Bayesian segmentation of brainstem structures in MRI

Juan Eugenio Iglesias[1], **Koen Van Leemput**, **Priyanka Bhatt**, **Christen Casillas**, **Shubir Dutt**, **Norbert Schuff**, **Diana Truran-Sacrey**, **Adam Boxer**, **Bruce Fischl**, and **for the Alzheimer's Disease Neuroimaging Initiative**[2]

Basque Center on Cognition, Brain and Language, Spain

Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Memory and Aging Center, University of California, San Francisco, CA, USA

Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), Denmark

Departments of Information and Computer Science and of Biomedical Engineering and Computational Science, Aalto University, Finland

Center for Imaging of Neurodegenerative Dieases (CIND), Department of Radiology, University of California, San Francisco, CA, USA

Abstract

In this paper we present a method to segment four brainstem structures (midbrain, pons, medulla oblongata and superior cerebellar peduncle) from 3D brain MRI scans. The segmentation method relies on a probabilistic atlas of the brainstem and its neighboring brain structures. To build the atlas, we combined a dataset of 39 scans with already existing manual delineations of the whole brainstem and a dataset of 10 scans in which the brainstem structures were manually labeled with a protocol that was specifically designed for this study. The resulting atlas can be used in a Bayesian framework to segment the brainstem structures in novel scans. Thanks to the generative nature of the scheme, the segmentation method is robust to changes in MRI contrast or acquisition hardware. Using cross validation, we show that the algorithm can segment the structures in previously unseen T1 and FLAIR scans with great accuracy (mean error under 1 mm) and

robustness (no failures in 383 scans including 168 AD cases). We also indirectly evaluate the algorithm with a experiment in which we study the atrophy of the brainstem in aging. The results show that, when used simultaneously, the volumes of the midbrain, pons and medulla are significantly more predictive of age than the volume of the entire brainstem, estimated as their sum. The results also demonstrate that that the method can detect atrophy patterns in the brainstem structures that have been previously described in the literature. Finally, we demonstrate that the proposed algorithm is able to detect differential effects of AD on the brainstem structures. The method will be implemented as part of the popular neuroimaging package FreeSurfer.

Graphical abstract
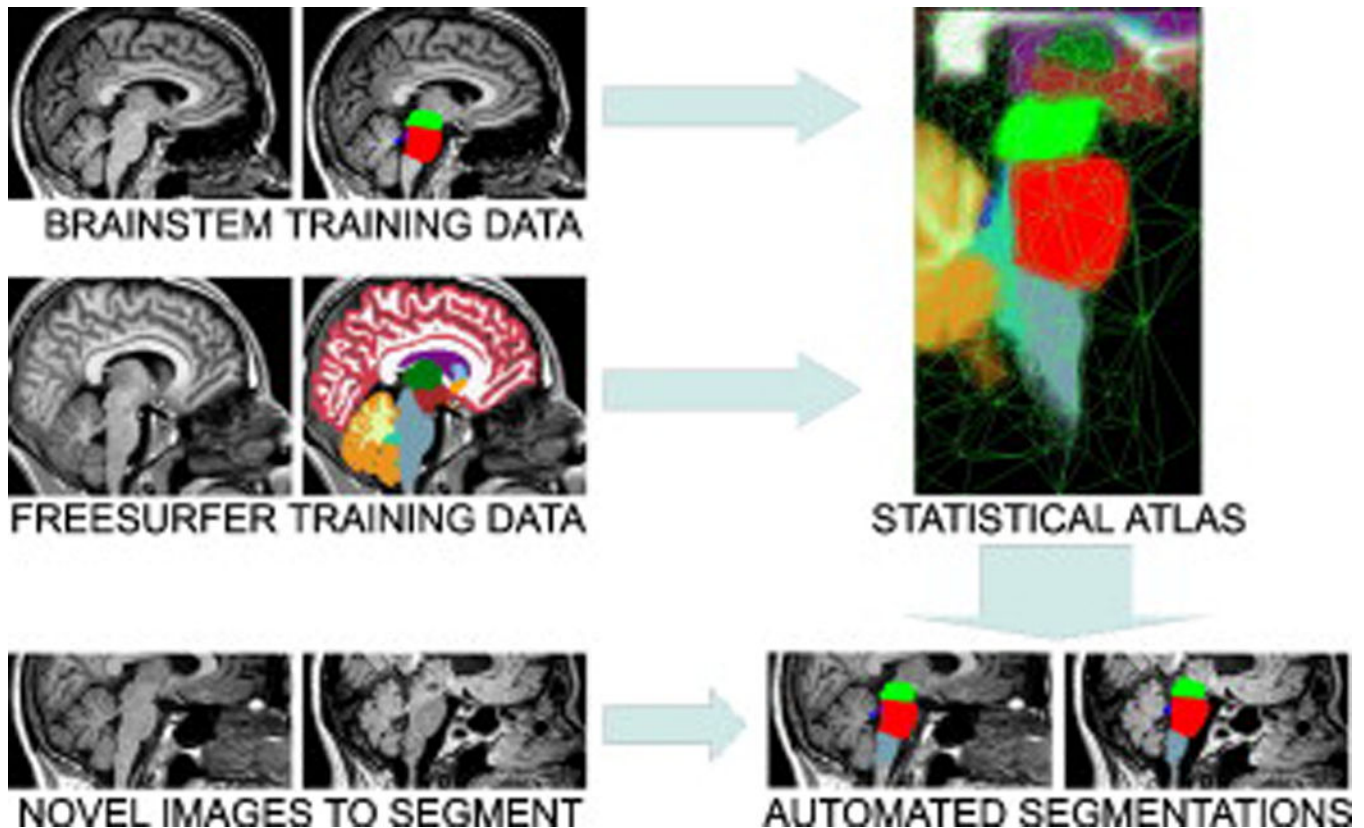
1 Introduction

The human brainstem is a complex brain structure consisting of long axons and scattered nuclei. At a high level, the brainstem is divided in three structures; from superior to inferior: midbrain, pons and medulla oblongata. These structures support different functions: while the midbrain is associated with vision, hearing, sleep and motor control, the pons mostly consists of white matter tracts that connect the cerebrum with the medulla. The pons is also connected with the cerebellum through nerve tracts knows as the cerebellar peduncles, and contains nuclei associated with functions such as respiration and facial expression. The medulla oblongata connects the rest of the brain to the spinal chord, and regulates cardiac and respiratory functions, as well as reflexes such as swallowing.

Automated segmentation of the brainstem structures can potentially improve our understanding of the role that they play in different functions and how they are affected by neurodegenerative pathologies, by circumscribing neuroimaging analyses (e.g., volumetry, functional MRI, tractography) to these specific regions. The brainstem is especially relevant to diseases with pure underlying tau pathology such as progressive supranuclear palsy and corticobasal degeneration, also called primary tauopathies. In progressive supranuclear palsy, brain atrophy occurs in the midbrain, pons and superior cerebellar peduncle, due to neuronal loss associated with accumulation of insoluble deposits of abnormal tau protein [1]. New therapies designed to prevent or decrease tau accumulation are rapidly entering human clinical trials, and longitudinal brainstem atrophy measurements with MRI – in which automated methods yield reproducible results and allow for much larger sample sizes – have been demonstrated to be useful outcome measures in these studies [2]. Other neurodegenerative diseases in which the brainstem structures are also differentially affected include Parkinson's [3] and Alzheimer's [4].

In addition to studies of neurodegenerative diseases, automated segmentation algorithms for the brainstem structures would also find application in other areas. For instance, the pedunculopontine nucleus is a target for the implantation of deep brain stimulators in Parkinson's disease [5]. The pons is often used as a reference region in positron emission tomography (PET) data, since there is no effect of interest in it [6]. Neuroimaging studies of pain [7], [8] have also relied on segmenting brainstem structures.

Despite all its possible applications, the segmentation of the brainstem structures remains largely unexplored in the medical image analysis literature, and none of the widely-distributed neuroimaging analysis package performs it so far. Instead, most works have aimed at segmenting the brainstem as a whole. *Bondiau et al.* [9] used a single labeled template that was deformed towards the novel scan to produce the automated segmentation. *Lee et al.* [10] proposed a semi-automatic algorithm in which fuzzy connectedness and morphological operations are used to generate a preliminary segmentation, which is subsequently refined with active contours. The same authors [11] later proposed a similar, though fully automated method in which AdaBoost [12] was used to generate the initial coarse region containing the brainstem.

There are also brain parcellation methods that include the whole brainstem. The popular package FreeSurfer [13], [14] has it as a label in the atlas that it uses to segment T1 MRI data. The segmentation algorithm [15] in FSL [16] also includes the brainstem in its parcellation, which is based on active shape and appearance models. Multi-atlas methods that segment a large number of structures have also included the whole brainstem; see for instance [17], which uses majority voting to fuse the deformed segmentations propagated from 30 manually labeled scans.

To our best knowledge, only two works have addressed the issue of parcellating the brainstem in MRI data. *Nigro et al.* [18] proposed a method to automatically segment the pons and the midbrain using thresholds and geometric criteria defined upon heuristic rules, which makes their method sensitive to variations in MRI acquisition protocol or scanning platform. *Lambert et al.* [19] used multimodal MRI data to produce probability maps for

four tissue classes using an unsupervised segmentation algorithm. While these maps can be used to segment novel scans, they do not necessarily correspond to the underlying brainstem structures, due to the lack of expert manual delineations.

In this paper, we present a supervised segmentation method for the midbrain, pons, medulla and superior cerebellar peduncle (SCP). The method is based on a probabilistic atlas and Bayesian inference. To build the atlas, we used the training data that was used to build the atlas in FreeSurfer (which has labels for the whole brainstem) and enhance it with an additional dataset of 10 scans in which the four brainstem structures were manually labeled with a delineation protocol that was specifically designed for this study. Using Bayesian inference, the probabilistic atlas can be used to efficiently segment a novel scan, and due to the generative nature of the framework, the segmentation is robust to changes in MRI scanning platform and/or MRI pulse sequence. An implementation of the segmentation algorithm will be made publicly available as part of FreeSurfer.

The rest of this paper is organized as follows. In Section 2, we describe the MRI data used in this study and the manual delineation protocol for the brainstem structures of interest; and we briefly revise the methods to build the atlas with heterogeneously labeled data (i.e., the FreeSurfer dataset and our newly labeled dataset) and to segment a novel scan with a probabilistic atlas and Bayesian inference. In Section 3, we evaluate the performance of the segmentation algorithm with experiments on three different datasets. Finally, Section 4 concludes the paper.

2 Materials and Methods

2.1 MRI data

Three datasets of MRI scans were used in this study. The first dataset, which we will refer to as the **"brainstem dataset"**, consists of T1-weighted and FLAIR brain scans of 10 clinically normal subjects (age range 58–77, mean age 67.8 years, four males, six females). The data were acquired with a 3 Tesla Siemens TIM Trio scanner at the UCSF Neuroscience Imaging Center. The T1 scans were acquired with a MP-RAGE sequence with the following parameters: TR = 2300 ms, TE = 2.98 ms, TI = 900 ms, flip angle = 9°, 1 mm isotropic resolution. The FLAIR sequence used the following parameters: TR = 6000 ms, TE = 388 ms, TI = 2100 ms, 1 mm isotropic resolution. The midbrain, pons and SCP were independently delineated by PB and CC on the 10 scans using the protocol detailed in Section 2.2 below. This dataset will be used with two purposes: first, to build the probabilistic atlas of the brainstem (in combination with the FreeSurfer dataset, described below) ; and second, to directly evaluate the segmentation method, by comparing the labels automatically derived from the T1 and FLAIR scans with one another (to evaluate robustness against changes in MRI sequence) and with the gold standard (to evaluate accuracy) using metrics such as Dice overlap and Hausdorff distance. In addition, the independent annotations from two different labelers allow us to compute a more reliable gold standard for the segmentation than using a single delineation, and also allow us to estimate the inter-observer variability of the manual tracings.

The second dataset, which we will refer to as the **"FreeSurfer dataset"**, consists of T1-weighted brain MRI scans from 39 subjects (age range 18–87, mean age 56.3 years). These scans were acquired on a Siemens 1.5T platform with a MP-RAGE sequence with the following parameters: TR = 9.7 ms, TE = 4 ms, flip angle = 10°, TI = 20 ms, in-plane resolution 1 mm (sagittal), slice thickness 1.25 mm. These scans were resampled to 1 mm isotropic resolution with trilinear interpolation. Thirty-six brain structures, including the whole brainstem, were labeled by an expert neuroanatomist using the delineation protocol in [20]. We note that these are the subjects that were used to train the probabilistic atlas in FreeSurfer[13]. This dataset was used for two purposes: building the atlas (in conjunction with the brainstem dataset) and indirectly evaluating the segmentation algorithm with an aging experiment.

The third dataset, which we will refer to as the **"ADNI dataset"**, consists of 383 baseline T1 scans from elderly controls (*n* = 215) and Alzheimer's disease (AD) subjects (*n* = 168) from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

The list of subjects, along with the corresponding demographics, can be found in the supplementary material (Tables 2–7). The mean age of the subjects was 75.8 years (range: 56–91 years). The images were acquired with MP-RAGE sequences at 1 mm isotropic resolution. Since ADNI is a multi-site effort, different scanning platforms were used for acquiring the images; for further detail on the acquisition parameters and up-to-date information, we refer the reader to the website http://www.adni-info.org.

The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The main goal of ADNI is to test whether MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to analyze the progression of MCI and early AD. Markers of early AD progression can aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as decrease the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California - San Francisco. ADNI is a joint effort by co-investigators from industry and academia. Subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. These three protocols have recruited over 1,500 adults (ages 55–90) to participate in the study, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the corresponding protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2.

2.2 Delineation protocol for brainstem dataset

Rather than delineating the brainstem structures in the native space of the scans directly, these scans were first rigidly registered to the FreeSurfer reference space ("fsaverage"). The manual annotations were made on the registered scans, which helps reduce the variability in the annotations, and then warped back to the original space using the inverse transform and

nearest neighbor interpolation. The order in which the brainstem structures were delineated was: pons, midbrain and SCP; the corresponding delineation protocols are detailed in 5, 6 and 7, respectively. All the annotations were made on the T1 scans; the FLAIR images were not used in the delineation process. The labeling protocol is illustrated in Figure 1, which displays slices of a sample scan of the brainstem dataset with its corresponding annotations.

Note than the delineation protocol does not include the medulla. Instead, this structure is implicitly defined through the combination of the labeling protocols of the brainstem and FreeSurfer datasets. Specifically, the medulla is defined as the portion of the whole brainstem (as defined in the FreeSurfer dataset) that is not labeled as midbrain, pons or SCP in the brainstem dataset.

2.3 Atlas construction

The manually labeled training data (i.e., the brainstem and FreeSurfer datasets) are used to build a probabilistic atlas of the brainstem and its surrounding structures. This atlas, which encodes the frequency with which the labels occur at each spatial location, will be used as a prior distribution in a Bayesian framework to produce automated segmentation of novel scans in Section 2.4 below. The prior is based on a generalization of probabilistic atlases [21], [22], [23] that was presented in [24]. For the sake of completeness, we summarize the framework here.

Let $\mathbf{l} = \{l_i, i = 1,2, \ldots, I\}$ be a 3D discrete label image (i.e., a segmentation) defined at I spatial locations (voxels), such that each voxel has a label belonging to one of L possible classes, i.e., $l_i \in \{1, \ldots, L\}$. The prior assumes that this segmentation was generated through the following process: [(i)]

1. A tetrahedral mesh covering the region of interest (a bounding box containing the brainstem with a 15 mm margin in each direction) is defined by the reference position $\mathbf{x}^{ref}$ of its N nodes and their connectivity $\mathscr{K}$. Each node n has an associated set of probabilities for the different possible neuroanatomical labels

   $\alpha_n = (\alpha_n^1, \ldots, \alpha_n^L)$.

2. The mesh is deformed from its reference position by sampling from the following prior probability distribution, which was introduced in [25]:

$$p(\mathbf{x}|K, \mathbf{x}^{ref}, \mathscr{K}) \propto \exp\left[-K\sum_{t=1}^{T}U_t^{\mathscr{K}}(\mathbf{x}|\mathbf{x}^{ref})\right],$$

   where T is the number of tetrahedra in the mesh, K is its stiffness, and $U_t^{\mathscr{K}}(\mathbf{x}|\mathbf{x}^{ref})$ is a term that goes to infinity if the Jacobian determinant of the deformation of the tth tetrahedron approaches zero, ensuring that the topology of the mesh is preserved.

3. Using the deformed position, the label probabilities at each voxel location in the region of interest are computed from the values at the vertices of the tetrahedron using barycentric interpolation.

$$p_i(l|\alpha, \mathbf{x}, \mathcal{K}) = \sum_{n=1}^{N} \alpha_n^l \phi_n(\mathbf{r}_i),$$

where $\alpha = (\alpha_1, \ldots, \alpha_N)$ groups the label probabilities of all mesh nodes, $\mathbf{r}_i$ represents the spatial coordinates of voxel i, and φ_n is an interpolation basis function linked to node n. We use linear barycentric interpolation for simplicity, but more complex models may be useful, based for example on a *softmax* function [26], [27].

4. At each voxel location, the corresponding label is independently sampled from the categorical distribution parametrized by the interpolated probability vector, such that:

$$p(\mathbf{l}|\alpha, \mathbf{x}, \mathcal{K}) = \prod_{i=1}^{I} p_i(l|\alpha, \mathbf{x}, \mathcal{K}).$$

Given this generative model, learning an atlas from a set of training data (manual segmentations) amounts to estimating the mesh (reference position $\mathbf{x}^{ref}$ and connectivity $\mathcal{K}$) and associated probability vectors α that most likely generated the label images. As shown in [24], learning the atlas is equivalent to minimizing the number of bits needed to encode the training data, which yields sparse atlases with adaptive resolution, i.e., few nodes are used to describe flat regions of the atlas, while nodes are more dense in convoluted areas.

In this study, we wish to combine the manual annotations from the FreeSurfer and brainstem datasets, which carry complementary information: the former provides information on the whole brainstem and surrounding structures, but not on the internal brainstem parcellation, while the latter describes the midbrain, SCP, pons and medulla, but carries no information on the structures surrounding the brainstem. By combining the two datasets, we can build a probabilistic atlas that includes both the brainstem structures (midbrain, SCP, pons, medulla) and surrounding anatomy (cerebellum, cerebral white matter, etc). For such scenarios, we previously proposed a modification [28] of the atlas construction algorithm [24] that can cope with heterogeneously labeled datasets.

Specifically, we assume that the probabilistic atlas generated M segmentations $\mathbf{l}_m$, $m = 1, \ldots, M$ (where M is the number of labeled scans in the FreeSurfer and brainstem datasets combined), at a *fine* level of detail, in which pons, midbrain, medulla and SCP coexist with all the surrounding structures defined in the FreeSurfer dataset. These segmentations are not observed; instead, we have access to a different set of *coarse* label volumes $\mathbf{c}_m$, $m = 1, \ldots, M$, which are obtained by merging all the surrounding structures into a single, generic background label (brainstem dataset) or by merging pons, midbrain, medulla and SCP into a single brainstem structure (FreeSurfer dataset). These coarse label volumes correspond to the manual delineations from which we build the atlas, and are related to the fine labels by two protocol functions f_{FS} and f_{BS}, such that f_{FS} collapses all brainstem structures into a generic brainstem label, and f_{BS} collapses all the structures surrounding the brainstem into a

single, generic background label. Therefore, the probability of observing a collapsed label at a given spatial location is:

$$p_i(c_{i,m}|\alpha, \mathbf{x}_m, \mathscr{K}) = \sum_{k|f_{(\cdot)}(k)=c_{i,m}} p_i(k|\alpha, \mathbf{x}_m, \mathscr{K}),$$

where $f_{(\cdot)}$ is the protocol function corresponding to training volume m (f_{FS} or f_{BS}, depending on whether it belongs to the FreeSurfer or brainstem dataset, respectively). The sum over all classes compatible with $c_{i,m}$, reflects the uncertainty in the underlying fine labels at each voxel i.

The whole generative process is summarized in Figure 2. The final atlas, which is defined at the fine level of detail, describes (at least partially) the following structures: midbrain, pons, medulla, SCP, third ventricle, fourth ventricle, left / right lateral ventricle, left / right choroid plexus, left / right cerebellar cortex, left / right cerebellar white matter, left / right thalamus, left / right cerebral cortex, left / right cerebral white matter, left / right hippocampus, left / right amygdala, left / right pallidum, left / right putament, left / right thalamus and left / right accumbens area.

2.4 Segmentation

Given the probabilistic atlas of brainstem anatomy, the segmentation of a novel scan can be carried out with the algorithm described in [24]. This algorithm builds on the generative model of the data described above: first, we assume that the probabilistic atlas generates an underlying segmentation (at the fine level of detail) following the four-step process described in Section 2.3. Given the segmentation $\mathbf{l}$, an intensity image $\mathbf{y} = \{y_i, i = 1,2, \ldots, I\}$ is generated from the labels by independently drawing at each voxel a sample from a Gaussian distribution, whose parameters (mean and variance) depend on the label of the voxel. Because the appearance of the brainstem is relatively flat in the MRI scans of all the datasets used in this study, a single Gaussian was found to suffice to model the intensities within each tissue type (although more complex mixture models can also be used [22], [29]). Rather than allowing each label to have its own Gaussian parameters, we assume that all white matter structures (cerebral and cerebellar white matter; medulla; pons; midbrain; and SCP) belong to a global white matter class, in order to reflect the fact that there is little image contrast between such structures, increasing the robustness of the segmentation. Likewise, CSF structures (third, fourth and lateral ventricles) share a global class, and so do the gray matter structures (cerebellar and cerebral cortex, hippocampus and amygdala). The rest of structures in the atlas (pallidum, accumbens, putamen, thalamus, choroid plexus and background) have their own global classes, i.e., their own sets of Gaussian parameters. The probability of observing an intensity image is therefore:

$$p(\mathbf{y}|\mathbf{l}, \theta) = \prod_{i=1}^{I} p_i(y_i|\mu_{G(l_i)}, \sigma^2_{G(l_i)}) = \prod_{i=1}^{I} \mathscr{N}(y_i; \mu_{G(l_i)}, \sigma^2_{G(l_i)}),$$

where $\mathcal{N}$ is the Gaussian distribution, θ groups the Gaussian parameters of all global classes, and $G(l_i)$ is the global class corresponding to label l_i.

Given this generative model, segmentation can be cast as Bayesian inference problem: given the probabilistic atlas and the observed image intensities, what is the most likely segmentation? This problem can be solved by first estimating the model parameters (mesh deformation and Gaussian means and variances) from the data, and using the computed point estimates $\hat{\mathbf{x}}$ and θ to determine the most likely segmentation. Assuming a flat prior for the Gaussian parameters and using Bayes rule, the point estimates are given by:

$$\{\hat{\mathbf{x}}, \theta\} = argmax_{\mathbf{x},\theta} p(\mathbf{x}, \theta | \mathbf{y}, \alpha, \mathbf{x}^{ref}, K, \mathcal{K})$$
$$= argmax_{\mathbf{x},\theta} \log p(\mathbf{x} | K, \mathbf{x}^{ref}, \mathcal{K})$$
$$+ \sum_{i=1}^{I} \log \left[\sum_G p_i(y_i | \mu_G, \sigma_G^2) \sum_{k \in G} p_i(k | \alpha, \mathbf{x}, \mathcal{K}) \right].$$

This problem is solved with with a coordinate ascent scheme, alternately optimizing the mesh deformation $\mathbf{x}$ with a conjugate gradient optimizer and the Gaussian parameters θ with an expectation maximization (EM) algorithm [30]. Once the optimal parameters have been computed, the final segmentation can be computed for each voxel independently as:

$$\hat{l}_i = argmax_k p_i(y_i | \mu_{G(k)}, \sigma_{G(k)}^2) p_i(k | \alpha, \mathbf{x}, \mathcal{K}),$$

and the expected value of the volume of a given structure is (in voxels):

$$V(k) = \sum_{i=1}^{I} \frac{p_i(y_i | \mu_{G(k)}, \sigma_{G(k)}^2) p_i(k | \alpha, \mathbf{x}, \mathcal{K})}{\sum_{k'=1}^{L} p_i(y_i | \mu_{G(k')}, \sigma_{G(k')}^2) p_i(k' | \alpha, \mathbf{x}, \mathcal{K})}, \quad (1)$$

where k is the label corresponding to the structure.

Further details on the segmentation algorithm can be found in [24], [31].

3 Experiments and results

3.1 Experimental setup

The brainstem segmentation algoritmh was evaluated in three different sets of experiments, one with each dataset. In all experiments, the brain MRI scans were preprocessed as follows. First, the T1 data were processed with the FreeSurfer pipeline, which includes resampling to 1 mm isotropic resolution, bias field correction [32], skull stripping [33], intensity normalization and segmentation of subcortical structures [13]. The FLAIR scans (in the brainstem dataset) were bias field corrected and rigidly aligned with the corresponding T1 images using mutual information in order to ensure that the gold standard, T1 and FLAIR images were in the same coordinate frame. In addition, the brain masks computed by FreeSurfer from the T1 data were applied to the FLAIR scans of this dataset.

After preprocessing, the skull-stripped, bias-field-corrected images (T1 or FLAIR) were then fed to the segmentation algorithm, which was initialized by aligning the probabilistic atlas to the whole brainstem segmentation produced by FreeSurfer ("aseg.mgz") with an affine transform. The stiffness of the mesh was set to $K = 0.05$ in all experiments. The mesh was rasterized (i.e., interpolated to a regular voxel grid) at 0.5 mm isotropic resolution, which produces a segmentation at that voxel size.

3.1.1 Direct evaluation with brainstem dataset—In this set of experiments, we used a leave-one-out scheme to automatically segment the subjects in the brainstem dataset using the T1 and FLAIR scans as input. First, we fused the two manual delineations of each T1 scan of the brainstem dataset into a single gold standard segmentation using the multi-label version of the STAPLE algorithm [34] with flat label priors. Then, the leave-one-out atlases were built upon the gold standard segmentations and the manual delineations of the FreeSurfer dataset. The T1 and FLAIR scans of each subject were finally analyzed with the proposed segmentation algorithm using the corresponding leave-one-out atlas (i.e., built upon the annotations made on the images from the other nine subjects, in addition to the FreeSurfer dataset).

The automated segmentations computed from the T1 and FLAIR scans of each subject were compared with each other (in order to estimate the robustness of the algorithm against changes in MRI contrast) and with the gold standard (in order to evaluate the accuracy of the segmentation). Segmentations were compared with three different metrics: Dice overlap, symmetric maximal surface-to-surface (Hausdorff) distance and symmetric mean surface-to-surface distance (see definitions in 8). We also computed the correlation of the volume estimates derived from the T1 and FLAIR scans of each subject.

3.1.2 Indirect evaluation through aging study on FreeSurfer dataset—We also evaluated the segmentation method indirectly with an aging analysis. First, we tested whether the algorithm could detect the effects of aging in the volume of specific brainstem subregions. Such effects have been previously reported by studies based on manual delineations [35], [36]. We segmented the scans of the FreeSurfer dataset in a leave-one-out fashion, i.e., each scan was segmented with an atlas created upon the other 38 (in addition to the 10 gold standard segmentations of the brainstem dataset). Then, the volumes of the brainstem structures of each scan were computed with Equation (1). Next, for each of the brainstem structures, we fitted a general linear model (GLM) predicting the volume of the structure at hand as a linear combination of a bias, the age of the subject and his/her intracranial volume (ICV, as estimated by FreeSurfer). Then, we tested whether the slope corresponding to age was significantly different from zero. We chose the FreeSurfer dataset – rather than ADNI – for the aging experiment because of its wider age range (69 vs. 35 years).

In order to demonstrate the value of working with the volumes of the midbrain, pons, medulla and SCP (rather than using only the volume of the whole brainstem), we conducted another experiment in which we used a GLM to predict the age of a subject as a linear combination of a bias, his/her ICV and either the volume of the whole brainstem or the volumes of the four brainstem structures. Then, we used an F-test to assess whether the

improvement of the fit yielded by the additional variables (the volumes of the brainstem structures) was significant. Moreover, we also predicted ages from both models using a leave-out-one scheme (such that the regression coefficients used to predict the age of each subject are computed upon all other subjects), in order to compare the correlations of the predictions given by both models with the real age. The statistical significance of the difference between the two correlations was assessed with Meng's test [37].

3.1.3 Evaluation with pathological dataset (ADNI)—The third set of experiments was based on the ADNI dataset, which includes scans of elderly controls and AD subjects acquired at different sites with different platforms, and therefore exhibits a larger degree of variability in image contrast and anatomy than the brainstem and FreeSurfer datasets. We segmented the ADNI scans with an atlas built upon all 39 manual delineations of the FreeSurfer dataset and all 10 gold standard segmentations of the brainstem dataset. In a first experiment, we assessed the impact of AD on the volumes of the brainstem structures in a quantitative fashion. To do so, we first corrected the data for age and ICV by fitting a GLM predicting the volume of each structure from these two variables, and then using a two-sample, one-tailed t-test to compare the residuals from the AD and control groups. In a second experiment, we evaluated the robustness of the segmentation qualitatively. Since no ground truth was available for this dataset, the robustness was assessed by visually inspecting the outputs and grading each segmentation as satisfactory or unsatisfactory; this task was performed by JEI.

3.2 Results

3.2.1 Direct validation: Dice scores and surface-to-surface distances on brainstem dataset—Figure 3 shows box plots for the Dice overlap, symmetric mean surface-to-surface distance and symmetric maximal surface-to-surface (i.e., Hausdorff) distance. The plots compare the agreement of the automatic segmentations of T1 and FLAIR between themselves and with the gold standard. They also display the agreement between the two human raters (i.e., the inter-observer variability), which puts the other metrics in context – since it represents an upper bound of the performance than an automated method can achieve.

For the midbrain and the pons, the automated segmentation based on T1 images is very accurate (mean Dice: 88% and 94%; mean surface distance: 0.7 mm and 0.5 mm; Hausdorff distance 3.7 mm and 3.5 mm, respectively), and so is the segmentation based on FLAIR scans, which produces almost identical results (mean Dice: 88% and 94%; mean surface distance: 0.7 mm and 0.5 mm; Hausdorff distance 3.9 mm and 3.8 mm). Compared with the inter-observer variability (with paired t-tests), the performance is not significantly inferior according to the Dice scores (T1 and FLAIR) and the Hausdorff distances (T1) ; however, the mean surface-to-surface distance is significantly larger for both the T1 and FLAIR segmentations ($p < 0.05$ and $p < 0.01$, respectively).

For the SCP, which is a small and thin structure, the gap between the automated method and the inter-observer variability is wider and statistically significant ($p < 0.01$) according to all metrics, except for the Hausdorff distance in FLAIR. The Dice score is particularly

penalized by the thin shape of the structure, since its width is comparable to the voxel size. Therefore, surface distances are more informative for this structure. Specifically, the mean and maximal surface-to-surface distances are comparable to those obtained for the midbrain and pons, which indicates that the performance of the automated algorithm in the SCP is on par with the larger structures. The mean surface distance is 0.6 mm for both T1 and FLAIR (compared with 0.3 mm for intra-observer variability) and the mean Hausdorff distance is 4.0 mm for T1 and 3.5 mm for FLAIR (the intra-observer variability is 3 mm).

The robustness of the method against changes in MRI contrast is demonstrated by how close the similarity metrics are when the T1 and FLAIR segmentations are compared with the gold standard. The similarity of the two automated segmentations with each other is also large, particularly when measured with Dice. Moreover, the volumes derived from them are highly correlated (see Figure 4): the correlation coefficient is 0.999 for the pons, 0.987 for the midbrain, 0.968 for the medulla and 0.815 for the SCP, which is once more penalized by its thin shape.

Finally, Figure 5 shows sample automatic segmentations and compares them with the manual delineations. The agreement between the two is strong, except for the SCP, which is typically undersegmented by the automated method – especially in T1.

3.2.2 Indirect validation with FreeSurfer dataset: aging study—Figure 6 shows scatter plots and the linear fit of the ICV-corrected volumes of the brainstem structures of the subjects from the FreeSurfer dataset against their ages; in all four structures, the dependence of the volume on ICV is statistically significant ($p < 10^{-4}$). However, the only structure for which there is significant atrophy (i.e., statistically significant dependence of volume on age) is the midbrain ($p = 0.01$, yearly decline 0.12%); the pons, medulla and SCP are spared. This is consistent with previous MRI studies based on manual delineations [35], [36].

In the age prediction experiment, the simultaneous use of all brainstem structures in the estimation produces a significant improvement of the fit of the GLM (i.e., age prediction) compared with using only the volume of the whole brainstem ($p = 5.3 \times 10^{-5}$). Moreover, when age is predicted in a leave-one-out framework, the standard error of the prediction error decreases from 24.95 to 18.64 years, and the correlation coefficient increases from 0.14 to 0.60 ($p = 8.5 \times 10^{-3}$). The scatter plots and linear fits of the true and predicted ages are shown in Figure 7.

3.2.3 Effect of AD and robustness of segmentation against pathology—Table 1 summarizes the differences in volume between the AD and control groups for the different brainstem structures. The largest effect is found in the midbrain, as in the aging experiment. Moderate effect sizes were also obtained for the pons, SCP and whole brainstem, whereas no difference between the groups was found in the medulla.

The table shows the mean difference in volume between the two classes for each structure (as a percentage of the mean volume), the effect size of the difference, and the corresponding p value (two-sample, one-tailed t-test).

Finally, Figure 8 shows the segmented midsagittal slices of the first 132 scans in the ADNI dataset; segmentations for the remaining 251 scans are displayed in Figures 9 and 10 in the supplementary material. Despite the anatomical heterogeneity of the images, visual inspection of the complete 3D labelings did not reveal any poorly segmented scan withing the whole dataset. The volume estimates for the brainstem structures of these subjects can be found in Tables 2–7 (also in the supplementary material).

4 Conclusion and discussion

In this paper we have described the construction of a probabilistic atlas of four brainstem structures (midbrain, pons, medulla and SCP) and evaluated the segmentations derived from it on three different datasets. The segmentation is efficient and runs in approximately 15 minutes on a desktop computer. The results have shown that the method can accurately segment the midbrain and pons. The segmentation of the SCP yields lower Dice scores due to its thin shape (its thickness is comparable to the voxel size), but approximately the same surface-to-surface and Hausdorff distances as the midbrain and pons. The segmentation of the medulla could not be evaluated directly due to the lack of ground truth segmentations. In the indirect evaluation through the aging experiment, the medulla did not shown the mild decline reported in [36]; however, this could be due to the noise introduced by the inferior part of the medulla's being left out by the field of view of the scan or the brain extraction. This could also explain why no difference was found between the AD and control groups for this structure.

The results on the age prediction experiment have also shown that the volumes of the different brainstem structures contain more information than the volume of the brainstem as a whole: the GLM based on all the volumes produces a much more accurate prediction than the GLM that only uses the volume of the whole brainstem. However, the differences found between AD patients and controls in the ADNI dataset were modest compared with the values reported by *Nigro et al.* [18]. Further exploration will be required to assess whether this is due to differences in the chosen subset of ADNI or in the segmentation methods.

The experiments have also shown that the segmentation method is robust against changes in MRI acquisition platforms and protocols: it produces consistently satisfactory results on three different datasets, including one with two types of MRI contrast (brainstem dataset, T1 and FLAIR) and another that contains scans from elderly subjects and AD patients scanned at different sites (ADNI). The segmentations of the FLAIR scans were only marginally less accurate than those of the T1 scans. This is in spite of the fact that manual delineations were made in the space of the T1 images, implying that errors in the registration of the FLAIR volumes directly affect the similarity metrics computed for their segmentations.

In order to model the relationship between the segmentations and the intensities, we used a simple Gaussian likelihood. While this model sufficed in our study, MRI sequences designed to maximize the contrast of brainstem structures might require more flexible distributions, such as Gaussian mixture models. More complex likelihood terms will also be necessary to incorporate other MRI modalities into the algorithm in order to increase its performance. For instance, diffusion MRI promises to improve the accuracy of the method

in the SCP, since cerebellar tracts provide a salient feature for its segmentation. Exploring these directions, along with including brainstem substructures (e.g., raphe nuclei, red nuclei) in the atlas, remains as future work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Appendix

5 Guidelines used in the manual delineation of the pons

1. Tracing of the superior boundary: in the midsagittal plane, first trace the line passing through the superior pontine notch and the inferior edge of the quadriminal plate, then the line to the quadriminal plate and, after all regions have been traced, erase the extraneous portions of the lines. Then, continue tracing in lateral slices. Once the oculomotor nerve (CN III) is visible, make sure that the anterior boundary point is below CN III. Once the inferior colliculus is no longer visible, switch view to the most lateral slice in which the midbrain and pons begin to separate. On this slice, trace a diagonal line along the notch that appears between the midbrain and pons. Repeat this procedure in the medial slices. If the posterior notch is not visible in a given slice, identify where it would be based on the posterior notch position in both medial and lateral adjacent slices.

2. Tracing of the inferior boundary: in sagittal view, identify the slice in the left hemisphere where the anatomical boundary between the medulla and pons is most prominently visible as a bright white line. Trace a straight line from the anterior to the posterior point of the anatomical boundary. Even if the anatomical boundary is not straight, the line defining the inferior boundary should be a straight line. Then, in axial view, identify the most superior slice where the voxel from the sagittally drawn line appears. In this slice, trace the posterior boundary based on the tissue-CSF (cerebrospinal fluid) boundary between the fourth ventricle and the pons. Trace along the CSF-tissue boundary just past the vestibular nuclei (CN VIII), which can be visualized by the notch of the fourth ventricle boundary, which becomes a vertical line. Then, in sagittal slices, trace the inferior boundary as the straight diagonal line that extends anteriorly from the inferior pontine notch to the posterior voxel created by the axial-defined boundary. In the most lateral (sagittal) slice where the axial-defined voxel boundary is visible, move the cursor to the inferior pontine notch and switch to coronal view. In this one coronal slice, trace around the curvature of the bright pons and middle cerebellar peduncle regions and fill in the region. Finally, return to sagittal view and verify that in the next lateral sagittal slice a vertical line appears extending from the fourth ventricle. This line will define the posterior boundary in subsequent lateral sagittal slices.

3. Tracing of the posterior boundary: in sagittal view, first trace the line along the tissue-CSF boundary. Once the middle and superior cerebellar peduncles make contact with the pons, draw a straight line from the superior point where the peduncle first branches from the pons to the most inferior point where the peduncle branches from the pons. Then, repeat this step in subsequent sagittal slices. If there is incidentally any CSF space covered by the labeling, make sure it is not included in the final segmentation.

4. Tracing of thee anterior and anterior-inferior boundaries: first trace the line along the tissue-CSF boundary in sagittal view. Then, in lateral slices, trace the inferior boundary as defined by the tissue-CSF interface, without including the blood vessels and nerves that extend from the middle of the pons. Finally, identify the most inferior axial slice where the posterior boundary of the segmentation appears to protrude posteriorly from the line that defines the posterior boundary. On that slice, draw a straight diagonal line from the most lateral point of the medially protruding segmentation to the most lateral voxel of the line defining the posterior boundary.

6 Guidelines used in the manual delineation of the midbrain

1. In sagittal view: in the most lateral slice of the right hemisphere where the CSF boundary is clearly visible between the thalamus and midbrain, trace the superior boundary of the midbrain as defined by the CSF boundaries. In order to make sure structures above the midbrain are not included, do not segment any voxels above the line of superior-most line of the superior colliculus. Trace the anterior boundary

as the straight vertical line just posterior of the mammillary bodies. Repeat on left side.

2. Identify the superior-most axial slice in which the outline of midbrain is visible based on tissue-CSF boundary. Make sure that the CSF boundary is clearly visible on the anterior boundary. In this slice, the midbrain should appear clearly separated from other structures; this may be different slices for each side of the brain. Trace around this shape.

3. Identify the most posterior coronal slice where the "neck/bridge" portion that is thinner than both the midbrain and thalamus is clearly visible. In this slice, trace a straight diagonal line from the lateral inferior corner of the third ventricle to the inferior notch between the midbrain and thalamus. This will likely be in different slices in each side of the brain.

4. Continue tracing posteriorly in coronal view using the technique described in step 3, i.e., tracing a straight diagonal line from the lateral inferior corner of the third ventricle to the most lateral voxel of the line created by the axially traced slice.

5. Once the midbrain and thalamus are separated by CSF space, the superior midbrain boundaries are defined by the tissue-CSF boundaries. Trace around the colliculi and midbrain in coronal view until the colliculi are no longer visible.

6. Continue tracing anteriorly in coronal view using the technique described in step 5. If two voxels from the sagittal tracing are visible, use the most superior to define the superior midbrain boundary. Once the sagitally traced voxels are no longer visible in coronal, stop drawing the superior boundary in coronal and trace around the inferior portion bounded by CSF.

7. Identify the sagittal slice described in step 1. Then, draw a line from the superior voxel of the line created in the coronally-traced slice from step 3 to the anterior voxel of the horizontal line given by the segmentation at this point. This will create a right triangle that must be filled in. Repeat this procedure in all lateral sagittal slices. Also, make sure that the small area of midbrain tissue bounded by CSF space below the horizontal line is filled in.

8. In sagittal slices medial to the slice described in step 4, make sure that the thin midbrain portion posterior to the mammillary body are segmented by tracing a straight vertical line upwards from the most posterior voxels of the mammillary body. Of this line only include the voxels that are superior to the most inferior point of the thin midbrain bridge.

7 Guidelines used in the manual delineation of the SCP

1. In axial view, identify the most inferior slice where the parabrachial recess is clearly visible. In this slice draw a vertical line extending down from the lateral boundary of the fourth ventricle. The sagittal slice where this line appears will be the most lateral slice for tracing the SCP. The recess will appear in different slices on the left and right sides. Erase extraneous portions of the axially drawn line.

2. Do all tracings in sagittal view. In the midsagittal plane, trace around the thread-like structure that extends from the bottom of the tectum into the cerebellum. If the upper and lower parts of the SCP are not connected, trace around both parts separately.

3. The superior boundary is formed by the inferior boundary of the midbrain tectum. The upper part of the SCP will be defined as the non-black voxels that are excluded from the pons and midbrain.

4. The posterior boundary is defined as a straight vertical line extending down from the superior point where the SCP merges with the cerebellum, at the vertex of the dark right triangle.

5. In lateral sagittal slices, where the SCP makes contact with the pons, the anterior boundary is defined by the posterior boundary of the pons.

8 Metrics used to compare two segmentations

In this study, we have used three different metrics to measure the (dis-)similarity of two segmentations. The first one is the Dice overlap. If *A* and *B* are two binary masks corresponding to a brain structure, their Dice overlap is:

$$DICE = \frac{2|A \cap B|}{|A| + |B|},$$

where |·| represents the size (number of voxels) of a mask.

The other two measures are based on the distances between surfaces. If δA and δB are the surfaces of masks *A* and *B*, the symmetric Hausdorff distance is:

$$SHD = \frac{1}{2} \sup_{a \in \delta A} \inf_{b \in \delta B} d(a, b) + \frac{1}{2} \sup_{b \in \delta B} \inf_{a \in \delta A} d(b, a),$$

where sup is the supremum, inf is the infimum, and $d(a, b) = d(b, a)$ is the Euclidean distance between two points *a* and *b*. The symmetric mean surface-to-surface distance is:

$$SMSTSD = \frac{1}{2} \frac{1}{|\delta A|} \sum_{a \in \delta A} \inf_{b \in \delta B} d(a, b) + \frac{1}{2} \frac{1}{|\delta B|} \sum_{b \in \delta B} \inf_{a \in \delta A} d(b, a).$$

References

1. Williams DR, Lees AJ. Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges. The Lancet Neurology. 2009; 8(3):270–279. [PubMed: 19233037]

2. Boxer AL, Lang AE, Grossman M, Knopman DS, Miller BL, Schneider LS, Doody RS, Lees A, Golbe LI, Williams DR, et al. Davunetide in patients with progressive supranuclear palsy: a randomised, double-blind, placebo-controlled phase 2/3 trial. The Lancet Neurology.

3. Hawkes CH, Del Tredici K, Braak H. A timeline for parkinson's disease. Parkinsonism & related disorders. 2010; 16(2):79–84. [PubMed: 19846332]

4. Grinberg L, Rüb U, Ferretti R, Nitrini R, Farfel J, Polichiso L, Gierga K, Jacob-Filho W, Heinsen H. The dorsal raphe nucleus shows phospho-tau neurofibrillary changes before the transentorhinal region in alzheimer's disease. a precocious onset? Neuropathology and applied neurobiology. 2009; 35(4):406–416. [PubMed: 19508444]

5. Stefani A, Lozano AM, Peppe A, Stanzione P, Galati S, Tropepi D, Pierantozzi M, Brusa L, Scarnati E, Mazzone P. Bilateral deep brain stimulation of the pedunculopontine and subthalamic nuclei in severe parkinson's disease. Brain. 2007; 130(6):1596–1607. [PubMed: 17251240]

6. Minoshima S, Frey KA, Foster NL, Kuhl DE. Preserved pontine glucose metabolism in alzheimer disease: a reference region for functional brain image (pet) analysis. Journal of computer assisted tomography. 1995; 19(4):541–547. [PubMed: 7622680]

7. Dunckley P, Wise RG, Fairhurst M, Hobden P, Aziz Q, Chang L, Tracey I. A comparison of visceral and somatic pain processing in the human brainstem using functional magnetic resonance imaging. The Journal of neuroscience. 2005; 25(32):7333–7341. [PubMed: 16093383]

8. Hadjipavlou G, Dunckley P, Behrens TE, Tracey I. Determining anatomical connectivities between cortical and brainstem pain processing regions in humans: a diffusion tensor imaging study in healthy controls. Pain. 2006; 123(1):169–178. [PubMed: 16616418]

9. Bondiau P-Y, Malandain G, Chanalet S, Marcy P-Y, Habrand J-L, Fauchon F, Paquis P, Courdi A, Commowick O, Rutten I, et al. Atlas-based automatic segmentation of MR images: validation study on the brainstem in radiotherapy context. International Journal of Radiation Oncology* Biology* Physics. 2005; 61(1):289–298.

10. Lee, J-D.; Wang, N-W.; Huang, C-H.; Liu, L-C.; Lu, C-S. A segmentation scheme of brainstem and cerebellum using scale-based fuzzy connectedness and deformable contour model; Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the, IEEE; 2005. p. 459-462.

11. Lee, J-D.; Tseng, Y-x; Liu, L-c; Huang, C-H. A 2-d automatic segmentation scheme for brainstem and cerebellum regions in brain MR imaging; Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on, Vol. 4, IEEE; 2007. p. 270-274.

12. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features; Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, Vol. 1, IEEE; 2001. p. I–511

13. Fischl B, Salat D, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale A. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron. 2002; 33:341–355. [PubMed: 11832223]

14. Fischl B. Freesurfer. Neuroimage. 2012; 62(2):774–781. [PubMed: 22248573]

15. Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage. 2011; 56(3):907–922. [PubMed: 21352927]

16. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, et al. Advances in functional and structural MR image analysis and implementation as fsl. Neuroimage. 2004; 23:S208–S219. [PubMed: 15501092]

17. Heckemann R, Hajnal J, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage. 2006; 33(1):115–126. [PubMed: 16860573]

18. Nigro S, Cerasa A, Zito G, Perrotta P, Chiaravalloti F, Donzuso G, Fera F, Bilotta E, Pantano P, Quattrone A, et al. Fully automated segmentation of the pons and midbrain using human T1 MR brain images. PloS one. 2014; 9(1):e85618. [PubMed: 24489664]

19. Lambert C, Lutti A, Helms G, Frackowiak R, Ashburner J. Multiparametric brainstem segmentation using a modified multivariate mixture of gaussians. NeuroImage: clinical. 2013; 2:684–694. [PubMed: 24179820]

20. Caviness V Jr, Filipek P, Kennedy D. Magnetic resonance technology in human brain science: blueprint for a program based upon morphometry. Brain Dev. 1989; 11(1):1–13. [PubMed: 2646959]

21. Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based tissue classification of MR images of the brain. Medical Imaging, IEEE Transactions on. 1999; 18(10): 897–908.

22. Ashburner J, Friston KJ. Unified segmentation. Neuroimage. 2005; 26(3):839–851. [PubMed: 15955494]

23. Pohl KM, Fisher J, Grimson WEL, Kikinis R, Wells WM. A bayesian model for joint segmentation and registration. NeuroImage. 2006; 31(1):228–239. [PubMed: 16466677]

24. Van Leemput K. Encoding probabilistic brain atlases using bayesian inference. Medical Imaging, IEEE Transactions on. 2009; 28(6):822–837.

25. Ashburner J, Andersson JL, Friston KJ. Image registration using a symmetric prior—in three dimensions. Human brain mapping. 2000; 9(4):212–225. [PubMed: 10770230]

26. Pohl KM, Fisher J, Bouix S, Shenton M, McCarley RW, Grimson WEL, Kikinis R, Wells WM. Using the logarithm of odds to define a vector space on probabilistic atlases. Medical Image Analysis. 2007; 11(5):465–477. [PubMed: 17698403]

27. Ashburner J, Friston KJ. Computing average shaped tissue probability templates. Neuroimage. 2009; 45(2):333–341. [PubMed: 19146961]

28. Iglesias JE, Augustinach JC, Nguyen K, Player CM, Player A, Wright M, Roy N, Frosch MP, McKee AC, Wald LL, Fischl B, Van Leemput K. A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI. Neuroimage. (submitted).

29. Puonti, O.; Iglesias, JE.; Van Leemput, K. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer; 2013. Fast, sequence adaptive parcellation of brain mr using parametric models; p. 727-734.

30. Dempster AP, Laird NM, Rubin DB, et al. Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soc. 1977; 39(1):1–38.

31. Van Leemput K, Bakkour A, Benner T, Wiggins G, Wald LL, Augustinack J, Dickerson BC, Golland P, Fischl B. Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. Hippocampus. 2009; 19(6):549–557. [PubMed: 19405131]

32. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. Medical Imaging, IEEE Transactions on. 1998; 17(1):87–97.

33. Ségonne F, Dale A, Busa E, Glessner M, Salat D, Hahn H, Fischl B. A hybrid approach to the skull stripping problem in MRI. Neuroimage. 2004; 22(3):1060–1075. [PubMed: 15219578]

34. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. Medical Imaging, IEEE Transactions on. 2004; 23(7):903–921.

35. Luft AR, Skalej M, Schulz JB, Welte D, Kolb R, Bürk K, Klockgether T, Voigt K. Patterns of age-related shrinkage in cerebellum and brainstem observed in vivo using three-dimensional MRI volumetry. Cerebral Cortex. 1999; 9(7):712–721. [PubMed: 10554994]

36. Raininko R, Autti T, Vanhanen S-L, Ylikoski A, Erkinjuntti T, Santavuori P. The normal brain stem from infancy to old age. Neuroradiology. 1994; 36(5):364–368. [PubMed: 7936176]

37. Meng X-L, Rosenthal R, Rubin DB. Comparing correlated correlation coefficients. Psychological bulletin. 1992; 111(1):172.
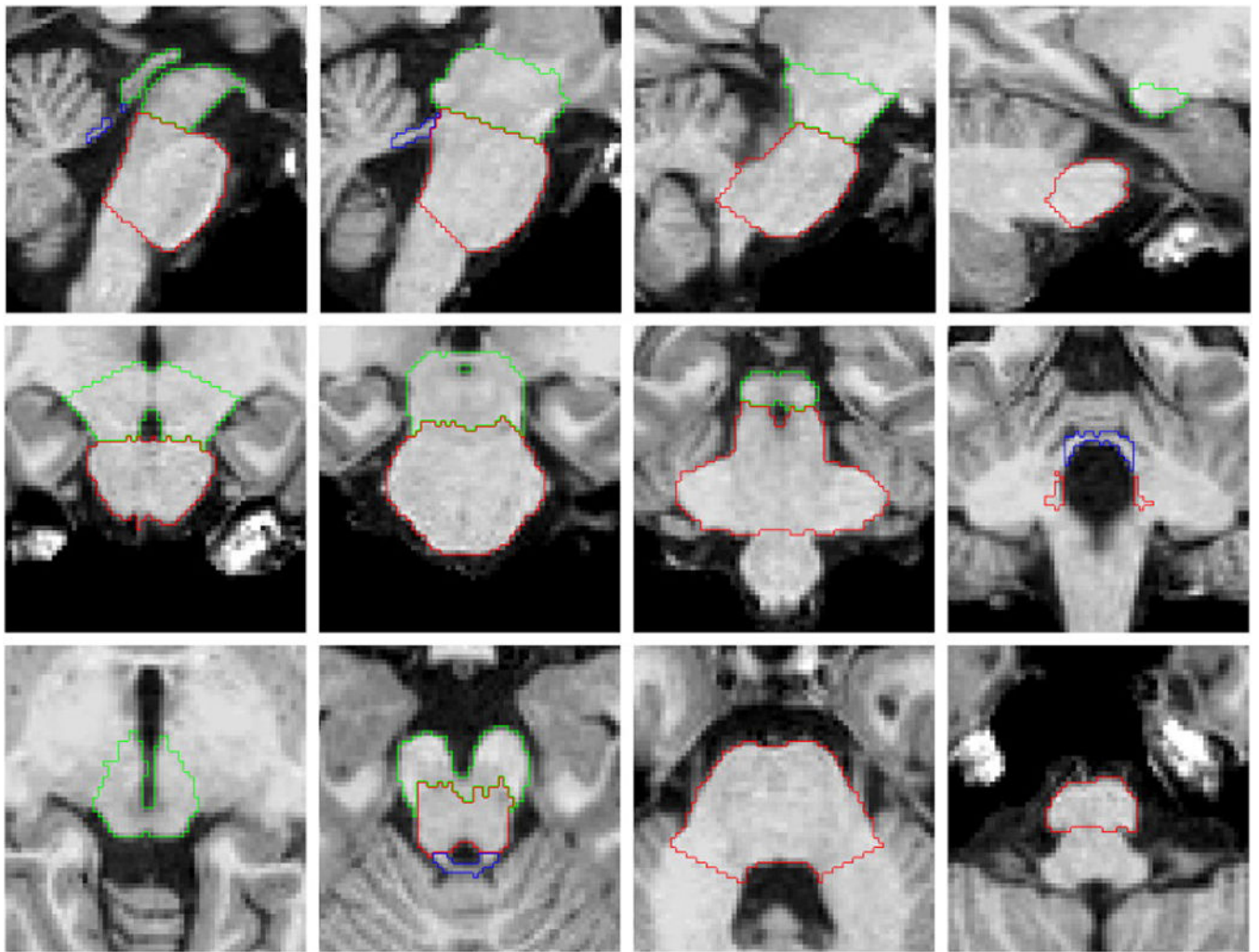
Figure 1.
Manual delineations of a sample subject from the brainstem dataset. Top row: sagittal slices, from medial (left) to lateral (right). Middle row: coronal slices, from anterior (left) to posterior (right). Bottom row: axial view, from superior (left) to inferior (right). The pons is labeled in red, the midbrain in green, and the SCP in blue.
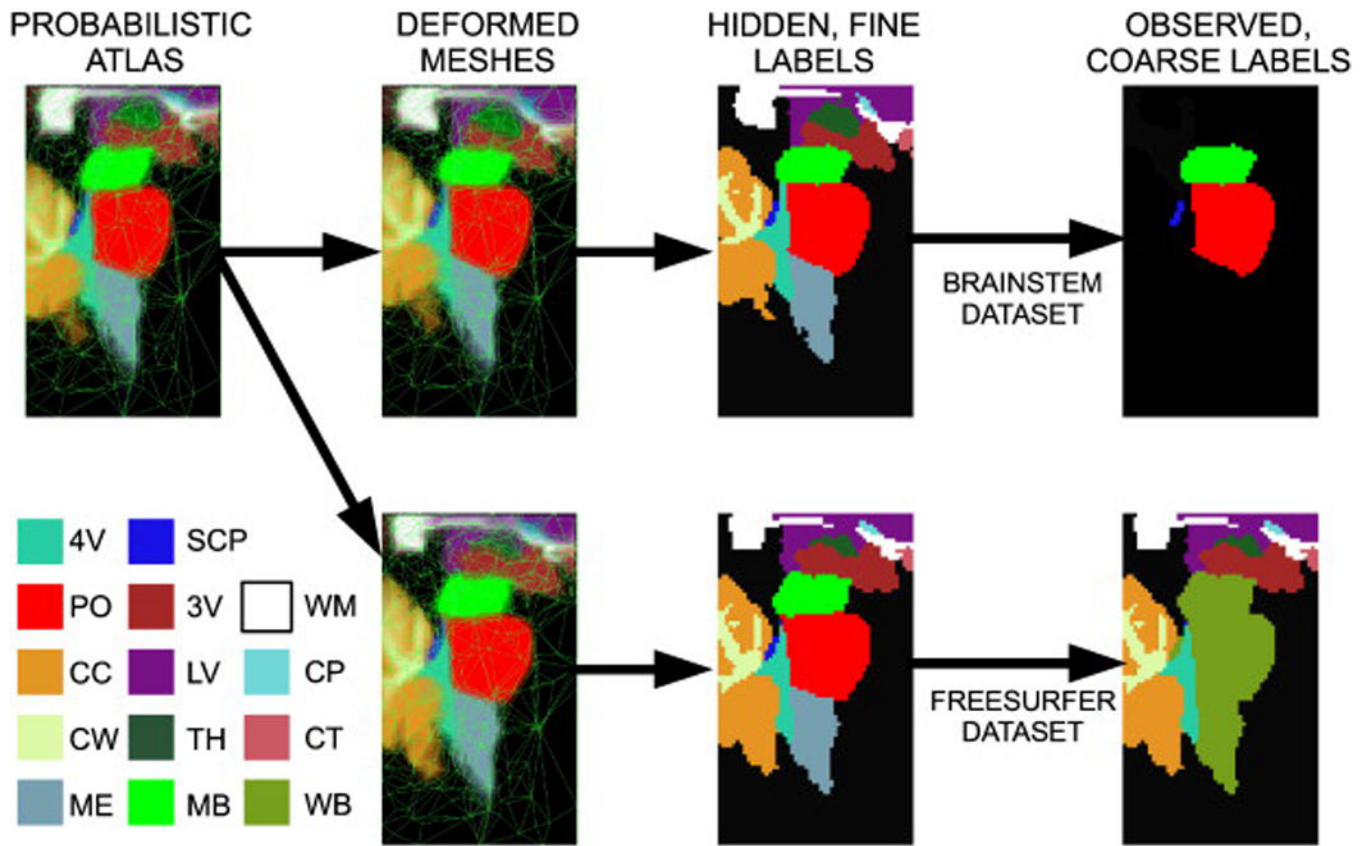
Figure 2.
Generative model of training data. The abbreviations for the structures are the following: 4V = fourth ventricle, PO = pons, CC = cerebellar cortex, CW = cerebellar white matter, ME = medulla, SCP = superior cerebellar peduncle, 3V = third ventricle, LV = lateral ventricle, TH = thalamus, MB = midbrain, WM = white matter, CP = choroid plexus, CT = cortex, WB = whole brainstem. The background is represented in black.
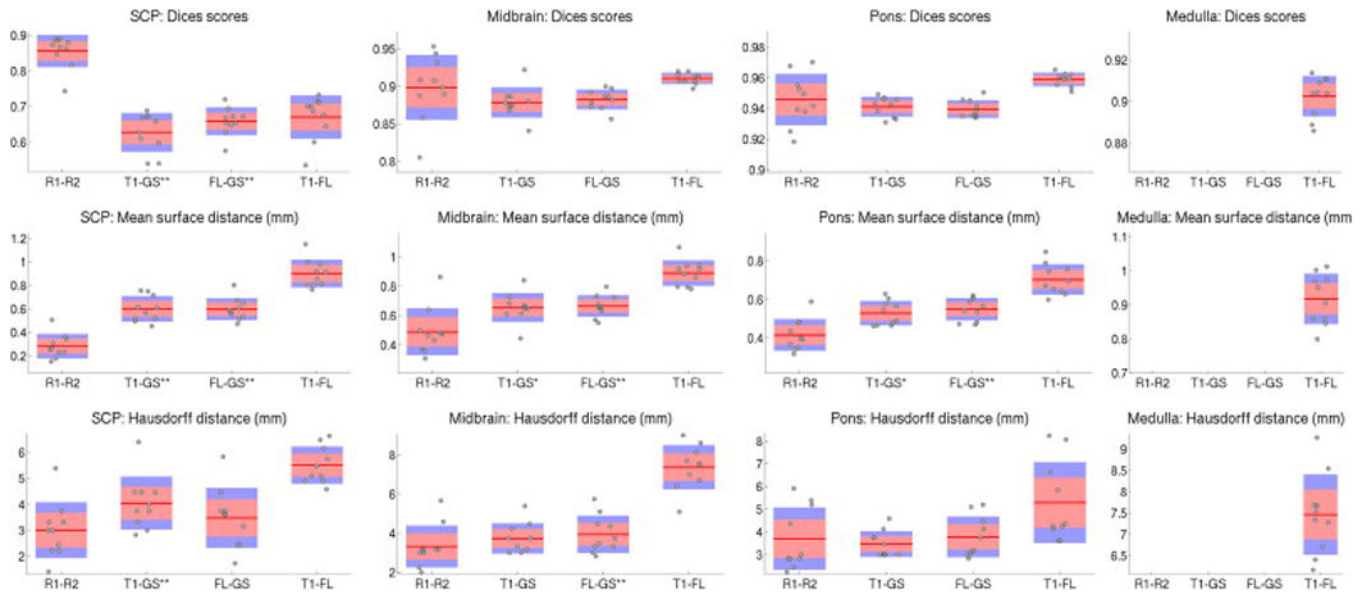
Figure 3.
Box plots for the Dice overlap, symmetric mean surface-to-surface distance and symmetric Hausdorff (maximal) distance for the SCP, midbrain, pons and medulla. R1–R2 represents the agreement between the two human raters (inter-observer variability), T1-GS is between the T1 segmentation and the gold standard, FL-GS is between the FLAIR segmentation and the gold standard, and T1-FL is between the T1 and FLAIR segmentations. Statistically significant differences (as measured by a paired t-test) between T1-GS and R1–R2, as well as between FL-GS and R1–R2, are marked with an asterisk (when $p < 0.05$) or two (when $p < 0.01$). The light red box spans the 95% confidence interval of the mean, which is marked by the red line. The blue box spans one standard deviation of the data. The circles mark the raw data points – slightly jittered along the x axis for clarity. Note that, since there is no ground truth for the medulla, only T1-FL can be computed for this structure.
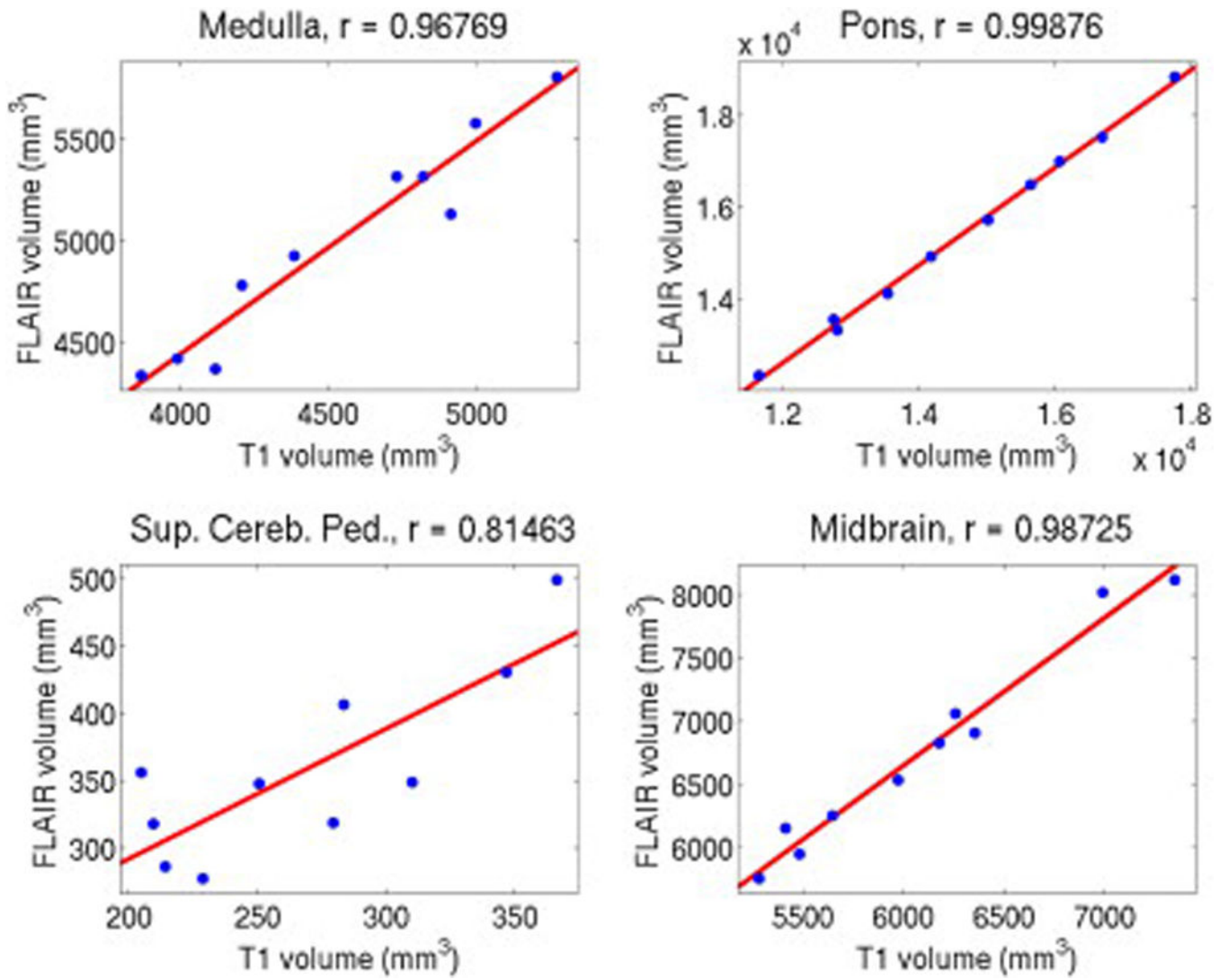
Figure 4.
Scatter plots and linear fits for the volumes of the brainstem structures derived from the segmentations of the T1 and FLAIR scans of the brainstem dataset.
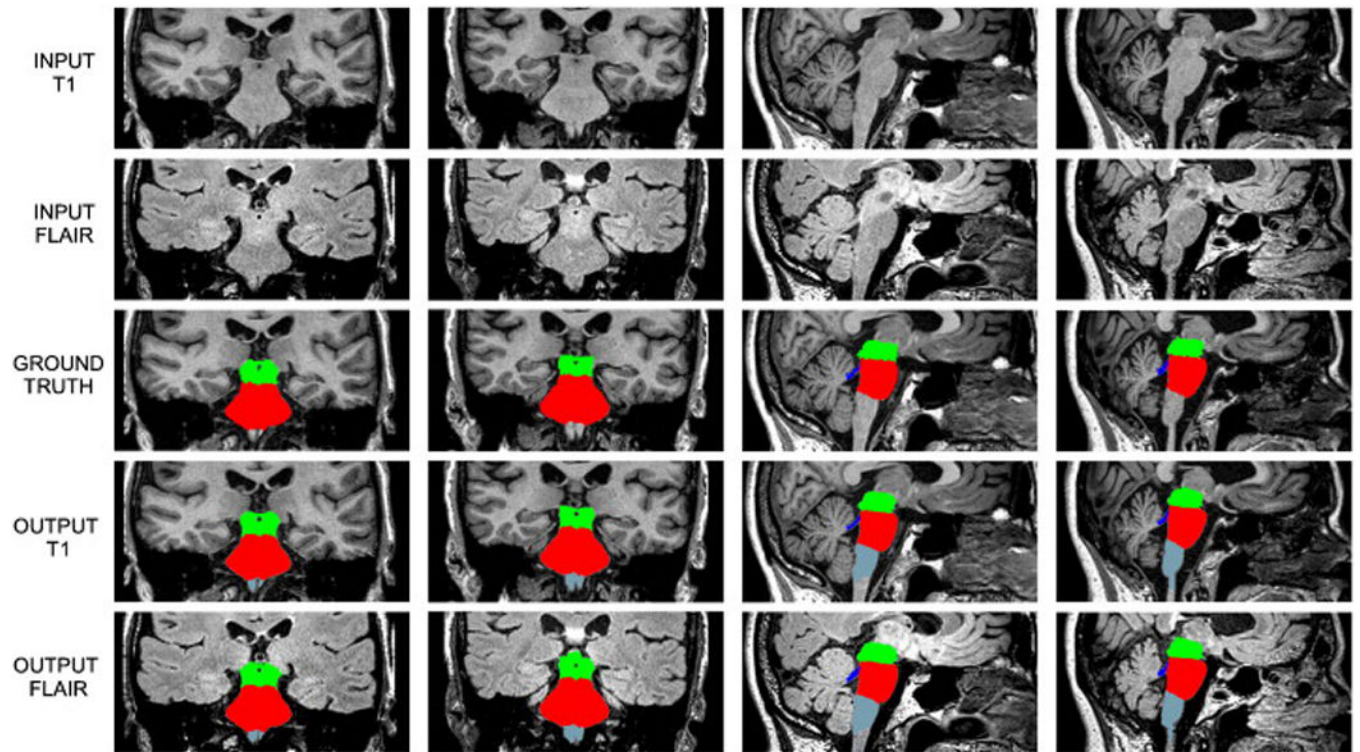
Figure 5.
Sample slices (top two rows), manual delineations (middle row) and automated segmentations (bottom two rows) from the brainstem dataset. The color code is the following: red is pons, green is midbrain, blue is SCP, and gray is medulla. Note that there is no manual segmentations for the medulla.
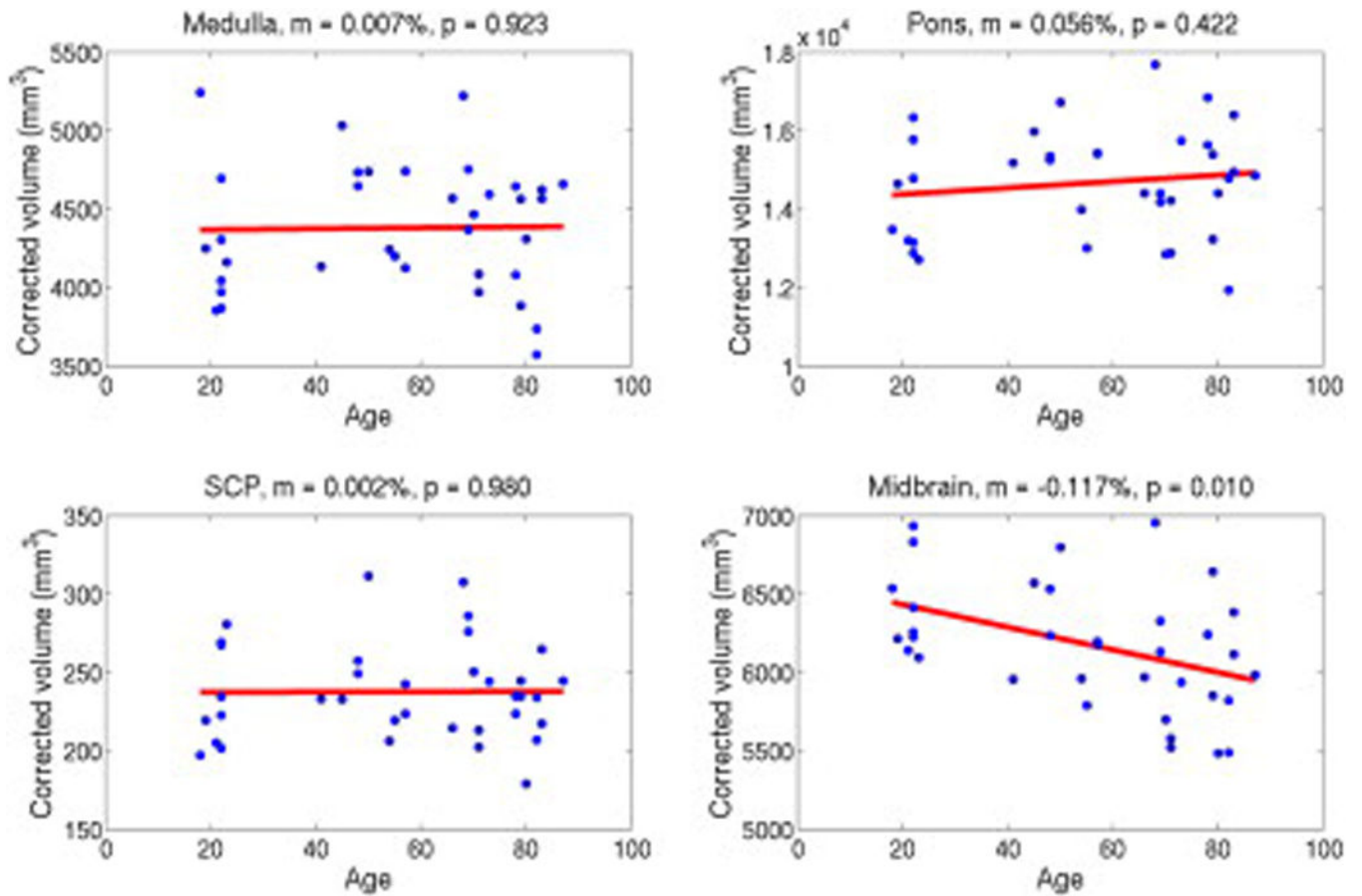
Figure 6.
Scatter plots for the ICV-corrected volumes of the brainstem structure versus age (FreeSurfer dataset). The linear fit is superimposed. The p-value for the hypothesis that the slope of this fit is zero is displayed in the title of each subfigure, along with the yearly atrophy (in %).
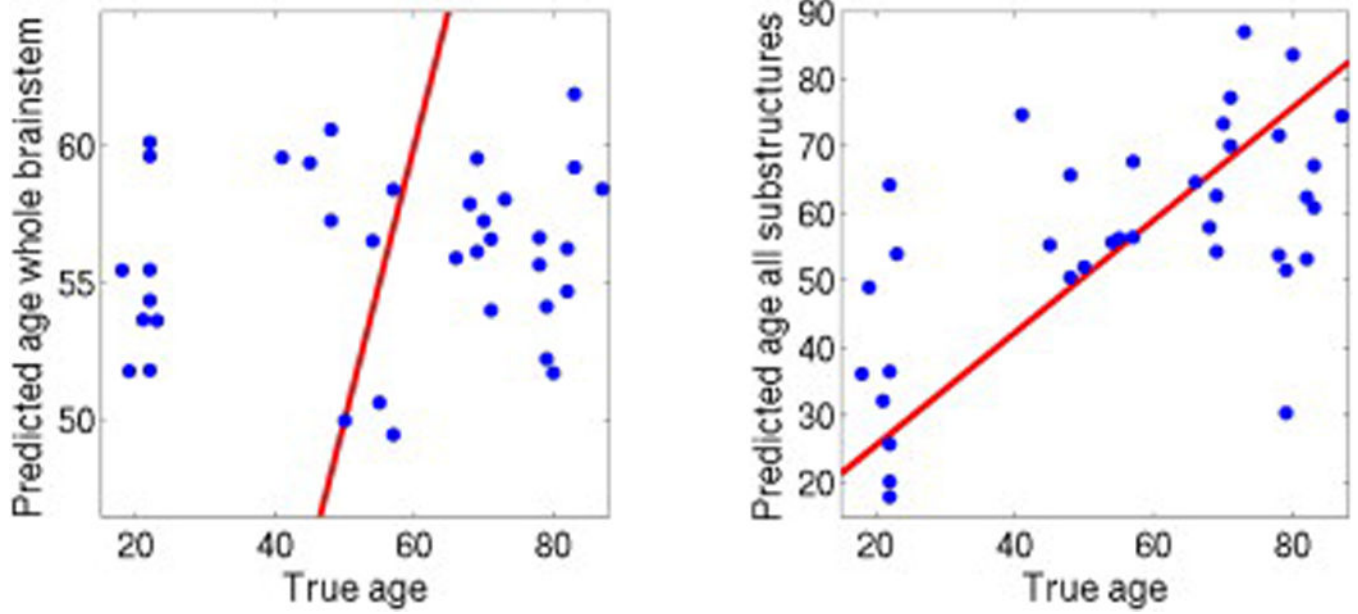
Figure 7.
Scatter plots for real and predicted ages in the FreeSurfer dataset, using only the volume of the whole brainstem (left, $r = 0.14$) and the volumes of all the brainstem structures (right, $r = 0.60$).

Figure 8.
Region of interest covering the brainstem in the midsagittal slice of the first 132 scans from the ADNI dataset. The segmentation is superimposed with 50% transparency. See caption of Figure 5 for the color code.

Table 1

Volumetric study of brainstem structures in ADNI: elderly controls vs. AD patients.

Structure	Vol.diff.(%)	Effect size	p value
Pons	2.6	0.25	0.0072
SCP	4.4	0.24	0.011
Midbra in	2.2	0.32	0.00091
Medull a	0.4	0.04	0.67
Whole BS	2.0	0.24	0.011