCS	Volumetric	SA	–	T2; PD	WML
			[Warfield, 2000]	KNN	–	A	X	T2	WML
			[Zijdenbos, 2002]	ANN	Volumetric	A	–	T1; T2; PD	WML
			[Wei, 2002]	Self-adaptive EM and PVEC	Volumetric	A	X	T2; PD	WMSA
			[Ashton, 2003]	Bayesian	Volumetric	SA	–	T1; T2; PD	WML
			[Meier, 2003]	Self-adaptive EM and PVEC	Temporal analysis	A	X	T2	WMSA
			[Antel, 2003]	Bayesian	–	A	–	T1	FCDL
			[Anbeek, 2004]	kNN	–	A	–	T1; T2; PD; FLAIR; IR	WML
			[Wu, 2006]	kNN	–	A	X	T1; T2; PD	WML and GEL
			[Duan, 2008]	PVEC and thresholding and manual	Volumetric	SA	X	T2; PD	–
			[Zacharaki, 2008]	SVM	Volumetric	A	–	T1; T2; PD; FLAIR	WML
			[Shen, 2008]	FCM	Volumetric	A	X	T1	IL
			[Shiee, 2010]	FCM	–	A	X	T1; T2; FLAIR	WML
			[Yamamoto, 2010]	Level sets and SVM	–	A	–	T1; T2; FLAIR	WML
			[Cerasa, 2011]	ANN	–	A	–	FLAIR	WML
	Unsupervised		[Ettinger, 1994]	EM and subtraction	Volumetric (4D-CCA)	A	–	T2; PD	WML
			[Lee, 1998]	Thresholding and subtraction	Volumetric	SA	–	T2; T1	GEL
			[Guttmann, 1999]	EM and PVEC	Volumetric (4D-CCA)	A	–	T1; PD	WML
			[Kikinis, 1999]	EM and PVEC	Volumetric (4D-CCA)	A	–	T1; PD	WML
			[Weiner, 2000]	EM and PVEC	Volumetric (4D-CCA)	A	–	T1; PD	GEL
			[Hillary, 2009]	ISODATA	Volumetric	A	–	T1; FLAIR	–
			[Duan, 2008]	Thresholding and manual	Volumetric	SA	–	T2; PD	–
			[Juang, 2010]	Histogram-based classification	VI	A	–	T2; T1	Tumor
Change detection	Intensity	Det.	[Curati, 1996]	2D subtraction	VI (manual)	SA	–	T1	–
			[Tan, 2002]	2D subtraction	VI (manual)	A	–	T2	–
			[Moraal, 2009]	2D subtraction	VI (manual)	A	–	T1; T2; PD	WML
		Stat.	[Moraal, 2010]	3D subtraction	VI (manual)	A	–	FLAIR; DIR; MP-RAGE	WML
			[Lemieux, 1998]	2D subtraction and SDF and SNM	VI (SCD)	A	–	T1	–
			[Bosc, 2003]	2D subtraction and GLRT	VI (SCD)	A	–	T1; RARE; FLAIR	–
		Temp.	[Gerig, 2000]	Temporal analysis	Volumetric	A	–	FMRI	WML and GML
			[Welti, 2001]	Spatio-temporal analysis	–	A	–	T1; T2; Pd; FLAIR	WML and GML
	Deformation	VDF	[Thirion, 1999]	Norm and divergence of vector fields	Warping	A	–	T2	–
			[Rey, 2002]	Flow field and Jacobian operator	Warping	A	–	T2; PD	–
		DFM	[Pieperhoff, 2008]	Flow field and LVR	Warping	A	–	PD	–

The different acronyms refer to: detection methods: *FCS* fuzzy-connectedness segmentation, *kNN* k-nearest neighbor, *ANN* artificial neural network, *EM* expectation maximization, *PVEC* partial volume effect correction, *SVM* support vector machines, *FCM* fuzzy C-mean, *SDF* structure difference filtering, *SNM* structure noise map, *GLRT* generalized likelihood ratio test, *LVR* local volume ratio; quantification methods: *VI* visual inspection, *SCD* statistical change detection, 4DCCA 4D connected component analysis. Lesion types: *WML* white matter lesion, *WMSA* white matter signal abnormalities, *GML* gray matter lesion, *GEL* gadolinium enhancing lesion, *FCDL* focal cortical dysplasia lesion

visually by a user or expert in order to render a decision. Although this is a very subjective method, some improvements can be used to reduce the misinterpretations made by the expert. For instance, statistical change detection techniques based on using statistical correction [5] or using structure noise maps [31] in order to reduce false positives in the subtracted images may be applied. In a different way, the volumetric approaches typically use already segmented lesions in order to quantify the lesion evolution by means of their volume changes. These volumetric quantification approaches have been proven to be useful for detecting positive and negative disease activity [66]. Notice that this quantification process can be done by either subtracting single lesion volumes or subtracting total lesion volumes 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. As a result, comparing lesion volumes individually seems 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 and introduce errors in the quantification.

Note that we could also add in this quantification classification the temporal analysis and the warping methods which were also included as detection approaches. In fact, these methods produce in a single step the detection and quantification of the MS active lesions. For instance, the main property of the warping algorithms (also known as deformation field-based approaches [5]) is that they are based on one-to-one tissue correspondence and, as well as providing lesion detection, they allow the lesion mass effect to be quantified from the registration process between temporal studies of a patient.

General problems in MRI

Image intensities between the corresponding tissues or structures of successive scans may differ from each other. Thus, normalization algorithms are used to compensate global intensity changes between successive images before and/or after registration processes [5]. This normalization process improves the alignment between the images if used before the registration step and also allows a better comparison between the analyzed tissues, structures, and lesions if used after registration. For instance, Bosc et al. [5] used a linear intensity normalization algorithm before each registration step and a non-linear joint normalization algorithm after registration.

Another well-known issue when processing MRI images is that noise and artifacts may be present due to scanner performance, and this may affect the detection and quantification accuracy [31]. For instance, Guttman et al. [21] and Kikinis et al. [27] used a non-linear anisotropic diffusion filtering, which is an edge-preserving noise reduction method, to overcome this problem. With a similar strategy Bosc et al. [5] applied a low-pass Gaussian filter to the images obtained by subtracting registered and normalized successive images to eliminate residual artifacts such as radio-frequency artifacts.

Besides these difficulties, partial volume effects, where a single voxel contains a mixture of multiple tissue values, generally occurs in medical imaging. This situation is particularly true for voxels on the boundaries [12] or brain surfaces that contain both brain tissue (skull bone) and cerebrospinal fluid [27] due to the particular intensity characteristics of PD-w and T2-w images. Thus, such regions which have similar intensity values as the lesions may introduce errors in the quantification process. Several approaches have been proposed to deal with this issue. For instance, some methods use a priori anatomical knowledge [21] to eliminate spurious lesions selectively [27].

Lesion detection approaches

Image segmentation is the process of assigning a label to every voxel in a single image such that voxels with the same label share certain visual characteristics typically indicating a particular object such as tissue or lesion. To our knowledge, segmentation-based approaches cover the largest area of methods for MS lesion detection and are still the largest active area of research. Different methods have been proposed for this purpose [1, 16, 21, 26, 49, 75], and some attempts to classify these automated MS lesion segmentation approaches have been made. Mortazavi et al. [40] have recently presented a review of segmentation of multiple sclerosis lesions in MR images, providing a classification of the reviewed approaches into four different categories: data-driven, statistical, intelligent, and deformable methods. Even though numerous attempts have been done to solve this segmentation problem, due to the arbitrary lesion shapes and locations, automated segmentation is still an open issue and a challenging task [32].

Segmentation-based approaches can be classified as manual outlining methods, semi-automated methods, and fully automated methods. In this paper, we explore the fully automated methods, which do not require user interaction and reduce the intra/inter-operator variability [52]. However, to provide a wider analysis, we also include some of the most relevant semi-automated methods in our review [4, 16, 30, 61]. Figure 3 shows a flowchart with the general idea of the segmentation-based approaches for brain MRI time-series analysis.

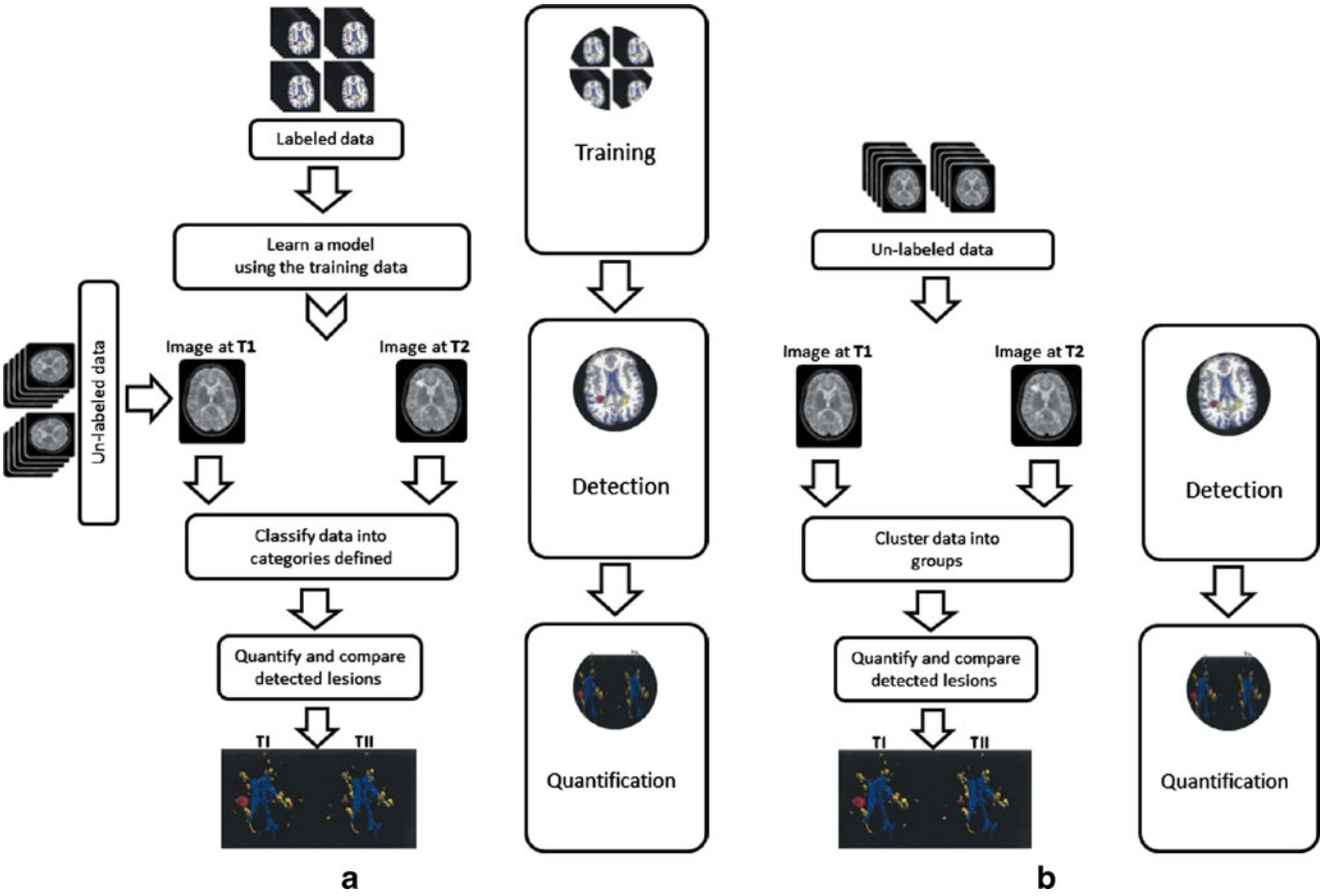

Fig. 3 Flowchart of the lesion detection approaches: **a** supervised, **b** unsupervised. As it is clearly shown, the main difference between both strategies is the use or not of an initial training step

The automated MS lesion segmentation is a difficult task due to intensity similarity between lesions and normal tissues. For instance, gray matter lesions may share intensities with gray matter (GM) or cerebrospinal fluid (CSF) [49]. Thus, traditional segmentation methods, like region-based methods where the voxels are directly analyzed by means of a region growing strategy, or methods using thresholding techniques, may not provide the desired results. Noise and residual artifacts also make lesion segmentation difficult, even for the white matter lesions (WML).

Analyzing the literature, we have seen that tissue class segmentation-based approaches (clustering methods) are commonly used for automated MS lesion segmentation. Tissue class segmentation, which uses the tissue-class weights to consider the presence of lesion, can be considered as an estimation problem for determining intensity inhomogeneities [67]. These techniques use spatial information (position of the tissues and lesions) and inconsistency of lesions (intensity differences between lesions and the normal tissue distributions) to detect and then quantify the lesions. Note that they are statistical segmentation approaches that use knowledge about tissue properties and therefore rely on the fact that the same tissues have the same intensity values.

Several techniques have been used to improve segmentation accuracy. It is well known that the use of prior knowledge of normal tissue distribution improves the capability of segmentation methods [65]. The main strategy is to use an anatomical template (atlas) to introduce spatial information into the statistical segmentation. Although this information can be introduced in different ways [6], the most common approach is based on TDS, which mainly consists of a non-linear registration step [53, 63] to match MR images to the atlas. As reported by Warfield et al. [63], statistical classification and non-linear registration are often complementary since pathologic structures such as lesions are not modeled in an anatomical template. Lesions cannot be segmented directly with an anatomical template. Therefore, statistical methods are performed to compensate this problem.

In addition to TDS, multi-spectral approaches are used to improve the segmentation accuracy since different modalities of MR images (T1-w, T2-w, PD-w, FLAIR, etc.) have different signal characteristics that provide different information. However, multi-spectral anatomical images are not always available in clinical practice since the acquisition of all these images is cost-intensive and requires more

processing time [49]. Methods using multi-spectral information also require a registration step, which may be assumed as an affine [49] or a deformable registration [63].

Furthermore, some works also use a partial volume effect correction (PVEC) method to eliminate false-positive detected lesions [16, 21, 27, 65]. For instance, Guttmann et al. [21] and Kikinis et al. [27] applied a PVEC algorithm and improved their previous results. Moreover, Wei et al. [65] concluded that the PVEC algorithm eliminated only false-positive errors while TDS corrected false-negative misclassifications and some of the false-positive misclassifications. They also pointed out that using TDS with PVEC together showed the highest accuracy in the segmentation of the white matter signal abnormalities [65].

Supervised methods

We consider as supervised approaches those methods that use mainly the image intensities of different MR images to train a classifier by using labeled tissues and manually identified lesions and those methods which use information from a template (atlas) to classify tissues and to segment lesions as deviations from normal human brains. It can be seen in Table 1 that several techniques have been used to perform supervised classification. For instance, k-nearest neighbors (kNN), artificial neural networks (ANNs), and support vector machines (SVM) are typical supervised approaches for the tissue segmentations. Furthermore, Yamamoto et al. [71] have recently proposed a false-positive reduction step which uses a level set method and a SVM classifier to substantially reduce the false MS lesion detections.

According to Udupa et al. [61], human experts usually outperform automated algorithms in the recognition task, and, therefore, in their approach, the brain tissues such as WM, GM and CSF are manually determined by an operator. They claimed that the automated algorithms conversely perform better in the delineation. Hence, they used a fully automated algorithm for the delineation process from which they segmented the MS lesions based on the principle of fuzzy-connectedness [60] using the manually recognized brain tissues (WM, GM, CSF) as fuzzy connected regions. After the detection of CSF, WM, and GM as 3D fuzzy objects, lesions appear as "holes". The approach of Udupa et al. [61] can also be considered as an early multi-spectral approach, since they used both T2-w and PD-w images to classify brain tissues. They state that CSF tissue is better recognized in the T2-w image whereas WM and GM tissues are better recognized in PD-w images.

Another semi-automated supervised and multi-spectral method was proposed by Ashton et al. [4]. They compared the regional-based methods (GEORG) with a directed multi-spectral segmentation approach and conclude that both methods were acceptable in terms of speed and precision. They used statistical characteristics of background tissues supplied by a Bayesian classifier and target statistics supplied by the exemplar. This approach is also multispectral since they mapped the three T1-w, T2-w, and PD-w images to the red, green, and blue channel, respectively. Nevertheless, both algorithms need user interaction: A single mouse click was used to place a seed for a region growing algorithm, and a manually traced exemplar was needed for the classification method.

Warfield et al. [63] applied TDS segmentation and spatially varying statistical classification based on a multiple feature kNN classification process. Also based on the kNN classifier, Wu et al. [70] proposed an 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. Assuming 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. On the other hand, Wei et al. [65] and Meier and Guttman [35] included a template-driven strategy to perform the tissue class segmentation based on an expectation maximization algorithm. Meier and Guttman [35] also applied subtraction and partial volume correction to identify lesion load change and to eliminate false-positive lesions. After the lesion segmentation, they combined space and time into the MS lesion characterization process via direct quantitative analysis of signal intensity in the time domain obtained from serial MR images. In this way, they showed the signal dynamics of active and chronic MS lesions [35].

Zijdenbos et al. [75] proposed a supervised MS lesion segmentation method using multi-spectral information (T1-w, T2-w, and PD-w) using an ANN. In particular, they used a back-propagation ANN method to classify the MS lesions because of the reliability of the method under different image conditions. Similarly, Cerasa et al. [9] propose a technique for segmenting white matter lesions in MS patients by using a cellular neural network (CNN)-based approach. Unlike ANN, in a CNN, interconnections among cells are local, that is, each processing unit directly interacts only with the neighboring cells, located within a prescribed sphere of influence. The authors applied this CNN-based technique to automatically segment MS lesions on FLAIR images, comparing the performance of their approach with the manual segmentation provided by two expert radiologists. Moreover, Anbeek et al. [1] combined a supervised classification algorithm with a multi-spectral approach for WML detection. They used five different modalities (T1-w, T2-w, PD-w, IR, and FLAIR) and applied a kNN

classification technique. Likewise, Zacharaki et al. [73] recently presented a supervised WML segmentation method based on SVM. They applied an AdaBoost algorithm to each of the scans. As they reported, WMLs had intensities similar to GM tissue in T1-w images and similar to CSF in T2-w and PD-w images, so they applied a multi-spectral approach.

Besides these techniques, Antel et al. [2] used texture feature maps obtained by co-occurrence matrices together with a supervised classification based on two-step Bayesian classifier to perform the MS lesion detection. More recently, Shen et al. [49] identified the MS lesions using their inconsistency by a defined threshold. They combined the fuzzy c-mean (FCM) algorithm and TDS which was used to create tissue probability maps. There are more examples of atlas-based approaches. The method proposed by Shiee et al. [50] 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 FCM algorithm which performs the classification. As reported by Shiee et al. [50], 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.

One of the drawbacks of the supervised segmentation is that segmentation accuracy may depend highly on the selection of the training set and the control groups [49] used to compare individual patient images to a normal control group (model-based strategy) [53]. Gerig et al. [19] compared an unsupervised classification (ISODATA) with a supervised classification (parametric maximum likelihood classification and Parzen window technique) for the brain MR images and found similar estimated parameters. Furthermore, although supervised clustering methods are more efficient for the segmentation purpose, they require some user interaction for the training steps. Besides, different users or trainings at different times on the same data may produce different results (Fig. 4 shows an example of a MS patient volume segmented by two different experts). Thus, unsupervised methods are less subjective and completely automated and more reproducible methods with respect to supervised classifications.

Unsupervised methods

As illustrated in Table 1, many of the unsupervised classification methods [17, 21, 27, 66] use the expectation maximization (EM) algorithm [68]. For instance, Ettinger et al. [17] combined statistical tissue classification based on the EM algorithm and subtraction in order to detect positive and negative changes. In a similar way, Guttmann et al. [21], Kikinis et al. [27], and Weiner et al. [66] have used a similar strategy to segment MS lesions based on tissue classification and expectation maximization.

Lee et al. [30] used a local threshold defined by a single observer in order to segment MS lesions. Areas of new lesions and areas of resolving lesions were defined by subtracting normalized and co-registered images. They labeled the lesion areas with color and subtracted two successive images. The outcome image yields a colored subtraction map which indicates areas of new lesions and areas of resolving lesions.

More recently, Duan et al. [16] compared two different approaches called conventional image segmentation (CSEG) and segmentation of subtraction image (SSEG). The first one was a supervised approach due to the use of TDS, while the second one was an unsupervised approach using the intensities of the subtracted images to detect the MS lesions. However, both segmentation methods are refined by applying an automated Otsu threshold and manual editing. The authors concluded that the SSEG method provided significantly higher measurement of reproducibility and enhanced sensitivity to cortical and subcortical lesions.

Hillary et al. [23] used an ISODATA technique consisting of a multi-parametric unsupervised classification method. As a different approach to classification methods, Juang

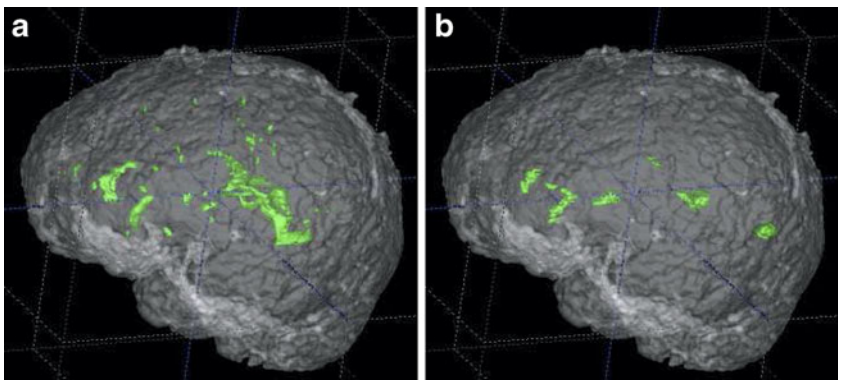

Fig. 4 Generated 3D volume with MS lesions segmented by two different experts showing a large inter-rater variability. Note here the importance of using more than one manual annotation when evaluating the automatic algorithms (example extracted from the MICCAI challenge)

and Wu [66] applied color-based segmentation with k-means clustering based on applying a histogram-based metric to produce colored images indicating the tissues and lesions.

Change detection approaches

As we have already stated in the introduction, the patient follow-up over time is crucial to determine the evolution of the disease. Therefore, change detection techniques are needed to compare the brain evolution over time. As shown in Fig. 2, we distinguish among three main different strategies to perform these tasks, which are described in the following subsections. Figure 5 shows the different flowcharts of each category.

Intensity-based approaches

Intensity-based approaches for change detection use voxel-to-voxel intensity comparison to distinguish evolving lesions. Therefore, a lesion without changes in the follow-up scan, such as static lesions, cannot be detected using this strategy.

Voxel-to-voxel comparison methods usually suffer from repositioning errors due to patient movement, inconsistent objects over time such as blood and cerebrospinal fluid flow artifacts [38], noise in the images, and partial volume effects [12]. Therefore, image registration, bias field correction, and intensity normalization [48] are necessary to compensate for these problems. Furthermore, the selection of image type (T1-w, magnetization prepared rapid gradient echo (MP-RAGE), T2-w, PD-w) and the interpolation method during the registration process are also important criteria for the accuracy and robustness of the subtraction methods.

Deterministic approaches

We include in this group those intensity-based approaches that are based on subtracting two successive images in order to find intensity differences due to evolving lesions. Typically, after the subtraction of two consecutive temporal images, unchanged areas (normal tissue) appear as gray areas, while changed areas are due to the appearance or disappearance of lesions. Hence, the positive activity (new or enlarging lesions) appears as a bright area while the negative activity (resolving or shrinking lesions) appears as a dark area against the gray background [55].

The roots of the subtraction approach to detect MS lesions were made by Curati et al. [12], who investigated contrast enhancement with registered difference images. They reported that the recognition of small changes, changes at the boundaries, and tissues and fluids with very high or very low signals were more difficult to determine. Furthermore, they noted that, while the use of thin slices decreased the partial volume effects, it increased the mis-registration. Thus, they stated that an accurate alignment was necessary to assess the changes. They also claimed that using 3D scans of the MP-RAGE images might increase the accuracy of the results since these types of scans have better contrast.

Tan et al. [55, 56] suggested that using only the variation in the intensity signal to determine negative or positive activity was not sufficient, since change in intensity signal may also be due to different conditions such as the use of a different scanner or a high level of noise. Thus, they determined regional activity by also comparing if there was a change in the lesions size or shape. They concluded that using subtracted images for lesion detection showed better agreement for the positive activity than for the negative one. Besides, they reported that success of this approach depends

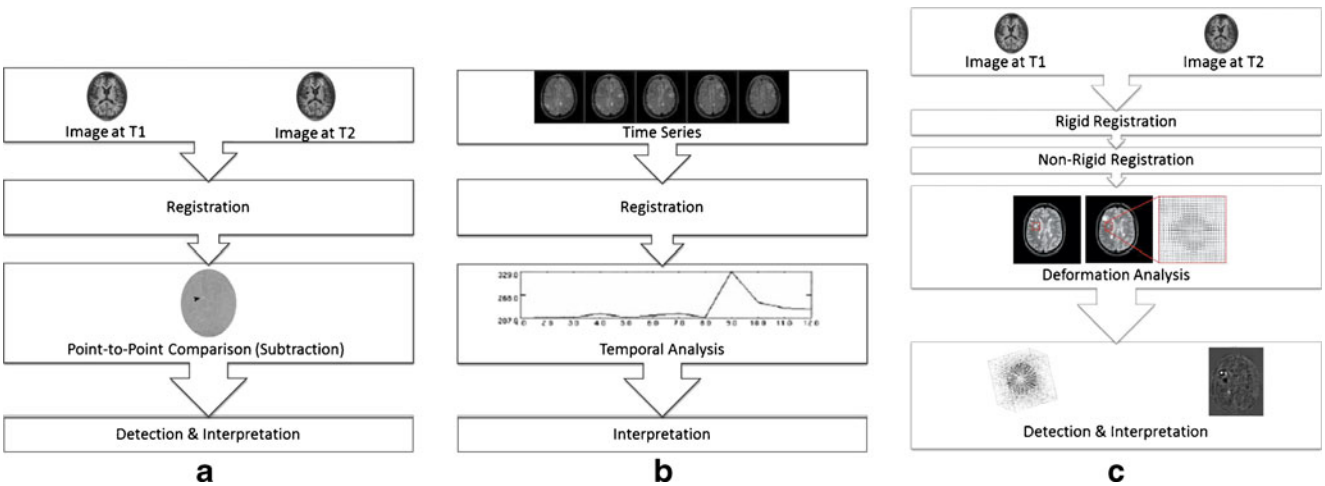

Fig. 5 Flowchart of the main change detection categories. **a** Intensity-based techniques, **b** temporal analysis, and **c** deformable approaches

highly on the lesion size. To detect enlarging lesions smaller than 5 mm in diameter, they must increase their size more than 100% and appear in at least two consecutive slices. On the other hand, the detection of shrinking lesions with a diameter smaller than 5 mm was not reliable.

Following a similar approach, Moraal et al. [38] also concluded that subtracted images provide a sufficient measure for the quantification of positive disease activity. The authors found a good inter-observer agreement in the quantification of positive disease activity and compared their results with the previous works in terms of inter-observer agreement, concluding that their success was due to the improvements in the registration and intensity correction methods used. They also noted that results obtained for the negative activity were not as good as the results obtained for the positive disease activity. In a different study, Moraal et al. [39] evaluated the performance of 2D and 3D subtraction methods and conclude that 3D subtraction techniques, after image registration, provided greater inter-observer agreement. Furthermore, they compared several image sequences (3D DIR, 3D FLAIR, 3D T2-w, 3D MP-RAGE) and found that negative active lesions, even the small ones, could be detected using the 3D MP-RAGE images, owing to good anatomical detail and clear GM-WM contrast.

Statistical-based approaches

Statistical change detection techniques for interpreting intensity differences aims at reducing the noisy results obtained by direct point-to-point subtraction [5]. This group of methods is based on building a statistical model of intensity changes between successive scans in order to detect active lesions and their evolutions. Those methods rely on the changes of the lesions and not on the changes of the individual voxels.

For instance, after the image subtraction, Lemieux et al. [31] classified each voxel as changed versus unchanged according to a threshold value and subsequently grouped together the changed voxels. They called these grouped voxels structured differences objects, which can be caused by either biological processes or image artifacts. Afterward, in order to quantify changes in the image difference, these structured difference objects were thresholded by applying the structure difference filtering which was used to estimate the Gaussian noise level. After the normalization, the outcome image was a map of the classification of voxels as no signal change, signal increase, signal decrease, and outside of the brain. The authors also compared this map with the one obtained by a set of normal volunteers in order to assess the significance of the changes. By using this full scheme, they avoided the structured noise, and they were able to determine the real changes more correctly. However, note that this statistical method cannot give the total count of active lesions directly, although a set of statistics, such as the total genuine change voxels and total number of normal structured voxels, can be easily obtained.

On the other hand, Bosc et al. [5] presented both a single-modal and multi-spectral (FLAIR, RARE, and GE 3D) change detection approach. They registered the images into a common reference according to their modality, instead of choosing a baseline reference from the serial images of a patient, since the registered images undergo geometrical transformations while the reference image does not. In this sense, all images undergo equivalent processing steps, like it is done in the well-known half-way registration [24, 25]. Affine registration was used for registering the single modality matching while affine and deformable registration was used for the multi-modality matching. Afterwards, they computed the voxel probability ratio of change and grouped together those neighboring changed voxels. Thus, clustered voxels (also sorted in decreasing likelihood) were presented to the experts instead of individual voxels. Notice that evaluating individual voxel changes is more difficult, and also, manually delimiting the lesion evolution is more subjective. They evaluated their results with simulated lesions and found that lesions with a radius greater than 0.6 voxels could be detected. Furthermore, they found that the multi-modality detection increased the detection probability from 79% to 95% due to the richer information, which avoids more false-positive detection.

Temporal analysis approaches

Temporal analysis is based on the analysis of long time-series of MR images, i.e., more than two explorations. Note that, in these cases, the subtraction techniques should not be employed. Hence, in temporal analysis, the intensity of each voxel is regarded as a function of time, and the aim is to see how the brightness of these voxels varies over the time. This analysis is useful for either lesion segmentation [20] or characterization [35].

Gerig et al. [20] combined space and time into a 4D volume in order to track the brightness of each voxel. They first applied a supervised method to segment normal tissues based on parametric maximum likelihood classification and Parzen windows [19]. Afterwards, they distinguished active lesions by computing the mean and variance of the voxel time-series, since voxels belonging to active lesions show higher variance compared with static tissues. Note, however, that this temporal analysis relying on voxel level comparison assumes a perfect registration among the different volumes, which cannot be true in most of the cases. This drawback can be minimized by taking the spatial correlation between neighboring voxels into account [69]. Therefore, the voxels' gray-value information and their surrounding tissue for all serial scans were stored in the database,

implicitly assuming that the mean spatio-temporal evolution of all lesions in the database can be regarded as a characteristic model of a typical MS lesion.

Recently, Srivastava et al. [53] presented a statistical segmentation method based on building a lesion specific feature map. They incorporated a template-driven segmentation of the three main tissues (CSF, WM, and GM) and then used the ratio of cortical thickness over absolute image intensity gradient. The statistical parametric map was thresholded in order to detect lesions. They stated that their method can be applied to almost all lesions satisfying the thickening and the blurring model; hence lesions with volume smaller than 3.8 cm^3 could be detected.

Deformation field-based approaches

An MS lesion is generally seen as the combination of two different effects, tissue transformation and tissue deformation [57]. Tissue transformation refers to the intensity change in the tissue of the lesion, while tissue deformation refers to the modification of its surrounding tissue, due to lesion expansion or contraction. Therefore, using only approaches based on intensity changes between serial scans to evaluate the evolution of lesions may not give satisfactory results, since the surrounding tissue deformation due to the presence of the lesion is not taken into account. In order to consider the mass effect of the lesions, deformation-based approaches should be employed.

In deformation field-based approaches, a non-linear registration is performed between successive scans, and the structural changes are determined based on the local deformation of voxels. Note, however, that, due to the fact that this approach looks for the differences between the successive scans, static lesions cannot be detected.

Vector displacement fields

Thirion and Calmon [57] proposed a semi-automatic approach using vector displacement fields obtained by a non-rigid registration of two successive scans to track MS lesions. They proposed to use both the divergence and the norm of the displacement vector fields in order to be sensitive to deformation and intensity change. Therefore, high values of the norm indicated large deformation areas, while high divergence indicated evolving lesions, where the sign of the divergence operator showed whether the lesion was growing or shrinking. Moreover, they also observed that noise was characterized by high divergence and low norm while the norm was large and the divergence low in the case of a translation. Hence, a region of interest encompassing the lesion and the surrounding tissues should be selected to perform this analysis. In their evaluation, the authors showed that this method worked better than intensity-based methods when there was a mass effect without change in enhancement, although intensity-based methods performed slightly better when there was no mass effect.

Rey et al. [44] improved the approach of Thirion and Calmon [57] by using the Jacobian operator to determine local volume changes instead of using the divergence and norm of the vector fields. Furthermore, they used multi-resolution levels to avoid the influence of the motion in the center of a lesion by the vectors in the boundary. Using the Jacobian operator, it is possible to distinguish the lesion's evolution. As it is commonly accepted, the authors stated that a Jacobian operator larger than 1 indicates a local expansion, while smaller values indicate local shrinking. Furthermore, they can segment the lesions by using a threshold defined on the Jacobian operator (for instance, a threshold of 0.3 indicates significant shrinking). Actually, in their work, they only analyze shrinking lesions, due to the richer information when looking at the shrinking field and expanding areas more influenced by the spatial smoothing. Note that this is not a main drawback, since they use both information of the deformation field from old to new images as well as from new to old images. Comparing this algorithm with image subtraction, they demonstrated that the Jacobian operator was invariant to registration errors, although the algorithm gave poor results for segmentation.

Deformation field morphometry

Recently, Pieperhoff et al. [42] applied deformation field morphometry for the detection of local volume changes in Parkinson patients, although this algorithm could also be used to detect the evolution of MS lesions. The authors considered MR images as a 3D set of grid points, and they calculated the deformation vectors related to the grid points between the images which indicate shifted voxels in the source image (a deformed image to target image). Hence, they defined local volume ratio (LVR) as the volume of the deformed voxels in the source image divided by the volume of the non-deformed voxels in the target image. A local volume ratio greater than 1 shows a local increase and vice-versa. Subsequently, they created LVR-maps, which comprised the LVR values of all voxels. An LVR-map can be used in a ROI by adding up the LVR values of all the voxels. Furthermore, they compared LVR and the Jacobian determinant, and they reported that LVR gave smoother volume measures since the latter only considers four to six deformation vectors, whereas LVR is computed from 27 deformation vectors. Moreover, the Jacobian operator requires the calculation of partial derivatives, which usually introduces approximation computation problems.

Quantification approaches

As well as of performing the MS lesion detection and the change detection in MR images, the quantification process is also essential for radiologists and neurologists to analyze the patient follow-up. As already presented in "Lesion quantification," there are different ways to quantify the lesion evolution: *visual inspection*, *statistical change detection*, and *volumetric approaches*.

As shown in Table 1, the approaches based on lesion detection typically use volumetric approaches to quantify the lesion evolution. Metcalf's 4D connected component analysis [36], which uses a time domain on registered segmented images, may be the most common approach for this purpose (see, for instance, the following works which use this quantification approach [17, 21, 27, 66]). 4D connected component analysis provides the size and position of the lesions in a time line and is commonly used to identify individual lesions in time series.

In order to perform the quantification with temporal-based approaches, the outcome of the obtained images must be interpreted. Observing Table 1, it is clear that point-to-point subtraction methods commonly use visual inspection to detect active lesions and interpret the lesion evolution. For instance, Moraal et al. [39] detect positive activity by analyzing the bright area against a gray background. Furthermore, statistical intensity-based approaches use additional techniques to interpret the outcome images, for example, Lemieux et al. [31], who used a structured noise map (SNM) to identify lesion evolutions by comparing the outcome image with the SNM, and Bosc et al. [5], who used the generalized likelihood ratio test to avoid the drawbacks of direct manual visual inspection.

Regarding the deformation-based approaches, both Thirion and Calmon [57] and Rey et al. [44] used vector fields obtained from the non-linear registration step to identify the lesion evolution. Vector fields allow the displacement of the tissues and lesions to be more readily visible. For instance, Rey et al. [44] showed how the displacement field emphasizes a shrinking lesion while Thirion and Calmon showed the 3D deformation field measured between two volumetric MRI's of the same patient at the level of the lesion. Moreover, Thirion and Calmon also used the measurements of the volume variations to validate their method's accuracy by comparing with the conventional segmentation result. The approach of Rey et al. could also be used to segment lesions by defining a threshold. Therefore, volumetric analysis was also used to quantify the lesion evolutions.

Experimental validation

The experimental validation of brain MRI serial methods is not an easy task. The main problem when evaluating serial brain analysis remains the difficulty of obtaining a solid ground truth. Also, some of the automated methods do not provide a final quantitative result, but a processed image that is later displayed to the experts who provide the final diagnosis. In these cases, the experimental results usually evaluate the performance of the radiologists with and without using the software. In what follows, we explain the main steps that researchers follow to prepare the data and to evaluate their approaches.

Data preparation

The initial step needed to perform a validation of any algorithm is the selection of the cases. Depending on the aim of the validation a different subset of images may be necessary. For instance, if the accuracy of a segmentation algorithm is evaluated, only lesioned volumes are necessary, while lesioned volumes along with healthy controls are necessary to evaluate the performance of a lesion detection algorithm. Moreover, it is usually interesting to cluster the data according to the total lesion load in order to correlate this with the performance of the algorithms.

Reviewing the literature, a variety of MRI scanner machines is used, like the 2-T Bruker [5], the 1.5-T Phillips [1], the 1.5-T Siemens [55], or the most common one, the 1.5-T GE machine [16, 21, 37]. All these systems provide different fields of view (25.6 cm, 230 mm, 196×310 mm, 230×310 mm, etc.), different slice thicknesses (usually between 2 and 6 mm), and different sizes of the final image volume (256×256×54, 256×256×38, 162×256×20, 128×256×22, etc.). Moreover, different MRI modalities are acquired for each patient, typically T1-w, T2-w, PD-w, and FLAIR images (2D or 3D), which can be acquired from different views, usually axial or sagittal. This variety of inputs should be covered by the developed algorithm, which cannot be an easy task in terms of computational speed or in amount of memory used. The most common way to deal with this data is to construct a (virtual) 3D volume. Hence, in serial analysis where two or more volumes are analyzed at the same time, researchers use the term *4D dataset*, assuming that time is the fourth dimension.

Once the 3D volumes are obtained, they are still not ready for direct processing. As explained in "General problems in MRI," some inherent problems of the MRI data should be addressed before tracking the lesions. Bias-field correction, spatial co-registration, and intensity normalization are applied to correct for inter-scan intensity variations (due to scanner drift or other technical sources) and are usually applied for each 3D volume individually. Once these artifacts are minimized in both volumes, the registration step between the volumes is performed. New problems arise here, like different intensity normalization between the different volumes and issues caused by deformation artifacts that may be related

with the registration itself (repositioning) or with the voxel interpolation. Note that the brain extraction is usually performed after the registration step, in order to take advantage of the fact that the skull should be invariant in the different scans.

Ground-truth preparation

In general, there are two different ways of evaluating the approaches: with experiments using synthetic data or with experiments using real data. The use of a phantom brain (like the BrainWeb one [8]) provides an excellent framework to quantitatively evaluate the algorithms. However, it is well-known that synthetic data do not reproduce all the complex factors involved with real data, and algorithms working in these environments may fail when tested in real data. In contrast, the use of simulated lesions into real MRI scans provides a controlled ground-truth in a more realistic environment. For example, Thirion and Calmon [57] introduced spherical lesions with blurred contours that were obtained by averaging the intensities of real lesions, while Bosc et al. [5] used cubic lesions with Gaussian profiles obtained from real lesions in all the modalities they used. In contrast to these works that only introduce lesions in the new volumes, Rey et al. [44] suggested the addition of lesions in both old and new volumes to obtain a more realistic evaluation.

The common way of obtaining the ground-truth of real data is the accurate manual segmentation performed by at least one expert. If more than one expert segments the images, the final ground-truth will be more reliable [64]. For instance, the ground-truth used in Anbeek et al. [1] was first segmented by an expert, and then the manual segmentations were independently reviewed and corrected by two other experts, who were blinded for the clinical symptoms of the patients. Finally, the manual segmentation was re-evaluated in a consensus meeting and considered as a gold standard. Molyneux et al. [37] also noted that the potential for any memory of the images may introduce a systematic bias. Therefore, they suggested minimizing it by randomizing the scan order and ensuring a delay of at least 1 week between repeated measurements of the same scan.

One of the key points usually not considered in the approaches is related to the degree of difficulty of the data, which can be measured using the coefficient of variation (COV) between the annotations. The COV is the ratio of the standard deviation of the measurements to the mean and provides a measure to indicate the reproducibility of one strategy [74]. It is common to differentiate between:

- Inter-rater COV: variation of the results between different experts.
- Intra-rater COV: variation of the results at different times with same expert.

As an example, Zijdenbos et al. [75] presented a COV of 44% for the evaluation done with experts from seven different institutes. This value indicates that the image data they used was complicated and resulted in large variability even among the experts.

Moraal at al. [39] also noted the necessity of training the radiologists when performing the ground truth. First, the radiologists checked the image differences in healthy patients. Subsequently, they checked the difference of brains with MS lesions (not present in the testing set). Therefore, when they provided the ground-truth using their software, they looked for lesions individually, and finally they arrived at a consensus opinion.

Validation with ground-truth

The validation of an algorithm using ground-truth depends on its final aim. In many computer aided diagnosis systems [14, 15], the output is not the accurate segmentation of the lesion but the capacity of the algorithm to detect lesions. In these systems, the performance is computed using receiver operating characteristic (ROC) and free-response receiver operating characteristic (FROC) analysis. ROC analysis is performed at a case-level and is used to evaluate the capacity of the algorithm to distinguish between normal or abnormal (containing lesions) cases [10]. In contrast, FROC analysis is performed at region-level and plots the percentage of detected lesions against the number of false-positive regions detected. This analysis is useful when evaluating the performance of the algorithm to detect lesions [72]. In this latter analysis, a region of interest should be defined. For instance, Yamamoto et al. [71] assume that a lesion is detected when a single voxel is marked inside the lesion. On the other hand, to evaluate the performance of a segmentation algorithm, the most common computed measurements are the sensitivity, the specificity, and the Dice similarity coefficient [13], all of them computed at voxel-level. The sensitivity measures the percentage of well-detected voxels among all the lesions in the volume, the specificity is related to the capacity of an algorithm to avoid false positive voxels, while the Dice coefficient indicates the overlap between the automated and the manually delineated lesions (this measure is also known as similarity index [3]). Again, the COV coefficient may be used to compare both the automated and the manual obtained results.

However, in serial analysis, this quantitative evaluation is less important, since it does not quantify the effectiveness of the algorithm to track the lesion evolution. In this sense, the comparison between the result of the automated algorithm and the ground-truth in terms of absolute [75] and changed lesion [16, 37] volume may provide a more realistic evaluation of the algorithm. A reliable qualitative evaluation was performed by Bosc et al. [5], who visually evaluated their

algorithm using two experts and classified the automatic detected lesions in three different categories: valid lesion evolution, valid non-lesion evolution, or false detection. Although subjective, this evaluation provides a clear indication of how well the algorithm tracks the evolution of the lesions.

Validation without ground-truth

Due to the difficulty of obtaining ground-truth in their experiments, researchers developed different ways to demonstrate the consistency of their approaches. One of the most common ways to show the robustness of an algorithm is the scan–rescan validation, where the experiments are repeated several times to show the differences in the final result, which can be done with the COV coefficient (COV for reproducibility or also known as inter-scan COV). Note that, to correctly perform this validation, patients are removed from the MR room after the first scan and then repositioned in the MR machine by a different technologist [16, 21, 35].

Other ways to show the robustness of the approaches is through temporal coherency and sequence coherency, although some specific features are needed in both cases. Temporal coherency consists of checking the differences in lesion volume through the different explorations [57]. The idea is that the lesioned volume should not drastically change between two consecutive explorations in time (assuming no relapses in that time). This is analog to the evaluation of SIENAx [51], where the authors computed the error of their method for atrophy quantification in three-time exploration by checking if the tissue loss in t1–t2 added to the loss in t2–t3 were equal to the loss in t1–t3. On the other hand, the sequence coherency aims to compare the results of an algorithm when detecting lesions through the different MRI sequences independently [37]. However, those algorithms that rely on the analysis of a single sequence (i.e., FLAIR) or the use of two or more sequences together cannot perform this evaluation.

Discussion

Analysis of the reported results

Table 2 summarizes the results of the reviewed lesion detection algorithms in terms of reproducibility (comparison without ground-truth) and agreement with the experts (comparison with ground-truth). Note that the automatic segmentation methods obtain good reproducibility results, even more if we consider the fact that the MRI sequences are acquired again. Regarding the comparison with ground-truth, we can see that the work of Anbeek et al. [1] provides the highest performances in terms of DSC and sensitivity (voxel-wise computed). Notice also that most of the works provide specificity values close to 1. This is due to the fact that this measure evaluates the ratio between the numbers of voxels correctly classified as healthy divided by the total number of healthy voxels. Therefore, considering that lesions are small spots within the whole volume, the specificity value always tends to be close to 1 [32]. A different way to evaluate the performance of an algorithm is to use a region-wise measure instead of voxel-wise, as it is done in the work of Yamamoto et al. [71]. In this case, the sensitivity is computed as the number of detected lesions divided by the total number of lesions (81.5% in their work), and it is compared with the total number of false-positive lesions per volume or slice (2.9 per slice in [71]).

Looking at the results of the algorithms, clustering techniques perform better than conventional segmentation methods [16], and the use of additional strategies like PVEC or TDS [65] leads to increased accuracy of the approaches. Note that these strategies are based on introducing the experience of the expert into the algorithms, and hence, supervised segmentation methods perform better than unsupervised methods. Nevertheless, it should be considered that this additional information which either comes from a training set or from an anatomical template will bias the accuracy of the results.

On the other hand, Fig. 6 provides a comparison of the results obtained by different subtraction methods in terms of inter-observer agreement, detailing the results for positive and negative lesion detection. Tan et al. [55] investigated the lesion evolution from 26 patients using a 2D subtraction-based approach, while Moraal et al. [38] tested also the use of a 2D subtraction-based method using 46 pairs of MR images from 40 patients. In a later work, Moraal et al. [39] proposed and evaluated a 3D subtraction-based approach using controls of 14 patients. Comparing the results obtained by the 2D subtraction approaches, Moraal et al. [38] outperformed the results of Tan [55], mainly thanks to both the improvement in the registration algorithms and the use of an initial normalization step. However, the results of a similar strategy used with different data [38, 39] were drastically decreased. On the other hand, comparing 2D and 3D subtraction, one can see that the 3D subtraction outperforms the 2D approach, especially in the detection of negative activity. Furthermore, analyzing the results for each MRI sequence made it possible to see that the FLAIR sequence provided the best overall performance, while the use of the MP-RAGE sequence improves the detection of MS cortical lesions.

Improvements and further trends

Regarding the imaging modalities, the analysis of the approaches has shown that FLAIR discriminates well

Table 2 Summary of the results obtained by different lesion detection approaches

References	Methods	Data acquisition	Dataset	Measure	Results
[Udupa, 1997]	FCS	DE FSE T2-w/PD-w	20 MS patients	Avg. COV with FCM vs COV of 3 Experts without FCM	0.9% 22.6%
[Guttmann, 1999]	EM+PVEC	SE/DE-SE T1-w GE Signa 1.5 T	20×2 RR MS patients	Avg. LVE	0.05 cm3
[Kikinis, 1999]	EM+PVEC	SE/DE/longTR T1-w GE Signa 1.5 T	1 RR MS patient	COV of WML	39.5% vs 52.0%
[Wei, 2002]	EM+PVEC	DE SE PD-w/T2-w	11×2 CP MS	Avg. Inter-Scan COV	7.50%
		GE Signa 1.5 T	9×2 RR MS patients	Zscore	−2.84
	EM+TDS			Avg. inter-scan COV	2.57%
				Zscore	1.84
	EM+TDS+ PVEC			Avg. inter-scan COV	4.98%
				Z-score	−0.99
[Zijdenbos, 2002]	ANN-BP	T1-w/2D SE T2-w/PD-w	500×3;100×4	Avg. inter-scan COV	0%
		14 Hospitals North America	MS patients	Avg. CC with 7 rater	0.93
				Avg. Kappa (Dice)	0.60
[Ashton, 2003]	Bayesian (DMSS)	SE VE T1-w/T2-w/PD-w	10 dataset for Intra	Intra-rater COV	5.1% vs 1.5%
			1 dataset for Inter	Avg. inter-rater COV	16.5% vs 5.2%
[Ashton, 2003]	GEORG	SE VE T1-w/T2-w/PD-w	10 dataset for intra	Intra-rater COV	5.1% vs 1.4%
			1 dataset for inter	Avg. inter-rater COV	16.5% vs 2.3%
[Antel, 2003]	Bayesian	FFE T1-w	18 MS patients with FCD	Region-wise sensitivity	0.85%
				Voxel-wise sensitivity	0.2%
[Anbeek, 2004]	KNN	T1-w/T2-w/PD-w/ FLAIR/IR	18 MS patients	Avg. DSC	0.81%
				Avg. sensitivity	0.971%
				Avg. specifity	0.974%
[Wu, 2006]	KNN+TDS+ PVEC	DE-SE PD-w/T2-w	6 MS patients	Avg. sensitivity(T2L-BH)	0.70%–0.623%
		SE/T1c-w		Avg. specifity(T2L-BH)	0.987%–0.997%
[Duan, 2008]	SSGE	DE PD-w/T2-w MR	10×2 RR MS patients	Avg. inter-scan COV	0.98%
		GE Signa 1.5 T		Avg. LVE	1.50%
[Duan, 2008]	CSEG+PVEC	DE PD-w/T2-w MR	10×2 RR MS patients	Avg. inter-ican COV	8.64%
		GE Signa 1.5 T		Avg. LVE	11.40%
[Shiee, 2010]	FCM	T1-MPRAGE/FLAIR	10 MS patients	Avg. DSC	0.633%
				Avg. sensitivity	0.712%
[Yamamoto, 2010]	LS+SVM	T1-w FSE/T2-w/ FLAIR	3×2 MS patients	Avg. DSC	0.77%
[Cerasa, 2011]	CNN	FLAIR GE Signa 1.5 T	11 RR MS patients	Avg. DSC	0.64%

The datasets are defined by (number of patients)×(number of controls). If not specified, the measures are computed voxel-wise

The acronyms refer to: MRI sequences: *DE* dual echo, *SE* spin echo, *GE* gradient echo, *VE* variational echo, *FSE* fast spin echo, *FFE* fast field echo; patients: *CP* chronic progressive, *FCD* focal cortical dysplasia, *RR* relapsing remitting TPI traumatic brain injury; measures: *CC* correlation coefficient, *COV* coefficient of variation, *DSC* Dice similarity coefficient

between lesions and healthy tissue and is used in numerous approaches to perform the automated lesion segmentation and lesion evolution analysis [40]. 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.

As the MR images suffer from various image acquisition issues, pre-processing and post-processing steps play an important role for MS diagnosis and follow-up MS patients.

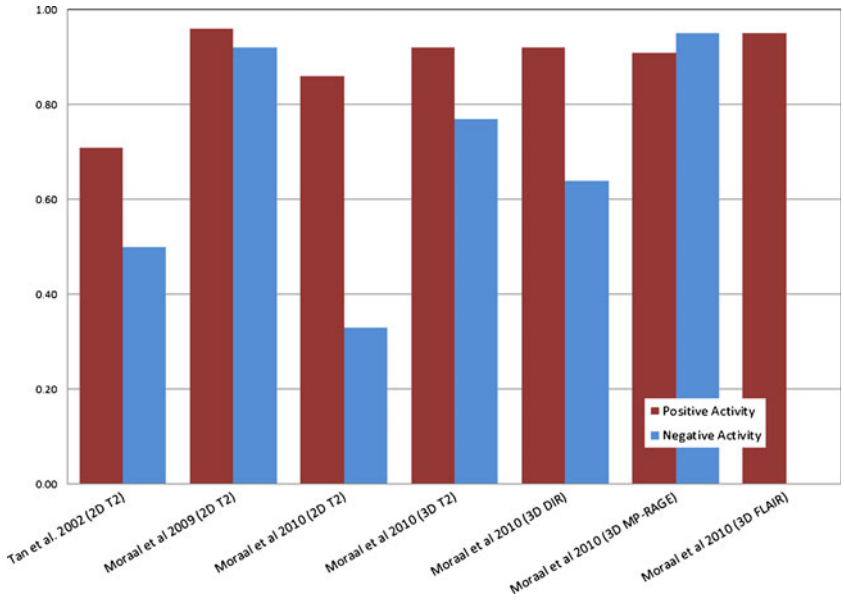

Fig. 6 Inter-observer agreement of the subtraction-based approaches. The performance of the algorithms according to the lesion activity is shown. 2D and 3D refers to the way the subtraction is performed

Therefore, bias field correction algorithms and global scaling of the images are commonly employed before registration. Besides, most of the approaches use a normalization algorithm particularly implemented for MR images (as for example the recent N4 algorithm [59]).

To perform a better comparison between images of different controls and particularly for the change detection algorithms, the registration is without any doubt the most important step. However, the registration procedure includes a re-sampling an interpolation process which may affect the images and the posterior measure of the lesion volume. Moreover, the lesions themselves, for instance, enlarged lesions or shrunken or resolved lesions, may affect the registration accuracy negatively. One possible way to reduce this miss-alignment caused by the lesion evolution is to use a similarity metric robust to local differences. For instance, mutual information or normalized mutual information, which are the most commonly used measures in multi-modal registration [22, 45, 54], can be used for the serial MRI registration to reduce the effects of the lesion evolution and other variations on the images which are caused by misalignments. The correlation ratio used for this purpose can be also a good choice for the serial MRI registration, since it can deal with intensity differences [76] and has been shown in some cases to be more robust than MI with respect to the initialization of registration [45]. In order to avoid residual artifacts caused by the registration, we have seen that some approaches used also the half-way registration method [5, 39], which is a robust way to avoid interpolation artifacts and consists of applying the same interpolation effect on both the fix and moving images. Notice that the type of interpolation method used is also important. For instance, using a spline interpolation will provide better results than using a linear interpolation method. Some authors also suggested the *sinc* interpolation for registering MR images while using a 3D pipeline [12, 41] since the frequency content of the MR images is strictly band-limited [12]. Therefore, it is suitable for a *sinc* interpolation. However, using a high-level interpolation method drastically increases the processing time with respect to the number of iterations and resolution. Thus, a linear interpolation method may be used within the iterations of the registration, while the principal interpolation method could be used within the last iterations or just for the final re-sampling process.

By analyzing the reported approaches, we have seen that the lesion detection and change detection techniques can be combined. In fact, this may help to carry out the diagnosis and follow-up of the patients at the same time and compensate for their inherent weaknesses. For instance, Duan et al. [16] combined a change detection algorithm based on subtraction of registered serial MR images with a detection algorithm based on a direct segmentation of the lesions. Rey et al. [44] proposed a uniform threshold over the Jacobian operator obtained from a deformation analysis to perform the lesion segmentation. Though they provided an experimental evaluation, the results were still far from a desired segmentation. Regarding these strategies which merge different methods, we believe that the quantification of the mass effect in vivo for the MS will be a new challenge for the near future. Besides, the selection of one MR image sequence (i.e. T1-w, T2-w, PD-w, and FLAIR) for specific purposes such as registration, detection, or segmentation, or the combination of some of them will have an important effect on the obtained results. In fact, combining the advance characteristics of the different MR image types is another important factor, which was also pointed out by Mortazavi et al. [40]. Some of the reviewed approaches have already applied multi-spectral algorithms which benefit from the different signal characteristics of the MR images.

We want to stress also that performing an exhaustive evaluation and comparison of the existing works is a very difficult task. The use of different data sets and different evaluation measures has been a major obstacle to reviewing these methods. Ideally, approaches should be applied to a common database and compared with a ground truth. This is, however, very difficult due to the lack of common public databases of real images along with several controls and their ground truth and the fact that the methods are not publicly available. Implementation of some significant works and a comparison with a common database will, without any doubt, provide a more objective comparison. However, integration of the expert knowledge and a proper setting of the algorithm parameters will be another important issue when trying to reproduce those results. As an example, Klein et al. [28] recently evaluated 14 different nonlinear deformation algorithms applied to human brain MRI registration. However, the work just focused on deformable registration, comparing a set of protocols rather than independent algorithms.

Conclusion

A review and classification of the classical and up-to-date approaches for automatic monitoring of MS lesion evolution has been proposed and discussed in this paper. These techniques, which have been classified according to their nature, are essential for the diagnosis and follow-up of MS patients from the MR images. Assessment of MS lesion evolution involves both detection and quantification of the lesion change. In accordance, we have also distinguished between lesion detection and lesion change detection techniques.

The lesion detection-based methods rely on using just a scan of a patient to detect the lesions, and a posterior quantification method may be used to determine the lesion evolution, which is usually carried out by using the total lesion volume between the image time-series. In this category, we have distinguished between supervised and unsupervised techniques, based on the use or not of a priori training of the algorithm. On the other hand, lesion change detection techniques make it possible to detect active lesions and interpret the lesion evolution at the same time. However, these algorithms cannot detect static lesions, since they need changed or deformed regions between the time-series. We have further sub-divided those strategies into two main categories: intensity-based and deformation-field based techniques, the former based on performing subtraction of successive scans while the latter allowing also to detect the mass effect of the lesions, which is an aspect overlooked by lesion detection and intensity-based methods, and which may be crucial for the MS patients.

Comparing different approaches and highlighting a single strategy is a difficult task due to the lack of a common database and a proper gold standard which prevents doing an exhaustive analysis. Furthermore, the setting of all the algorithm parameters and the integration of the expert knowledge are also important aspects to consider for a proper experimental validation. In this work, we have studied the reported results of all the analyzed automated MS lesion detection and quantification methods. We have seen that, for the lesion detection methods, the work of Anbeek et al. [1] was the most remarkable approach in terms of precision since they provided the highest values of DSC and sensitivity (voxel-wise computed). Other approaches have used a different way to evaluate the performance using region-wise measures instead of voxel-wise [71]. We have also seen that the precision of a proposal may be analyzed by considering the reproducibility and repeatability. In this case, the COV measure is a good way to indicate these two aspects. For example, among the lesion detection methods, the work of Zijdenbos et al. [75] had the best reproducibility and reliability since it provided the best COV value. On the other hand, among the change detection techniques the approach of Moraal et al. [39] provided the highest performances with respect to inter-observer agreements.

Summarizing, from the analysis done, we have seen that the lesion detection approaches are required for detecting static lesions and for diagnostic purposes, while either quantification of detected lesions or change detection algorithms are needed to follow up MS patients. In this latter case, deformation field-based algorithms allow the mass effect of the lesions to be detected, although analyzing all individual detected lesions is a time-consuming task and may not

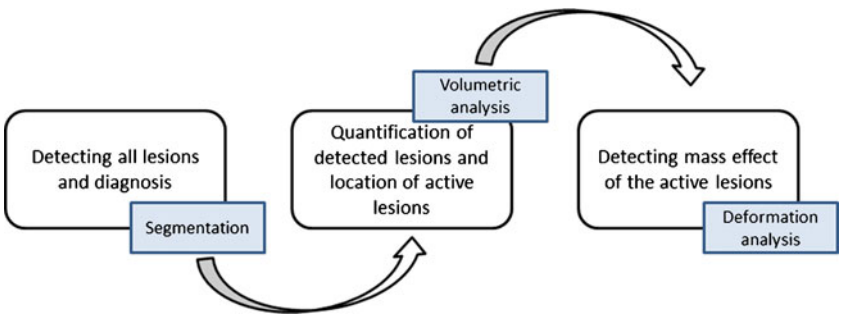

Fig. 7 A possible general framework for lesion detection and quantification. This framework allows the detection and tracking of evolving lesions computing their volumetry and mass effect

be necessary for the expert radiologists. Figure 7 shows an overall view of what we think could be a general framework for the time-series analysis of MS patients. Observe that this process involves the lesion detection, the volumetric quantification, and the deformation analysis, respectively. Notice that all MS lesions would be detected before deforming the MR images, while the regions of interest in the time-series (i.e., active lesions) and the quantification would be done by volumetric analysis and deformation analysis. Although this proposal may miss lesions which include only tissue deformation without any change in intensity, we believe it would be suitable for clinical practice.

Acknowledgments This study has been supported by the Instituto de Salud Carlos III Grant PI09/91018 and Grant VALTEC09-1-0025 from the Generalitat de Catalunya, and by the Fundació Esclerosi Múltiple de Catalunya through the CEM-Cat 2011 Grant Miquel Martí i Pol. O. Ganiler holds a FI grant.

Conflict of interest We declare that we have no conflict of interest.

References

1. Anbeek P, Vincken KL, van Osch MJ (2004) Automatic segmentation of different-sized white matter lesions by voxel probability estimation. Med Image Anal 8:205–215
2. Antel SB, Collins DL, Bernasconi N, Andermann F, Singhal R, Kearney RE, Arnold D, Bernasconi A (2003) Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis. IEEE Trans Med Imaging 19(4):1748–1759
3. Arimura H, Magome T, Yamashita Y, Yamamoto D (2009) Computer-aided diagnosis systems for brain diseases in magnetic resonance images. Algorithms 2(3):925–952
4. Ashton EA, Takahashi C, Berg MJ, Goodman A, Totterman S, Ekholm S (2003) Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI. J Magn Reson Imaging 17:300–308
5. Bosc M, Heitz F, Armspach J, Namer I, Gounot D, Rumbach L (2003) Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. NeuroImage 20 (2):643–656
6. Cabezas M, Oliver A, Lladó X, Freixenet J, Bach-Cuadra M (2011) A review of atlas-based segmentation for magnetic resonance brain images. Comput Methods Programs Biomed 104(3): e158–e177
7. Calcagno G, Staiano A, Fortunato G, Brescia-Morra V, Salvatore E, Liguori R, Capone S, Filla A, Longo G, Sacchetti L (2010) A multilayer perceptron neural network-based approach for the identification of responsiveness to interferon therapy in multiple sclerosis patients. Inf Sci 180(21):4153–4163
8. Brain Imaging Center M (2010) BrainWeb: simulated brain database. http://mouldy.bic.mni.mcgill.ca/brainweb/, last visit: 29/10/2011
9. Cerasa A, Bilotta E, Augimeri A, Cherubini A, Pantano P, Zito G, Lanza P, Valentino P, Gioia MC, Quattrone A (2011) A cellular neural network methodology for the automated segmentation of multiple sclerosis lesions. J Neurosci Methods In press
10. Chakraborty DP, Yoon HJ, Mello-Thomas C (2007) Localization accuracy of radiologists in free-response studies: inferring perceptual FROC curves from mark-rating data. Med Phys 14(1):4–18
11. Compston A, Coles A (2006) Multiple sclerosis. Lancet 359 (9313):1221–1231
12. Curati WL, Williams EJ, Oatridge A, Hajnal JV, Saeed N, Bydder GM (1996) Use of subvoxel registration and subtraction to improve demonstration of contrast enhancement in MRI of the brain. Neuroradiology 38:717–723
13. Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302
14. Doi K (2006) Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. Phys Med Biol 51(13):R5–R27
15. Doi K (2007) Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph 31(4–5):198–211
16. Duan Y, Hildenbrand PG, Sampat MP (2008) Segmentation of subtraction images for the measurement of lesion change in multiple sclerosis. Am J Neuroradiol 29:340–346
17. Ettinger GJ, Grimson WEL, Lozano-Perez T, III WMW, White SJ, Kikinis R (1994) Automatic registration for multiple sclerosis change detection. Int Proc. IEEE Work. Biomed. Image Anal., pp 297–306
18. Ge Y (2006) Multiple sclerosis: the role of MR imaging. Am J Neuroradiol 27(6):1165–1176
19. Gerig G, Martin J, Kikinis R, Kübler O, Shenton M, Jolesz FA (1992) Unsupervised segmentation of 3-D dual-echo MR head data. Image Vis Comput 10:349–360
20. Gerig G, Welti D, Guttmann CRG, Colchester ACF, Székely G (2000) Exploring the discrimination power of the time domain for segmentation and characterization of active lesions in serial MR data. Med Image Anal 4:31–42
21. Guttmann CRG, Kikinis R, Anderson MC, Jakab M, Warfield SK, Killiany RJ, Weiner HL, Jolesz FA (1999) Quantitative follow-up of patients with multiple sclerosis using MRI: reproducibility. J Magn Reson Imaging 9:509–518
22. Hill DLG, Batchelor PG, Holden M, Hawkes DJ (2001) Medical image registration. Phys Med Biol 46(3):R1–R45
23. Hillary FG, Biswal BB (2009) Automated detection and quantification of brain lesions in acute traumatic brain injury using MRI. Brain Imaging Behav 3:111–112
24. Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. Med Image Anal 5(2):143–156
25. Jenkinson M, Bannister PR, Brady JM, Smith S (2002) Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17(2):825–841
26. Juang LH, Wu MN (2010) MRI brain lesion image detection based on color-converted k-means clustering segmentation. Measurement 43:941–949
27. Kikinis R, Guttmann CRG, Metcalf D (1999) Quantitative follow-up of patients with multiple sclerosis using MRI: technical aspects. J Magn Reson Imaging 9:519–530
28. Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV (2009) Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46(3):786–802
29. Lao Z, Shen D, Liu D, Jawad AF, Melhem ER, Launer LJ, Bryan RN, Davatzikos C (2008) Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. Acad Radiol 15(3):300–313
30. Lee MA, Smith S, Palace J, Matthews PM (1998) Defining multiple sclerosis disease activity using MRI T2-weighted difference imaging. Brain 121:2095–2102

31. Lemieux L, Wieshmann U, Moran N, Fish D, Shorvon S (1998) The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization. Med Image Anal 2(3):227–242
32. Lladó X, Oliver A, Cabezas M, Freixenet J, Vilanova JC, Quiles A, Valls L, Ramió-Torrentà L, Rovira A (2012) Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches. Inf Sci 186(1):164–185
33. Maintz JBA, Viergever MA (1998) A survey of medical image registration. Med Image Anal 2(1):1–37
34. Martola J, Bergström J, Fredrikson S, Stawiarz L, Hillert J, Zhang Y, Flodmark O, Lilja A, Ekbom A, Aspelin P, Wiberg MK (2010) A longitudinal observational study of brain atrophy rate reflecting four decades of multiple sclerosis: a comparison of serial 1D, 2D, and volumetric measurements from MRI images. Neuroradiology 52(2):109–117
35. Meier DS, Guttmann CRG (2003) Time-series analysis of MRI intensity patterns in multiple sclerosis. NeuroImage 20:1193–1209
36. Metcalf D, Kikinis R, Guttmann CRG, Vaina L, Jolesz F (1988) 4D connected component labelling applied to quantitative analysis of MS lesion temporal development. In: Proc. IEEE Eng. Med. Biol. Society, pp 945–946
37. Molyneux P, Tofts P, Fletcher A, Gunn B, Robinson P, Gallagher H, Moseley I, Barker G, Miller D (1998) Precision and reliability for measurement of change in MRI lesion volume in multiple sclerosis: a comparison of two computer assisted techniques. J Neurol Neurosurg Psychiatry 65:42–47
38. Moraal B, Meier DS, Poppe PA (2009) Subtraction MR images in a multiple sclerosis multicenter clinical trial setting. Radiology 250:506–514
39. Moraal B, Wattjes MP, Geurts JJG (2010) Improved detection of active multiple sclerosis lesions: 3D subtraction imaging. Radiology 255(1)
40. Mortazavi D, Kouzani AZ, Soltanian-Zadeh H (2011) Segmentation of multiple sclerosis lesions in MR images: a review. Neuroradiology (in press)
41. Patriarche J, Erickson B (2004) A review of the automated detection of change in serial imaging studies of the brain. J Digit Imaging 17(3):158–174
42. Pieperhoff P, Sudmeyer M, Homke L, Zilles K, Schnitzler A, Amunts K (2008) Detection of structural changes of the human brain in longitudinally acquired MR images by deformation field morphometry: methodological analysis, validation and application. NeuroImage 43(2):269–287
43. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, Fujihara K, Havrdova E, Wilde MH, Kappos L, Lublin FD, Montalban X, O'Connor P, Sandberg-Wollheim M, Thompson AJ, Waubant E, Weinshenker B, Wolinsky JS (2011) Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 69(2):292–302
44. Rey D, Subsol G, Delingette H (2002) Automatic detection and segmentation of evolving processes in 3D medical images: applications to multiple sclerosis. Med Image Anal 6:163–179
45. Roche A, Malandain G, Pennec X, Ayache N (1998) The correlation ratio as a new similarity measure for multimodal image registration. In: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv., pp 1115–1124
46. Rovira A, León A (2008) MR in the diagnosis and monitoring of multiple sclerosis: an overview. Eur J Radiol 67(3):409–414
47. Rovira A, Swanton J, Tintor M, Huerga E, Barkhof F, Filippi M, Frederiksen JL, Langkilde A, Miszkiel K, Polman C, Rovaris M, Sastre-Garriga J, Miller D, Montalban X (2009) A single, early magnetic resonance imaging study in the diagnosis of multiple sclerosis. Arch Neurol 66(5):587–592
48. Shah M, Xiao Y, Subbanna N, Francis S, Arnold DL, Collins DL, Arbel T (2011) Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. Med Image Anal 15 (2):267–282
49. Shen S, Szameitat A, Sterr A (2008) Detection of infarct lesions from brain MRI images using inconsistency between voxel intensity and spatial location. A 3D automatic approach. IEEE Trans Inf Technol Biomed 12(4):532–540
50. Shiee N, Bazin PL, Ozturk A, Reich DS, Calabresi PA, Pham DL (2010) A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. NeuroImage 49 (2):1524–1535
51. Smith S, Zhang Y, Jenkinson M, Chen J, Matthews P, Federico A, Stefano ND (2002) Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage 17(1):479–489
52. Solomon J, Sood A (2004) 4-D lesion detection using expextation-maximization and hidden markov model. In: Proc. IEEE Int. Symp. Biomed. Imag., pp 125–128
53. Srivastava S, Maes F, Vandermeulen D, Paesschen WV, Dupont P, Suetens P (2005) Automatic detection of focal cortical dysplastic lesions. NeuroImage 27:253–266
54. Studholme C, Hill DLG, Hawkes DJ (1999) An overlap invariant entropy measure of 3D medical image alignment. Pattern Recogn 32(1):71–86
55. Tan IL, van Schijndel RA, Fazekas F, Filippi M, Freitag P, Miller DH, Yousry TA, Pouwels PJW, Adèr HJ, Barkhof F (2002) Image registration and subtraction to detect active T2 lesions in MS: an interobserver study. J Neurol 249(5):767–773
56. Tan IL, van Schijndel RA, van Walderveen MAA, Quist M, Bos R, Pouwels PJW, Desmedt P, Adèr HJ, Barkhof F (2002) Magnetic resonance image registration in multiple sclerosis: comparison with repositioning error and observer-based variability. J Magn Reson Imaging 15(5):505–510
57. Thirion JP, Calmon G (1999) Deformation analysis to detect and quantify active lesions in three-dimensional medical image sequences. IEEE Trans Med Imaging 18:429–441
58. Tian W, Zhu T, Zhong J, Liu X, Rao P, Segal BM, Ekholm S (2011) Progressive decline in fractional anisotropy on serial DTI examinations of the corpus callosum: a putative marker of disease activity and progression in SPMS. Neuroradiology (in press)
59. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC (2010) N4itk: improved N3 bias correction. IEEE Trans Med Imaging 29(6):1310–1320
60. Udupa J, Samarasekera S (1996) Fuzzy connectedness and object definition: theory, algorithms and applications in image segmentation. Graph Model Image Proc 58(3):246–261
61. Udupa JK, Wei L, Samarasekera S, Miki Y, van Buchem MA, Grossman RI (1997) Multiple sclerosis lesion quantification using fuzzy-connectedness principles. IEEE Trans Med Imaging 16 (5):598–609
62. van den Elskamp IJ, Boden B, Dattola V, Knol DL, Filippi M, Kappos L, Fazekas F, Wagner K, Pohl C, Sandbrink R, Polman CH, Uitdehaag BMJ, Barkhof F (2010) Cerebral atrophy as outcome measure in short-term phase 2 clinical trials in multiple sclerosis. Neuroradiology 52(10):875–881
63. Warfield SK, Kaus M, Jolesz FA, Kikinis R (2000) Adaptive, template moderated, spatially varying statistical classification. Med Image Anal 4:43–55
64. Warfield SK, Zou KH, Wells WM III (2004) Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 23 (7):903–921
65. Wei X, Warfield SK, Zou KH, Wu Y, Li X, Guimond A, Mugler JP III, Benson RR, Wolfson L, Weiner HL, Guttmann CRG (2002) Quantitative analysis of MRI signal abnormalities of brain white matter with high reproducibility and accuracy. J Magn Reson Imaging 15:203–209

66. Weiner HL, Guttmann CRG, Khoury SJ, Orav EJ, Hohol MJ, Kikinis R, Jolesz FA (2000) Serial magnetic resonance imaging in multiple sclerosis: correlation with attacks, disability, and disease stage. J Neuroimmunol 104:164–173
67. Wells WM III, Grimson WEL (1996) Adaptive segmentation of MRI data. IEEE Trans Med Imaging 15:429–443
68. Wells III WM, Grimson WEL, Kikinis R, Jolesz FA (1994) Statistical intensity correction and segmentation of MRI data. In: Proc. SPIE Conf. Visualization Biomed. Computing, pp 13–24
69. Welti D, Gerig G, Radü EW, Kappos L, Székely G (2001) Spatio-temporal segmentation of active multiple sclerosis lesions in serial MRI data. In: Proc. Int. Conf. Inform. Proc. Medical Imaging, pp 438–445
70. Wu Y, Warfield SK, Tan IL, Wells WM III, Meier DS, van Schijndel RA, Barkhof F, Guttmann C (2006) Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. NeuroImage 32(3):1205–1215
71. Yamamoto D, Arimura H, Kakeda S, Magome T, Yamashita Y, Toyofuku F, Ohki M, Higashida Y, Korogi Y (2010) Computer-aided detection of multiple sclerosis lesions in brain magnetic resonance images: false positive reduction scheme consisted of rule-based, level set method, and support vector machine. Comput Med Imaging Graph 34(5):404–413
72. Yoon HJ, Zheng B, Sahiner B, Chakraborty DP (2007) Evaluating computer-aided detection algorithms. Med Phys 34(6):2024–2038
73. Zacharaki EI, Kanterakis S, Bryan RN, Davatzikos C (2008) Measuring brain lesion progression with a supervised tissue classification system. Proc Int Conf Med Image Comput Comput Assist Interv 11:620–627
74. Zar J (1984) Measures of dispersion and variability. In: Zar J (ed) Biostatistical analysis, Prentice Hall, Englewood Cliffs, NJ. pp 27–39
75. Zijdenbos AP, Forghani R, Evans AC (2002) Automatic pipeline analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. IEEE Trans Med Imaging 21(10):1280–1291
76. Zitova B, Flusser J (2003) Image registration methods: a survey. Image Vis Comput 21:977–1000