A review on brain structures segmentation in magnetic resonance imaging

González-Villà, Sandra; Oliver, Arnau; Valverde, Segi; Wang, Liping; Zwiggelaar, Reyer; Lladó, Xavier

Published in:
Artificial Intelligence in Medicine

DOI:
10.1016/j.artmed.2016.09.001

Publication date:
2016

Citation for published version (APA):
González-Villà, S., Oliver, A., Valverde, S., Wang, L., Zwiggelaar, R., & Lladó, X. (2016). A review on brain structures segmentation in magnetic resonance imaging. *Artificial Intelligence in Medicine*, *73*, 45-69. https://doi.org/10.1016/j.artmed.2016.09.001

Title: A review on brain structures segmentation in magnetic resonance imaging

Author: Sandra González-Vill*à* Arnau Oliver Sergi Valverde Liping Wang Reyer Zwiggelaar Xavier Lladó

PII: S0933-3657(16)30115-4
DOI: http://dx.doi.org/doi:10.1016/j.artmed.2016.09.001
Reference: ARTMED 1480

To appear in: *ARTMED*

Received date: 1-4-2016
Revised date: 27-7-2016
Accepted date: 5-9-2016

Please cite this article as: Sandra González-Vill*gravea*, Arnau Oliver, Sergi Valverde, Liping Wang, Reyer Zwiggelaar, Xavier Lladó, A review on brain structures segmentation in magnetic resonance imaging, *<![CDATA[Artificial Intelligence In Medicine]]>* (2016), http://dx.doi.org/10.1016/j.artmed.2016.09.001

Identifying the brain structures is a key aspect for cognitive disease diagnosis.

We present a review of automatic brain structures segmentation methods.

We provide a classification according to the strategy used to segment the structures.

Qualitative and quantitative results are presented and discussed.

Future trends should combine multi-atlas with learning-based or deformable approaches

A review on brain structures segmentation in magnetic resonance imaging

Sandra González-Villà[a], Arnau Oliver[a], Sergi Valverde[a], Liping Wang[b], Reyer Zwiggelaar[b], Xavier Lladó[a]

[a]*Institute of Computer Vision and Robotics, University of Girona.*
Ed. P-IV, Campus Montilivi, University of Girona, 17071 Girona (Spain).
[b]*Department of Computer Science, Aberystwyth University, Aberystwyth SY23 3DB, Wales (UK).*

Abstract

Background and objectives: Automatic brain structures segmentation in magnetic resonance images has been widely investigated in recent years with the goal of helping diagnosis and patient follow-up in different brain diseases. Here, we present a review of the state-of-the-art of automatic methods available in the literature ranging from structure specific segmentation methods to whole brain parcellation approaches.

Methods: We divide first the algorithms according to their target structures and then we propose a general classification based on their segmentation strategy, which includes atlas-based, learning-based, deformable, region-based and hybrid methods. We further discuss each categorys strengths and weaknesses and analyze its performance in segmenting different brain structures providing a qualitative and quantitative comparison.

Results: We compare the results of the analyzed works for the following brain structures: hippocampus, thalamus, caudate nucleus, putamen, pallidum, amygdala, accumbens, lateral ventricles, and brainstem. **The structures on which more works have focused on are the hippocampus and the caudate nucleus. In general, the accumbens (0.69 mean DSC) is the most difficult structure to segment whereas the structures that seem to get the best results are the brainstem, closely followed by the thalamus and the putamen with 0.88, 0.87 and 0.86 mean DSC, respectively. Atlas-based approaches achieve good results when segmenting the hippocampus (DSC between 0.75 - 0.90), thalamus (0.88 - 0.92) and lateral ventricles (0.83 - 0.93), while deformable methods perform good for caudate nucleus (0.84 - 0.91) and putamen segmentation (0.86 - 0.89).**

Conclusions: There is not yet a single automatic segmentation approach that can emerge as a standard for the clinical practice, providing accurate brain structures segmentation. Future trends need to focus on combining multi-atlas methods with learning-based or deformable approaches. Employing atlases to provide spatial robustness and modeling the structures appearance with supervised classifiers or Active Appearance Models could lead to improved segmentation results.

Keywords: review, automated segmentation methods, brain structures, magnetic resonance imaging

1. Introduction

Magnetic resonance imaging (MRI) of the brain has become standard tool in medical practice for diagnosis [1], disease follow up [2], treatment evaluation [3] and brain development monitoring [4]. As it is non-intrusive, painless, fast to acquire and provides good contrast between tissues it provides the best choice for a range of clinical application areas.

Brain segmentation is very useful for clinical analysis since a qualitative evaluation of brain morphological characteristics is very subjective and therefore quantified techniques are needed. Segmentation is one of the fundamental problems in biomedical image analysis, which refers to the process of assigning a label with biological meaning to each pixel or voxel in such a way that the pixels or voxels with the same label share certain characteristics or belong to the same anatomical region. Image segmentation is a well known and extensively studied problem in computer vision for which a huge amount of approaches have been proposed in recent years, e.g. edge based approaches [5], region growing approaches [6], clustering techniques [7] or graph cuts [8]. All these segmentation methods have many applications in medical imaging, but they have also been useful in other disciplines such as recognition [9], video surveillance [10] and machine vision [11].

Although manual segmentation of brain MRI is still a highly used practice, it suffers from important short-

comings. First of all, manual segmentations are poorly reproducible and are subject to inter- and intra- operator variability. More importantly, it is a very time-consuming task and, given the increasing amount of data to analyze, the capacity of expert visual analysis is being exceeded. As such, the need for accurate automatic segmentation methods has emerged in recent years [12–16].

Initial automatic methods were focused on segmenting the brain MRI into three different tissues, namely white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF). Most of these methods relied only on signal intensity in T1-weighted images, where there is a clear difference in its intensity distributions of these tissues. However, the task of brain structure segmentation is not that trivial, since its segmentation cannot be performed based only on image intensities because there is too much overlap between the class distributions and the structures boundaries are not always strong enough. Hence, more information such as shape, location in the brain or the relative position among structures has to be incorporated in the segmentation algorithms.

The automatic segmentation of the hippocampus has received significant attention in the last years since it has been demonstrated to be an important biomarker in many clinical applications such as Alzheimer's disease. In the recent work of Dill et al. [17] a review of the evolution of automated methods for the segmentation of the hippocampus in MRI was presented, whereas [18] covered a quantitative comparison of four automatic methods to segment the hippocampus in patients with mesial temporal lobe epilepsy. Another group of structures to which the community has paid attention is the deep gray matter or subcortical structures. Several approaches [15, 19, 20] have been presented focused on automatically segmenting this group of structures, and a recent study reviewed automatic and semi-automatic methods [21]. In addition, Babalola et al. [22] quantitatively evaluated four different algorithms for the task of subcortical brain structure segmentation. Other works in brain structure segmentation have been presented recently such Devi et al. [23], who reviewed several works for neonatal brain segmentation in MRI and Iglesias et al. [24] who gave an overview of the state-of-the-art in multi-atlas segmentation of biomedical images. Besides, Klein et al. [25] quantitatively compared 14 non-linear registration algorithms which were evaluated based on brain structures segmentation. As far as we know, there is no article which reviews the state-of-the-art and describes automatic methods not concentrating on an exclusive segmentation strategy and that either segment a single structure or the whole brain in MR images.

In this paper, we present a review of such methods and classify them according to the segmentation strategy used, for which we propose a classification that includes atlas-based, learning-based, deformable, region-based and hybrid categories. Furthermore, to the best of our knowledge, this paper is the first attempt to review the most relevant works in brain structures segmentation that also presents an analysis of the state-of-the-art results, showing different evaluation measures, databases and number of testing cases, both from the point of view of segmentation strategies and segmented structures.

We searched the literature in the following databases: PubMed, Scholar, IEEE Xplore and Scopus. The main search strategy combined four concepts: Brain structures, Magnetic Resonance Imaging, segmentation and automatic methods. From the initial search we discarded the works that were out of the scope of this review, such as tissue segmentation methods, approaches that used contrast MRI or diffusion tensor images, while keeping only the works that strictly proposed an automatic approach for brain structures segmentation in structural MRI. To retrieve other relevant publications, we also examined the reference lists of the selected publications, and we included those works that were related to our aim.

2. Clinical applications

Psychiatric and neurodegenerative disorders are frequently associated with structural changes in the brain, such as variations in the volume or shape of the deep gray matter structures or in the thickness, area and folding pattern of the cortical regions [26]. Because of that, the morphometric analysis of brain structures can be used as an important biomarker of the disease or even as a diagnostic test [27]. Other applications of MRI brain structures segmentation may include pre-operative evaluation and surgical planning [28] for situations in which the procedure requires high accuracy, such as deep brain stimulation [29] or ablation of the appropriate functional regions; longitudinal monitoring for disease progression or remission [30]; or radiotherapy treatment planning [31]. In this section we briefly review the clinical applications of some brain structures segmentation such as the hippocampus, caudate nucleus, thalamus, pallidum or brainstem. Table 1 provides a summary of the relation of such structures with different diseases and the consequent structure abnormalities.

Table 1: Clinical applications. Brain structure abnormalities associated with various diseases.

Structure	Implied Disease	Abnormality
Hippocampus [32, 35]	Alzheimer	Atrophy [34]
	Temporal lobe epilepsy	Asymmetric atrophy [33]
	Posttraumatic stress disorder	Reduced volume [36, 37]
	Major depression	Reduced volume [38]
	Schizophrenia	Reduced volume [39]
	Bipolar disorder	Non-conclusive volume difference [40, 41]
Brainstem [45, 46]	Progressive supranuclear palsy	Atrophy in midbrain and pons [42–44]
	Parkinson	Reduced nigral volume [47]
	Alzheimer	Reduced volume and structure deformation [48]
Caudate [50–52]	Huntington's disease	Atrophy [49]
	Tourette syndrome	Reduced volume [53]
	Autism	Increased right volume [52]
	Attention deficit hyperactivity disorder	Reduced right volume [54]
	Fragile X syndrome	Increased volume [55]
Thalamus [57, 58]	Multiple sclerosis	Atrophy [56]
	Alzheimer	Atrophy [59]
	Schizophrenia	Non-conclusive volume difference [60]
	Parkinson	Reduced volume [61]
Corpus Callosum [63, 64]	Multiple sclerosis	Atrophy [62]
	Schizophrenia	Reduced volume [65]
	Autism	Reduced volume [63]
	Alzheimer	Atrophy [63]
	Multi-Infarct dementia	Atrophy [66]
Amygdala [35, 68]	Schizophrenia	Reduced volume [67]
	Anxiety disorders	Reduced left volume [69]
	Bipolar disorder	Non-conclusive volume difference [39, 41, 70, 71]

The hippocampus plays an important role in human memory and orientation. Its atrophy has been shown to be a predictive biomarker for patients with mild cognitive impairment and Alzheimer's disease, but it has also been related to other diseases such as schizophrenia, major depression, bipolar disorder, post-traumatic stress disorder, etc. [32]. Asymmetric atrophy of this structure has also been demonstrated to be a good predictor of epilepsy [33].

The brainstem, which is usually described as including the medulla oblongata, pons, and midbrain (red nucleus and substantia nigra), is especially relevant to primary tauopathies such as progressive supranuclear palsy [45], in which brain atrophy occurs in the midbrain and pons, or corticobasal degeneration. The importance of the brainstem substructures in other diseases has also been reported, such as in [72] where the authors proposed a technique for supporting the clinical diagnosis of Parkinson's disease; they claimed that the initial assessment of the neurological condition of a patient should be performed by estimating the area of the substantia nigra [73]. Alzheimer's disease, is another degenerative disease that also affects brainstem structures [46].

The diminished right caudate volume is one of the most replicated findings among attention deficit hyperactivity disorder patients and hence, the ratio between right caudate volume and the bilateral caudate volume is applied as a diagnostic test [27]. Aberrant morphology and function of the caudate nucleus have also been implicated in a number of important brain disorders, including Huntington's disease [50], Tourette syndrome [51], autism [52, 74, 75], attention deficit hyperactivity disorder [54, 76], and fragile X syndrome [55, 77].

The thalamus is associated with a wide range of clinical manifestations including cognitive decline, motor deficits, fatigue, painful syndromes, and ocular motility disturbances in patients with multiple sclerosis. It has also been demonstrated that the atrophy of deep gray nuclei is closely related to the magnitude of inflammation [57]. As stated before, the surgical treatment for many movement disorders, such as essential tremor, Parkinson's disease, drug-resistant epilepsy as well as chronic pain syndromes, involves ablation or electric stimulation of the appropriate functional region within inner-brain structures such as the subthalamic nucleus and globus pallidus [78]. Surgical planning for these procedures is often based on preoperatively acquired MR images, thus segmenting the implied regions would improve the planning and guidance of the surgery.

The corpus callosum is also an important structure due to its vulnerability to environmental toxins,

white matter diseases (such as multiple sclerosis) and schizophrenia [79]. Effects on regional callosal structure have been reported in attention deficit hyperactivity disorder, Alzheimer's disease, multi-infarct dementia, and a range of neurodevelopmental disorders and dysplasias [63].

3. Methods

Looking at the literature, we have seen that among automatic brain structures segmentation methods, some aim at parcellating the whole brain but, the vast majority are centered on segmenting a few or even only one specific structure. **Tables 2 to 4 show in their 'C' column, for all the reviewed methods that include segmentation results, either graphically or numerically, the type of method we are referring according to the following criteria: (1) methods which parcellate the whole brain, (2) methods that segment a group of structures such as subcorticals, basal ganglia or those that, according to their results, have been demonstrated that their method can be extended to several structures, (3) structure specific methods, and (4) methods that segment a specific structure and its substructures.**

In this section, we provide a classification of the state-of-the-art methods according to the strategy used to segment the target structures, for which we mainly distinguish four categories: atlas-based, learning-based, deformable and region-based strategies. A fifth category which combines some of these four approaches is also included in our classification.

3.1. Atlas-based methods

In the context of image segmentation, an atlas is defined as the combination of two image volumes: one intensity image (or template) and one segmented image (or labeled image). As stated in Cabezas et al. [80], an atlas can be either topological or probabilistic. Topological or deterministic atlases consist of a single subject volume together with its corresponding, often manual, segmentation. Probabilistic or statistical atlases are constructed on the basis of populations, co-registering all the segmented cases to a standard space and computing the frequency of each voxel to belong to a specific structure.

While the methods presented in this section are all based on topological atlases, statistical atlases can also be used as prior information in statistical image segmentation algorithms, as we will see later in section 3.2.2.

Table 2 shows a summary of the methods found in this category that provided either graphical or quantitative experimental results. The table indicates also the target structures of each method based on the results presented.

3.1.1. Label propagation

Probably the most straightforward principle to automatically segment an image using a single training data set (henceforth an atlas) is label propagation. The basic idea of this technique is to spatially map or deform (i.e. register) the atlas image to the volume we want to segment (target image). This registration produces a deformation field that can be used to propagate (i.e. warp) the atlas labels to this new volume in order to get the final segmentation.

To register two images, it is necessary to find a spatial transformation, mapping the content of one image to the corresponding area of the other in such a way the image similarity is maximized [81]. According to this definition, several works have been presented in the last years focused on label propagation that differ on: the image parts considered to perform the alignment (feature-based vs intensity-based), the transformation function employed (parametric vs non-parametric) or the measure used (sum of squared differences, cross correlation, mutual information (MI), landmark distances, etc.) which defines how similar both images are. Collins and Evans [12] proposed a popular intensity-based registration strategy, called ANIMAL (Automatic Nonlinear Image Matching and Anatomical Labeling), which performed at different spatial scales, starting with very blurred data (where only major structures are apparent, such as temporal lobe, ventricles and longitudinal fissure) and increasing details at each step by using less blurred images, refining the registration at each step. The cross-correlation metric was used as a similarity measure and the deformation was constrained to consist of a linear combination of smooth basis warps that were defined by discrete cosine transforms, performing local translations. Alternatively, Wang and Vemuri [82] employed B-splines to perform the deformation with a previously introduced metric [83], called cross cumulative residual entropy, as a similarity measure.

Combining geometric and intensity features for registration should result in more robust methods. This is actually of current interest and we have seen several methods combining intensity-based and feature-based criteria to establish more accurate correspondences in difficult registration problems [84]. Shen and Davatzikos [85] presented an elastic registration algorithm, called HAMMER (Hierarchical Attribute Matching Mecha-

nism for Elastic Registration), that applied deformation to image sub-volumes rather than voxels, based on the similarity of attribute vectors over the whole subvolume. These attribute vectors consisted of three individual components: edge type (tissue), intensity and geometric moment invariants, being a more robust way to establish anatomical correspondences in the deformation procedure than considering only a measure derived directly from the intensity.

According to the transformation function, which defines how an image is deformed to match the other, we can mainly distinguish between rigid or non-rigid transformation, which range from smooth regional variation described by a small number of parameters to dense displacement fields defined at voxel level [84]. Klein et al. [25] presented an evaluation of thirteen non-rigid registration algorithms and stated, corroborating Hellier's evaluation [86], that there was a modest correlation between the number of degrees of freedom of the deformation algorithm and the registration accuracy. Similarly, Carmichael et al. [87] stated in their work that registration methods that produced higher degrees of geometric deformation, produced automated segmentations with higher agreement with manual annotations.

Another common technique is to perform first a coarse registration, and refine in a second step the result with another registration method [20, 78]. This is the case for BrainSuite [88]. In its brain labeling tool (SVReg [89]), subject and atlas surfaces are first smoothed and coarsely aligned in 3D space [90]. After that, a curvature-based alignment, followed by volumetric spatial alignment is performed and, once cortical features are aligned, as subcortical features trend to be misaligned, an intensity-based registration refinement is done. In a similar way, Postelnicu et al. [91] combined a feature-based with an intensity-based non-rigid registration method. They first aligned cortical folding patterns and using the resulting deformation as initialization they aligned subcortical regions, while preserving the cortical alignment. With a similar idea, Luo and Chung [92] presented a method to segment the subcortical structures, for which they first obtained a coarse structure-by-structure segmentation by means of affine registrations, exploiting the spatial dependency relations between the deep brain structures to determine the segmentation order. In a second step, the segmentation result was refined performing a non-rigid registration, that used information about the histogram of the gradient magnitudes lying on the structure boundaries. Following also a two step registration (affine + non-rigid surface-based), Iacono et al. [78] presented a method to segment the internal part of the globus pallidus by means of registering an ultra-high resolution atlas (7T MRI), in which this structure was well defined, to the target image. Yousefi et al. [93] compared different strategies to segment subcortical structures based on different registration methods, combining affine and non-rigid registration applied to all brain and subcortical area. They concluded that the best results were those obtained by means of an affine transformation applied to the entire brain area, followed by a deformable transformation applied only to the subcortical structures.

3.1.2. Label fusion

In general, label propagation suffers from two main drawbacks, which are the fact that a simple atlas cannot sufficiently represent the whole population of potential testing data and that the performance and quality of the obtained results are limited by the accuracy of the pairwise registration method. As an attempt to solve the inherent problems associated with label propagation, label fusion techniques have been extensively developed in recent years. This approach, also known as classifier fusion or multi-atlas segmentation, consists of registering each training subject (i.e. atlas) to the test subject separately so that each atlas label is propagated to the target image space in the same way as in label propagation. Once all these transferred labels are obtained, they are fused to generate a segmentation result of the target image. Across-subject anatomical variability is better captured here than with a single atlas, and the registration error for a particular propagated atlas is less likely to affect the final segmentation when combined with other atlases [94].

Significant research has been done on multi-atlas segmentation with regard to the influence of several factors that affect the final segmentation such as the atlas selection, the best number of atlases involved in the segmentation or the fusion strategy used. Lötjönen et al. [95] developed and compared different similarity measures, atlas-selection strategies and methods to combine multi-atlas segmentation and intensity modeling. They demonstrated that all these factors play an important role in multi-atlas segmentation and optimizing them is clearly reflected in the brain structures segmentation accuracy.

Several atlas selection strategies have been studied in recent years. Wang et al. [96] first built a graph including all the atlases and the target image and once the graph was built, they grouped the atlases in different clusters by searching the shortest path from each atlas to the target. Finally, they chose from each resulting cluster the atlas with the shortest path to the target to perform the label fusion. Collins and Pruess-

ner [97] also selected the best samples for a given subject from the atlas database, but they used normalized mutual information. Aljabar et al. [94] compared different atlas selection strategies such as image similarity, segmentation similarity or demographics and concluded that image-based selection provided better segmentations than random subsets. Regarding the number of atlases used to perform the fusion, Heckemann et al. [98] showed that the segmentation accuracy increased in a logarithmic way when new random atlases were included in the label fusion up to a limiting value, while Aljabar et al. [99] found that beyond a certain number of ranked atlases (based on a similarity criteria such as image similarity or age-based selection) involved in the segmentation (15-25 depending on the structure) the accuracy of the resulting segmentation decreased. On the other hand, Pipitone et al. [100] came across in their experiments that when the number of templates was set to an even number, the segmentation performance slightly decreased.

Regarding the fusion strategy, we can mainly distinguish between global and local weighting methods. In global combination strategies, the weight of the contribution of each atlas to the segmentation is the same for every voxel. The simplest global strategy is majority voting, which weights each candidate segmentation equally and assigns to each voxel the label that most segmentations agree on. In spite of its simplicity, it has been shown to result in highly robust segmentations [94, 97, 98, 100, 101]. Another commonly used fusion method is weighted voting, which can be performed either globally [102] or locally [103]. In this strategy, larger weights are assigned to the atlases that show higher similarity to the target image.

Local weighting methods exploit the fact that different atlases may have achieved a good registration at different parts of the target image, and so it makes sense to borrow labels from different atlases at different target locations. A widely used local combination strategy is the Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm [13], which weights each segmentation based upon its estimated performance level with respect to other available candidate segmentations. It treats image label fusion as a maximum-likelihood problem, which is solved using Expectation Maximization (EM). Several authors have published STAPLE reformulations [104–107] that include different advances over the original framework. In [108], Artaechevarría et al. studied the performance of different weighting methods: either globally, using similarity measures for the whole volume, or locally, using a small neighborhood area, concluding that local methods should be preferred in regions that show high contrast with neighbor areas, while global methods should be used in regions that show similar intensities to the surrounding structures. They also stated that there is not a single method that emerged as the best for all regions and images.

In non-local label fusion [109–112], the labels of all the atlas voxels in the neighborhood of the target voxel have a weight in its label assignment based on their similarity. By means of local search windows, the one-to-one mapping constraint existing in traditional local weighting methods is relaxed. Due to the fact that they explore the neighborhood of each voxel, the registration does not need to be precise, hence it is possible to perform a linear registration instead of non-rigid deformation. Following this idea, Rousseau et al. [109] proposed a patch-based framework based on the construction of a weighted graph of non-local similarities that linked together voxels in the target image and the corresponding neighbor voxels in the atlases. They studied several patch aggregation strategies and also tested the influence of several parameters (neighborhood and patch sizes, number of atlases) in the final segmentation accuracy. Coupé et al. [110] also presented a non-local patch-based segmentation strategy which was conceptually very similar to that of Rousseau et al., differing basically on the patch preselection. In their approach, they first performed an atlas preselection based on the sum of squared differences. Once the best atlases were selected, for each voxel of the target image they discarded the corresponding atlas voxels that were not going to contribute in the label fusion based on their dissimilarity (luminance and contrast). The remaining voxels contributed to the weighting based on their intensity similarity. In a similar way, Wu et al. [111] recently presented a patch-based multi-atlas segmentation strategy in which they introduced a multi-scale image patch that combined both local and global information. They proposed to dynamically adjust the patch size from large to small during the label fusion procedure, using the global image information to remove the misleading candidate atlas patches and then gradually using more local information to refine the label fusion result.

Recently, several brain structures segmentation methods using graph-based or tree-based intermediate templates guided registration have been presented, achieving effective segmentation results [96, 98]. This strategy is based on the principle that it is generally difficult to obtain accurate registration between images with large shape differences and thus, these methods try to decompose a large registration into smaller ones with the help of intermediate templates. Jia et al. [103] introduced a multi-atlas-based multi-image segmentation

Table 2: Atlas-based methods. Acronyms from left to right are: hippocampus (HIP), thalamus (THA), caudate nucleus (CAU), putamen (PUT), pallidum (PAL), amygdala (AMY), accumbens (ACC), lateral ventricle (LV), brainstem (BS), corpus callosum (CC), cerebellum (CB), white matter (WM), cortical gray matter (CGM) and cerebrospinal fluid (CSF). Measures in order of appearance: Kappa index (KI), Jaccard index (JC), Relative overlap (RO), Dice similarity coefficient (DSC), Relative mean squared error (MSE), Relative volume (RV), False negatives (FN), False positives (FP), Hausdorff distance (HD), Similarity index (SI), Mean absolute distance (MAD), Precision (P) and Recall (R). Diseases in order of appearance: Normal controls (NC), Alzheimer's disease (AD), Clinical dementia (CD), Mild cognitive impairment (MCI), Probable Alzheimer's disease (PAD) and First episode psychosis (FEP). The ✓ and × symbols stand for numerical and graphical results, respectively, whereas the ⊛ symbol means that the method perform the internal substructures segmentation of the indicated structure. The - symbol indicates that no results have been reported for that particular structure. Column 'C' indicates the segmentation target: (1) whole brain; (2) a group of structures; (3) a single structure; (4) a single structure and its substructures. Column 'Ref' shows the reference work from which the results are obtained (in case they do not come from the original work).

				Segmented structures																	
	Article	Ref	C	HIP	THA	CAU	PUT	PAL	AMY	ACC	LV	BS	CC	CB	WM	CGM	CSF	Others	Measure	Database	Disease
Label propagation	Collins (1997) [12]	[97]	1	✓	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	KI;JC	80v (ICBM)	NC
		[114]		✓	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	KI	30v (ICBM); 10v (ICBM)	NC
	Shen (2002) [85]	[91]	1	✓	✓	✓	✓	✓	✓	-	✓	✓	-	-	✓	✓	-	✓	JC	11v (IBSR); 36v (Desikan et al.)	AD+NC
	Shattuck (2002) [88]		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Wang (2005) [82]		2	×	-	-	-	-	-	-	×	-	×	-	-	-	-	-	-	-	-
	Postelnicu (2009) [91]		1	✓	✓	✓	✓	✓	✓	-	✓	✓	-	-	✓	✓	-	✓	JC	11v (IBSR); 36v (Desikan et al.)	AD+NC
	Lin (2010) [20]		2	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	KI;RO	15v (IBSR)	NC
	Luo (2011) [92]		2	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	DSC	9v (IBSR)	NC
	Iacono (2011) [78]		4	-	-	-	-	⊛	-	-	-	-	-	-	-	-	-	-	RMSE	1v (MNI152)	NC
	Yousefi (2012) [93]		2	×	×	×	×	×	×	×	×	×	-	-	-	-	-	-	DSC;RV;FN;FP	18v (IBSR)	NC
				×	-	×	×	-	-	-	-	-	-	-	-	-	-	-	DSC;RV;FN;FP	40v (LPBA40)	NC
	Joshi (2012) [89]		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	HD	6v	-
Label fusion	Warfield (2004) [13]	[115]	1	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	✓	✓	-	-	DSC	39v (FS atlas)	AD+CD+NC
		[108]		✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	SI; MAD	17v (IBSR)	NC
	Heckemann (2006) [98]		1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	✓	SI	30v	NC
	Aljabar (2007) [94]		2	×	×	-	-	-	-	-	×	-	-	-	-	-	-	-	DSC	275v (CMA)	-
	Aljabar (2009) [99]		2	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	DSC	275v (CMA)	-
	Artaechevarría (2009) [108]		1	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	SI; MAD	17v (IBSR)	NC
		[104]		✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	DSC	30v (Hammers)	NC
		[104]		✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC	30v (ADNI)	AD+MCI+NC
		[109]		✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	DSC	18v (IBSR)	NC
	Lötjönen (2010) [95]		2	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	SI; P; R	18v (IBSR)	NC
				✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	SI	60v (ADNI)	AD+MCI+NC
		[109]		✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	DSC	18v (IBSR)	NC
	Collins (2010) [97]		2	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	KI; JC	80v (ICBM)	NC
	Heckemann (2010) [116]		1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	✓	JC	30v (Hammers)	NC
		[104]		✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	DSC	30v (Hammers)	NC
	Wolz (2010) [113]		3	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC	796v (ADNI)	AD+MCI+NC
	Rousseau (2011) [109]		1	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	DSC	18v (IBSR)	NC
	Coupé (2011) [110]		2	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	KI	80v (ICBM)	NC
				-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	KI	80v	AD
		[117]		✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC	80v (ICBM); 202v (ADNI)	AD+MCI+NC
	Zhang (2012) [112]		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC	5v (NA0-NIREP)	NC
	Jia (2012) [103]		1	-	-	-	-	-	-	-	✓	-	-	-	✓	✓	✓	-	DSC	50v (ADNI)	MCI+NC
	Cardoso (2013) [104]		1	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC	30v (ADNI)	AD+MCI+NC
				✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	DSC	30v (Hammers)	NC
	Wang (2013) [102]		1	✓	✓	✓	✓	-	✓	✓	-	✓	-	-	✓	✓	✓	-	DSC	20v (MICCAI'12)	CD(PAD)+NC
	Asman (2013) [105]		1	×	×	×	×	×	×	-	-	×	-	×	×	×	×	×	DSC	15v (MICCAI'12)	CD(PAD)+NC
	Wang (2014) [96]		1	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	-	-	✓	DSC; MAD; HD	20v (MICCAI'12)	CD(PAD)+NC
	Pipitone (2014) [100]		3	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC	60v (ADNI1); 81v (FEP)	AD+MCI+NC+FEP
				⊛	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC	5v (Winterburn et al.)	NC
	Wu (2015) [111]		1	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	DSC	30v (Hammers)	NC

(MABMIS) framework to perform simultaneous segmentation of a group of target images based on the construction of a combinative tree. Similarly, Wolz et al. [113] presented a graph based framework called LEAP (Learning Embedding for Atlas Propagation), where the newly segmented images also became candidate atlases to segment the remaining target images. Pipitone et al. [100] proposed the MAGeT-Brain (Multiple Automatically Generated Templates) algorithm, that performed multi-atlas segmentation using a template library built from a subset of target images, constructed via label propagation with each of the available atlases. The authors stated that MAGeT minimized the number of atlases needed whilst still achieving similar agreement to conventional multi-atlas approaches.

3.2. Learning-based methods

The goal of learning-based methods or machine learning strategies is to predict the segmentation label S given the input features I. From a probabilistic perspective, the goal is to find the conditional distribution $p(S|I)$, that can be either learned from a training set of labeled images, in which case we have a discriminative approach, or alternatively, the joint distribution $p(I,S)$ can be found and used to evaluate the conditional $p(S|I)$, where the approach is known as generative [118]. These two approaches differ in that discriminative models provide a model only for the target variables conditional on the observed variables, whereas generative models are full probabilistic models of all variables. In what follows, we present different strategies based on such both approaches which we have called supervised and Bayesian methods.

A summary of the reviewed learning-based algorithms can be found in Table 3.

3.2.1. Supervised methods

Supervised methods, also known as discriminative models, attempt to directly estimate a label for each voxel given the local appearance of the image around it. For this purpose, these methods extract image features with rich information and use them to train a classification model using supervised learning algorithms. Among these learning algorithms we can find Artificial Neural Networks (ANNs) such as in [119], where a segmentation framework that can be applied to extract various brain structures, composed by two architectures that act in two phases, was presented. The first network classified the textures of the target image while the second one took the output of the first classifier and refined the segmentation correcting possible errors of the initial stage via local shape/texture analysis. Moreover in [120], the authors proposed a two-stage method that combined ANNs with Geometric Moment Invariants (GMIs). At the first stage, a set of multiple Multi-Layer Perceptron (MLP) networks were used for function approximation. There was one MLP for each scale of the GMIs, whose outputs, together with voxel intensities and coordinates, were the input features of the ANN of the second stage. At that stage, the ANN worked as a classifier instead of a function approximator, classifying each voxel as inside or outside the structure of interest.

In [121], Morra et al. introduced a learning approach in which they iteratively learned the marginal distribution for each image voxel towards the final segmentation. The classifiers were trained not only on the features from the image patch, but also on the probability patch, and hence, their AdaBoost weak learners were decision stumps on both image and probability maps. In that context, they used the previously trained classifier to compute new classification maps that were used to train the next classifier, repeating this procedure until convergence. More recently, Tu and Bai [122] presented the Auto-Context algorithm, which was also based on this principle. Following this framework, Kim et al. [123] proposed a method to extract the hierarchical feature representation of image patches, from 7.0 T MRI images, based on deep learning. These features were further incorporated into a multi-atlas version of the Auto-Context segmentation framework to improve hippocampus segmentation on such high-resolution images. Moreover, Wels et al. [124] presented a method for segmenting brain subcortical structures based on the concept of marginal space learning. At each level of abstraction they built a discriminative model from a labeled set of images, training a probabilistic boosting tree from high-dimensional vectors of Haar and steerable features derived from the image intensities. These models were used to narrow the range of possible solutions until the final shape can be inferred.

Support Vector Machines (SVM) have also been used as a learning strategy to perform brain structures segmentation. In [27], Igual et al. proposed a method for internal caudate nucleus segmentation, which first delineated the external boundary of the structure by means of the previously proposed algorithm, Caudate-Cut [125]. After an automatic geometric criterion classification, a SVM classifier based on shape features of the caudate regions was used, to separate head and body caudate regions. Finally a post-processing step based on a decision stump to improve the global classification was applied.

Other learning approaches relying on dictionary

learning [117, 126] or Genetic Algorithms (GAs) [127, 128] have also emerged in the last years. Tong et al. [117] introduced a segmentation strategy based on the minimization of patch reconstruction errors, in which they learned a dictionary and a linear classifier simultaneously for every voxel of the target image from predefined neighborhood patches around that voxel and across the training atlases. Benkarim et al. [126] extended this method in their proposed multi-class dictionary learning approach. They learned a single discriminative dictionary and a multi-class linear classifier simultaneously for each target voxel, which allowed segmenting multiple structures at the same time. On the other hand, Deoni et al. [127] presented a method to segment fifteen thalamus internal structures by means of a GA approach that incorporated characteristics of the k-means clustering method. Regarding the fitness function, it gave greater values to those candidate segmentation solutions with maximized cluster (structure) sizes while the variance of the T1 and T2 image modalities was minimized. More recently, Traynor et al. [128] re-evaluated this method under much broader operating conditions.

3.2.2. Bayesian methods

Probabilistic segmentation methods try to infer the most likely segmentation given the observed image, which, according to the Bayes rule, can be approximated to the probability of the image occurring given a certain segmentation $p(I|S)$, together with the prior probability of the segmentation $p(S)$. This can be achieved via Maximum A Posteriori (MAP) estimation. $p(S)$, henceforth the prior, encodes the spatial organization of anatomical structures in the image domain, whereas $p(I|S)$, is a likelihood distribution that predicts how a label image, where each voxel is assigned a unique anatomical label, translates into an intensity image.

There is a large amount of work that relies on this general framework, differing mainly in the way the priors and the likelihood are specified as well as in the optimization method chosen to estimate the model parameters. In brain structures segmentation, priors come very often in the form of probabilistic atlases while likelihood is commonly modeled as a Mixture of Gaussians (MoG), where the parameters (mean and variance) are usually estimated by means of the EM algorithm. In [129, 130] the prior was a mesh-based probabilistic atlas where the mesh deformation was estimated in a coordinated ascent scheme with the Levenberg-Marquardt algorithm in combination with the Gaussian parameters (EM). On the other hand, Makropoulos et al. [131] obtained their priors from several atlases, whose labels were propagated to the target image and averaged to form, in combination with tissue probability maps, a probabilistic spatial prior for each structure, while their likelihood term was approximated by a MoG. Following this framework, Riklin-Raviv et al. [132] introduced a method for group-wise segmentation of brain structures that avoided the use of statistical atlas by introducing latent atlases, generated from an image ensemble. They proposed to alternate between estimating the MAP segmentations and refining the model parameters, replacing the expectation step by a gradient descent process using a probabilistic level-set formulation.

Markov Random Field (MRF) modelization, that introduces local spatial dependencies between voxels, has also been broadly used among probabilistic methods. Fischl et al. [26] presented a method to segment the whole brain which forms the basis for the well-known software FreeSurfer [133]. In this method the intensity distribution of each structure at each location was modeled as a Gaussian, while the priors, that came in the form of global spatial information given by an atlas and local spatial relationship between anatomical classes, were approximated by an anisotropic non-stationary MRF. Another approach, proposed by Scherrer et al. [134], is the LOcal Cooperative Unified Segmentation (LOCUS) algorithm, that performed tissue and subcortical structure segmentation, which cooperate gradually to improve the accuracy. They performed the segmentation partitioning the target image into a set of local subvolumes and distributed one local MRF per sub-volume. A-priori knowledge in the form of generic fuzzy spatial relations was introduced in the MRF model to segment the structures. More recently, Razlighi et al. [135] introduced a segmentation method where they used quadrilateral MRF to model both the priors and the likelihood probabilities. In such a model, not only the neighborhood labels were taken into account, but also their intensities, by contrast to the classical MRF model.

Other methods based on this probabilistic principle have been proposed in recent years. Askelrod-Ballin et al. [136] presented a multi-scale algorithm that used a graph representation of the target image which was recursively coarsened to obtain the final segmentation. The posterior probabilities that two nodes were aggregated was estimated by means of Bayesian formulation, where the priors were given in a form of probabilistic atlas, while the likelihood was estimated from a set of manually labeled atlases. On the other hand, Pohl et al. [14] proposed a hierarchical algorithm guided by a prior information represented within a tree structure. They followed a recursive segmentation process that started

Table 3: Learning-based methods. Acronyms from left to right are: hippocampus (HIP), thalamus (THA), caudate nucleus (CAU), putamen (PUT), pallidum (PAL), amygdala (AMY), accumbens (ACC), lateral ventricle (LV), brainstem (BS), corpus callosum (CC), cerebellum (CB), white matter (WM), cortical gray matter (CGM) and cerebrospinal fluid (CSF). Measures in appearance order: Classification rate (CR), Reproducibility (Rep), Precision (P), Recall (R), Relative overlap (RO), Similarity index (SI), Hausdorff distance (HD), Dice similarity coefficient (DSC), Mean absolute distance (MAD), Overlap error (OE), MICCAI'07 score (Score), Overlap by pairs (OBP), Tanimoto coefficient (TAO), Sensitivity (SN), Specificity (SP), Accuracy (Acc), Jaccard index (JC) and Volume (Vol). Diseases in order of appearance: Normal controls (NC), Alzheimer's disease (AD), Mild cognitive impairment (MCI), Schizotypical personality disorder (SPD), Autism (AU), Parkinson disease (PD), Attention deficit hyperactivity disorder (ADHD), Clinical dementia (CD), Probable Alzheimer's disease (PAD), Schizophrenia (SZ), Bipolar disorder (BD), Major depressive disorder (MDD), Elderly controls (EC) and Neonatal normal controls (NNC). The ✓ and × symbols stand for numerical and graphical results, respectively, whereas the ⊛ symbol means that the method perform the internal substructures segmentation of the indicated structure. The - symbol indicates that no results have been reported for that particular structure. Column 'C' indicates the segmentation target: (1) whole brain; (2) a group of structures; (3) a single structure; (4) a single structure and its substructures. Column 'Ref' shows the reference work from which the results are obtained (in case they do not come from the original work).

				Segmented structures																
	Article	Ref	C	HIP	THA	CAU	PUT	PAL	AMY	ACC	LV	BS	CC	CB	WM	CGM	CSF	Others	Measure	Database	Disease
Supervised	Pitiot (2002) [119]		2	✓	-	✓	-	-	-	-	-	-	✓	-	-	-	-	-	CR	10v	-
	Deoni (2007) [127]		4	-	⊛	-	-	-	-	-	-	-	-	-	-	-	-	-	Rep	4v (T1&T2)	NC
	Morra (2008) [121]		2	✓	-	×	-	-	-	-	-	-	-	-	-	-	-	-	P; R; RO; SI; HD	83v (AD)	AD+MCI+NC
	Moghaddam (2009) [120]		2	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	DSC; MAD; HD	6v (IBSR)	NC
	Wels (2009) [124]		2	✓	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	DSC; MAD	18v (IBSR)	NC
				-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	OE; MAD	MICCAI'07	SPD+AU+PD+NC
	Tu (2010) [122]		1	-	-	×	-	-	-	-	-	-	-	-	-	-	-	-	Score	MICCAI'07	SPD+AU+PD+NC
	Traynor (2011) [128]		4	-	⊛	-	-	-	-	-	-	-	-	-	-	-	-	-	OBP; TAO	16v; 18v	NC
	Igual (2012) [27]		4	-	-	⊛	-	-	-	-	-	-	-	-	-	-	-	-	SN; SP; Acc	39v (VH)	ADHD+NC
	Tong (2013) [117]		3	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC	202v (ADNI); 80v (ICBM)	AD+MCI+NC
		[126]		-	-	✓	✓	✓	-	✓	-	-	-	-	-	-	-	-	DSC	35v (MICCAI'12)	CD(PAD)+NC
	Kim (2013) [123]		3	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P; R; RO; SI	20v (7T)	-
	Benkarim (2014) [126]		2	-	-	✓	✓	✓	-	✓	-	-	-	-	-	-	-	-	DSC	35v (MICCAI'12)	CD(PAD)+NC
Bayesian	Fischl (2002) [26]	[132]	1	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	DSC	39v (Sabuncu et al.)	AD+NC
		[126]		-	-	✓	✓	✓	-	✓	-	-	-	-	-	-	-	-	DSC	35v (MICCAI'12)	CD(PAD)+NC
		[115]		✓	✓	✓	✓	✓	✓	-	✓	-	-	-	✓	✓	-	-	DSC	39v (FS atlas)	CD+AD+NC
		[137]		✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	DSC	30v	-
		[138]		✓	-	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	P; R	14v (LONI28)	NC
		[109]		✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	DSC	18v (IBSR)	NC
	Scherrer (2007) [134]		2	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	DSC	BrainWeb	-
	Scherrer (2007) [139]		2	-	-	-	-	-	-	-	-	-	-	-	×	×	×	-	JC	BrainWeb	-
	Akselrod-Ballin (2007) [136]		1	✓	✓	✓	✓	✓	✓	-	-	✓	-	-	✓	✓	✓	✓	DSC; MAD; HD	18v (IBSR)	NC
		[109]		✓	✓	✓	✓	✓	✓	-	-	✓	-	-	✓	✓	✓	✓	DSC	18v (IBSR)	NC
	Pohl (2007) [14]		1	✓	-	-	-	-	✓	-	-	-	-	-	-	-	-	✓	DSC	50v	SZ+BD+MDD+NC
	Scherrer (2009) [140]		2	-	-	-	-	-	-	-	-	-	-	-	×	×	×	-	DSC	18-20v (IBSR)	NC
				-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	DSC	BrainWeb	-
	VanLeemput (2009) [129]		4	⊛	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC; MAD	10v	NC
	Riklin-Raviv (2010) [132]		2	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	DSC	39v (Sabuncu et al.)	AD+NC
	Razlighi (2012) [135]		1	×	×	×	×	×	×	-	×	×	-	×	×	×	×	-	DSC	20v (OASIS)	-
	Iglesias (2013) [130]		4	⊛	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Vol	383v (ADNI)	AD+EC
	Makropoulos (2014) [131]		1	✓	✓	✓	-	-	✓	-	✓	✓	✓	✓	-	-	-	-	DSC	20v (ALBERTs)	NNC
	Iglesias (2015) [141]		4	-	-	-	-	-	-	-	-	⊛	-	-	-	-	-	-	DSC; HD; MAD	10v (BS)	NC

at the root, segmenting the image into its children, using the propagated structure-specific information in a classical Bayesian framework and estimating the solution through EM.

Recently, Iglesias et al. [130] presented a Bayesian segmentation framework that extended the work of Van Leemput et al. [129] based on the statement that traditional Bayesian methods do not fully consider the uncertainty in the model parameters, relying just on point estimates. To overcome this issue, they proposed a Monte Carlo sampling to account for the uncertainty in prior and likelihood free parameters which resulted in an improved approximation of the segmentation posterior.

3.3. Deformable methods

Deformable methods start with an initial contour placed in the image, either manually or automatically, which is then iteratively deformed, generating a new contour at each iteration. In the primitive version of a deformable model, known as snakes or Active Contour Models (ACM) [142], the initial contour was deformed under the influence of internal and external forces. The internal forces were related to the surface features and aim to maintain a smooth contour, while the external forces were related to the image features of the adjacent regions to the surface and were responsible for attracting the model towards the structure surface. Ghanei et al. [143] presented an improved discrete deformable model to segment the hippocampus, that addressed some of the associated problems such as optimizing the internal force weight, contour stability and extraction of image features for external energy calculations. They introduced a new external force, which was based on searching for local minima of the image energy in the contour normal direction, to produce better results near multiple and discontinuous edges. Colliot et al. [144] also presented a framework in which they use ACM. They added spatial relations between the different structures (direction and distance) represented as fuzzy subsets which were integrated in the deformable model as a new external force that attracted the model to the edges of the structure being segmented. Fouquier et al. [145] extended that work proposing a criteria to optimize the structure segmentation order and introducing a strategy to evaluate the obtained segmentation quality and detect errors to prevent their propagation. Zarpalas et al. [146] also presented a method for segmenting several structures by means of a mixture of different ACMs, all balanced by the gradient distribution boundaries which tried to differentiate regions that need greater support of prior knowledge from those which can be segmented only by their gray-scale information. They used geodesic ACM on boundary parts with strong gradients and a Chan-Vese model with prior knowledge in parts where the boundary was not well formed or contained weak parts. On the other hand, Shariatpanahi et al. [147] used m-Rep (medial representation) deformable models in a multi-agent framework to segment several brain structures, namely, the thalamus, the caudate nucleus and the putamen.

An evolution of ACM is the Active Shape Models (ASM) [148], in which the internal energy, besides keeping the contour smooth, avoids also deformations that go beyond the average geometric variation of the structure being segmented, by using shape constraints learned from a collection of training samples. A Point Distribution Model (PDM) is used to build a shape model of the structure of interest, in which shapes are represented by a set of points or landmarks. Kelemen et al. [149] presented a framework that closely followed the seminal work of Cootes and Taylor [150] on ASM, but based on a hierarchical parametric object description rather than a PDM, under the statement that for a large training set containing several anatomical structures, the generation of the PDM parameterization became very tedious and could be a source of errors. Ettaïeb et al. [151] also proposed a segmentation model based on the ASM and a spatial distance relation. At the segmentation stage, the contours evolved iteratively in two steps: they first evolved independently of each other, according to the constraints imposed by the corresponding shape models and then, applying the constraint imposed by the statistical distance model, which was also estimated during the training phase. In [152], the authors introduced an approach referred to as a active hierarchical shape model, that was able to characterize the different inter-structure relationships and to model the particular local variations of each single structure. Gao et al. [153] also proposed a multi-scale representation for the shape using the wavelet transform. Given the initial shape obtained from label fusion, they proposed a segmentation method that alternated data-driven and a multi-scale shape-based process, iteratively evolving the contour until convergence. Alternatively, Olveres et al. [73] evaluated several ways to segment the midbrain including the combination of ASM with LBP (Local Binary Pattern) descriptors and compared them with the classical ASM segmentation, concluding that incorporating information about the texture surrounding the edge (with LBP), the segmentation performance increased, converging also in less iterations.

Another commonly used version of deformable model is the Active Appearance Model (AAM) [154],

which incorporates constraints of image intensity variation to the ASM. This appearance model is usually based on the normalized first derivative of fixed-size gray profiles, normal to the surface of the object and centered at each landmark. In their work, Babalola et al. [155] presented an AAM based method, called profile appearance models, that instead of modeling the intensities across an entire region containing the model, as in the original AAM approach, they only modeled the intensities along the profiles that were normal to the boundary of the structure. Duchesne et al. [114] also reformulated the original AAM approach, that initially was not suitable for 3D images. Instead of using the original PDM to characterize the shape, which is impractical in 3D, they proposed to utilize a warp distribution model, that centered on 3D deformation fields from ANIMAL [12]. On the other hand, Patenaude et al. [15] utilized the principles of the AAM but placed them within a Bayesian framework. The so called Bayesian Appearance Model (BAM) incorporated both shape and intensity information from a training set but, opposed to the original AAM, it used a probabilistic framework to estimate the relationship between shape and intensity making use of conditional probabilities. Based on the learned model, BAM searched the linear combination for the most probable shape given the observed image intensities to find the best fit for the segmentation. This model was implemented as part of the FSL[1] package under the name FIRST.

Other energy minimizing strategies have been proposed in recent years. Pitiot et al. [31] presented a segmentation method that relied on the deformable templates framework which incorporated available a-priori anatomical expertise, either in the form of implicit knowledge (structures shape and appearance) or of explicit information (relative distance between structures, non-intersection rules) as constraints of the model. Al-Shaikhli et al. [156], on the other hand, proposed an approach for multi-region segmentation using a multilevel set formulation which included a topological graph prior and topological information of an atlas. This topological representation was embedded in the multilevel set energy equation and together with a curvature term, constrained the curve evolution.

Deformable models have to be initialized, either manually or automatically. Atlas registration is a typical practise, but other techniques are also used. In [157] the results of a classifier provided a localization point for the initiation of a deformable template model, whereas García et al. [158] proposed to use a Chan-Vese model to initialize the structure contour.

A summary of the deformable methods presented here is presented in Table 4.

3.4. Region-based methods

Region based brain structures segmentation methods have also been proposed in recent years. These methods rely on the similarity of different properties of the voxels belonging to the same region. Probably, one of the most well established region based techniques is region growing, which is the most frequently used in brain structures segmentation. Based on this technique, Xue et al. [159] proposed a method that performed regionwise labeling by means of GAs followed by voxelwise refinement using parallel region growing. They first over-segmented the target image into three brain tissues (WM, GM and CSF) and also got a coarse location of the structures by registering an atlas to the image. Afterwards, they built a fuzzy model of the regions of interest that represented useful structural knowledge from the atlas (shape, distance and relationship of structures), which was then used to design the objective function of GAs and to guide the region growing. Xia et al. [77] also presented an algorithm to segment the caudate nucleus that performed region growing constrained by anatomical knowledge. They first identified the lateral ventricles, which are easily locatable due to the high tissue contrast of the CSF, and based on their position they determined an initial caudate location by applying region growing from gray matter voxels adjacent to them. Bounding boxes were defined to reduce potential region growing leakage to other structures. After obtaining this coarse segmentation, caudate boundaries were fine-tuned to be smooth, recognizable and valid, based on anatomical knowledge.

However, there are other strategies within region-based techniques such as successive erosion and dilation operations or the use of the watershed algorithms. For instance, Gui et al. [160] proposed an approach to segment neonatal brain based on the use of general knowledge of neonatal brain morphology, integrating information about tissue connectivity, structure and relative positions. They performed a sequential segmentation of the brain structures that combined well-established segmentation methods (marker-based and similarity-based watershed, region growing and region-based active contours), guided by anatomical knowledge, with morphological operations (openings/closings).

The region-based works reviewed here are summarized in Table 4.

[1] http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/ (Accessed: 8 June 2016)

Table 4: Deformable methods. Region-based methods. Hybrid methods. Acronyms from left to right are: hippocampus (HIP), thalamus (THA), caudate nucleus (CAU), putamen (PUT), pallidum (PAL), amygdala (AMY), accumbens (ACC), lateral ventricle (LV), brainstem (BS), corpus callosum (CC), cerebellum (CB), white matter (WM), cortical gray matter (CGM) and cerebrospinal fluid (CSF). Measures in appearance order: Relative agreement (RA), Mean absolute distance (MAD), Kappa index (KI), Hausdorff distance (HD), Similarity (Sim), True positive fraction (TPF), Similarity index (SI), Jaccard index (JC), Comparison (Comp), Dice similarity coefficient (DSC), Overlap (Over), Hausdorff distance 95 (HD95), Mean squared error (MSE), False positive rate (FPR), False negative rate (FNR), Volume difference (VD), Precision (P), Recall (R), Mean distance (MD), Relative volume difference (RVD), Average symmetric surface distance (ASSD) and Root mean square distance (RMSD). Diseases in order of appearance: Schizophrenia (SZ), Normal controls (NC), Elderly controls (EC), Prenatal cocaine exposure (PC), Clinical dementia (CD), Probable Alzheimer's disease (PAD), Alzheimer's disease (AD), Attention deficit hyperactivity disorder (ADHD), Parkinson disease (PD), Fragile X syndrome (FXS), Mild cognitive impairment (MCI) and Autism (AU). The ✓ and × symbols stand for numerical and graphical results, respectively, whereas the ⊛ symbol means that the method perform the internal substructures segmentation of the indicated structure. The - symbol indicates that no results have been reported for that particular structure. Column 'C' indicates the segmentation target: (1) whole brain; (2) a group of structures; (3) a single structure; (4) a single structure and its substructures. Column 'Ref' shows the reference work from which the results are obtained (in case they do not come from the original work). (*)DSC; JC; P; R; HD; HD95; MAD; ASSD; RMSD.

				Segmented structures																	
	Article	Ref	C	HIP	THA	CAU	PUT	PAL	AMY	ACC	LV	BS	CC	CB	WM	CGM	CSF	Others	Measure	Database	Disease
Deformable	Ghanei (1998) [143]		3	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	RA	11slices	-
	Kelemen (1999) [149]		2	×	×	-	×	×	-	-	-	-	-	-	-	-	-	-	MAD	21v	SZ+NC
	Duchesne (2002) [114]		2	✓	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	KI	30v;10v (ICBM)	NC
	Ashton (2003) [157]		3	×	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10v	NC
	Pitiot (2004) [31]		2	✓	-	✓	-	-	-	-	✓	-	✓	-	-	-	-	-	HD; MAD	20v	EC
	Shariatpanahi (2006) [147]		2	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	Sim; TPF; HD	4v (IBSR)	NC
	Colliot (2006) [144]		2	-	×	✓	-	-	-	-	×	-	-	-	-	-	-	✓	HD; MAD; SI	10v	-
	Babalola (2007) [155]		3	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	JC	24v	EC+PC+NC
	Zarpalas (2011) [146]		2	×	-	-	-	-	×	-	-	-	-	-	-	-	-	-	Comp	13v (OASIS)	CD(PAD)+NC
	Patenaude (2011) [15]		2	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-	-	DSC	37v;42v;17v;87v;14v;139v	SZ+AD+PC+ADHD+NC
	Cerrolaza (2012) [152]		2	-	-	×	×	-	-	-	×	-	-	-	-	-	-	-	Over	87v	NC
	Gao (2012) [153]		2	✓	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	DSC; HD95	24v (HIP); 24v (CAU)	-
	Fouquier (2012) [145]		2	-	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	MAD	30v (IBSR+OASIS)	NC
	Olveres (2013) [73]		3	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	DSC; HD	10slices	NC
	García (2014) [158]		3	-	×	-	-	-	-	-	-	-	-	-	-	-	-	-	MSE	4v (DB-UTP)	PD
	Al-Shaikhli (2014) [156]		1	-	-	-	-	-	-	-	×	×	-	×	×	×	×	×	DSC; MAD; HD	BrainWeb; MedPix; NCSUIAL	-
	Ettaïeb (2014) [151]		3	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	HD	10slices	-
Region	Xue (2000) [163]		2	-	✓	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	FPR; FNR; SI; KI	-	-
	Xue (2001) [159]		2	-	✓	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	FPR; FNR; SI; KI	-	-
	Xia (2007) [77]		3	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	Over	55v	FXS+NC
	Gui (2012) [160]		1	-	-	-	-	-	-	-	-	✓	-	✓	✓	✓	✓	✓	DSC	10v (newborns)	NC
Hybrid	Zhou (2005) [19]		2	✓	✓	✓	✓	-	✓	-	-	-	-	-	-	-	-	-	VD; Over; MAD	17v (IBSR)	NC
	Tu (2008) [164]		2	✓	-	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	P; R; HD; MD	14v	-
	Karsch (2009) [165]		2	✓	-	-	-	-	-	-	✓	-	✓	-	-	-	-	-	DSC; Over	-	-
	Sabuncu (2009) [137]		1	✓	✓	✓	✓	✓	✓	-	×	-	-	-	×	×	-	-	DSC	30v	-
	Sabuncu (2010) [115]		1	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	✓	✓	-	-	DSC	39v (FS atlas)	AD+CD+NC
		[104]		✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	DSC	30v (Hammers)	NC
		[104]		✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC	30v (ADNI)	AD+MCI+NC
	He (2011) [166]		2	-	✓	-	-	-	-	-	✓	-	✓	-	-	-	-	-	DSC; Over	20v+25v	AU+NC
	Weisenfeld (2011) [107]		1	-	×	×	×	-	-	-	-	-	-	-	×	×	×	-	DSC	14v	NC
	Iglesias (2012) [161]		1	×	×	×	×	×	×	-	×	-	-	×	×	×	-	-	DSC	8v (multimodal)	NC
	van der Lijn (2012) [167]		2	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	DSC; JC; RVD	18v	-
				✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	DSC; JC; RVD	18v	-
	Iglesias (2013) [162]		1	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	✓	✓	-	-	DSC	8v (PD)	-
	Liu (2013) [138]		2	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-	-	-	DSC	6v (IBSR)	NC
				✓	-	✓	✓	-	-	-	-	-	-	-	-	-	-	-	DSC; HD	15v (LPBA40)	NC
				✓	-	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	P; R	14v (LONI28)	NC
				✓	-	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	HD; MAD	28v (LONI28)	NC
	Hao (2014) [16]		2	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(*)	30v+30v (ADNI); 57v	AD+MCI+NC

3.5. Hybrid methods

Several combinations of the previous categories have also been described in the literature. These methods try to combine the strengths of the different strategies in order to improve the segmentation accuracy. A common example is the combination of label fusion and learning-based strategies. For instance, in [115] the authors presented a probabilistic framework that lead to label fusion style segmentation algorithms. Under the assumption that each voxel of the target image is generated from one of the atlases, they constructed the conditional probability of generating the target image and label map where the final segmentation was achieved via MAP estimation. Recently, Iglesias et al. [161, 162] extended this framework to multi-modal data. Another approach combining these strategies is the one proposed by Weisenfeld and Warfield [107] in which a multi-classifier fusion algorithm called Learning Likelihoods for Labeling (L3), which combined label fusion and statistical classification, was introduced. They employed each atlas from the training set to train a classifier that was used to generate a classification of the target image. These resulting classifications, generated by a Bayesian segmentation strategy, were later fused with the STAPLE algorithm to produce the final segmentation. Hao et al. [16] also proposed a learning-based label fusion method to segment the hippocampus. In their approach, for each voxel in the target image, candidate training samples were obtained from voxels of atlases within a spatial neighborhood of the voxel considered, and the image feature vectors were then computed. Once the image features were extracted, a k-NN strategy based SVM classification algorithm was adopted to build a classifier for each voxel, which was applied to each voxel feature vector of the target image to obtain the final segmentation.

Combining discriminative and generative models is also common practice. Van der Lijn et al. [167] presented a segmentation method that combined structures' spatial and appearance information in a posterior probability function which was maximized using graph cuts. The spatial information came in the form of a probability map and the structure appearance was described by a k-NN voxel classifier based on Gaussian scale-space features. Liu et al. [138] proposed a hybrid method that combined a generative and a discriminative model, with feature augmentation and adaptation. Their approach was based on using the estimated segmentation and the parameters of a Bayesian segmentation to normalize the image intensities and extract robust, invariant local features. Afterwards, they used the auto-context algorithm [122] to obtain the final segmentation. Tu et al. [164] also introduced a hybrid model that combined both a discriminative approach to model the appearance and a generative model to describe shape. For appearance modeling they adopted a probabilistic boosting tree framework to learn a multi-class discriminative model while shape information was incorporated through principal component analysis. Once the system was trained, structures segmentation was obtained performing surface evolution by minimizing an energy function associated with the proposed hybrid model.

Other combinations have also been proposed, such as the ones in [165] and [166] that presented a hybrid method combining both region-based and boundary-based procedures. In these approaches they first applied a clustering technique (k-Means) to generate an initial seed contour and afterwards, the seed was deformed based on a level-set PDE. On the other hand, Zhou and Rajapakse [19] proposed a segmentation method based on fuzzy templates. From a set of labeled samples they obtained three fuzzy maps (intensity, spatial location and relative spatial relations to other structures) based on information obtained from structure histograms. These fuzzy maps were calculated for each structure and in each training image and were fused to obtain a total fuzzy template involving all features for different structures. This total fuzzy template was then registered to the target image, and combining its information with a probabilistic tissue segmentation [168], a fuzzy membership map of each structure was created. Final segmentation was performed by applying alpha-cut thresholding to this final fuzzy map.

Table 4 includes an overview of the segmentation targets and evaluation criteria of the methods summarized in this section.

4. Pros and cons of the strategies

In section 3 we have presented five main strategies to perform brain structures segmentation, namely atlas-based, learning-based, deformable, region-based and hybrid methods. The main advantages and drawbacks of these strategies are presented here and summarized in Table 5.

Atlas-based methods insert robustness to the segmentation strategy, as they overcome the deficiencies of contrast and MRI resolution. Label propagation is the most straightforward atlas-based technique, based on the propagation of the atlas labels to the target image after registration. It is a good technique when only one atlas is available and is also quite fast, but it suffers from two main drawbacks. The fist one is the fact that a simple atlas cannot sufficiently represent the whole popu-

Table 5: Pros & Cons. Advantages and disadvantages of the segmentation strategies presented.

	Method	Advantages	Disadvantages
Atlas-based	Label propagation	-Quite fast -Only one atlas is needed -Useful as initialization	-Dependent on the atlas anatomical similarity -Dependent on the registration method
	Label fusion	-Anatomical variation is better captured -Registration error is less likely to affect the segmentation	-Computationally expensive -Usually rely on registration accuracy -Prior segmented images (atlases) needed
Learning-based	Supervised methods	-Good results if test/training data variations are small -Can accurately capture local appearance variations	-Prior segmented images needed (training) -Not easily adapted to capture global shape information -Changes in MRI contrast reduce their performance -Highly dependent on the training set
	Bayesian methods	-Require small amounts of training data -Allows explicit incorporation of prior information -Robust against image artifacts -Flexibility and adaptability	-Can be difficult to define and learn -Can be slow
Deformable	ACM	-No training required -Minimized number of parameters -Perform an accurate adjustment that registration cannot	-Sensitive to initialization -Low contrast boundaries can cause them to fail -Initialization required
	ASM & AAM	-Robust against noise -Avoids deformations that go beyond the average -Handles discontinuities along structure boundaries -Perform an accurate adjustment that registration cannot	-Construction of an explicit model required (training) -Prior segmented images needed -Less flexible -Larger number of parameters
	Region-based methods	-Quite fast -Do not need training data	-Low contrast boundaries can cause them to fail -Additional information to guide the growing is needed -Initialization required

Hybrid methods are not included in the table since they combine the strategies presented here, with their corresponding strengths and weaknesses.

lation of potential test data and the second is that the quality of this approach is limited by the accuracy of the pairwise registration method. This technique is commonly used as a starting point of other methods, which use the propagated labels as a initialization or as a prior.

In order to overcome the problems presented in label propagation, multi-atlas or label fusion methods have emerged in recent years. Several atlases are used to improve the capture of anatomical variability between different scans but at the expense of a high computational time. These methods are less dependent on the registration as the effect of errors associated with any single atlas propagation is reduced in the combination process.

Supervised methods provide good results when the differences between the training data and the target image are small, capturing accurately local appearance variations. However, they are very dependent on these training set, hence changes in MRI contrast or strong anatomical differences among training and testing images highly reduce their performance, limiting their applicability to images acquired with the same protocol as the images used for training. As such, larger training data sets trend to be beneficial. On the other hand, Bayesian approaches are more flexible and adaptable, as they permit explicitly modelling image artifacts such as the bias field or other image acquisition parameters, making these methods more robust. They also allow the explicit incorporation of a-priori information by means of the prior term, that frequently comes in the form of a probabilistic atlas, which captures global shape information. However, these methods can be slow and difficult to design and learn, especially for complex structures with inhomogeneous textures.

Deformable methods give dynamics to the segmentation, performing an accurate adjustment that the registration methods cannot perform, but they require robust initialization. In the case of ACM, as they represent a local search, this initialization must be done near the structure of interest in order to avoid falling in a local minima. Furthermore, low contrast boundaries may prevent them to provide the desired solution. By contrast to ACMs, ASMs and AAMs need to be trained to learn an explicit model (shape or appearance constraints) that captures the variation of shape and gray level across a training set. This model guides the contour evolution and avoids deformations that go beyond the average geometric variation of the structure being segmented, making these methods robust against noise, but at the same time less flexible and with a larger number of parameters/constraints than ACMs.

Contrary to these methods, region-based strategies do not require to be trained and are usually quite fast in

finding a solution. However, these methods must usually be combined with some other kind of anatomical knowledge and require an initial seed to initialize the growing. Furthermore, and similarly to ACMs, noisy structure boundaries can easily cause them to fail.

Tables 2 to 4 summarize the approaches reviewed in section 3 and show, for each work, the brain structures segmented as well as the measures and databases used for evaluation indicating also if they were tested with normal controls or diseased subjects. As shown from these tables, atlas-based methods are the most commonly used to segment the whole brain whereas they are rarely used to segment a single structure. They use to register the atlas/es to the whole volume, which is the hardest and more computationally expensive task, and once it is done, performing label propagation and fusion is as trivial for one label as for all of them. Even among the hybrid methods, the ones which segment the whole brain are all based on combinations of atlas-based approaches with another segmentation strategy. According to these tables, region-based strategies are the less popular in segmenting brain structures. As stated before, these methods on its own are not sufficient in most of the cases and usually need additional anatomical information or to be combined with another segmentation strategy, which makes them not very attractive. Furthermore, it can be stated that there are not so many methods requiring a training phase, such as supervised, ASM and AAM, which segment the whole brain. This is due to the fact that these trainings use to be task specific instead of generalistic, with the aim of segmenting a single structure or a reduced group of them.

From these tables, we can also observe that the vast majority of research evaluate their algorithms with non-lesioned brain databases. As an exception, Fouquier et al. [145] used a database composed by 30 healthy cases and 14 pathological (brain tumor) cases. However, they only provide quantitative segmentation results for the healthy cases, which does not provide any information of how these lesions affect the structure segmentation.

To the best of our knowledge, how white matter lesions, such as the ones produced in multiple sclerosis or lupus, affect these algorithms has not yet been evaluated. Nevertheless, it is a well known problem among automatic tissue segmentation methods [169], where lesion filling techniques [170–172] have already been applied improving the accuracy of tissue volume [173, 174]. As far as we know, these techniques have not yet been evaluated in combination with automatic brain structure segmentation algorithms and making these algorithms robust against lesions remains still an open challenge to the research community.

Automatic segmentation of severely atrophied brains, which produces morphological changes in the brain structures, can also make some methods performance, such as the ones relying on registration, oscillate. This could be prevented by means of large atlas databases with high anatomical variability reducing the dissimilarity between the atlases and the target image. However, there are not many public databases with ground truth and furthermore most of them only contain healthy subjects, which makes this solution hard to accomplish.

In order to give an overall quantitative performance estimate of the state-of-the-art methods, in the following section we present an evaluation of the methods reviewed here, grouping the results both per segmentation strategies and target structures and reporting the different evaluation measures and databases used. Furthermore, the existing challenges in the evaluation of automatic brain structure segmentation methods are discussed and a quantitative evaluation of three well known software tools, each relying on a different category of our classification, is performed.

5. Validation and quantitative evaluation

5.1. Public databases

Comparing the results of the available brain structures segmentation methods is not a trivial task. There are a few publicly available manually labeled images [175–177] that serve as a ground truth and some authors use their own images for evaluation, which makes quantitative comparison of the structures segmentation algorithms difficult. As we have noticed from the reviewed papers, the most frequently used public databases for evaluation include the Internet Brain Segmentation Repository[2] (IBSR18), the LONI Probabilistic Brain Atlas[3] (LPBA40) [178] and the Hammers Adult atlases[4] [179–181]. Table 6 shows the main features of these databases, that include the name and webpage, the number of images it contains, the image modalities, the number of structures labeled, the scanner used for acquisition, the image resolution, the voxel size and some demographics such as subjects' age and sex.

[2] https://www.nitrc.org/projects/ibsr (Accessed: 8 June 2016)

[3] http://www.loni.usc.edu/atlases/Atlas_Detail.php?atlas_id=12 (Accessed: 8 June 2016)

[4] http://brain-development.org/brain-atlases/ (Accessed: 8 June 2016)

Table 6: Most commonly used public databases for evaluation in brain structure segmentation.

Name	Subjects	Ages	Modality	Structures	Scanner	Volume (mm)	Voxel (mm)
IBSR18[2]	18 (14 ♂, 4 ♀)	7 - 71	T1-w	43	12: GE (1.5T) 6: Siemens (1.5T)	256×256×128	8: 0.94×0.94×1.5 6: 0.84×0.84×1.5 4: 1×1×1.5
LPBA40[3]	40 (20 ♂, 20 ♀)	19.3 - 39.5	T1-w	56	GE Signa (1.5T)	256×256×124	38: 0.86×0.86×1.5 2: 0.78×0.78×1.5
Hammers[4]	30 (15 ♂, 15 ♀)	20 - 54	T1-w	83	GE Signa (1.5T)	192×256×124	0.937×0.937×1.5

Volume and voxel dimensions in native space.

5.2. *Evaluation measures*

Apart from the difficulty of having good datasets, there is not a standard measure for evaluation and results are presented using different metrics. Analyzing the literature we have seen that the most commonly used measures are based either on volume overlap or contour distance. Measures based on volume compute how much the ground truth and the obtained segmentation volume overlap. The most common metric in this category is the Dice similarity coefficient (DSC) and variants such as the Kappa index (KI), the Similarity index (SI), the relative overlap (RO) or the Jaccard index (JC), but other metrics are also used such as the specificity (SP), sensitivity (SN), accuracy (Acc), precision (P) and recall (R). Regarding contour distances, which rely on computing how close the ground truth and the obtained segmentation contours are, the most recurrent are the Hausdorff distance (HD) and its variants, and the mean absolute distance (MAD).

As an attempt to standardize this evaluation procedure, three MICCAI Challenges[5] on brain structures segmentation have been proposed during recent years. The first one was the CAUdate SEgmentation 2007 (CAUSE07)[6], which was a competition held as part of the workshop '3D Segmentation in the Clinic: A Grand Challenge' [182], in conjunction with MICCAI 2007. The goal of this competition was to compare different algorithms when segmenting the caudate nucleus from brain MRI scans, for which they provided 67 images (33 for training and 34 for testing) collected from different databases. As a comparative measure, they devised a scoring system that combined several metrics into a single overall score. The second one, was the MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling[7] [183]. The challenge was on whole-brain labeling, assuming the majority of the participant methods to be multi-atlas but accepting any method as long as the approach were described in a reproducible manner. They provided 15 images for training and 20 for testing, obtained from the Open Access Series of Imaging Studies (OASIS) project[8] and the primary metric for evaluation was the mean DSC across all brain labels and all subjects in the testing cohort. Finally, the MICCAI 2013 Segmentation: Algorithms, Theory and Applications (SATA) challenge[9] [184], was created to test the limits of the applicability of multi-atlas segmentation. In this case, they proposed two sub-challenges: the free-for-all sub-challenge, which allowed any segmentation framework to be applied, and the standardized registration sub-challenge in which pairwise registrations were provided to remove the impact of the registration algorithm. A collection of 47 images was provided, 45 of which were from the OASIS project and the remaining two were part of the Child and Adolescent NeuroDevelopment Initiative (CANDI)[10], whereas the image labels, that included seven subcortical structures (accumbens, amygdala, caudate, hippocampus, pallidum, putamen and thalamus), were provided by Neuromorphometrics, Inc.[11] As an evaluation metric for strategies comparison, they used DSC and the symmetric HD.

5.3. ***Quantitative analysis of the reviewed literature***

We present in what follows a quantitative comparison of the works reviewed in this article, separately for each brain structure, including: hippocampus, thalamus, caudate nucleus, putamen, pallidum, amygdala, accumbens, lateral ventricles, and brainstem. Table 7 summarizes the results, as well as the data and the evaluation measures obtained from the analyzed approaches. Note that the

[5] http://grand-challenge.org/ (Accessed: 8 June 2016)

[6] http://cause07.grand-challenge.org/ (Accessed: 8 June 2016)

[7] https://masi.vuse.vanderbilt.edu/workshop2012/index.php (Accessed: 8 June 2016)

[8] http://www.oasis-brains.org/ (Accessed: 8 June 2016)

[9] https://masi.vuse.vanderbilt.edu/workshop2013/index.php (Accessed: 8 June 2016)

[10] http://www.nitrc.org/projects/candi_share (Accessed: 8 June 2016)

[11] http://www.neuromorphometrics.com/ (Accessed: 8 June 2016)

Table 7: **Quantitative results of the reviewed methods. Acronyms from left to right are: hippocampus (HIP), thalamus (THA), caudate nucleus (CAU), putamen (PUT), pallidum (PAL), amygdala (AMY), accumbens (ACC), lateral ventricle (LV) and brainstem (BS). Measures in order of appearance: Kappa index (KI), Jaccard index (JC), Relative overlap (RO), Dice similarity coefficient (DSC), Similarity index (SI), Mean absolute distance (MAD), Hausdorff distance (HD), Classification rate (CR),Similarityy (Sim), Hausdorff distance 95 (HD95) and Overlap (Over). Column 'Ref' shows the reference work from which the results are obtained (in case they do not come from the original work).**

			Segmented structures										
	Article	Ref	HIP	THA	CAU	PUT	PAL	AMY	ACC	LV	BS	Measure	Database
Label propagation	Collins (1997) [12]	[97]	0.86	-	-	-	-	0.82	-	-	-	KI	80v (ICBM)
		[97]	0.76	-	-	-	-	0.70	-	-	-	JC	80v (ICBM)
		[114]	0.71	-	-	-	-	0.65	-	-	-	KI	30v (ICBM)
		[114]	0.69	-	-	-	-	0.64	-	-	-	KI	10v (ICBM)
	Shen (2002) [85]	[91]	0.49	0.64	0.55	0.52	0.41	0.46	-	0.65	0.72	JC	11v (IBSR)
		[91]	0.62	0.74	0.65	0.72	0.60	0.61	-	0.58	0.79	JC	36v (Desikan et al.)
	Postelnicu (2009) [91]		0.45	0.60	0.53	0.48	0.23	0.41	-	0.67	0.77	JC	11v (IBSR)
			0.63	0.75	0.67	0.75	0.60	0.57	-	0.66	0.73	JC	36v (Desikan et al.)
	Lin (2010) [20]		-	0.81	0.73	0.78	-	-	-	-	-	KI	15v (IBSR)
			-	0.68	0.57	0.64	-	-	-	-	-	RO	15v (IBSR)
	Luo (2011) [92]		-	0.84	0.78	0.80	-	-	-	-	-	DSC	9v (IBSR)
Label fusion	Warfield (2004) [13]	[115]	0.81	0.89	0.83	0.88	0.82	0.79	-	0.86	-	DSC	39v (FS atlas)
		[108]	0.52	0.85	0.68	0.80	0.70	0.65	0.49	0.51	0.85	SI	17v (IBSR)
		[108]	3.73	0.96	1.64	1.04	1.36	1.33	3.74	3.08	1.08	MAD	17v (IBSR)
	Heckemann (2006) [98]		0.82	0.91	0.90	0.90	0.80	0.81	0.71	0.90	0.94	SI	30v
	Aljabar (2009) [99]		0.83	0.91	0.88	0.90	0.82	0.78	0.76	0.91	0.94	DSC	275v (CMA)
	Artaechevarría (2009) [108]		0.75	0.88	0.83	0.86	0.79	0.72	0.67	0.83	0.91	SI	17v (IBSR)
			0.79	0.75	0.64	0.67	0.72	0.85	0.68	0.69	0.69	MAD	17v (IBSR)
		[104]	0.84	0.88	0.88	0.89	0.77	0.80	0.69	-	-	DSC	30v (Hammers)
		[104]	0.87	-	-	-	-	-	-	-	-	DSC	30v (ADNI)
		[109]	0.75	0.88	0.83	0.86	0.79	0.72	0.67	0.83	0.91	DSC	18v (IBSR)
	Lötjönen (2010) [95]		0.81	0.90	0.87	0.91	0.84	0.77	-	-	-	SI	18v (IBSR)
			0.88	-	-	-	-	-	-	-	-	SI	60v (ADNI)
		[109]	0.80	0.89	0.85	0.90	0.83	0.75	-	-	-	DSC	18v (IBSR)
	Collins (2010) [97]		0.89	-	-	-	-	0.83	-	-	-	KI	80v (ICBM)
			0.80	-	-	-	-	0.70	-	-	-	JC	80v (ICBM)
	Heckemann (2010) [116]		0.71	0.80	0.81	0.81	0.63	0.65	0.52	0.83	0.88	JC	30v (Hammers)
		[104]	0.83	0.89	0.89	0.89	0.77	0.79	0.68	-	-	DSC	30v (Hammers)
	Wolz (2010) [113]		0.85	-	-	-	-	-	-	-	-	DSC	796v (ADNI)
	Rousseau (2011) [109]		0.83	0.89	0.89	0.89	0.79	0.75	0.67	0.93	0.93	DSC	18v (IBSR)
	Coupé (2011) [110]		0.88	-	-	-	-	-	-	-	-	KI	80v (ICBM)
			-	-	-	-	-	-	-	0.96	-	KI	80v
		[117]	0.88	-	-	-	-	-	-	-	-	DSC	80v (ICBM)
		[117]	0.85	-	-	-	-	-	-	-	-	DSC	202v (ADNI)
	Jia (2012) [103]		-	-	-	-	-	-	-	0.91	-	DSC	50v (ADNI)
	Cardoso (2013) [104]		0.90	-	-	-	-	-	-	-	-	DSC	30v (ADNI)
			0.84	0.89	0.89	0.89	0.80	0.81	0.70	-	-	DSC	30v (Hammers)
	Wang (2013) [102]		0.87	0.92	0.88	0.91	-	0.82	0.80	-	0.95	DSC	20v (MICCAI'12)
	Wang (2014) [96]		0.80	0.89	0.75	0.88	0.84	0.76	0.71	0.84	0.90	DSC	20v (MICCAI'12)
			0.58	0.52	0.79	0.38	0.42	0.56	0.55	0.61	0.55	MAD	20v (MICCAI'12)
			5.95	4.10	5.36	3.12	2.74	3.65	4.00	8.76	6.85	HD	20v (MICCAI'12)
	Pipitone (2014) [100]		0.87	-	-	-	-	-	-	-	-	DSC	60v (ADNI1)
			0.89	-	-	-	-	-	-	-	-	DSC	81v (FEP)
	Wu (2015) [111]		0.85	0.90	0.90	0.89	0.80	0.82	0.71	-	-	DSC	30v (Hammers)
Supervised	Pitiot (2002) [119]		0.91	-	0.90	-	-	-	-	-	-	CR	10v
	Morra (2008) [121]		0.70	-	-	-	-	-	-	-	-	RO	83v (AD)
			0.82	-	-	-	-	-	-	-	-	SI	83v (AD)
			4.15	-	-	-	-	-	-	-	-	HD	83v (AD)
	Moghaddam (2009) [120]		-	0.89	0.83	0.88	-	-	-	-	-	DSC	6v (IBSR)
			-	0.90	0.75	0.70	-	-	-	-	-	MAD	6v (IBSR)
			-	2.21	2.40	1.92	-	-	-	-	-	HD	6v (IBSR)
	Wels (2009) [124]		0.73	-	0.80	0.82	0.75	-	-	-	-	DSC	18v (IBSR)
			0.91	-	0.67	0.72	0.79	-	-	-	-	MAD	18v (IBSR)
			-	-	0.66	-	-	-	-	-	-	MAD	MICCAI'07
	Tong (2013) [117]		0.87	-	-	-	-	-	-	-	-	DSC	202v (ADNI)
			0.89	-	-	-	-	-	-	-	-	DSC	80v (ICBM)
		[126]	-	-	0.87	0.90	0.86	-	0.74	-	-	DSC	35v (MICCAI'12)
	Kim (2013) [123]		0.82	-	-	-	-	-	-	-	-	RO	20v (7T)
			0.89	-	-	-	-	-	-	-	-	SI	20v (7T)
	Benkarim (2014) [126]		-	-	0.87	0.91	0.87	-	0.76	-	-	DSC	35v (MICCAI'12)
Bayesian	Fischl (2002) [26]	[132]	0.84	0.88	0.85	0.85	0.80	0.75	-	-	-	DSC	39v (Sabuncu et al.)
		[126]	-	-	0.82	0.79	0.74	-	0.55	-	-	DSC	35v (MICCAI'12)
		[115]	0.85	0.88	0.85	0.84	0.79	0.80	-	0.88	-	DSC	39v (FS atlas)
		[137]	0.79	0.88	0.79	0.81	0.71	0.71	-	-	-	DSC	30v
		[109]	0.75	0.86	0.82	0.81	0.71	0.68	0.58	0.78	-	DSC	18v (IBSR)
	Scherrer (2007) [134]		-	0.80	0.76	0.79	-	-	-	-	-	DSC	BrainWeb
	Akselrod-Ballin (2007) [136]		0.69	0.84	0.80	0.79	0.74	0.63	-	-	0.84	DSC	18v (IBSR)
			1.88	1.44	1.44	1.60	2.43	1.67	-	-	1.62	MAD	18v (IBSR)
			4.57	2.90	3.07	3.36	3.75	3.38	-	-	3.42	HD	18v (IBSR)
		[109]	0.69	0.84	0.80	0.79	0.74	0.63	-	-	0.84	DSC	18v (IBSR)
	Pohl (2007) [14]		0.81	-	-	-	-	0.86	-	-	-	DSC	50v
	Scherrer (2009) [140]		-	0.72	0.83	0.77	-	-	-	-	-	DSC	BrainWeb
	Riklin-Raviv (2010) [132]		0.76	0.85	0.82	0.85	0.78	0.79	-	-	-	DSC	39v (Sabuncu et al.)
	Makropoulos (2014) [131]		0.79	0.90	0.85	-	-	0.83	-	0.84	0.92	DSC	20v (ALBERTs)
Deformable	Duchesne (2002) [114]		0.68	-	-	-	-	0.63	-	-	-	KI	30v (ICBM)
			0.67	-	-	-	-	0.61	-	-	-	KI	10v (ICBM)
	Pitiot (2004) [31]		3.00	-	2.00	-	-	-	-	2.60	-	HD	20v
			2.10	-	1.60	-	-	-	-	1.80	-	MAD	20v
	Shariatpanahi (2006) [147]		-	0.78	0.72	0.74	-	-	-	-	-	Sim	4v (IBSR)
			-	1.54	1.17	1.10	-	-	-	-	-	HD	4v (IBSR)
	Colliot (2006) [144]		-	-	2.20	-	-	-	-	-	-	HD	10v
			-	-	1.00	-	-	-	-	-	-	MAD	10v
			-	-	0.87	-	-	-	-	-	-	SI	10v
	Babalola (2007) [155]		-	-	0.73	-	-	-	-	-	-	JC	24v
	Patenaude (2011) [15]		0.80	0.87	0.84	0.89	0.78	0.73	0.72	-	0.86	DSC	37v (NC+SZ)
			0.83	0.87	0.84	0.87	0.78	0.77	0.71	-	0.86	DSC	42v (NC+AD)
			0.84	0.87	0.87	0.86	0.76	0.76	0.67	-	0.86	DSC	17v (NC+AD)
			0.80	0.85	0.85	0.88	0.72	0.73	0.73	-	0.81	DSC	87v (NC+SZ)
			0.80	0.87	0.84	0.86	0.76	0.74	0.67	-	0.85	DSC	14v (NC+PC)
			0.81	0.86	0.84	0.89	0.79	0.74	0.70	-	0.83	DSC	139v (NC+ADHC+SZ)
	Gao (2012) [153]		0.82	-	0.91	-	-	-	-	-	-	DSC	24v (HIP); 24v (CAU)
			3.32	-	2.36	-	-	-	-	-	-	HD95	24v (HIP); 24v (CAU)
	Fouquier (2012) [145]		-	2.13	3.16	3.25	-	-	-	-	-	MAD	30v (IBSR+OASIS)

	Article	Ref	Segmented structures									Measure	Database
			HIP	THA	CAU	PUT	PAL	AMY	ACC	LV	BS		
Region	Xue (2000) [163]		-	0.94	0.91	0.95	-	-	-	0.98	-	SI	-
			-	0.92	0.90	0.93	-	-	-	0.96	-	KI	-
	Xue (2001) [159]		-	0.94	0.91	0.95	-	-	-	0.98	-	SI	-
			-	0.92	0.90	0.93	-	-	-	0.96	-	KI	-
	Xia (2007) [77]		-	-	0.87	-	-	-	-	-	-	Over	55v
	Gui (2012) [160]		-	-	-	-	-	-	-	-	0.90	DSC	10v (newborns)
Hybrid	Zhou (2005) [19]		0.71	0.84	0.81	0.83	-	0.65	-	-	-	Over	17v (IBSR)
			0.56	0.32	0.32	0.29	-	0.67	-	-	-	MAD	17v (IBSR)
	Tu (2008) [164]		10.50	-	7.35	10.15	-	-	-	6.60	-	HD	14v
			2.40	-	1.45	2.50	-	-	-	1.10	-	MD	14v
	Karsch (2009) [165]		0.70	-	-	-	-	-	-	0.80	-	DSC	-
			0.66	-	-	-	-	-	-	0.77	-	Over	-
	Sabuncu (2009) [137]		0.81	0.84	0.84	0.89	0.83	0.80	-	-	-	DSC	30v
	Sabuncu (2010) [115]		0.87	0.91	0.87	0.89	0.84	0.82	-	0.91	-	DSC	39v (FS atlas)
		[104]	0.82	0.89	0.89	0.87	0.77	0.78	0.67	-	-	DSC	30v (Hammers)
		[104]	0.87	-	-	-	-	-	-	-	-	DSC	30v (ADNI)
	He (2011) [166]		-	0.70	-	-	-	-	-	0.80	-	DSC	20v+25v
			-	0.66	-	-	-	-	-	0.77	-	Over	20v+25v
	van der Lijn (2012) [167]		0.87	-	-	-	-	-	-	-	-	DSC	18v (Set I)
			0.77	-	-	-	-	-	-	-	-	JC	18v (Set I)
			0.87	-	-	-	-	-	-	-	-	DSC	18v (Set II)
			0.76	-	-	-	-	-	-	-	-	JC	18v (Set II)
	Iglesias (2013) [162]		0.80	0.88	0.85	0.89	0.83	0.70	-	0.81	-	DSC	8v (PD)
	Liu (2013) [138]		0.78	0.89	0.84	0.87	0.81	0.73	-	0.81	-	DSC	6v (IBSR)
			0.83	-	0.81	0.84	-	-	-	-	-	DSC	15v (LPBA40)
			7.82	-	4.21	6.85	-	-	-	-	-	HD	15v (LPBA40)
			4.90	-	4.89	5.56	-	-	-	42.68	-	HD	28v (LONI28)
			1.25	-	0.91	0.96	-	-	-	0.72	-	MAD	28v (LONI28)
	Hao (2014) [16]		0.89	-	-	-	-	-	-	-	-	DSC	30v (ADNI 1.5T)
			3.26	-	-	-	-	-	-	-	-	HD	30v (ADNI 1.5T)
			0.27	-	-	-	-	-	-	-	-	MAD	30v (ADNI 1.5T)
			0.91	-	-	-	-	-	-	-	-	DSC	30v (ADNI 3T)
			1.88	-	-	-	-	-	-	-	-	HD	30v (ADNI 3T)
			0.21	-	-	-	-	-	-	-	-	MAD	30v (ADNI 3T)
			0.91	-	-	-	-	-	-	-	-	DSC	57v (3T)
			2.91	-	-	-	-	-	-	-	-	HD	57v (3T)
			0.25	-	-	-	-	-	-	-	-	MAD	57v (3T)

results are provided with different measures which have been obtained on different databases with different number of volumes. Ideally, a comparison of the methods should have been done using the same dataset and measures, but as we can see from the table only a few of those methods share these properties. The choice of the cited brain structures has been done because there were the ones for which a sufficient number of quantitative results to perform analysis was available. The results are shown as averages for each structure (left and right pair combined), except the brainstem (BS) which is a unique structure.

From the table we can observe that the most commonly used evaluation measures are those based on volume overlap, in particular DSC. For this reason, we perform a first analysis based on this measure regardless of the data used, to have an overall quantitative estimate of the state-of-the-art methods reviewed here. Another reason of choosing an overlap measure is that the distance measures are calculated differently in each of the works, even if the name of the measure is the same, it can be symmetric or asymmetric and the results can be given in pixels, millimeters or even being not specified, thus the provided distance measure is not always comparable.

At first glance, we can see that the structures on which more works have focused on are the hippocampus and the caudate nucleus, whereas the structures on which less attention has been paid are the brainstem and the nucleus accumbens. Furthermore, we can say that the nucleus accumbens is the most difficult structure to segment, with a mean DSC of 0.69, whereas the structures that seem to get the best results in terms of volume overlap are the brainstem with a mean DSC of 0.88, closely followed by the thalamus (with mean DSC of 0.87) and the putamen (mean DSC of 0.86). The lower results for the accumbens are reasonable since it is the smallest structure, thus small errors in overlap give the highest changes in the DSC. On the other hand, the brainstem presents relatively strong contrast boundaries therefore its segmentation should be easier compared to the rest of the structures.

If we perform an analysis by structure, for hippocampus segmentation we can highlight the work of Hao et al. [16] which we have classified in the hybrid category. Their method has been evaluated with three different databases with a total number of 117 volumes, achieving DSC values ranging from 0.89 to 0.91. Moreover, the works of Cardoso et al. [104] and Pipitone et al. [100], both atlas-based strategies, which have been tested with 60 and 141 volumes respectively (two different databases each) obtained DSC values of 0.84-0.90 and 0.87-0.89. Finally, the learning-based approach proposed by Tong et al. [117], achieved DSC values of 0.87 and 0.89 in two different databases, with a total number of 282 cases.

Regarding the thalamus, some important works have to be mentioned, such as the ones by Aljabar et al. [99] and by Wang and Yushkevich [102], which achieved DSC coefficients of 0.91 and 0.92 respectively, when tested with 275 and 20 volumes (both classified in the atlas-based category). Another remarkable work for segmenting this structure is the deformable strategy proposed by Patenaude et al. [15], which was tested with 6 different databases with a total number of 336 volumes and obtained DSC values between 0.85 and 0.87. Moreover, the hybrid approach presented by Sabuncu et al. [115] achieved a DSC of 0.89 and 0.91 in two different datasets with a total of 69 testing images.

In caudate nucleus segmentation, the work of Gao et al. [153], classified as a deformable strategy, obtained a DSC of 0.91 when tested with 24 cases. Moreover, the atlas-based approach presented by Wu et al. [111] and tested with 30 cases, achieved a DSC of 0.90. Finally, the hybrid proposal of Sabuncu et al. [115] obtained DSC values of 0.87 and 0.89 with a total of 69 cases.

Segmenting the putamen, some atlas-based strategies have provided good results such as the ones by Aljabar et al. [99], Lötjönen et al. [95] and Wang and Yushkevich [102] that achieved DSC values of 0.90, 0.90 and 0.91 respectively when tested with 275, 18 and 20 volumes. Some learning-based strategies have also obtained good results in putamen and globus pallidum segmentation, which include the approaches of Tong et al. [117] and Benkarim et al. [126] , that obtained DSC values for putamen segmentation of 0.90 and 0.91 respectively, and 0.86 and 0.87 for pallidum segmentation, both with a database composed of 35 images. Finally, the deformable approach proposed by Patenaude et al. [15] is also remarkable for the large amount of data used for testing (336v). When segmenting the putamen, the authors achieved DSC values ranging from 0.86 to 0.89.

For amygdala segmentation we can highlight two learning-based approaches, which are the ones proposed by Pohl et al. [14] and by Makropoulos et al. [131] which achieved DSC values of 0.86 and 0.83 respectively, tested with 50 and 20 cases. In the category of atlas-based strategies we can remark the works of Wang and Yushkevich [102] and Wu et al. [111], both obtaining a DSC of 0.82 when tested with 20 and 30 volumes respectively.

With regard to nucleus accumbens segmentation, some notable works are those presented by Aljabar et al. [99], Wang and Yushkevich [102] and Benkarim et al. [126]. The first two are atlas-based strategies, while the third is a learning-based approach. The DSC values achieved when segmenting this structure are 0.76, 0.8 and 0.76, respectively (275, 20 and 35 testing volumes).

In segmenting the lateral ventricles and the brainstem, atlas-based approaches provide good results. The work presented by Rousseau et al. [109] obtained a DSC of 0.93 for both structures with a testing dataset of 18 volumes. Furthermore, Aljabar et al. [99] and Jia et al. [103] achieved both a DSC value of 0.91 when segmenting the lateral ventricles with 275 and 50 volumes respectively, whereas Wang and Yushkevich [102] obtained a DSC of 0.95 in brainstem segmentation with a testing cohort of 20 cases. It has to be highlighted that in the other categories there are not too many works that segment these particular structures, however the hybrid approach proposed by Sabuncu et al. [115] achieved good results when segmenting the lateral ventricles (DSC of 0.91 with 39 testing volumes).

If we perform a second analysis based only on the works that use the same database and measure (DSC) for evaluation, we can state that for the IBSR18 database, the approach that seems to perform best in terms of volume overlap for almost all the structures is the one presented by Rousseau et al. [109], which is a patch-based label fusion strategy and therefore classified in the atlas-based category. On the other hand, if we look only at the works that use the Hammers adult atlases database for evaluation, it seems that the strategy that achieved the highest DSC for all the evaluated structures is the one proposed by Wu et al. [111], which is also a patch-based approach (atlas-based) that dynamically adjusts the patch size during the label fusion procedure. Finally, when using the 35 volumes of the MICCAI12 database for evaluation, the approach that performed best is the multi-class dictionary learning approach presented by Benkarim et al. [126].

5.4. Quantitative analysis of available software

We have evaluated three publicly available and well-known software tools, each relying on a different category of our classification, namely MABMIS [103] (atlas-based), FreeSurfer [26] (learning-based) and FIRST [15] (deformable). **The three software tools have been evaluated on the 30 subjects of the Hammers Adult atlases database [179, 180] with default parameters. See Table 6 for the details of this database.** The Dice similarity coefficient **and the**

Table 8: Software tools evaluation on the Hammers adult atlases database. **The results show the Dice similarity coefficient (DSC) and the Hausdorff distance (HD) for each structure separately (mean±std).** Acronyms from left to right are: hippocampus (HIP), thalamus (THA), caudate nucleus (CAU), putamen (PUT), pallidum (PAL), amygdala (AMY), accumbens (ACC), lateral ventricle (LV) and brainstem (BS). The results are averages for each structure (left and right pair combined), except for brainstem. The - symbol indicates that the FIRST software does not segment the lateral ventricles. (*) The low DSC values reported for the hippocampus are due to the ground truth segmentations of the Hammers atlases database, in which only the hippocampus head is labeled, whereas the evaluated tools segment the whole structure.

Method	Ms	HIP (*)	THA	CAU	PUT	PAL	AMY	ACC	LV	BS
MABMIS	**DSC**	0.65±0.04	0.85±0.02	0.80±0.03	0.84±0.03	0.71±0.07	0.58±0.08	0.50±0.08	0.82±0.04	0.73±0.03
(atlas)	**HD**	**17.75±6.35**	**6.73±3.03**	**8.40±2.28**	**5.11±1.11**	**4.98±1.08**	**7.08±2.04**	**7.76±2.00**	**37.00±2.59**	**16.76±1.89**
FreeSurfer	**DSC**	0.61±0.03	0.82±0.03	0.81±0.02	0.80±0.02	0.70±0.06	0.73±0.03	0.48±0.06	0.73±0.07	0.83±0.01
(learning)	**HD**	**19.00±6.17**	**7.96±2.74**	**7.00±1.50**	**5.50±0.82**	**5.21±1.61**	**4.37±0.81**	**8.04±1.97**	**39.30±2.44**	**14.30±1.87**
FIRST	**DSC**	0.64±0.02	0.85±0.03	0.85±0.02	0.87±0.02	0.75±0.05	0.74±0.04	0.54±0.08	-	0.77±0.10
(deform.)	**HD**	**19.55±5.72**	**7.81±2.87**	**6.34±1.90**	**4.40±1.13**	**4.65±1.38**	**4.18±0.76**	**9.29±6.91**	**-**	**15.63±3.94**

Hausdorff distance have been computed for the following structures: hippocampus, thalamus, caudate nucleus, putamen, pallidum, amygdala, accumbens, lateral ventricles and brainstem. The obtained results are shown in Table 8 for each of the strategies.

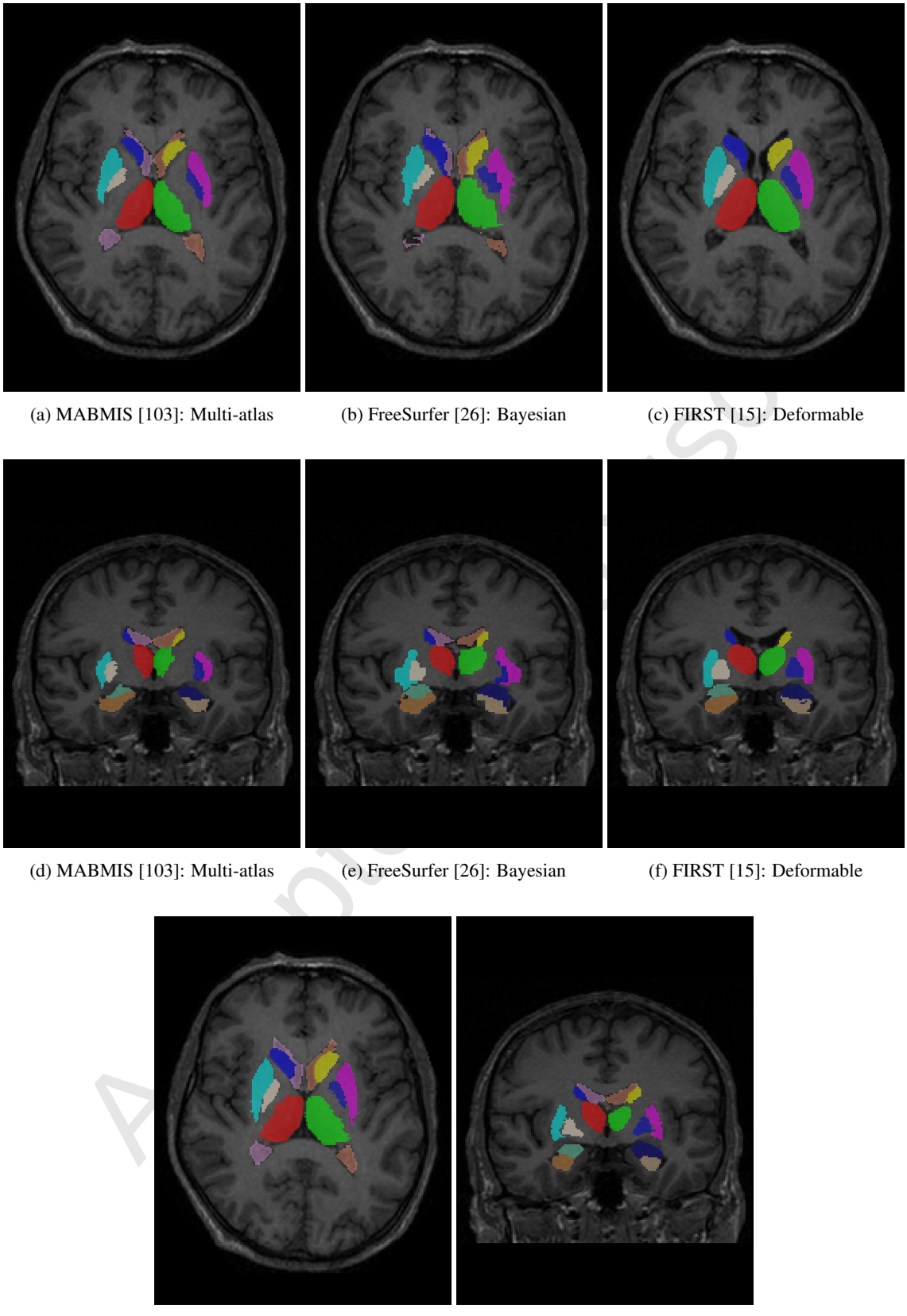
Analyzing each strategy separately, we can observe that in terms of both volume overlap (DSC) and contour distance (HD), the deformable method is the one that performs better for caudate nucleus (0.85 mean DSC / 6.34 mean HD, respectively), putamen (0.87 / 4.40) and globus pallidum (0.75 / 4.65), whereas the atlas-based strategy is the one that performs better for thalamus (0.85 / 6.73) and lateral ventricles (0.82 / 7.76) segmentation. We can see from the table that the learning-based strategy provides the lowest performance in segmenting the thalamus (0.82 / 7.96), the putamen (0.80 / 5.50), the pallidum (0.70 / 5.21) and the lateral ventricles (0.73 / 39.30), while the atlas-based strategy provides the poorest results when segmenting the caudate nucleus (0.80 / 8.40). On the other hand, looking only at the HD measure for comparison, we can see that for hippocampus (17.75) and accumbens (7.76) segmentation, the atlas-based method performs better than the others.

We can see that the results obtained here mostly follow the trend observed in the works highlighted in section 5.3 for each particular structure. For hippocampus segmentation, two of the four enhanced works were atlas-based (Cardoso et al. [104] and Pipitone et al. [100]) while the same happened in thalamus segmentation (Aljabar et al. [99] and Wang and Yushkevich [102]). In caudate nucleus segmentation, the approach that achieved the highest DSC (Gao et al. [153]), with a total of 24 testing cases, was a deformable strategy. On the other hand, in putamen segmentation, the work of Patenaude et al. [15], which is the deformable method analyzed in this section, even not being the one that achieved the highest DSC results, was highlighted due to its results robustness with a large testing dataset (336 volumes). Finally, either for accumbens or lateral ventricles segmentation, almost all of the highlighted methods were atlas-based [99, 102, 103, 109].

As shown in Table 8, when using MABMIS for segmentation, the results for the amygdala and the brainstem are much poorer than the ones of the learning-based and deformable strategies. The reason for this arises from the database used to build the atlases tree in the MABMIS algorithm. In this case we used as atlases the MICCAI'12 database, whose labels for the amygdala and brainstem are not exactly the same as the ones in the testing dataset (Hammers adult atlases). A common practice in the evaluation of multi-atlas algorithms is to use the atlases database, leaving one atlas out for testing and repeating this strategy for all the atlases in the database. This can be a good practice although it can also provide biased results.

Another relevant issue from the performed quantitative evaluation is that the DSC values obtained for the hippocampus when using the same tools are much lower than the ones gathered from the analyzed works in the literature. This is also due to the ground truth segmentations of the Hammers atlases database, in which the hippocampus is labeled following a different protocol. Therefore, the non-existence of a standardized labeling protocol for brain structures segmentation constitutes another open problem in order to evaluate the performance of automatic brain structure segmentation methods.

Figure 1 shows some illustrative results of automatic brain structures segmentation obtained with these different strategies. As can be seen from this figure, the result of the atlas-based approach is similar to the ground-truth, in spite of the atlases used were collected from a different database to that of the target image, whereas



(a) MABMIS [103]: Multi-atlas (b) FreeSurfer [26]: Bayesian (c) FIRST [15]: Deformable

(d) MABMIS [103]: Multi-atlas (e) FreeSurfer [26]: Bayesian (f) FIRST [15]: Deformable

(g) Ground-truth axial (h) Ground-truth coronal

Figure 1: Example of automatic structures segmentation using various publicly available softwares. Image a03 from Hammers's work [179][180]

Table 9: MICCAI'07 Challenge (CAUSE07) results. Acronyms from left to right are: volumetric overlap error (OE), relative absolute volume difference (VD), average symmetric surface distance (AD), root mean squared symmetric surface distance (RMD) and maximum symmetric surface distance (MD). OE and VD results are shown in percentage while AD, RMSD and MD are measured in millimeters.

Method	OE	VD	AD	RMD	MD	Score	Strategy
ISICAD [185]	31.09	4.98	0.62	1.03	7.03	79.16	Multi-atlas
MIAL-SFU [186]	25.49	-5.01	0.56	1.41	12.77	75.99	Hybrid
Moghaddam [120]	29.98	-0.96	0.64	1.37	11.79	75.77	Learning-based

Table 10: MICCAI'12 Grand Challenge results. Dice similarity coefficient (DSC) results for cortical structures (CS) and non-cortical structures (NCS). The fourth column of the table shows the mean DSC across all brain labels and all subjects in the testing cohort.

Method	DSC (CS)	DSC (NCS)	Overall mean DSC	Strategy
Wang and Yushkevich [102]	0.73	0.83	0.76	Multi-atlas
Non Local STAPLE [105]	0.73	0.82	0.75	Multi-atlas
MALP_EM [187]	0.73	.82	0.75	Hybrid

the learning-based approach, as already mentioned, is the one that seems to provide the least accurate results, not being able to provide well defined structure boundaries, and including misclassifications such as the ones in the lateral ventricles (purple label). The deformable method, contrary to the other two, performs a well-defined segmentation, as it is topologically constrained.

5.5. Quantitative analysis of the MICCAI Challenges

Analyzing the results of the MICCAI challenges we see that for caudate nucleus segmentation (CAUSE07 challenge) the method that gave the best overall result was the ISICAD [185]. This method performed an adaptive local multi-atlas segmentation that locally decided how many and which atlases were needed to segment a target image and registered only the selected parts of those atlases. The second place in this challenge was for the MIAL-SFU [186] algorithm, that can be classified in the hybrid category, since it used FreeSurfer segmentation for initialization (which has been classified as a learning-based algorithm) and then looked for the best transformation to perform label propagation. Finally, the third position in this caudate segmentation challenge was for the method proposed by Moghaddam and Soltanian-Zadeh [120], a learning-based strategy. The numerical results obtained in the challenge for these three methods are detailed in Table 9.

For the MICCAI 2012 Grand Challenge, we observe that the best segmentation strategy was the one presented by Wang and Yushkevich [102] which performed label fusion followed by 'corrective learning'. The second position was for the Non Local STAPLE (NLS) algorithm [105], which is also a label fusion strategy, while the third position was for the MALP_EM [187] algorithm. MALP_EM combined both multi-atlas label propagation and probabilistic segmentation, performing a locally weighted fusion to obtain probabilistic labels that were used as priors in an EM refinement step. It is important to remark that the representation of each category in this challenge may be biased since, as said before, although any reproducible method was accepted, it was assumed the majority of the participant methods to be multi-atlas based. Therefore, it should come as no surprise that the best strategies relied on the atlas-based strategy. The obtained results for these three methods in terms of Dice similarity coefficient are shown in Table 10.

Finally, in the free-for-all subchallenge of the MICCAI SATA challenge the mean/median DSC results ranged from 0.61/0.63 to 0.86/0.87 and the mean/median Symmetric HD results ranged from 3.30/3.10 mm to 7.68/8.00 mm. The best entry in terms of both DSC and Symmetric HD was the UPENN_SBIA_MAM algorithm [188], which was based on a multi-atlas strategy that performed atlas selection.

6. Concluding remarks and future trends

6.1. Overview

In this article we have reviewed the problem of automatic brain structures segmentation in MR images, presenting a classification of the state-of-the-art methods based on the segmentation strategy used, where we have identified five main categories. The first category includes the approaches that rely on topological atlas registration, either using a single atlas (label propagation) or a set of them (label fusion). The second category in this classification comprises the learning-based strategies which have been subdivided into supervised methods such as classifiers, like ANNs

or SVM, and methods based on Bayesian inference. Finally, the last three categories include deformable methods (which include ACM, ASM, AAM and other energy minimizing strategies), region-based approaches and strategies that combine some of the methods of the previous categories. To conclude this classification, we have discussed about the strengths and weaknesses that each category presents.

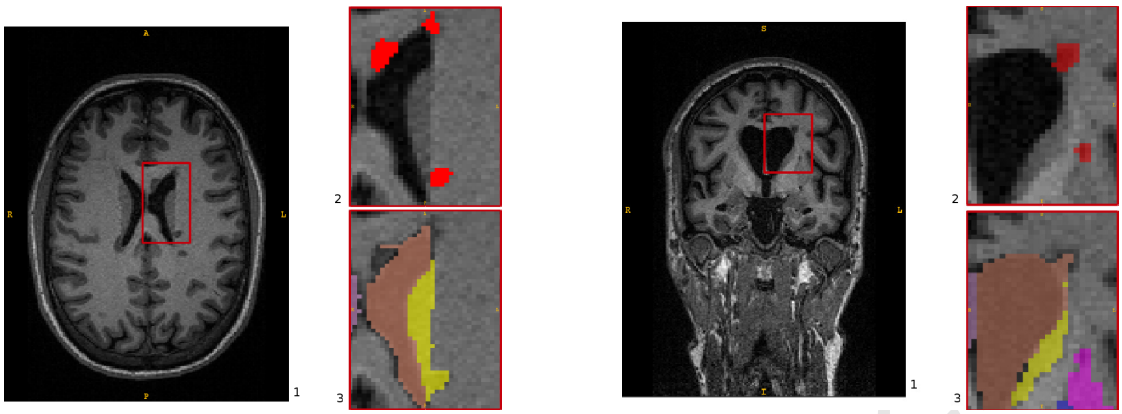
As we have seen from Tables 2 to 4, a comparison of the state-of-the-art methods is not an easy task, since there is a lack of standard in both databases and metrics for evaluation. In order to solve this problem, some challenges [182–184] have been proposed in recent years, but they are either target specific (caudate nucleus) [182] or methodology oriented (multi-atlas) [183, 184], which makes the participant methods in the challenge to be not representative of the overall strategies available in the literature. Furthermore, the quantitative results of the challenges are often given in terms of average measures for the whole structures instead of giving evaluation results for each structure independently, which means that an analysis of which method performs better for a given structure cannot be deduced.

We have analyzed the results of the different works reviewed in this article, providing an overview of the state-of-the-art in terms of different evaluation measures, databases and number of cases used, showing the results both from the point of view of the segmentation strategies and segmented target structures. Furthermore, a quantitative evaluation of three publicly available software tools, each relying on a different category of our classification, have been performed. Lastly, we have commented on the results and the methods presented to the three MICCAI challenges in brain structures segmentation.

6.2. Future trends

In recent years, it seems there exists a trend on the use of multi-atlas strategies for brain structures segmentation, either as a segmentation method on its own or combined with other strategies. Given the fact that its greatest weakness is that it is computationally very expensive due to the big amount of registrations it has to perform, some strategies have included either atlas selection [94, 97, 102] to reduce the number of registrations or non-local weighting label fusion [109, 110] in which only affine registrations are needed, which reduces significantly the computational time. All that, in combination with computers becoming more powerful, seems to make multi-atlas approaches a good strategy to follow [102, 111]. **Furthermore, hybrid methods have shown to achieve good results [16, 19, 115] in segmenting brain structures as they can combine the advantages of each category of methods and try to overcome their weaknesses with the strengths of the other methods in the amalgamation.** Because of that, combining multi-atlas methods with any other method in the classification could be a good line of investigation. Atlases give robustness to the method and they are good in providing spatial information for structures, so merging atlases with another method which is able to model the structures appearance such as supervised classifiers or AAM could lead to improved segmentation results. Van der Lijn et al. [167] presented work relying on this strategy for hippocampus and cerebellum segmentation, achieving mean Dice similarity indexes of 0.87 and 0.95 respectively.

Most of the methods reviewed here are tested with images of non-lesioned brains, either healthy subjects or patients with schizophrenia, Alzheimer, epilepsy or attention deficit hyperactivity disorder. One of the biggest problems is that, as far as we know, there is not any publicly available database of lesioned brains with ground truth of structures segmentation, to test or train the methods. However, when performing structures segmentation in brain MRI of patients with demyelinating lesions (like in multiple sclerosis or lupus) or space-occupying lesions (such as tumors), the performance of some of these methods is affected, providing less accurate results. To the best of our knowledge, how these lesions affect the automatic brain structure segmentation algorithms has not yet been evaluated. In Figure 2 we show an example of how these lesions can introduce errors in the automatic brain structures segmentation. In Figure 2a we can see how either the caudate nucleus (yellow label) and the lateral ventricle (brown label) are oversegmented, due to the fact that the automatic method is interpreting the lesions as part of the structures. The same situation is seen in Figure 2b, in which multiple sclerosis lesions produce false positives in the putamen (pink label) and lateral ventricle segmentation. As far as we know there has not been a proposal tackling this issue and therefore there is a need to improve the performance of such methods trying to make them robust with brain lesions. Note that an accurate segmentation is necessary in order to perform disease follow-up, for instance, in multiple sclerosis patients where deep gray nuclei atrophy is closely related to the magnitude of inflammation. Interestingly, lesion filling techniques [170–172] to reduce the effect of hypo-intense T1-weighted multiple sclerosis lesions have already been applied to assess the progression of GM atrophy [173, 174] showing an improvement on the accuracy of tissue volume. Furthermore, integrat-



(a) (b)

Figure 2: Structures segmentation in brains with demyelinating lesions. FreeSurfer [133] segmentation results show that these lesions affect the structures segmentation performance, increasing the number of false positives. Images show (1) the original T1-weighted MRI, (2) the lesion ground truth and (3) the automatic segmentation. Figure 2a shows oversegmentation of the caudate nucleus (yellow) and the lateral ventricle (brown) whereas in Figure 2b the same situation is shown for the putamen (pink) and the lateral ventricle. Images from Vall d'Hebron Hospital, Barcelona.

ing automatic lesion segmentation and filling on automatic tissue segmentation pipelines has recently been studied [169], showing very similar results to that of manually segmenting the lesions. This has not yet been integrated as part of any automatic brain structures segmentation pipeline and indeed opens new challenges to the research community.

Acknowledgements

S. González-Villà holds a UdG-BRGR2015 grant from the University of Girona. This work has been partially supported by "La Fundació la Marató de TV3", by Retos de Investigación TIN2014-55710-R, and by MPC UdG 2016/022 grant."

References

[1] M. Filippi, M. A. Rocca, O. Ciccarelli, N. DeStefano, N. Evangelou, L. Kappos, A. Rovira, J. Sastre-Garriga, M. Tintorè, J. L. Frederiksen, C. Gasperini, J. Palace, D. S. Reich, B. Banwell, X. Montalban, F. Barkhof, MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines, The Lancet Neurology 15 (3) (2016) 292–303.

[2] C. Jacobsen, J. Hagemeier, K.-M. Myhr, H. Nyland, K. Lode, N. Bergsland, D. P. Ramasamy, T. O. Dalaker, J. P. Larsen, E. Farbu, R. Zivadinov, Brain atrophy and disability progression in multiple sclerosis patients: a 10-year follow-up study, Journal of Neurology, Neurosurgery & Psychiatry 85 (10) (2014) 1109–1115.

[3] N. C. Andreasen, D. Liu, S. Ziebell, A. Vora, B.-C. Ho, Relapse duration, treatment intensity, and brain tissue loss in schizophrenia: A prospective longitudinal MRI study, The American Journal of Psychiatry 170 (6) (2013) 609–615.

[4] R. C. Knickmeyer, S. Gouttard, C. Kang, D. Evans, K. Wilber, J. K. Smith, R. M. Hamer, W. Lin, G. Gerig, J. H. Gilmore, A structural MRI study of human brain development from birth to 2 years, The Journal of Neuroscience 28 (47) (2008) 12176–12182.

[5] T. F. Chan, L. A. Vese, Active contours without edges, IEEE Trans. Image Proc. 10 (2) (2001) 266–277.

[6] Z. Lin, J. Jin, H. Talbot, Unseeded region growing for 3d image segmentation, in: P. Eades, J. Weckert (Eds.), Selected papers from the Pan-Sydney workshop on Visualisation, Vol. 2, Australian Computer Society, Inc., Darlinghurst, Australia, 2000, pp. 31–37.

[7] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, T.-J. Chen, Fuzzy c-means clustering with spatial information for image segmentation, Computerized Medical Imaging and Graphics 30 (1) (2006) 9–15.

[8] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Machine Intell. 22 (8) (2000) 888–905.

[9] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, Face recognition: A literature survey, ACM Computing Surveys 35 (4) (2003) 399–458.

[10] S. Brutzer, B. Hferlin, G. Heidemann, Evaluation of background subtraction techniques for video surveillance, in: IEEE Conf. Comput. Vision Pattern Recog., Institute of Electrical and Electronics Engineers (IEEE), Colorado Springs, USA, 2011, pp. 1937–1944.

[11] A. ckermann, C. Elbrechter, R. Haschke, H. Ritter, 3D scene segmentation for autonomous robot grasping, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, Institute of Electrical and Electronics Engineers (IEEE), Vila Moura, Portugal, 2012, pp. 1734–1740.

[12] D. L. Collins, A. C. Evans, ANIMAL: Validation and applications of nonlinear registration-based segmentation, Int. J. Pattern Recogn. Artif. Intell. 11 (08) (1997) 1271–1294.

[13] S. Warfield, K. Zou, W. Wells, Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation, IEEE Trans. Med. Imag. 23 (7) (2004) 903–921.

[14] K. Pohl, S. Bouix, M. Nakamura, T. Rohlfing, R. McCarley, R. Kikinis, W. Grimson, M. Shenton, W. Wells, A hierarchical algorithm for MR brain image parcellation, IEEE Trans. Med. Imag. 26 (9) (2007) 1201–1212.

[15] B. Patenaude, S. M. Smith, D. N. Kennedy, M. Jenkinson, A Bayesian model of shape and appearance for subcortical brain segmentation, NeuroImage 56 (3) (2011) 907–922.

[16] Y. Hao, T. Wang, X. Zhang, Y. Duan, C. Yu, T. Jiang, Y. Fan, A. D. N. Initiative, Local Label Learning (LLL) for subcortical structure segmentation: Application to hippocampus segmentation, Hum. Brain Mapp. 35 (6) (2014) 2674–2697.

[17] V. Dill, A. R. Franco, M. S. Pinho, Automated methods for hippocampus segmentation: the evolution and a review of the state of the art, Neuroinformatics 13 (2) (2015) 133–150.

[18] M.-P. Hosseini, M. Nazem-Zadeh, D. Pompili, H. Soltanian-Zadeh, Statistical validation of automatic methods for hippocampus segmentation in MR images of epileptic patients, in: Engineering in Medicine and Biology Society, EMBC, Annual International Conference of the IEEE, Institute of Electrical and Electronics Engineers (IEEE), Chicago, IL, USA, 2014, pp. 4707–4710.

[19] J. Zhou, J. C. Rajapakse, Segmentation of subcortical brain structures using fuzzy templates, NeuroImage 28 (4) (2005) 915–924.

[20] X. Lin, T. Qiu, F. Morain-Nicolier, S. Ruan, A topology preserving non-rigid registration algorithm with integration shape knowledge to segment brain subcortical structures from MRI images, Pattern Recog. 43 (7) (2010) 2418–2427.

[21] J. Dolz, L. Massoptier, M. Vermandel, Segmentation algorithms of subcortical brain structures on MRI for radiotherapy and radiosurgery: A survey, IRBM 36 (4) (2015) 200–212.

[22] K. O. Babalola, B. Patenaude, P. Aljabar, J. Schnabel, D. Kennedy, W. Crum, S. Smith, T. Cootes, M. Jenkinson, D. Rueckert, An evaluation of four automatic methods of segmenting the subcortical structures in the brain, NeuroImage 47 (4) (2009) 1435–1447.

[23] C. N. Devi, A. Chandrasekharan, V. Sundararaman, Z. C. Alex, Neonatal brain MRI segmentation: A review, Comput. Biol. Med. 64 (2015) 163–178.

[24] J. E. Iglesias, M. R. Sabuncu, Multi-atlas segmentation of biomedical images: A survey, Med. Image Anal. 24 (1) (2015) 205–219.

[25] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, R. V. Parsey, Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration, NeuroImage 46 (3) (2009) 786–802.

[26] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, A. M. Dale, Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain, Neuron 33 (3) (2002) 341–355.

[27] L. Igual, J. Soliva, R. Gimeno, S. Escalera, O. Vilarroya, P. Radeva, Automatic internal segmentation of caudate nucleus for diagnosis of attention-deficit/hyperactivity disorder, Image Analysis and Recognition 7325 (2012) 222–229.

[28] R. Kikinis, M. E. Shenton, D. V. Iosifescu, R. W. McCarley, P. Saiviroonporn, H. H. Hokama, A. Robatino, D. Metcalf, C. G. Wible, C. M. Portas, R. M. Donnino, F. A. Jolesz, A digital brain atlas for surgical planning, model-driven segmentation, and teaching, IEEE Trans. Visual. Comput. Graphics 2 (3) (1996) 232–241.

[29] F. Castro, C. Pollo, R. Meuli, P. Maeder, O. Cuisenaire, M. Cuadra, J.-G. Villemure, J.-P. Thiran, A cross validation study of deep brain stimulation targeting: From experts to atlas-based, segmentation-based and automatic registration algorithms, IEEE Trans. Med. Imag. 25 (11) (2006) 1440–1450.

[30] J. L. Phillips, L. A. Batten, P. Tremblay, F. Aldosary, P. Blier, A prospective, longitudinal study of the effect of remission on cortical thickness and hippocampal volume in patients with treatment-resistant depression, International Journal of Neuropsychopharmacology 18 (8) (2015) pyv037.

[31] A. Pitiot, H. Delingette, P. M. Thompson, N. Ayache, Expert knowledge-guided segmentation system for brain MRI, NeuroImage 23, Supplement 1 (2004) S85–S96.

[32] E. Geuze, E. Vermetten, J. Bremner, MR-based in vivo hippocampal volumetrics: 1. Review of methodologies currently employed, Molecular Psychiatry 10 (2) (2005) 147–159.

[33] N. Bernasconi, S. Duchesne, A. Janke, J. Lerch, D. Collins, A. Bernasconi, Whole-brain voxel-based statistical analysis of gray matter and white matter in temporal lobe epilepsy, NeuroImage 23 (2) (2004) 717–723.

[34] N. Fox, E. Warrington, P. Freeborough, P. Hartikainen, A. Kennedy, J. Stevens, M. N. Rossor, Presymptomatic hippocampal atrophy in Alzheimer's disease, Brain 119 (6) (1996) 2001–2007.

[35] L. L. Altshuler, G. Bartzokis, T. Grieder, J. Curran, J. Mintz, Amygdala enlargement in bipolar disorder and hippocampal reduction in schizophrenia: an MRI study demonstrating neuroanatomic specificity, Archives of General Psychiatry 55 (7) (1998) 663–664.

[36] G. Villarreal, D. A. Hamilton, H. Petropoulos, I. Driscoll, L. M. Rowland, J. A. Griego, P. W. Kodituwakku, B. L. Hart, R. Escalona, W. M. Brooks, Reduced hippocampal volume and total white matter volume in posttraumatic stress disorder, Biological Psychiatry 52 (2) (2002) 119–125.

[37] N. Kitayama, V. Vaccarino, M. Kutner, P. Weiss, J. D. Bremner, Magnetic resonance imaging (MRI) measurement of hippocampal volume in posttraumatic stress disorder: a meta-analysis, Journal of Affective Disorders 88 (1) (2005) 79–86.

[38] J. D. Bremner, M. Narayan, E. R. Anderson, L. H. Staib, H. L. Miller, D. S. Charney, Hippocampal volume reduction in major depression, The American Journal of Psychiatry.

[39] L. L. Altshuler, G. Bartzokis, T. Grieder, J. Curran, J. Mintz, Amygdala enlargement in bipolar disorder and hippocampal reduction in schizophrenia: an MRI study demonstrating neuroanatomic specificity, Archives of General Psychiatry 55 (7) (1998) 663–664.

[40] S. Strakowski, M. Delbello, C. Adler, The functional neuroanatomy of bipolar disorder: a review of neuroimaging findings, Molecular Psychiatry 10 (1) (2005) 105–116.

[41] H. P. Blumberg, J. Kaufman, A. Martin, R. Whiteman, J. H. Zhang, J. C. Gore, D. S. Charney, J. H. Krystal, B. S. Peterson, Amygdala and hippocampal volumes in adolescents and adults with bipolar disorder, Archives of General Psychiatry 60 (12) (2003) 1201–1208.

[42] A. L. Boxer, M. D. Geschwind, N. Belfor, M. L. Gorno-Tempini, G. F. Schauer, B. L. Miller, M. W. Weiner, H. J. Rosen, Patterns of brain atrophy that differentiate corticobasal degeneration syndrome from progressive supranuclear palsy, Archives of Neurology 63 (1) (2006) 81–86.

[43] R. P. Bote, M. Fernández-Gil, Degeneration of the brainstem, in: H. W. Raymond, G. Gayer, J. D. Swartz (Eds.), Seminars in Ultrasound, CT and MRI, Vol. 34, Elsevier, 2013, pp. 142–152.

[44] K. A. Josephs, Key emerging issues in progressive supranuclear palsy and corticobasal degeneration, Journal of Neurology 262 (3) (2015) 783–788.

[45] D. R. Williams, A. J. Lees, Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges, The Lancet Neurology 8 (2009) 270–79.

[46] L. T. Grinberg, U. Rüb, R. E. L. Ferretti, R. Nitrini, J. M. Farfel, L. Polichiso, K. Gierga, W. Jacob-Filho, H. Heinsen, B. B. B. S. Group, The dorsal raphe nucleus shows phospho-tau neurofibrillary changes before the transentorhinal region in Alzheimer's disease. A precocious onset?, Neuropathology and Applied Neurobiology 35 (4) (2009) 406–416.

[47] W. Sako, N. Murakami, Y. Izumi, R. Kaji, MRI can detect nigral volume loss in patients with Parkinson's disease: evidence from a meta-analysis, Journal of Parkinson's disease 4 (3) (2014) 405–411.

[48] J. H. Lee, J. Ryan, C. Andreescu, H. Aizenstein, H. K. Lim, Brainstem morphological changes in Alzheimers disease, NeuroReport 26 (7) (2015) 411–415.

[49] E. H. Aylward, A. M. Codori, A. Rosenblatt, M. Sherr, J. Brandt, O. C. Stine, P. E. Barta, G. D. Pearlson, C. A. Ross, Rate of caudate atrophy in presymptomatic and symptomatic stages of Huntington's disease, Movement Disorders 15 (3) (2000) 552–560.

[50] A. Peinemann, S. Schuller, C. Pohl, T. Jahn, A. Weindl, J. Kassubek, Executive dysfunction in early stages of Huntington's disease is associated with striatal and insular atrophy: A neuropsychological and voxel-based morphometric study, Journal of the Neurological Sciences 239 (1) (2005) 11–19.

[51] J. W. Mink, Neurobiology of basal ganglia and Tourette syndrome: basal ganglia circuits and thalamocortical outputs, Advances in Neurology 99 (2006) 89.

[52] E. Hollander, E. Anagnostou, W. Chaplin, K. Esposito, M. M. Haznedar, E. Licalzi, S. Wasserman, L. Soorya, M. Buchsbaum, Striatal volume on magnetic resonance imaging and repetitive behaviors in autism, Biological Psychiatry 58 (3) (2005) 226–232.

[53] M. H. Bloch, J. F. Leckman, H. Zhu, B. S. Peterson, Caudate volumes in childhood predict symptom severity in adults with Tourette syndrome, Neurology 65 (8) (2005) 1253–1258.

[54] V. Tremols, A. Bielsa, J.-C. Soliva, C. Raheb, S. Carmona, J. Tomas, J.-D. Gispert, M. Rovira, J. Fauquet, A. Tobea, A. Bulbena, O. Vilarroya, Differential abnormalities of the head and body of the caudate nucleus in attention deficit-hyperactivity disorder, Psychiatry Research: Neuroimaging 163 (3) (2008) 270–278.

[55] S. Eliez, C. M. Blasey, L. S. Freund, T. Hastie, A. L. Reiss, Brain anatomy, gender and IQ in children and adolescents with fragile X syndrome, Brain 124 (8) (2001) 1610–1618.

[56] M. Houtchens, R. Benedict, R. Killiany, J. Sharma, Z. Jaisani, B. Singh, B. Weinstock-Guttman, C. Guttmann, R. Bakshi, Thalamic atrophy and cognition in multiple sclerosis, Neurology 69 (12) (2007) 1213–1223.

[57] A. Minagar, M. H. Barnett, R. H. Benedict, D. Pelletier, I. Pirko, M. A. Sahraian, E. Frohman, R. Zivadinov, The thalamus and multiple sclerosis: Modern views on pathologic, imaging, and clinical aspects, Neurology 80 (4) (2013) 210–219.

[58] N. C. Andreasen, S. Arndt, V. Swayze, T. Cizadlo, M. Flaum, D. O'Leary, J. C. Ehrhardt, W. Yuh, Thalamic abnormalities in schizophrenia visualized through magnetic resonance image averaging, Science 266 (5183) (1994) 294–298.

[59] L. De Jong, K. Van der Hiele, I. Veer, J. Houwing, R. Westendorp, E. Bollen, P. De Bruin, H. Middelkoop, M. Van Buchem, J. Van Der Grond, Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study, Brain 131 (12) (2008) 3277–3285.

[60] W. Byne, E. A. Hazlett, M. S. Buchsbaum, E. Kemether, The thalamus and schizophrenia: current status of research, Acta neuropathologica 117 (4) (2009) 347–368.

[61] S. Lee, S. Kim, W. Tae, S. Lee, J. Choi, S. Koh, D. Kwon, Regional volume analysis of the Parkinson disease brain in early disease stage: gray matter, white matter, striatum, and thalamus, Am. J. Neuroradiol. 32 (4) (2011) 682–687.

[62] A. Z. Kazi, P. C. Joshi, A. B. Kelkar, M. S. Mahajan, A. S. Ghawate, MRI evaluation of pathologies affecting the corpus callosum: A pictorial essay, The Indian journal of radiology & imaging 23 (4) (2013) 321–332.

[63] P. M. Thompson, K. L. Narr, R. E. Blanton, A. W. Toga, Mapping structural alterations of the corpus callosum during brain development and degeneration, The parallel brain: The cognitive neuroscience of the corpus callosum (2003) 93.

[64] N. Garg, S. Reddel, D. Miller, J. Chataway, D. Riminton, Y. Barnett, L. Masters, M. Barnett, T. Hardy, The corpus callosum in the diagnosis of multiple sclerosis and other CNS demyelinating and inflammatory diseases, Journal of Neurology, Neurosurgery & Psychiatry 86 (12) (2015) 1374–1382.

[65] J. E. Downhill, M. S. Buchsbaum, T. Wei, J. Spiegel-Cohen, E. A. Hazlett, M. M. Haznedar, J. Silverman, L. J. Siever, Shape and size of the corpus callosum in schizophrenia and schizotypal personality disorder, Schizophrenia research 42 (3) (2000) 193–208.

[66] I. K. Lyoo, A. Satlin, C. K. Lee, P. F. Renshaw, Regional atrophy of the corpus callosum in subjects with Alzheimer's disease and multi-infarct dementia, Psychiatry Research: Neuroimaging 74 (2) (1997) 63–72.

[67] S. M. Lawrie, H. C. Whalley, D. E. Job, E. C. Johnstone, Structural and functional abnormalities of the amygdala in schizophrenia, Annals of the New York Academy of Sciences 985 (1) (2003) 445–460.

[68] M. P. DelBello, M. E. Zimmerman, N. P. Mills, G. E. Getz, S. M. Strakowski, Magnetic resonance imaging analysis of amygdala and other subcortical brain regions in adolescents with bipolar disorder, Bipolar Disorders 6 (1) (2004) 43–52.

[69] M. P. Milham, A. C. Nugent, W. C. Drevets, D. S. Dickstein, E. Leibenluft, M. Ernst, D. Charney, D. S. Pine, Selective reduction in amygdala volume in pediatric anxiety disorders: a voxel-based morphometry investigation, Biological Psychiatry 57 (9) (2005) 961–966.

[70] L. L. Altshuler, G. Bartzokis, T. Grieder, J. Curran, T. Jimenez, K. Leight, J. Wilkins, R. Gerner, J. Mintz, An MRI study of temporal lobe structures in men with bipolar disorder or schizophrenia, Biological Psychiatry 48 (2) (2000) 147–162.

[71] S. M. Strakowski, M. P. DelBello, K. W. Sax, M. E. Zimmerman, P. K. Shear, J. M. Hawkins, E. R. Larson, Brain magnetic resonance imaging of structural abnormalities in bipolar disorder, Archives of General Psychiatry 56 (3) (1999) 254–260.

[72] A. Sakalauskas, A. Lukoševičius, K. Laučkaitė, Transcranial echoscopy for diagnostic of Parkinson disease: technical constraints and possibilities, Ultragarsas 65 (2010) 47–50.

[73] J. Olveres, R. Nava, B. Escalante-Ramírez, G. Cristóbal, C. M. García-Moreno, Midbrain volume segmentation using active shape models and LBPs, Proc. SPIE 8856 (2013) 88561F–88561F–11.

[74] A. Nayate, J. L. Bradshaw, N. J. Rinehart, Autism and Asperger's disorder: Are they movement disorders involving the cerebellum and/or basal ganglia?, Brain Research Bulletin 67 (4) (2005) 327–334.

[75] L. L. Sears, C. Vest, S. Mohamed, J. Bailey, B. J. Ranson, J. Piven, An MRI study of the basal ganglia in autism, Progress in Neuro-Psychopharmacology and Biological Psychiatry 23 (4) (1999) 613–624.

[76] L. J. Seidman, E. M. Valera, N. Makris, Structural brain imag-

ing of attention-deficit/hyperactivity disorder, Biological Psychiatry 57 (11) (2005) 1263–1272.

[77] Y. Xia, K. Bettinger, L. Shen, A. Reiss, Automatic segmentation of the caudate nucleus from human brain MR images, IEEE Trans. Med. Imag. 26 (4) (2007) 509–517.

[78] M. Iacono, N. Makris, L. Mainardi, J. Gale, A. van der Kouwe, A. Mareyam, J. Polimeni, L. Wald, B. Fischl, E. Eskandar, G. Bonmassar, Atlas-based segmentation for globus pallidus internus targeting on low-resolution MRI, in: Engineering in Medicine and Biology Society, EMBC, Annual International Conference of the IEEE, Institute of Electrical and Electronics Engineers (IEEE), Boston, MA, USA, 2011, pp. 5706–5709.

[79] F. Seixas, A. de Souza, A. dos Santos, D. Saade, Automated segmentation of the corpus callosum midsagittal surface area, in: Computer Graphics and Image Processing, XX Brazilian Symposium on, Belo Horizonte, Brazil, 2007, pp. 287–293.

[80] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, M. Bach Cuadra, A review of atlas-based segmentation for magnetic resonance brain images, Comput. Meth. Prog. Biomed. 104 (3) (2011) e158–e177.

[81] A. Valsecchi, S. Damas, J. Santamaría, L. Marrakchi-Kacem, Intensity-based image registration using scatter search, Artificial Intelligence in Medicine 60 (3) (2014) 151–163.

[82] F. Wang, B. Vemuri, Simultaneous registration and segmentation of anatomical structures from brain MRI, in: J. Duncan, G. Gerig (Eds.), Int. Conf. Med. Image Comput. Comput. Assist. Interv., Vol. 3749, Springer, Palm Springs, CA, USA, 2005, pp. 17–25.

[83] F. Wang, B. C. Vemuri, M. Rao, Y. Chen, A new & robust information theoretic measure and its application to image alignment, Information Processing in Medical Imaging (2003) 388–400.

[84] W. R. Crum, T. Hartkens, D. L. G. Hill, Non-rigid image registration: theory and practice, Brit. J. Radiol. 77 (suppl_2) (2004) 140–153.

[85] D. Shen, C. Davatzikos, HAMMER: hierarchical attribute matching mechanism for elastic registration, IEEE Trans. Med. Imag. 21 (11) (2002) 1421–1439.

[86] P. Hellier, C. Barillot, I. Corouge, B. Gibaud, G. Le Goualher, D. Collins, A. Evans, G. Malandain, N. Ayache, G. Christensen, H. Johnson, Retrospective evaluation of intersubject brain registration, IEEE Trans. Med. Imag. 22 (9) (2003) 1120–1130.

[87] O. T. Carmichael, H. A. Aizenstein, S. W. Davis, J. T. Becker, P. M. Thompson, C. C. Meltzer, Y. Liu, Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment, NeuroImage 27 (4) (2005) 979–990.

[88] D. W. Shattuck, R. M. Leahy, BrainSuite: An automated cortical surface identification tool, Med. Image Anal. 6 (2) (2002) 129–142.

[89] A. Joshi, D. Shattuck, R. Leahy, A method for automated cortical surface registration and labeling, Biomedical Image Registration 7359 (2012) 180–189.

[90] R. Woods, S. Grafton, C. Holmes, S. Cherry, J. Mazziotta, Automated image registration: I. general methods and intrasubject, intramodality validation, J. Comp. Assist. Tomo. 22 (1) (1998) 139.

[91] G. Postelnicu, L. Zollei, B. Fischl, Combined volumetric and surface registration, IEEE Trans. Med. Imag. 28 (4) (2009) 508–522.

[92] Y. Luo, A. Chung, An atlas-based deep brain structure segmentation method: from coarse positioning to fine shaping, in: Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on, Institute of Electrical and Electronics Engineers (IEEE), Prague, Czech Republic, 2011, pp. 1085–1088.

[93] S. Yousefi, N. Kehtarnavaz, A. Gholipour, Improved labeling of subcortical brain structures in atlas-based segmentation of magnetic resonance images, IEEE Trans. Biomed. Eng. 59 (7) (2012) 1808–1817.

[94] P. Aljabar, R. Heckemann, A. Hammers, J. Hajnal, D. Rueckert, Classifier selection strategies for label fusion using large atlas databases, in: N. Ayache, S. Ourselin, A. Maeder (Eds.), Int. Conf. Med. Image Comput. Comput. Assist. Interv., Vol. 4791, Springer, Brisbane, Australia, 2007, pp. 523–531.

[95] J. M. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, D. Rueckert, Fast and robust multi-atlas segmentation of brain magnetic resonance images, NeuroImage 49 (3) (2010) 2352–2365.

[96] J. Wang, C. Vachet, A. Rumple, S. Gouttard, C. Ouziel, E. Perrot, G. Du, X. Huang, G. Gerig, M. A. Styner, Multi-atlas segmentation of subcortical brain structures via the AutoSeg software pipeline, Frontiers in Neuroinformatics 8 (7).

[97] D. L. Collins, J. C. Pruessner, Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion, NeuroImage 52 (4) (2010) 1355–1366.

[98] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, A. Hammers, Automatic anatomical brain MRI segmentation combining label propagation and decision fusion, NeuroImage 33 (1) (2006) 115–126.

[99] P. Aljabar, R. Heckemann, A. Hammers, J. Hajnal, D. Rueckert, Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy, NeuroImage 46 (3) (2009) 726–738.

[100] J. Pipitone, M. T. M. Park, J. Winterburn, T. A. Lett, J. P. Lerch, J. C. Pruessner, M. Lepage, A. N. Voineskos, M. M. Chakravarty, Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates, NeuroImage 101 (2014) 494–512.

[101] A. Klein, B. Mensh, S. Ghosh, J. Tourville, J. Hirsch, Mindboggle: automated brain labeling with multiple atlases, BMC Med. Imag. 5 (1) (2005) 7.

[102] H. Wang, P. Yushkevich, Multi-atlas segmentation with joint label fusion and corrective learning - An open source implementation, Frontiers in Neuroinformatics 7 (27).

[103] H. Jia, P.-T. Yap, D. Shen, Iterative multi-atlas-based multi-image segmentation with tree-based registration, NeuroImage 59 (1) (2012) 422–430.

[104] M. J. Cardoso, K. Leung, M. Modat, S. Keihaninejad, D. Cash, J. Barnes, N. C. Fox, S. Ourselin, STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation, Med. Image Anal. 17 (6) (2013) 671–684.

[105] A. J. Asman, B. A. Landman, Non-local statistical label fusion for multi-atlas segmentation, Med. Image Anal. 17 (2) (2013) 194–208.

[106] N. Robitaille, S. Duchesne, Label fusion strategy selection, International Journal of Biomedical Imaging 2012 (2011) 431095–431095.

[107] N. I. Weisenfeld, S. K. Warfield, Learning likelihoods for labeling (L3): A general multi-classifier segmentation algorithm, in: G. Fichtinger, A. Martel, T. Peters (Eds.), Int. Conf. Med. Image Comput. Comput. Assist. Interv., Springer, Toronto, Canada, 2011, pp. 322–329.

[108] X. Artaechevarría, A. Muoz-Barrutia, C. Ortiz-de Solorzano, Combination strategies in multi-atlas image segmentation: Application to brain MR data, IEEE Trans. Med. Imag. 28 (8) (2009) 1266–1277.

[109] F. Rousseau, P. Habas, C. Studholme, A supervised patch-

based approach for human brain labeling, IEEE Trans. Med. Imag. 30 (10) (2011) 1852–1862.

[110] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, D. L. Collins, Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation, NeuroImage 54 (2) (2011) 940–954.

[111] G. Wu, M. Kim, G. Sanroma, Q. Wang, B. C. Munsell, D. Shen, Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition, NeuroImage 106 (2015) 34–46.

[112] D. Zhang, Q. Guo, G. Wu, D. Shen, Sparse patch-based label fusion for multi-atlas segmentation, Multimodal Brain Image Analysis 7509 (2012) 94–102.

[113] R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, D. Rueckert, LEAP: Learning embeddings for atlas propagation, NeuroImage 49 (2) (2010) 1316–1325.

[114] S. Duchesne, J. Pruessner, D. Collins, Appearance-based segmentation of medial temporal lobe structures, NeuroImage 17 (2) (2002) 515–531.

[115] M. Sabuncu, B. Yeo, K. Van Leemput, B. Fischl, P. Golland, A generative model for image segmentation based on label fusion, IEEE Trans. Med. Imag. 29 (10) (2010) 1714–1729.

[116] R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal, A. Hammers, Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation, NeuroImage 51 (1) (2010) 221–227.

[117] T. Tong, R. Wolz, P. Coupé, J. V. Hajnal, D. Rueckert, Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling, NeuroImage 76 (2013) 11–23.

[118] J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, M. West, Generative or discriminative? Getting the best of both worlds, Bayesian Statistics (8) (2007) 3–23.

[119] A. Pitiot, A. Toga, N. Ayache, P. Thompson, Texture based MRI segmentation with a two-stage hybrid neural classifier, in: Neural Networks, Proceedings of the International Joint Conference on, Vol. 3, Honolulu, Hawaii, 2002, pp. 2053–2058.

[120] M. Jabarouti Moghaddam, H. Soltanian-Zadeh, Automatic segmentation of brain structures using geometric moment invariants and artificial neural networks, Information Processing in Medical Imaging 5636 (2009) 326–337.

[121] J. Morra, Z. Tu, L. Apostolova, A. Green, A. Toga, P. Thompson, Automatic subcortical segmentation using a contextual model, in: D. Metaxas, L. Axel, G. Fichtinger, G. Szekely (Eds.), Int. Conf. Med. Image Comput. Comput. Assist. Interv., Vol. 5241, Springer, New York, NY, USA, 2008, pp. 194–201.

[122] Z. Tu, X. Bai, Auto-context and its application to high-level vision tasks and 3D brain image segmentation, IEEE Trans. Pattern Anal. Machine Intell. 32 (10) (2010) 1744–1757.

[123] M. Kim, G. Wu, D. Shen, Unsupervised deep learning for hippocampus segmentation in 7.0 Tesla MR images, Machine Learning in Medical Imaging 8184 (2013) 1–8.

[124] M. Wels, Y. Zheng, G. Carneiro, M. Huber, J. Hornegger, D. Comaniciu, Fast and robust 3-D MRI brain structure segmentation, in: G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, C. Taylor (Eds.), Int. Conf. Med. Image Comput. Comput. Assist. Interv., Vol. 5762, Springer, London, UK, 2009, pp. 575–583.

[125] L. Igual, J. C. Soliva, A. Hernández-Vela, S. Escalera, X. Jiménez, O. Vilarroya, P. Radeva, A fully-automatic caudate nucleus segmentation of brain MRI: application in volumetric analysis of pediatric attention-deficit/hyperactivity disorder, Biomedical Engineering Online 10 (1) (2011) 105.

[126] O. Benkarim, P. Radeva, L. Igual, Label consistent multiclass discriminative dictionary learning for MRI segmentation, Articulated Motion and Deformable Objects 8563 (2014) 138–147.

[127] S. C. Deoni, B. K. Rutt, A. G. Parrent, T. M. Peters, Segmentation of thalamic nuclei using a modified k-means clustering algorithm and high-resolution quantitative magnetic resonance imaging at 1.5 T, NeuroImage 34 (1) (2007) 117–126.

[128] C. R. Traynor, G. J. Barker, W. R. Crum, S. C. Williams, M. P. Richardson, Segmentation of the thalamus in MRI based on T1 and T2, NeuroImage 56 (3) (2011) 939–950.

[129] K. Van Leemput, A. Bakkour, T. Benner, G. Wiggins, L. L. Wald, J. Augustinack, B. C. Dickerson, P. Golland, B. Fischl, Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI, Hippocampus 19 (6) (2009) 549–557.

[130] J. E. Iglesias, M. R. Sabuncu, K. V. Leemput, Improved inference in Bayesian segmentation using Monte Carlo sampling: Application to hippocampal subfield volumetry, Med. Image Anal. 17 (7) (2013) 766–778.

[131] A. Makropoulos, I. Gousias, C. Ledig, P. Aljabar, A. Serag, J. Hajnal, A. Edwards, S. Counsell, D. Rueckert, Automatic whole brain MRI segmentation of the developing neonatal brain, IEEE Trans. Med. Imag. 33 (9) (2014) 1818–1831.

[132] T. Riklin-Raviv, K. V. Leemput, B. H. Menze, W. M. W. III, P. Golland, Segmentation of image ensembles via latent atlases, Med. Image Anal. 14 (5) (2010) 654–665.

[133] B. Fischl, FreeSurfer, NeuroImage 62 (2) (2012) 774–781.

[134] B. Scherrer, M. Dojat, F. Forbes, C. Garbay, LOCUS: LOcal Cooperative Unified Segmentation of MRI brain scans, in: N. Ayache, S. Ourselin, A. Maeder (Eds.), Int. Conf. Med. Image Comput. Comput. Assist. Interv., Vol. 4791, Springer, Brisbane, Australia, 2007, pp. 219–227.

[135] Q. Razlighi, A. Orekhov, A. Laine, Y. Stern, Causal Markov random field for brain MR image segmentation, in: Engineering in Medicine and Biology Society, EMBC, Annual International Conference of the IEEE, Institute of Electrical and Electronics Engineers (IEEE), San Diego, CA, USA, 2012, pp. 3203–3206.

[136] A. Akselrod-Ballin, M. Galun, J. M. Gomori, A. Brandt, R. Basri, Prior knowledge driven multiscale segmentation of brain MRI, in: N. Ayache, S. Ourselin, A. Maeder (Eds.), Int. Conf. Med. Image Comput. Comput. Assist. Interv., Springer, Brisbane, Australia, 2007, pp. 118–126.

[137] M. Sabuncu, B. Yeo, K. Van Leemput, B. Fischl, P. Golland, Supervised nonparametric image parcellation, in: G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, C. Taylor (Eds.), Int. Conf. Med. Image Comput. Comput. Assist. Interv., Vol. 5762, Springer, London, UK, 2009, pp. 1075–1083.

[138] C.-Y. Liu, J. Iglesias, Z. Tu, Deformable templates guided discriminative models for robust 3D brain MRI segmentation, Neuroinformatics 11 (4) (2013) 447–468.

[139] B. Scherrer, M. Dojat, F. Forbes, C. Garbay, MRF agent based segmentation: Application to MRI brain scans, Artificial Intelligence in Medicine 4594 (2007) 13–23.

[140] B. Scherrer, F. Forbes, C. Garbay, M. Dojat, Distributed local MRF models for tissue and structure brain segmentation, IEEE Trans. Med. Imag. 28 (8) (2009) 1278–1295.

[141] J. E. Iglesias, K. V. Leemput, P. Bhatt, C. Casillas, S. Dutt, N. Schuff, D. Truran-Sacrey, A. Boxer, B. Fischl, Bayesian segmentation of brainstem structures in MRI, NeuroImage 113 (2015) 184–195.

[142] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, International Journal of Computer Vision 1 (4) (1988) 321–331.

[143] A. Ghanei, H. Soltanian-Zadeh, J. P. Windham, Segmentation

of the hippocampus from brain MRI using deformable contours, Computerized Medical Imaging and Graphics 22 (3) (1998) 203–216.

[144] O. Colliot, O. Camara, I. Bloch, Integration of fuzzy spatial relations in deformable modelsApplication to brain MRI segmentation, Pattern Recog. 39 (8) (2006) 1401–1414.

[145] G. Fouquier, J. Atif, I. Bloch, Sequential model-based segmentation and recognition of image structures driven by visual features and spatial relations, Comput. Vis. Image Underst. 116 (1) (2012) 146–165.

[146] D. Zarpalas, A. Zafeiropoulos, P. Daras, N. Maglaveras, M. G. Strintzis, Brain structures segmentation using optimum global and local weights on mixing active contours and neighboring constraints, in: Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, no. 127, ACM, New York, NY, USA, 2011, pp. 1–5.

[147] H. Shariatpanahi, N. Batmanghelich, A. Kermani, M. Ahmadabadi, H. Soltanian-Zadeh, Distributed behavior-based multiagent system for automatic segmentation of brain MR images, in: Neural Networks, Proceedings of the International Joint Conference on, Vancouver, BC, Canada, 2006, pp. 4535–4542.

[148] T. Cootes, C. Taylor, D. Cooper, J. Graham, Active shape models-Their training and application, Comput. Vis. Image Underst. 61 (1) (1995) 38–59.

[149] A. Kelemen, G. Szekely, G. Gerig, Elastic model-based segmentation of 3-D neuroradiological data sets, IEEE Trans. Med. Imag. 18 (10) (1999) 828–839.

[150] T. Cootes, C. Taylor, Active shape models Smart snakes, in: D. Hogg, R. Boyle (Eds.), Proc. British Machine Vision Conference, BMVA Press, Leeds, UK, 1992, pp. 266–275.

[151] S. Ettaieb, K. Hamrouni, S. Ruan, Statistical models of shape and spatial relation-application to hippocampus segmentation, in: S. Battiato (Ed.), Computer Vision Theory and Applications, International Conference on, Vol. 1, Lisbon, Portugal, 2014, pp. 448–455.

[152] J. Cerrolaza, A. Villanueva, R. Cabeza, Hierarchical statistical shape models of multiobject anatomical structures: Application to brain MRI, IEEE Trans. Med. Imag. 31 (3) (2012) 713–724.

[153] Y. Gao, B. Corn, D. Schifter, A. Tannenbaum, Multiscale 3D shape representation and segmentation with applications to hippocampal/caudate extraction from brain MRI, Med. Image Anal. 16 (2) (2012) 374–385.

[154] T. Cootes, G. Edwards, C. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Machine Intell. 23 (6) (2001) 681–685.

[155] K. O. Babalola, V. Petrovic, T. F. Cootes, C. J. Taylor, C. J. Twining, A. Mills, Automatic segmentation of the caudate nuclei using active appearance models, in: T. Heimann, M. Styner, B. van Ginneken (Eds.), 3D Segmentation in the clinic: A grand challenge, Brisbane, Australia, 2007, pp. 57–64.

[156] S. D. S. Al-Shaikhli, M. Y. Yang, B. Rosenhahn, Multi-region labeling and segmentation using a graph topology prior and atlas information in brain images, Computerized Medical Imaging and Graphics 38 (8) (2014) 725–734.

[157] E. A. Ashton, J. K. Riek, L. Molinelli, M. J. Berg, K. J. Parker, A method for fully automated measurement of neurological structures in MRI, Proc. SPIE 5032 (2003) 1125–1134.

[158] H. García, M. Álvarez, Á. Orozco, Bayesian shape models with shape priors for MRI brain segmentation, Advances in Visual Computing 8888 (2014) 851–860.

[159] J.-H. Xue, S. Ruan, B. Moretti, M. Revenu, D. Bloyet, Knowledge-based segmentation and labeling of brain structures from MRI images, Pattern Recognit. Lett. 22 (34) (2001) 395–405.

[160] L. Gui, R. Lisowski, T. Faundez, P. S. Hüppi, F. Lazeyras, M. Kocher, Morphology-driven automatic segmentation of MR images of the neonatal brain, Med. Image Anal. 16 (8) (2012) 1565–1579.

[161] J. Iglesias, M. Sabuncu, K. Van Leemput, A generative model for probabilistic label fusion of multimodal data, Multimodal Brain Image Analysis 7509 (2012) 115–133.

[162] J. Iglesias, M. Sabuncu, K. Van Leemput, A probabilistic, non-parametric framework for inter-modality label fusion, in: K. Mori, I. Sakuma, Y. Sato, C. Barillot, N. Navab (Eds.), Int. Conf. Med. Image Comput. Comput. Assist. Interv., Vol. 8151, Springer, Nagoya, Japan, 2013, pp. 576–583.

[163] J.-H. Xue, S. Ruan, B. Moretti, M. Revenu, D. Bloyet, W. Philips, Fuzzy modeling of knowledge for MRI brain structure segmentation, in: B. Mercer, C. M. Services (Eds.), Image Processing, International Conference on, Vol. 1, Vancouver, Canada, 2000, pp. 617–620.

[164] Z. Tu, K. Narr, P. Dollar, I. Dinov, P. Thompson, A. Toga, Brain anatomical structure segmentation by hybrid discriminative/generative models, IEEE Trans. Med. Imag. 27 (4) (2008) 495–508.

[165] K. Karsch, Q. He, Y. Duan, A fast, semi-automatic brain structure segmentation algorithm for magnetic resonance imaging, in: Bioinformatics and Biomedicine, IEEE International Conference on, Institute of Electrical and Electronics Engineers (IEEE), Washington, DC, USA, 2009, pp. 297–302.

[166] Q. He, K. Karsch, Y. Duan, Semi-automatic 3D segmentation of brain structures from MRI, Int. J. Data Mining and Bioinformatics 5 (2).

[167] F. van der Lijn, M. de Bruijne, S. Klein, T. den Heijer, Y. Hoogendam, A. van der Lugt, M. Breteler, W. Niessen, Automated brain structure segmentation based on atlas registration and appearance models, IEEE Trans. Med. Imag. 31 (2) (2012) 276–286.

[168] J. Ashburner, K. Friston, Multimodal image coregistration and partitioningA unified framework, NeuroImage 6 (3) (1997) 209–217.

[169] S. Valverde, A. Oliver, E. Roura, D. Pareto, J. C. Vilanova, L. Ramió-Torrentà, J. Sastre-Garriga, X. Montalban, À. Rovira, X. Lladó, Quantifying brain tissue volume in multiple sclerosis with automated lesion segmentation and filling, NeuroImage: Clinical 9 (2015) 640–647.

[170] M. Battaglini, M. Jenkinson, N. De Stefano, Evaluating and reducing the impact of white matter lesions on brain volume measurements, Hum. Brain Mapp. 33 (9) (2012) 2062–2071.

[171] D. T. Chard, J. S. Jackson, D. H. Miller, C. A. Wheeler-Kingshott, Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes, J. Magn. Reson. Imaging 32 (1) (2010) 223–228.

[172] S. Valverde, A. Oliver, X. Lladó, A white matter lesion-filling approach to improve brain tissue volume measurements, NeuroImage: Clinical 6 (2014) 86–92.

[173] K. Nakamura, N. Guizard, V. S. Fonov, S. Narayanan, D. L. Collins, D. L. Arnold, Jacobian integration method increases the statistical power to measure gray matter atrophy in multiple sclerosis, NeuroImage: Clinical 4 (2014) 10–17.

[174] V. Popescu, N. Ran, F. Barkhof, D. Chard, C. Wheeler-Kingshott, H. Vrenken, Accurate GM atrophy quantification in MS using lesion-filling with co-registered 2D lesion masks, NeuroImage: Clinical 4 (2014) 366–373.

[175] K. A. Johnson, J. A. Becker, L. Williams, The whole brain atlas (1999).

[176] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thomp-

son, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, M. W. Weiner, The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods, J. Magn. Reson. Imaging 27 (4) (2008) 685–691.

[177] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, R. L. Buckner, Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults, Journal of Cognitive Neuroscience 19 (9) (2007) 1498–1507.

[178] D. W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K. L. Narr, R. A. Poldrack, R. M. Bilder, A. W. Toga, Construction of a 3D probabilistic atlas of human cortical structures, NeuroImage 39 (3) (2008) 1064–1080.

[179] A. Hammers, R. Allom, M. J. Koepp, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, J. S. Duncan, Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe, Hum. Brain Mapp. 19 (4) (2003) 224–247.

[180] I. S. Gousias, D. Rueckert, R. A. Heckemann, L. E. Dyet, J. P. Boardman, A. D. Edwards, A. Hammers, Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest, NeuroImage 40 (2) (2008) 672–684.

[181] A. Hammers, C.-H. Chen, L. Lemieux, R. Allom, S. Vossos, S. L. Free, R. Myers, D. J. Brooks, J. S. Duncan, M. J. Koepp, Statistical neuroanatomy of the human inferior frontal gyrus and probabilistic atlas in a standard stereotaxic space, Hum. Brain Mapp. 28 (1) (2007) 34–48.

[182] B. Van Ginneken, T. Heimann, M. Styner, 3D segmentation in the clinic: A grand challenge, in: T. Heimann, M. Styner, B. van Ginneken (Eds.), 3D Segmentation in the clinic: A grand challenge, Brisbane, Australia, 2007, pp. 7–15.

[183] B. Landman, S. Warfield, MICCAI 2012 workshop on multi-atlas labeling, in: B. A. Landman, S. K. Warfield (Eds.), MICCAI Grand Challenge and Workshop on Multi-Atlas Labeling, CreateSpace Independent Publishing Platform, Nice, France, 2012.

[184] A. Asman, A. Akhondi-Asl, H. Wang, N. Tustison, B. Avants, S. Warfield, B. Landman, MICCAI 2013 segmentation algorithms, theory and applications (SATA) challenge results summary, in: A. Asman, A. Akhondi-Asl, H. Wang, N. Tustison, B. Avants, S. Warfield, B. Landman (Eds.), MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA), Nagoya, Japan, 2013.

[185] E. M. van Rikxoort, I. Isgum, M. Staring, S. Klein, B. van Ginneken, Adaptive local multi-atlas segmentation: Application to heart segmentation in chest CT scans, Medical Imaging (2008) 691407–691407.

[186] A. R. Khan, L. Wang, M. F. Beg, FreeSurfer-initiated fully-automated subcortical brain segmentation in MRI using large deformation diffeomorphic metric mapping, NeuroImage 41 (3) (2008) 735–746.

[187] C. Ledig, R. A. Heckemann, P. Aljabar, R. Wolz, J. V. Hajnal, A. Hammers, D. Rueckert, Segmentation of MRI brain scans using MALP-EM, in: B. A. Landman, S. K. Warfield (Eds.), MICCAI Grand Challenge and Workshop on Multi-Atlas Labeling, CreateSpace Independent Publishing Platform, Nice, France, 2012, pp. 79–82.

[188] J. Doshi, G. Erus, Y. Ou, C. Davatzikos, Ensemble-based medical image labeling via sampling morphological appearance manifolds, in: A. Asman, A. Akhondi-Asl, H. Wang, N. Tustison, B. Avants, S. Warfield, B. Landman (Eds.), MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA), Nagoya, Japan, 2013.