Bayesian segmentation of brainstem structures in MRI

Juan Eugenio Iglesias[a,*], Koen Van Leemput[b,e,f], Priyanka Bhatt[c], Christen Casillas[c], Shubir Dutt[c], Norbert Schuff[g], Diana Truran-Sacrey[g], Adam Boxer[c], Bruce Fischl[b,d], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] *Basque Center on Cognition, Brain and Language, Spain*
[b] *Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA*
[c] *Memory and Aging Center, University of California, San Francisco, CA, USA*
[d] *Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology (MIT), Cambridge, MA, USA*
[e] *Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), Denmark*
[f] *Departments of Information and Computer Science and of Biomedical Engineering and Computational Science, Aalto University, Finland*
[g] *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.

Keywords: Brainstem, Bayesian segmentation, probabilistic atlas

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

*Corresponding author
Email address: e.iglesias@bcbl.eu (Juan Eugenio Iglesias)

[1] Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

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 E.2-E.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 Appendix A, Appendix B and Appendix C, 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) 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 $\mathcal{K}$. Each node n has an associated set of probabilities for the different possible neuroanatomical labels $\boldsymbol{\alpha}_n = (\alpha_n^1, \ldots, \alpha_n^L)$.

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

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

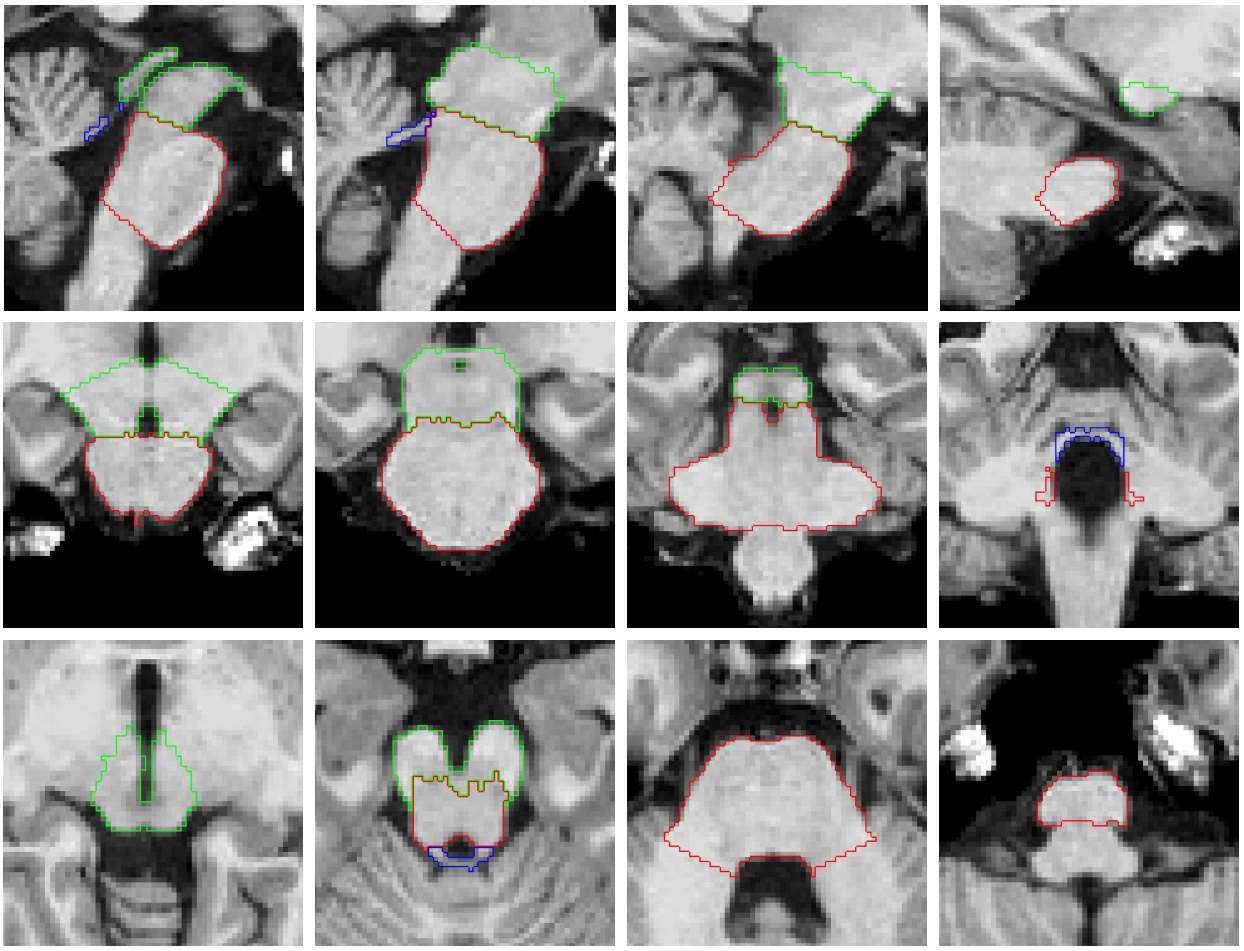

Figure 1: Manual delineations of a sample subject from the brainstem dataset. Top row: sagittal slices, from medial (left) to lateral (right). Middle row: coronal slices, from anterior (left) to posterior (right). Bottom row: axial view, from superior (left) to inferior (right). The pons is labeled in red, the midbrain in green, and the SCP in blue.

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

(iii) 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.

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

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N)$ 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].

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

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

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 $\boldsymbol{\alpha}$ 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:

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

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:

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

where $\mathcal{N}$ is the Gaussian distribution, $\boldsymbol{\theta}$ 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 $\hat{\boldsymbol{\theta}}$ to determine the most likely segmentation. Assuming a flat prior for the Gaussian parameters and using Bayes rule, the point estimates are given by:

$$\begin{aligned}\{\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}\} &= \underset{\mathbf{x}, \boldsymbol{\theta}}{\operatorname{argmax}}\, p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\alpha}, \mathbf{x}^{ref}, K, \mathcal{K}) \\ &= \underset{\mathbf{x}, \boldsymbol{\theta}}{\operatorname{argmax}}\, \log p(\mathbf{x}|K, \mathbf{x}^{ref}, \mathcal{K}) \\ &\quad + \sum_{i=1}^{I} \log \left[\sum_{G} p_i(y_i|\mu_G, \sigma^2_G) \sum_{k \in G} p_i(k|\boldsymbol{\alpha}, \mathbf{x}, \mathcal{K}) \right].\end{aligned}$$

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 $\boldsymbol{\theta}$ 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:

$$\hat{l}_i = \underset{k}{\operatorname{argmax}}\, p_i(y_i|\mu_{G(k)}, \sigma^2_{G(k)}) p_i(k|\boldsymbol{\alpha}, \mathbf{x}, \mathcal{K}),$$

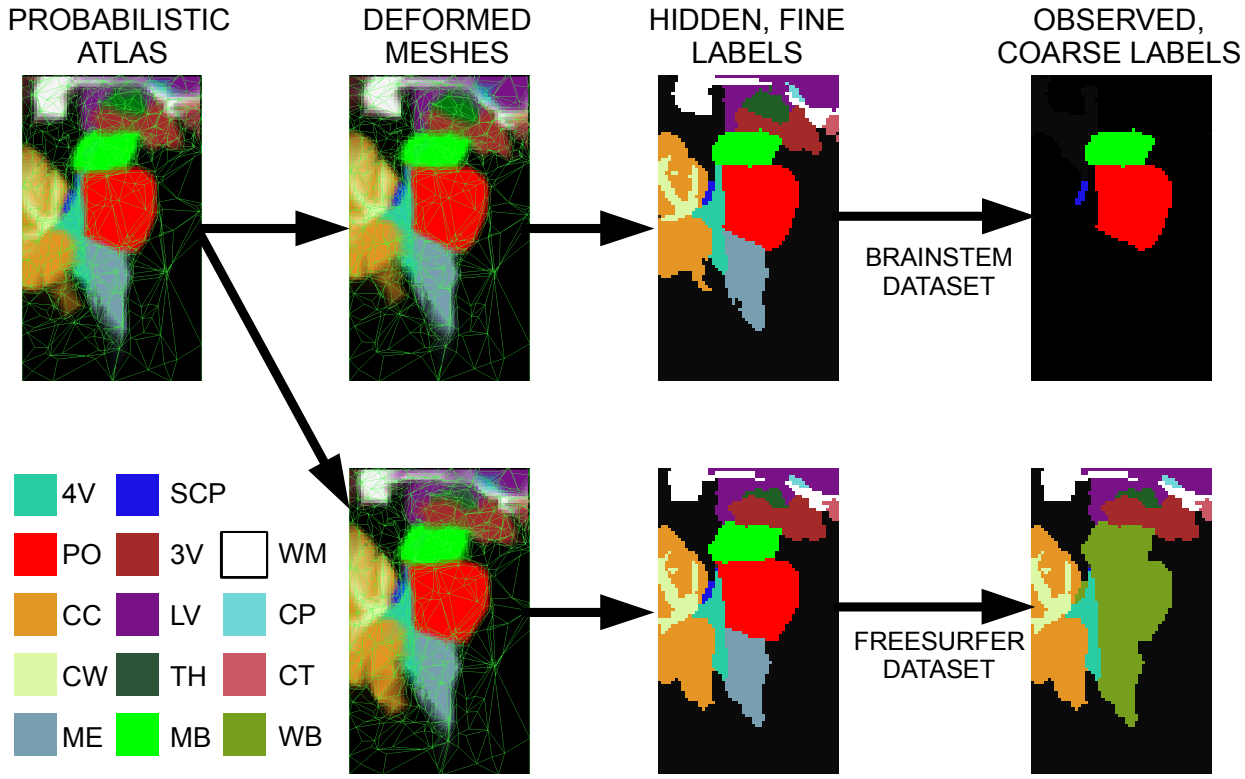

Figure 2: Generative model of training data. The abbreviations for the structures are the following: 4V = fourth ventricle, PO = pons, CC = cerebellar cortex, CW = cerebellar white matter, ME = medulla, SCP = superior cerebellar peduncle, 3V = third ventricle, LV = lateral ventricle, TH = thalamus, MB = midbrain, WM = white matter, CP = choroid plexus, CT = cortex, WB = whole brainstem. The background is represented in black.

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

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

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 anno-

tations 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 Appendix D). 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

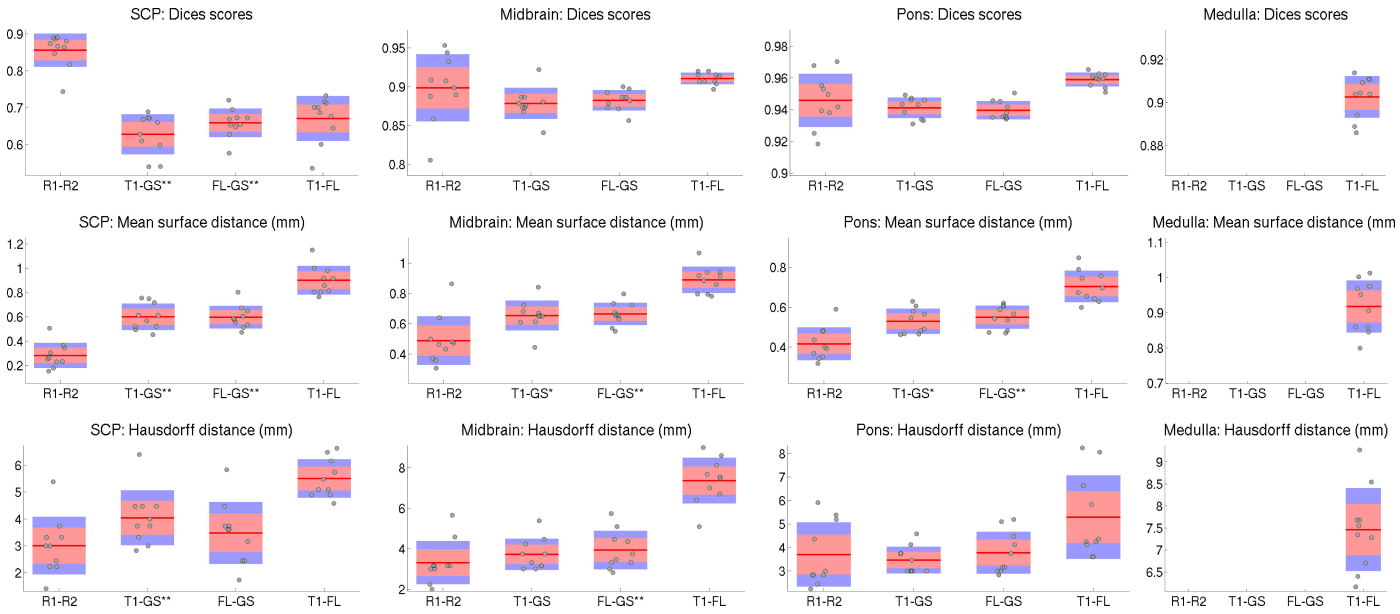

Figure 3: Box plots for the Dice overlap, symmetric mean surface-to-surface distance and symmetric Hausdorff (maximal) distance for the SCP, midbrain, pons and medulla. R1-R2 represents the agreement between the two human raters (inter-observer variability), T1-GS is between the T1 segmentation and the 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.

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.,

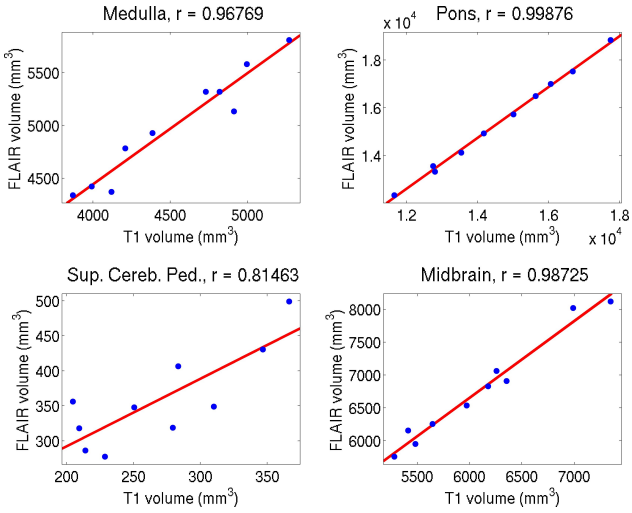

Figure 4: Scatter plots and linear fits for the volumes of the brainstem structures derived from the segmentations of the T1 and FLAIR scans of the brainstem dataset.

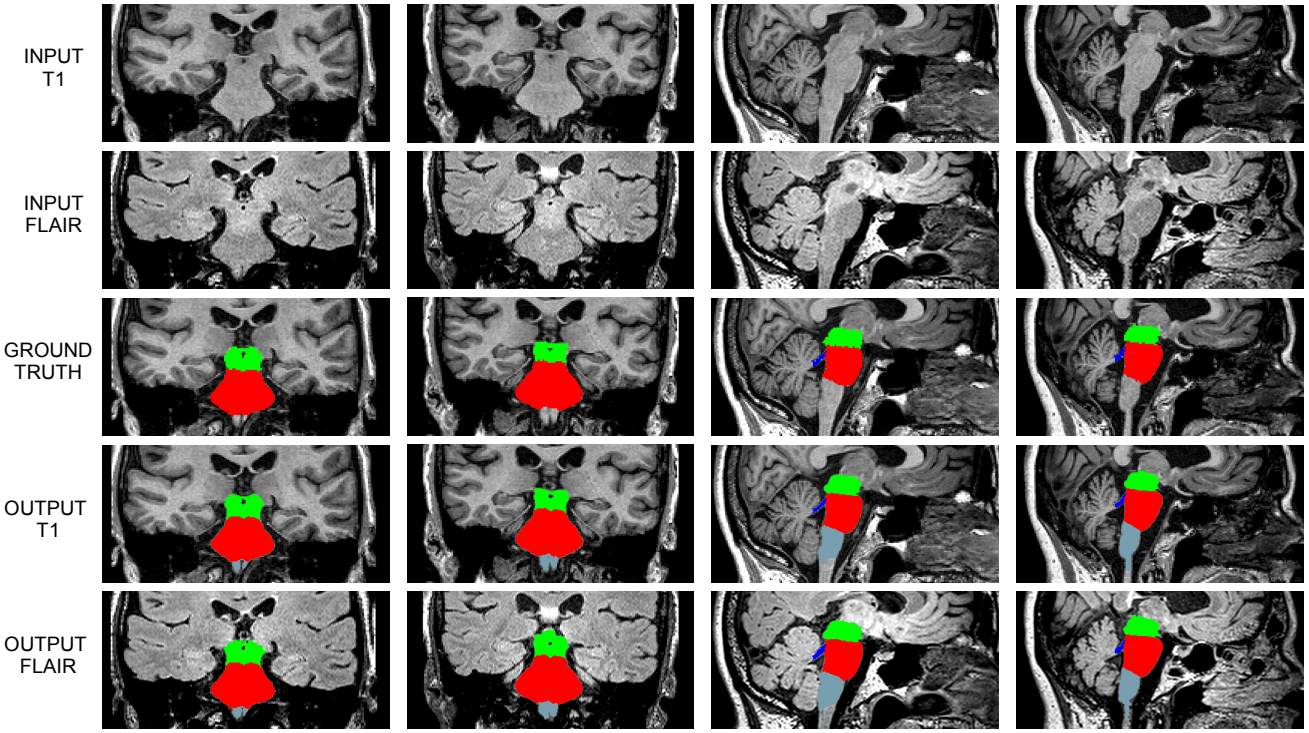

Figure 5: Sample slices (top two rows), manual delineations (middle row) and automated segmentations (bottom two rows) from the brainstem dataset. The color code is the following: red is pons, green is midbrain, blue is SCP, and gray is medulla. Note that there is no manual segmentations for the medulla.

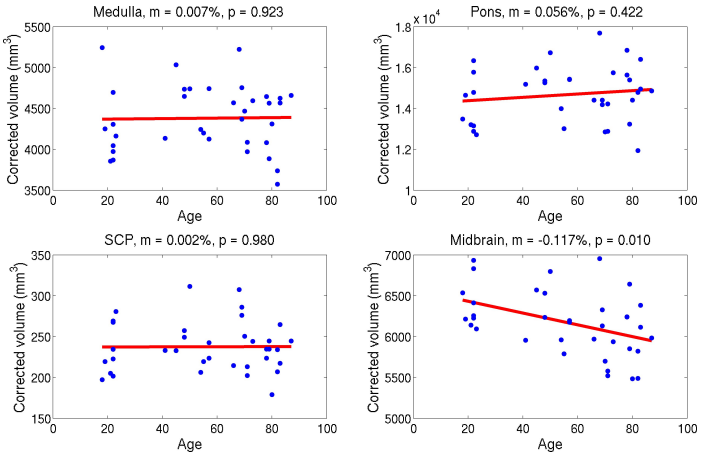

Figure 6: Scatter plots for the ICV-corrected volumes of the brainstem structure versus age (FreeSurfer dataset). The linear fit is superimposed. The p-value for the hypothesis that the slope of this fit is zero is displayed in the title of each subfigure, along with the yearly atrophy (in %).

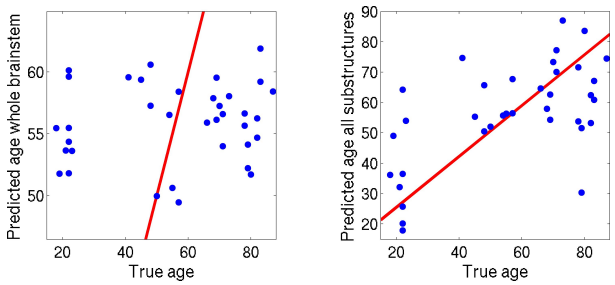

Figure 7: Scatter plots for real and predicted ages in the FreeSurfer dataset, using only the volume of the whole brainstem (left, $r = 0.14$) and the volumes of all the brainstem structures (right, $r = 0.60$).

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.

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

Table 1: Volumetric study of brainstem structures in ADNI: elderly controls vs. AD patients. 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 E.9 and E.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 E.2-E.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.

Acknowledgements

Support for this research was provided in part by the National Center for Research Resources (U24 RR021382, P41 RR14075, 1K L2RR025757-01), the National Institute for Biomedical Imaging and Bioengineering (P41 EB015896, R01 EB006758, R01 EB013565, 1K 25EB013649-01), the National Institute on Aging (AG022381, 5R01 AG008122-22, R01 AG016495-11, R01 AG038791), the National Center for Alternative Medicine (RC1 AT005728-01), the national Institute for Neurological Disorders and Stroke (R01 NS052585-01, 1R21 NS 072652-01, 1R01 NS070963, R01 NS083534, U54NS092089), the Tau Consortium, and was made possible by the resources provided by Shared Instrumentation Grants 1S10RR0 23401, 1S10RR0 19307, and 1S10RR0 23043. Additional support was provided by The Autism & Dyslexia Project funded by the Ellison Medical Foundation, and by the NIH Blueprint for Neuroscience Research (5U01 MH093765), part of the multi-institutional Human Connectome Project. This research was also funded by TEKES (ComBrain), Harvard Catalyst, and financial contributions from Harvard and affiliations. JEI was supported by the Gipuzkoako Foru Aldundia (Fellows Gipuzkoa Program). In addition, BF has a financial interest in CorticoMetrics, a company whose medical pursuits focus on brain imaging and measurement technologies. BF's interests were reviewed and are managed by

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.

Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies.

The collection and sharing of the MRI data used in the robustness experiment was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimers Association; Alzheimers Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Appendix A. 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.

Appendix B. 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.

Appendix C. 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.

Appendix D. 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 $|\cdot|$ 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] D. R. Williams, A. J. Lees, Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges, The Lancet Neurology 8 (3) (2009) 270–279.

[2] A. L. Boxer, A. E. Lang, M. Grossman, D. S. Knopman, B. L. Miller, L. S. Schneider, R. S. Doody, A. Lees, L. I. Golbe, D. R. Williams, et al., Davunetide in patients with progressive supranuclear palsy: a randomised, double-blind, placebo-controlled phase 2/3 trial, The Lancet Neurology.

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

[4] L. Grinberg, U. Rüb, R. Ferretti, R. Nitrini, J. Farfel, L. Polichiso, K. Gierga, W. Jacob-Filho, H. Heinsen, 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.

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

[6] S. Minoshima, K. A. Frey, N. L. Foster, D. E. Kuhl, Preserved pontine glucose metabolism in alzheimer disease: a reference region for functional brain image (pet) analysis., Journal of computer assisted tomography 19 (4) (1995) 541–547.

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

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

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

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

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

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

[13] B. Fischl, D. 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. Dale, Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain, Neuron 33 (2002) 341–355.

[14] B. Fischl, Freesurfer, Neuroimage 62 (2) (2012) 774–781.

[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] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, et al., Advances in functional and structural MR image analysis and implementation as fsl, Neuroimage 23 (2004) S208–S219.

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

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

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

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

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

[22] J. Ashburner, K. J. Friston, Unified segmentation, Neuroimage 26 (3) (2005) 839–851.

[23] K. M. Pohl, J. Fisher, W. E. L. Grimson, R. Kikinis, W. M. Wells, A bayesian model for joint segmentation and registration, NeuroImage 31 (1) (2006) 228–239.

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

[25] J. Ashburner, J. L. Andersson, K. J. Friston, Image registration using a symmetric priorin three dimensions, Human brain mapping 9 (4) (2000) 212–225.

[26] K. M. Pohl, J. Fisher, S. Bouix, M. Shenton, R. W. McCarley, W. E. L. Grimson, R. Kikinis, W. M. Wells, Using the logarithm of odds to define a vector space on probabilistic atlases, Medical Image Analysis 11 (5) (2007) 465–477.

[27] J. Ashburner, K. J. Friston, Computing average shaped tissue probability templates, Neuroimage 45 (2) (2009) 333–341.

[28] J. E. Iglesias, J. C. Augustinach, K. Nguyen, C. M. Player, A. Player, M. Wright, N. Roy, M. P. Frosch, A. C. McKee, L. L. Wald, B. Fischl, K. Van Leemput, A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI, Neuroimage (submitted).

[29] O. Puonti, J. E. Iglesias, K. Van Leemput, Fast, sequence adaptive parcellation of brain mr using parametric models, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, Springer, 2013, pp. 727–734.

[30] A. P. Dempster, N. M. Laird, D. B. Rubin, et al., Maximum likelihood from incomplete data via the em algorithm, J. R. Stat. Soc. 39 (1) (1977) 1–38.

[31] 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.

[32] J. G. Sled, A. P. Zijdenbos, A. C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, Medical Imaging, IEEE Transactions on 17 (1) (1998) 87–97.

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

[34] S. K. Warfield, K. H. Zou, W. M. Wells, Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation, Medical Imaging, IEEE Transactions on 23 (7) (2004) 903–921.

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

[36] R. Raininko, T. Autti, S.-L. Vanhanen, A. Ylikoski, T. Erkinjuntti, P. Santavuori, The normal brain stem from infancy to

old age, Neuroradiology 36 (5) (1994) 364–368.
[37] X.-L. Meng, R. Rosenthal, D. B. Rubin, Comparing correlated correlation coefficients., Psychological bulletin 111 (1) (1992) 172.

Appendix E. Supplementary material

Subj.	Group	Age	ICV	Med.	Pons	SCP	Midb.
0003	AD	81	1.913	5627	17349	261	7423
0005	EC	74	1.651	4913	15114	281	6104
0008	EC	85	1.396	4311	13216	215	5419
0010	AD	74	1.471	4647	16085	243	6336
0015	EC	81	1.512	4758	17216	333	6423
0019	EC	73	1.417	5182	15316	223	6048
0029	AD	64	1.905	5672	21405	394	7833
0031	EC	78	1.339	3460	12212	172	5191
0035	EC	77	1.495	4523	14595	236	6092
0040	EC	73	1.609	4676	14773	249	6184
0043	EC	76	1.674	4800	14304	303	5672
0047	EC	85	1.687	4483	15594	234	6647
0053	AD	80	1.753	5387	16019	311	6534
0055	EC	76	1.844	4045	15627	237	6962
0058	EC	70	1.433	4370	14900	258	6689
0066	EC	75	1.239	3617	12052	200	5215
0067	EC	75	1.527	4844	16611	282	6427
0068	EC	75	1.379	4343	13288	200	5357
0069	EC	73	1.730	4458	14418	241	6370
0070	EC	74	1.636	4931	16455	245	6512
0072	EC	71	1.552	4475	17147	267	6814
0074	EC	78	1.706	4554	14657	294	6295
0076	AD	78	1.523	4336	15411	332	6383
0081	EC	71	1.541	4199	13013	208	5848
0083	AD	73	1.515	4407	14039	247	6085
0084	AD	75	1.339	4196	13573	177	5287
0086	EC	80	1.431	4072	12271	277	5823
0088	AD	66	1.612	4410	13878	219	6497
0089	EC	65	1.568	4760	14849	245	6256
0090	EC	70	1.558	4893	17321	340	6695
0091	AD	62	1.365	4582	14632	260	6240
0093	AD	77	1.414	4247	15297	185	5753
0094	AD	71	1.565	3750	13651	198	5953
0095	EC	71	1.481	4773	15772	254	6346
0096	EC	80	1.815	4676	16349	235	6823
0097	EC	73	1.678	4809	15544	231	6643
0106	EC	73	1.806	4335	16122	218	6950
0109	AD	70	1.368	4268	12595	226	5269
0110	AD	83	1.698	4696	15973	270	5948
0112	EC	71	1.703	5353	17665	404	7562
0113	EC	75	1.760	4663	16665	364	7178
0118	EC	81	1.566	4417	12960	218	5395
0120	EC	72	1.560	5173	16122	317	6801
0123	EC	73	1.417	4186	12757	197	5184
0125	EC	73	1.856	5416	17348	264	7019
0127	EC	71	1.476	3934	12863	225	5616
0129	AD	80	1.342	3922	13864	215	5503
0130	EC	73	1.816	5487	18151	356	7008
0138	EC	86	1.664	4449	14739	248	6264
0139	AD	66	1.317	4649	13464	264	5697
0147	AD	60	1.685	4570	16369	245	6800
0149	AD	88	1.514	4559	14249	303	5508
0156	EC	74	1.614	4198	14475	234	6077
0159	EC	78	1.319	3745	14167	190	5656
0166	EC	76	1.452	3696	14213	203	5668
0168	EC	89	2.035	5106	18010	270	7606
0171	EC	78	1.404	4128	11213	198	5042
0172	EC	71	1.441	4005	13784	219	5388
0173	EC	73	1.649	4690	15264	336	6742
0177	EC	75	1.390	3919	13201	208	5353
0183	AD	73	1.532	4527	12761	160	6040
0184	EC	78	1.294	4792	15140	209	5466
0186	EC	81	1.593	4089	14776	285	5860
0188	EC	86	1.837	4958	16306	299	7028

Table E.2: List of ADNI subjects used in this study, along with their demographics (age, diagnosis) and automatically estimated volumes for the brainstem structures. All volumes are in cubic mm, except for the ICV, which is in liters. The ICV was estimated with FreeSurfer, whereas the volumes of the brainstem structures were estimated with the method presented in this article. EC stands for "elderly control."

Subj.	Group	Age	ICV	Med.	Pons	SCP	Midb.
0194	AD	81	1.297	3943	12500	182	4968
0196	EC	78	1.436	4146	12796	244	5227
0205	EC	67	1.655	4370	16167	332	6563
0210	EC	72	1.393	3857	13116	205	5280
0213	AD	63	1.600	4878	15845	259	6493
0216	AD	84	1.865	4875	15623	224	6848
0221	AD	68	1.776	4467	14191	194	6290
0228	AD	80	1.584	4403	14052	222	5533
0229	EC	74	1.346	3899	13196	174	5473
0230	EC	80	1.725	4841	15507	261	6640
0245	EC	74	1.527	4122	12430	242	5845
0257	EC	79	1.331	3702	12945	195	5139
0259	EC	76	1.420	4071	11814	216	5130
0260	EC	79	1.323	4408	13206	236	5794
0262	EC	86	1.468	5039	14483	233	6240
0266	AD	86	1.669	5092	16714	303	6384
0272	EC	71	1.591	4192	13661	251	5925
0283	EC	78	1.638	4216	15553	222	6718
0286	AD	66	1.869	5993	18149	361	7676
0295	EC	85	1.643	4511	15685	214	6444
0298	EC	77	1.501	4732	15908	214	6509
0299	AD	89	1.377	3525	13426	194	5573
0300	AD	56	1.847	4716	16402	290	6759
0301	EC	74	1.467	4218	15225	240	6236
0303	EC	84	1.505	4361	13800	257	5837
0304	EC	71	1.713	5508	16317	441	7175
0310	AD	76	1.521	5058	16974	331	6454
0311	EC	78	1.363	3791	11760	202	5273
0312	EC	83	1.709	5356	16401	295	6321
0315	EC	72	1.652	5066	19308	328	7323
0316	AD	76	1.618	4475	14353	224	6284
0319	EC	70	1.414	4680	15649	384	6282
0321	AD	68	1.438	4246	12649	218	5632
0327	EC	70	1.712	4563	14357	211	6487
0328	AD	77	1.576	4780	15417	333	6606
0335	AD	84	1.554	3877	12808	260	5759
0337	EC	76	1.827	4801	15888	283	6543
0341	AD	74	1.809	4981	14939	217	6689
0343	AD	72	1.802	5324	15298	267	6793
0352	EC	77	1.256	3332	10966	162	4830
0356	AD	80	1.702	4784	14176	185	6409
0359	EC	81	1.222	3587	11982	199	4998
0360	EC	73	1.656	4263	16813	291	7128
0363	EC	76	1.609	5221	14810	281	6457
0366	AD	57	1.287	4309	12944	260	5490
0369	EC	76	1.622	4809	13914	254	6676
0372	AD	79	1.725	4838	15980	238	6692
0374	AD	76	1.493	4996	16060	251	5846
0382	EC	76	1.545	4141	13586	237	5716
0384	EC	80	1.541	4838	17751	237	6750
0386	EC	72	1.522	4045	13446	230	6067
0392	AD	85	1.467	3987	13616	316	5528
0400	AD	69	1.720	5215	18582	295	7435
0403	EC	76	1.599	4779	16723	266	6457
0404	AD	88	1.290	3467	10990	204	4698
0405	EC	76	1.583	4403	15653	295	6161
0413	EC	76	1.610	4260	13885	285	5625
0416	EC	73	1.414	3979	11804	158	5575
0420	EC	74	1.504	5243	17334	230	6828
0422	EC	62	1.722	4539	15602	217	6578
0425	EC	86	1.716	5101	17184	265	7160
0426	AD	80	1.994	5607	16359	275	6761
0429	EC	63	1.800	5095	15646	460	6473
0431	AD	84	1.701	5066	16747	237	6311

Table E.3: List of ADNI subjects used in this study: continuation of Table E.2.

Subj.	Group	Age	ICV	Med.	Pons	SCP	Midb.
0433	EC	86	1.293	4290	16479	181	6066
0436	EC	85	1.680	5091	15099	293	6096
0438	AD	82	1.422	4516	15177	219	6711
0441	EC	73	1.473	3789	13526	218	5750
0443	EC	63	1.449	4622	15234	273	6120
0454	EC	82	1.447	4104	14156	199	5818
0457	AD	83	1.242	3771	12356	194	5313
0459	EC	73	1.594	4700	17653	352	6735
0467	EC	81	1.529	4161	14574	200	5712
0470	AD	87	1.480	4273	15467	202	5863
0472	EC	73	1.573	4353	15288	229	6152
0473	EC	73	1.657	4336	13327	217	5734
0474	AD	77	1.561	4225	12841	188	5454
0479	EC	74	1.502	5072	18812	247	7319
0484	EC	71	1.719	5416	16762	430	6977
0487	AD	77	1.469	4364	12526	265	5467
0488	EC	71	1.534	5121	17076	266	6669
0489	EC	74	1.229	3866	13016	185	4868
0492	AD	87	1.450	4323	14262	258	6152
0493	EC	78	1.759	6020	18048	235	7176
0497	AD	76	1.872	4928	16007	214	6805
0498	EC	71	1.621	5194	15556	256	6437
0500	EC	78	1.822	5170	18097	253	7774
0502	EC	75	1.629	4239	14358	252	6043
0506	EC	72	1.610	4450	15416	217	6419
0516	EC	88	1.637	4881	21728	417	7871
0517	AD	82	1.636	4380	14542	252	6205
0519	EC	74	1.548	4794	15570	346	6064
0520	EC	78	1.562	5097	17960	341	6545
0522	EC	70	1.757	4220	14695	241	6278
0525	EC	70	1.301	3833	13499	306	5664
0528	AD	71	1.334	3423	12644	199	5214
0534	EC	63	1.772	5078	16756	339	6551
0535	AD	77	1.671	4355	14568	219	6128
0538	EC	83	1.659	4853	16813	294	7370
0543	AD	72	1.227	4113	14063	192	5808
0545	EC	72	1.798	4581	16511	351	6949
0547	AD	76	1.597	3947	12037	233	5432
0551	EC	64	1.659	5156	17651	345	7100
0553	EC	85	1.602	4197	13751	263	5913
0555	EC	77	1.795	4863	16166	258	6504
0558	EC	80	1.688	5237	15787	301	6332
0559	EC	79	1.713	4821	17007	314	7336
0575	EC	87	1.471	4237	14817	254	5899
0576	EC	78	1.654	4285	13844	286	6468
0577	AD	72	1.406	4815	15912	285	6447
0578	EC	77	1.353	3860	11705	186	5280
0592	AD	78	1.607	5151	16082	306	6590
0601	EC	77	1.616	4284	13971	182	5865
0602	EC	71	1.461	5457	17224	280	6825
0605	EC	76	1.778	5339	17971	287	7219
0606	AD	69	1.440	4809	14600	198	5755
0610	EC	79	1.557	4457	14380	228	6133
0618	EC	75	1.514	4826	15622	240	6509
0619	AD	78	2.070	4759	15527	322	6985
0622	EC	75	1.409	3946	12767	238	5228
0627	AD	59	1.412	3900	14542	220	6117
0633	AD	83	1.374	4072	13254	186	5649
0637	EC	76	1.491	3949	12779	192	5457
0640	EC	73	1.450	4280	15653	241	6600
0642	AD	85	1.702	5434	18622	231	7136
0643	EC	71	1.531	4713	15798	284	6408
0647	EC	73	1.517	4182	14261	238	6713
0648	EC	72	1.692	5464	18111	351	6799

Table E.4: List of ADNI subjects used in this study: continuation of Table E.3.

Subj.	Group	Age	ICV	Med.	Pons	SCP	Midb.
0653	AD	74	1.455	4858	15424	249	5875
0657	EC	78	1.690	4690	16205	228	6778
0668	EC	74	1.523	5047	14449	243	6165
0669	EC	76	1.823	4823	16268	211	6596
0672	EC	62	1.424	4175	14303	253	5605
0677	EC	71	1.580	4423	15400	237	6466
0680	EC	78	1.455	4291	14397	338	6042
0681	EC	77	1.502	3970	13776	271	5888
0682	AD	68	1.421	5111	14764	266	6075
0685	EC	90	1.539	4385	14508	292	6422
0686	EC	72	1.831	4935	16323	151	7628
0689	AD	64	1.790	4930	16013	282	6708
0690	AD	76	1.572	4995	17304	272	6435
0691	AD	64	1.566	4699	14387	281	5740
0692	EC	77	1.503	4595	12845	196	5519
0702	EC	85	1.627	4197	11654	209	5577
0711	EC	77	1.597	4975	15017	256	6428
0712	AD	77	1.429	4049	12075	219	5227
0717	EC	76	1.793	4435	18181	270	7340
0720	AD	78	1.344	4778	14800	264	5864
0722	EC	70	1.897	4880	16042	372	6669
0724	AD	79	1.800	4833	15214	245	6684
0726	EC	81	1.604	4721	15114	277	6304
0730	AD	71	1.471	5294	16290	337	6382
0731	EC	72	1.501	4910	15952	394	6756
0733	AD	57	1.344	4575	13485	288	5967
0734	EC	73	1.671	4758	14623	269	6427
0740	AD	76	1.627	4784	15664	269	6485
0741	EC	80	1.535	4098	12303	234	5405
0751	EC	76	1.726	4182	13163	194	6218
0753	AD	66	1.770	4047	14559	264	5999
0754	AD	68	1.609	5595	15106	290	6341
0759	AD	82	1.442	4451	15747	245	5798
0760	AD	69	1.471	4138	13534	192	5573
0761	EC	71	1.449	4188	15394	253	6489
0767	EC	73	1.340	3921	15000	202	5539
0768	EC	77	1.799	5393	17601	282	7142
0778	EC	73	1.601	4100	13123	189	5383
0784	AD	76	1.395	4093	11085	173	5169
0786	AD	75	1.592	4156	14454	214	6206
0790	AD	83	1.351	4032	12009	171	5452
0793	AD	69	1.692	5546	16639	277	6836
0796	AD	80	1.893	4462	14581	301	6090
0803	AD	85	1.423	3962	13424	266	5792
0810	EC	84	1.669	4804	14123	271	6126
0812	AD	72	1.577	4418	14477	193	5940
0813	EC	73	1.659	4850	17073	311	6549
0814	AD	71	1.455	4056	12597	215	5455
0816	AD	71	1.618	5146	16625	357	6584
0818	EC	74	1.443	4970	15995	246	5916
0824	EC	76	1.518	4177	12587	312	5841
0828	AD	77	1.311	3743	12988	194	5176
0829	AD	65	1.318	4171	13762	213	5473
0836	AD	83	1.699	4501	18031	364	6328
0841	AD	76	1.323	3629	12244	222	5284
0842	EC	76	1.691	5205	16592	272	7091
0844	AD	85	1.552	4286	11273	213	5153
0845	EC	80	1.852	5690	17946	257	7305
0850	AD	78	1.458	4603	14195	269	5813
0852	AD	84	1.357	4108	14816	213	5841
0853	AD	87	1.553	4650	14700	209	6012
0862	EC	76	1.490	4553	14366	189	5817
0863	EC	79	1.652	4212	15319	253	6115
0866	EC	80	1.388	4278	16207	224	6051

Table E.5: List of ADNI subjects used in this study: continuation of Table E.4.

Subj.	Group	Age	ICV	Med.	Pons	SCP	Midb.
0883	EC	76	1.587	4886	14223	184	5721
0884	AD	79	1.449	4583	14258	249	5971
0886	EC	71	1.444	4500	13034	234	5404
0889	AD	73	1.450	3683	10201	214	4693
0891	AD	79	1.460	3710	12888	219	5510
0896	EC	77	1.789	4698	17880	290	6977
0898	EC	83	1.572	4295	14647	257	5873
0899	EC	80	1.363	4114	11527	224	5312
0907	EC	89	1.405	4967	14344	240	5521
0916	AD	80	1.537	4235	14228	209	6397
0920	EC	80	1.464	4821	15571	326	6433
0923	EC	83	1.571	5284	16709	200	6664
0926	EC	71	1.486	4035	15931	239	6102
0929	AD	82	1.591	5018	14481	226	5723
0931	EC	86	1.268	4074	13723	295	5467
0934	EC	70	1.397	4085	13920	234	5885
0938	AD	82	1.384	4373	13615	296	5643
0951	EC	85	1.690	5454	15299	233	6898
0955	AD	78	1.508	5324	14926	286	6583
0963	EC	73	1.533	4578	15582	403	6510
0969	EC	70	1.512	5134	15593	242	7027
0972	EC	78	1.477	4479	15354	265	6222
0979	AD	86	1.605	4740	14752	236	5804
0981	EC	84	1.388	3994	13761	192	5706
0984	EC	77	1.868	4338	14654	257	6669
0985	EC	82	1.354	4183	13467	235	5516
0991	AD	86	1.380	3962	12922	193	5185
0996	AD	91	1.392	3932	12116	224	5277
0999	AD	76	1.409	3920	11958	261	5666
1002	EC	76	1.622	3924	12535	216	5669
1016	EC	78	1.398	4466	15739	265	6203
1018	AD	71	1.522	5484	17119	248	6162
1021	EC	87	1.410	3450	11644	175	5349
1023	EC	77	1.392	3748	13239	215	5769
1024	AD	69	1.744	5265	16240	243	6678
1027	AD	69	1.323	5110	13812	235	5558
1035	EC	87	1.614	4816	17375	217	6446
1037	AD	74	1.785	4989	18034	301	6846
1041	AD	71	1.665	5364	17870	268	6965
1044	AD	66	1.609	4822	15770	317	6354
1055	AD	85	1.585	4573	16240	276	6754
1056	AD	71	1.481	4099	13525	156	5576
1059	AD	85	1.523	4349	13926	227	5416
1062	AD	82	1.656	4448	12095	205	5889
1079	AD	78	1.914	5174	17416	337	7404
1081	AD	85	1.564	4742	13834	202	5854
1082	AD	70	1.477	5112	14355	313	6022
1083	AD	72	1.484	3853	15755	200	6524
1086	EC	82	1.484	5049	12660	271	5735
1090	AD	71	1.715	5022	17010	254	6787
1092	EC	74	1.337	3854	14315	215	5729
1094	EC	76	1.774	4831	15138	252	6455
1095	AD	80	1.609	4600	15917	224	6417
1098	EC	72	1.341	3908	13551	180	5819
1101	AD	71	1.236	3730	10790	158	5159
1102	AD	73	1.269	3910	9836	141	4660
1109	AD	78	1.333	4136	12733	224	4944
1113	AD	79	1.540	4078	11856	220	5240
1123	EC	76	1.336	4156	12593	241	5210
1133	EC	80	1.385	3916	14859	243	5941
1137	AD	82	1.310	3755	13604	198	5413
1144	AD	78	1.605	4224	14061	243	5967
1152	AD	72	1.357	3775	10128	243	5243
1161	AD	80	1.640	4878	16327	205	6653

Table E.6: List of ADNI subjects used in this study: continuation of Table E.5.

Subj.	Group	Age	ICV	Med.	Pons	SCP	Midb.
1164	AD	70	1.329	4062	11685	234	5048
1168	EC	81	1.332	3541	12088	232	4976
1169	EC	72	1.566	4135	13527	275	5844
1170	AD	73	1.487	3968	13100	238	5487
1171	AD	72	1.811	5425	16595	273	6820
1184	AD	65	1.469	4087	11737	222	5097
1185	AD	62	1.627	4476	14400	325	5982
1188	EC	81	1.428	4382	15071	231	6379
1190	EC	77	1.425	3787	13092	282	5860
1191	EC	79	1.526	4567	12535	221	5773
1194	EC	85	1.741	5298	17123	266	6785
1195	EC	77	1.650	4298	13009	224	6031
1197	EC	82	1.407	4025	12892	208	5352
1200	EC	85	1.719	4421	16260	272	6891
1203	EC	83	1.419	3793	11410	231	5124
1205	AD	83	1.362	4392	14314	192	6008
1206	EC	73	1.747	5673	17982	253	6619
1209	AD	72	1.767	5169	18149	284	6879
1212	EC	75	1.559	4121	13437	226	5966
1221	AD	71	1.746	4456	13649	235	6473
1232	EC	72	1.431	5086	17392	250	6621
1245	EC	71	1.338	4070	13137	191	5299
1248	AD	80	1.415	4425	13983	211	6466
1249	EC	71	1.390	3997	14549	201	5812
1251	EC	74	1.374	4410	16565	259	6046
1253	AD	63	1.519	4908	15036	233	5962
1254	AD	84	1.653	4596	13872	258	6193
1256	EC	72	1.356	3876	12216	179	5446
1257	AD	85	1.959	5408	18128	229	7749
1261	EC	71	1.500	4513	13757	226	5460
1262	AD	73	1.385	4203	13438	203	5794
1263	AD	65	1.378	3702	12003	204	5074
1267	EC	73	1.746	4918	15377	275	6218
1276	EC	72	1.426	4334	13006	254	6020
1281	AD	78	1.420	4249	14449	218	5739
1283	AD	60	1.782	5116	14480	181	6635
1285	AD	80	1.625	4488	14921	242	6311
1286	EC	76	1.759	4707	14471	288	6339
1288	EC	60	1.676	5501	16578	321	7164
1289	AD	77	1.362	4762	16554	212	5901
1290	AD	79	1.440	4266	14089	249	5832
1296	AD	77	1.578	4542	14993	259	5907
1301	EC	72	1.631	5913	18437	358	6669
1304	AD	75	1.354	4232	12650	256	5977
1306	EC	76	1.321	3535	12278	240	5255
1307	AD	75	1.573	4769	14737	253	6296
1308	AD	80	1.601	4497	14713	293	6045
1334	AD	64	1.727	5028	16431	300	6475
1337	AD	71	1.639	4425	14746	208	5997
1339	AD	80	1.662	4605	15113	319	6414
1341	AD	72	1.307	4135	12717	163	5267
1368	AD	76	1.435	4259	12234	217	5283
1371	AD	84	1.541	4416	13627	329	6187
1373	AD	75	1.392	4220	13189	143	5797
1377	AD	83	1.952	4705	16207	275	6836
1379	AD	88	1.873	4230	13287	241	6006
1382	AD	64	1.943	5335	16278	447	6588
1391	AD	76	1.819	4500	13774	211	5775
1397	AD	76	1.537	4691	13749	221	5931
1402	AD	69	1.938	5178	18789	240	7607
1409	AD	66	1.807	5468	18930	288	7732
1430	AD	84	1.215	3685	12022	192	5351
1435	AD	82	1.572	4879	14922	287	6075

Table E.7: List of ADNI subjects used in this study: continuation of Table E.6.

Figure E.9: Segmentation of subjects 133-264 of the ADNI dataset. See caption of Figure 5 for the color code.

Figure E.10: Segmentation of subjects 265-383 of the ADNI dataset. See caption of Figure 5 for the color code.