

PBSI: A symmetric probabilistic extension of the Boundary Shift Integral

Christian Ledig^{1*}, Robin Wolz¹, Paul Aljabar^{1,2}, Jyrki Lötjönen³,
Daniel Rueckert¹

¹ Department of Computing, Imperial College London, London, UK

² Imaging Sciences and Biomedical Engineering, Kings College London, London, UK

³ VTT Technical Research Centre of Finland, Tampere, Finland

Abstract. The boundary shift integral (BSI) is a widely used technique to measure atrophy in dementia. In this work we extend the BSI method through the use of probabilistic posteriors obtained through a 4D expectation-maximization (EM) segmentation framework. Our method exploits the probabilistic information of these posteriors to calculate an accurate region of interest (ROI) on which the generalized BSI is evaluated. We present a complete framework which refines spatial probabilistic priors for a baseline and a follow-up magnetic resonance (MR) scan simultaneously and incorporates this spatial information in a probabilistic BSI (PBSI) measure. To ensure a consistent estimate, we follow a symmetric strategy by transforming baseline and follow-up images to a common, intermediate coordinate system. We use the resulting atrophy measure to compute group separation between normal subjects, patients with mild cognitive impairment (MCI) and patients with Alzheimer’s Disease (AD). We show the superiority of this approach over state-of-the-art methods on the ADNI cohort. We focus on a prediction of the conversion from MCI to AD, which is technically very challenging but clinically most useful. We present atrophy rates comparable to published manual rates and obtain a classification accuracy for separating stable and progressive MCI patient groups of 70.8%.

1 Introduction

The most common form of dementia is Alzheimer’s Disease (AD). While a conclusive diagnosis can only be obtained through pathological confirmation, hippocampal volume loss has been shown to be a good marker for the presence of AD [1]. Decline in hippocampal volume is also predictive at the stage of mild cognitive impairment (MCI) [1]. Accurate atrophy measurement based on MR images could thus assist the diagnosis of AD before a conventional clinical diagnosis is established or the assessment of the potential of disease-modifying therapies [2]. The anatomical structure of the hippocampus is complex. Manual labeling of the hippocampal region is thus time consuming [1] and leads to intra-

* This work is partially funded under the 7th Framework Programme by the European Commission (<http://cordis.europa.eu/ist/>).

and inter-rater variability. This motivates the need for accurate and automatic methods. For example, in [3] the scans of single timepoints are nonrigidly aligned with a probabilistic atlas which is then employed to segment the hippocampus. To quantify hippocampal atrophy automatically in longitudinal structural MR images several techniques have been proposed: These techniques are either based on the automatic segmentation of the hippocampal area, using for example graph cuts [4], or on registration-based methods such as the boundary shift integral (BSI) [1, 2].

In this work we present a model that exploits advantages of both segmentation- and registration-based approaches, by incorporating subject-specific probabilistic spatial information into the original BSI formulation [5]. Our method relies on probability maps that we enhance with a 4D expectation-maximization (EM) algorithm [6], which refines prior information based on the intensities of the images. We avoid bias towards either of the two timepoints by following a symmetric strategy [2, 7]. We evaluate the proposed method by comparing classification accuracies, atrophy rates and samples sizes on a subset of the ADNI cohort to a state-of-the-art method based on graph cuts [4].

2 Method

2.1 Notation and Prerequisites

In the following we consider a pair of T_1 -weighted MR images, \mathbf{I}^t and \mathbf{I}^{t+1} , of the same subject acquired at different timepoints. Let the n voxels in an MR image be indexed by $i = 1, \dots, n$. Denoting the intensity at a voxel as $y_i \in \mathbb{R}^m$ an image can be defined as $\mathbf{I}^t = \{y_1, y_2, \dots, y_n\}$. In this paper we focus on the single channel case ($m = 1$). For a given image \mathbf{I} the corresponding probabilistic segmentation is given as $\mathbf{S}_{\text{prob}} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$, where \mathbf{z}_i is a vector of size K and the k^{th} component represents the probability that a voxel belongs to a structural class k . Since hippocampal atrophy is a characteristic of AD, we focus in this work on the segmentation of the hippocampus. The component of \mathbf{S}_{prob} that is representing the hippocampus label is abbreviated, for a certain timepoint and thus for an image \mathbf{I}^t , as p^t . We denote the rigid transformation from the coordinate system of \mathbf{I}^t to that of \mathbf{I}^{t+1} as $\mathbf{R}_{\mathbf{I}^t, \mathbf{I}^{t+1}}$. Parameters are denoted as lower case Greek letters and sets with Ω .

We assume that individual brain masks are available to skull-strip all baseline images, \mathbf{I}^t , and that subject-specific probabilistic priors for the hippocampal region in the baseline scan, p_{prior}^t , exist. These priors can be calculated with, for example, multi-atlas propagation techniques such as LEAP [8]. Subsequent timepoints \mathbf{I}^{t+1} can be skull-stripped using a rigid transformation $\mathbf{R}_{\mathbf{I}^t, \mathbf{I}^{t+1}}$ of the baseline masks. All skull-stripped images are intensity normalized, since the BSI is defined on intensity differences on normalized images [5].

2.2 Symmetrizing the process

In order to process both timepoints simultaneously we transform each pair of baseline and follow up images to their common intermediate rigid space. This

is important to reduce bias introduced by asymmetric interpolation [7] or by randomly defining one of the coordinate systems, of \mathbf{I}^t or \mathbf{I}^{t+1} , as reference coordinate system [2]. The symmetrizing is carried out using the rigid transformation $\mathbf{R}_{\mathbf{I}^t, \mathbf{I}^{t+1}}^{0.5}$ obtained by taking the square root of the transformation matrix of $\mathbf{R}_{\mathbf{I}^t, \mathbf{I}^{t+1}}$. Probabilistic spatial hippocampus priors, which are required for each baseline scan, are then also rigidly mapped to this common space. Following this symmetric approach both baseline and follow-up image are mapped and resliced in the same space, which reduces bias towards one coordinate system [2, 7].

2.3 Posteriors through 4D EM optimization

The challenge is to estimate the underlying but unknown probabilistic segmentations $\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}$ by means of the observed intensities y_i^t, y_i^{t+1} and the available subject-specific probabilistic priors p_{prior}^t and p_{prior}^{t+1} . We adopt the common assumption [6] that the observed log-transformed intensities of voxels belonging to a certain class k are normally distributed with mean μ_k and standard deviation σ_k . The overall parameters for this intensity model are thus $\Phi = \{(\mu_1, \sigma_1), (\mu_2, \sigma_2), \dots, (\mu_K, \sigma_K)\}$. By assuming the same parameters Φ for both timepoints, we model and segment both timepoints simultaneously in a 4D EM framework. Smoothness within the segmentations is enforced with a global and stationary Markov Random Field (MRF) [6].

Our model thus consists of one Gaussian intensity distribution (μ_1, σ_1) , that models the intensity distribution within the hippocampus at both timepoints, and $K - 1$ Gaussian distributions modeling the intensities in the background. Prior spatial information is incorporated by a thresholded version of the available subject-specific soft segmentation of the hippocampus, where we use the same prior for both timepoints. We thus have:

$$p_{\text{prior}}^t(i) = p_{\text{prior}}^{t+1}(i) = \begin{cases} \alpha, & \text{if } p_{\text{prior}}^t(i) > \alpha \\ 0, & \text{if } p_{\text{prior}}^t(i) < \beta \\ p_{\text{prior}}^t(i), & \text{else} \end{cases} \quad (1)$$

A mask on which the 4D EM refinement is carried out, can be obtained by dilating a region defined by the spatial priors p_{prior}^t and p_{prior}^{t+1} .

2.4 Probabilistic Boundary Shift Integral

We calculate the volume change Δv between a structure in the follow-up scan \mathbf{I}^{t+1} and the corresponding structure in the baseline scan \mathbf{I}^t based on a modified version of the boundary shift integral (BSI), which was proposed by Freeborough et al. [5]. To incorporate the spatial information provided by the probabilistic segmentations $\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}$ into the measure, we define a probabilistic boundary shift integral (PSBI). We introduce a novel spatially dependent weighting function $\gamma(i) \equiv \gamma(\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}, i)$ that can be defined to adapt the model to

the application. The PBSI is calculated on a region of interest (ROI) Ω covering the boundary of the hippocampus to surrounding tissue. We define the PBSI as:

$$\Delta v = \frac{v_{\text{voxel}}}{\xi_{\text{high}} - \xi_{\text{low}}} \sum_{i \in \Omega} \gamma(i) (\text{clip}(\mathbf{I}^{t+1}(i), \xi_{\text{low}}, \xi_{\text{high}}) - \text{clip}(\mathbf{I}^t(i), \xi_{\text{low}}, \xi_{\text{high}})) \quad (2)$$

where the thresholds ξ_{low} and ξ_{high} are to be defined such that ξ_{low} represents intensities outside and ξ_{high} inside the hippocampus respectively. The clipping function is defined as:

$$\text{clip}(y_i, \xi_{\text{low}}, \xi_{\text{high}}) = \begin{cases} \xi_{\text{low}}, & \text{if } y_i < \xi_{\text{low}} \\ y_i, & \text{if } \xi_{\text{low}} \leq y_i \leq \xi_{\text{high}} \\ \xi_{\text{high}}, & \text{if } y_i > \xi_{\text{high}} \end{cases} \quad (3)$$

Since we have probabilistic segmentations ($\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}$) of the hippocampus in baseline and follow-up scans available, the ROI Ω can be obtained completely automatically. In order to calculate Ω , we generalize the definition given in [5] and compute Ω_{PBSI} based on probability maps. We use the fuzzy union and intersection of two probabilistic maps:

$$\cup(p_i, p_j) := \max(p_i, p_j) \quad \cap(p_i, p_j) := \min(p_i, p_j) \quad (4)$$

and the binarized versions depending on thresholds η and ζ :

$$\cap_{\eta}(p_i, p_j) := \begin{cases} 1, & \text{if } \eta \leq \cap(p_i, p_j) \\ 0, & \text{else} \end{cases} \quad \cup_{\zeta}(p_i, p_j) := \begin{cases} 1, & \text{if } \zeta \leq \cup(p_i, p_j) \\ 0, & \text{else} \end{cases} \quad (5)$$

The ROI Ω_{PBSI} is then given as the difference, $\bar{\Omega} \setminus \check{\Omega}$, of a region $\bar{\Omega}$ describing the outer boundary of the ROI and an interior region lying completely within the hippocampus $\check{\Omega}$.

$$\check{\Omega} = \varepsilon^{n_e} \left[\cap_{\eta}(\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}) \right] \quad \bar{\Omega} = \varrho^{n_d} \left[\cup_{\zeta}(\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}) \right] \quad (6)$$

The operators ε^{n_e} and ϱ^{n_d} denote the n_e - and n_d -times application of the erosion and dilation operators. A common choice is $n_e = n_d = 1$ for both the erosion and dilation operator [5]. In our model we can also achieve and control the erosion by varying η for $\check{\Omega}$. The model also allows to control the calculation of Ω via the parameter ζ . Fig. 1 shows an example of a ROI, computed based on either hard segmentations (Ω_{BSI}) or the proposed method (Ω_{PBSI}).

3 Experimental Results

We evaluated our method on 377 subjects from the Alzheimer’s Disease Neuroimaging Initiative (ADNI,adni.loni.ucla.edu) database [9], where we segmented the left hippocampus. The cohort consisted of the 105 healthy controls (CN), 189 subjects with MCI and 83 AD patients for which a 12 month follow-up

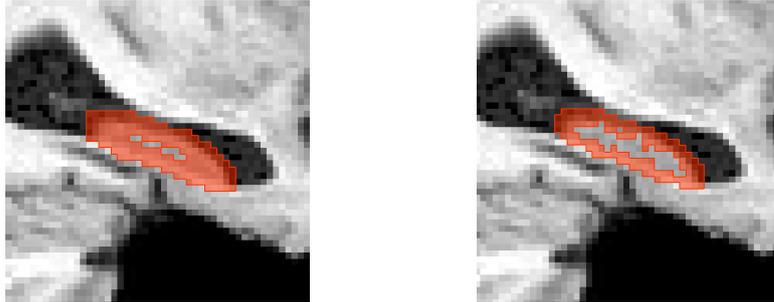


Fig. 1. ROI (red) for the calculation of BSI. Based on hard labels (Ω_{BSI} , left) and based on the proposed method with probabilistic labels (Ω_{PBSI} , right) [$\eta = 0.95$, $n_e = 0$ and $\zeta = 0.90$, $n_d = 1$].

scan, spatial priors [8], brain masks for the baseline scans [10] and results from a reference method [4] were available. The MCI group is further divided into 110 non-converting, stable (sMCI) and 79 converting, progressive MCI patients (pMCI) respectively. The main focus was on classification accuracy, but atrophy rate and sample size calculations were also performed.

In the following, we compare five different approaches against volume changes computed on the left hippocampus with a state-of-the-art method based on graph cuts (‘4D GC’) [4]. We denote the atrophy computation based on label volumes obtained by the 4D EM segmentation as ‘4D EM’. This approach is expected to perform similar to 4D GC since it represents also an intensity based refinement of spatial priors computed with LEAP [8]. Next to these we calculated atrophy rates based on the standard BSI, with a ROI based on hard labels (Ω_{BSI} , Fig. 1), and different versions of the PBSI with ROIs either Ω_{BSI} or Ω_{PBSI} . For this we defined γ as either $\gamma \equiv 1$ (‘PBSI₁’), which represents the standard BSI with the ROI determined according to the proposed method, or $\gamma = \max(0.5, p_{\text{prob}}^t, p_{\text{prob}}^{t+1})$ (‘PBSI _{γ} ’). The latter version decreases the weight of voxels that do not clearly represent hippocampal tissue.

3.1 Preprocessing Strategy and Parameters

For skull-stripping we employed brain masks obtained by a semi-automatic procedure as described in [10]. We then computed robustly intensity-scaled versions of the T_1 MR images and mapped the baseline \mathbf{I}^t , the 12 month follow-up scans \mathbf{I}^{t+1} and the probabilistic priors obtained through LEAP [8] to a common space as described in Section 2.2.

We thresholded the spatial prior for the hippocampus with parameters $\alpha = 0.8$ and $\beta = 0$. We consider an upper threshold of 0.8 as reasonable since it slightly relaxes the priors and thus allows non-hippocampal tissue in the follow-up scan to change the label. The MRF parameter controlling the connectivity over time was set to 0, since we do not want to penalize the detection of a decrease in label volume over time. Since different tissue types, including cerebrospinal fluid,

white matter and non-hippocampal gray matter, are expected within the region of interest we set the parameter $K = 4$ and thus model the background with 3 different distributions. The spatial priors for the 3 different background classes were built in a manner based on the probability fraction left undefined by the hippocampal prior and intensities of the images. For a maximum intensity value of 255, the lower and upper bounds for the clipping function were set to $\xi_{\text{low}} = 75$ and $\xi_{\text{high}} = 125$. As parameters for erosion and dilation we used $\eta = 0.95$, $n_e = 0$ and $\zeta = 0.90$, $n_d = 1$ respectively. We chose parameters without explicit training and kept them fixed for all evaluations.

3.2 Classification Accuracy

We employed a bootstrapping approach, as used in [4], to separate clinical groups based on atrophy rates. We randomly selected 75% of the subjects for each group and classified the remaining 25% based on their difference from the mean estimated on the training sets. Specificity, sensitivity and accuracy averaged over 1000 runs are shown in Table 1. As expected, classification rates for the 4D GC and the 4D EM approach are comparable. For the sMCI-pMCI comparison, 4D GC performs 1.8% better than the EM approach. When directly comparing these intensity-based approaches, one needs to consider that the preprocessing procedures were different while the same spatial LEAP-priors were used. All BSI based methods clearly outperform 4D GC and 4D EM in sMCI-pMCI classification. For classification we observed best results for $\text{PBSI}_{\gamma}^{\Omega_{\text{BSI}}}$ which is evaluated on a comparably rough ROI based on hard labels but incorporating spatial hippocampal information. It should be pointed out that the standard BSI is evaluated on a larger, potentially a too large, ROI Ω_{BSI} and thus not necessarily based on hippocampal structure alone.

Method	Statistics	CN-pMCI	CN-AD	sMCI-pMCI	sMCI-AD
4D GC	Spec./Sens.	69.9/65.4	75.8/68.5	69.4/62.0	74.2/65.8
	Accuracy	68.0	72.6	66.3	70.6
4D EM	Spec./Sens.	68.7/63.0	76.8/66.8	66.7/61.3	73.2/64.7
	Accuracy	66.3	72.4	64.5	69.6
$\text{BSI}^{\Omega_{\text{BSI}}}$	Spec./Sens.	70.8/71.4	75.4/69.1	71.5/69.6	76.1/63.1
	Accuracy	71.1	72.7	70.7	70.5
$\text{PBSI}_1^{\Omega_{\text{PBSI}}}$	Spec./Sens.	70.5/68.6	74.1/69.7	73.1/66.4	77.1/64.1
	Accuracy	69.7	72.2	70.3	71.6
$\text{PBSI}_{\gamma}^{\Omega_{\text{BSI}}}$	Spec./Sens.	72.7/70.7	77.1/66.5	74.3/66.0	80.6/61.6
	Accuracy	71.9	72.5	70.8	72.5
$\text{PBSI}_{\gamma}^{\Omega_{\text{PBSI}}}$	Spec./Sens.	72.7/67.1	76.5/64.5	75.3/61.5	78.9/58.4
	Accuracy	70.3	71.3	69.6	70.2

Table 1. Classification results based on the left hippocampus in % obtained using a 25% leave-one-out strategy for separating AD groups obtained with 4D GC, 4D EM, $\text{BSI}^{\Omega_{\text{BSI}}}$, $\text{PBSI}_1^{\Omega_{\text{PBSI}}}$, $\text{PBSI}_{\gamma}^{\Omega_{\text{BSI}}}$ and $\text{PBSI}_{\gamma}^{\Omega_{\text{PBSI}}}$.

3.3 Atrophy rates

Mean and standard deviation of the atrophy rates used for classification are shown in Table 2. Since volume loss depends on different factors, such as age, it is difficult to accurately predict atrophy rates [3]. However, compared to other methods and rates based on manual segmentations, as for example [1], PBSI_1 especially BSI seem to consistently overestimate atrophy rates. This may be due to the fact that without considering spatial information BSI potentially includes too many voxels at the boundaries, especially in the standard BSI where the ROI is defined using hard labels. Thus a decrease in intensity might contribute to the measured atrophy rate without considering if a voxel belongs to the hippocampus at all. In the presented framework we are able to calculate more realistic rates since we relax this problem by incorporating spatial information. This is confirmed by the atrophy rates presented for PBSI_γ .

		4D GC	4D EM	$\text{BSI}^{\Omega_{\text{BSI}}}$	$\text{PBSI}_1^{\Omega_{\text{PBSI}}}$	$\text{PBSI}_\gamma^{\Omega_{\text{BSI}}}$	$\text{PBSI}_\gamma^{\Omega_{\text{PBSI}}}$
Atrophy rates	Normal	-1.1 (2.0)	-2.0 (2.9)	-2.4 (5.2)	-1.9 (4.3)	-1.3 (3.1)	-1.1 (2.9)
	sMCI	-1.7 (2.3)	-2.3 (2.8)	-4.0 (6.1)	-3.1 (4.7)	-2.3 (3.7)	-1.9 (3.1)
	pMCI	-3.1 (2.4)	-4.0 (2.9)	-7.5 (6.3)	-6.0 (5.2)	-4.5 (4.0)	-3.7 (3.4)
	AD	-3.7 (2.5)	-4.9 (4.1)	-8.9 (8.0)	-7.2 (6.5)	-5.5 (5.1)	-4.6 (4.3)
Sample sizes	MCI (uncor.)	297	240	349	354	376	393
	MCI (CN-cor.)	1190	2007	1105	1117	1100	1102
	AD (uncor.)	110	176	203	208	209	217
	AD (CN-cor.)	228	504	378	381	366	368

Table 2. Mean atrophy rates of the left hippocampus in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom) for 4D GC, 4D EM, $\text{BSI}^{\Omega_{\text{BSI}}}$, $\text{PBSI}_1^{\Omega_{\text{PBSI}}}$, $\text{PBSI}_\gamma^{\Omega_{\text{BSI}}}$ and $\text{PBSI}_\gamma^{\Omega_{\text{PBSI}}}$. Standard deviation in brackets. Corrected sample sizes were computed on the excess change over normal aging.

3.4 Sample Sizes

In a hypothetical two-arm study, the required study population to measure a specified change in atrophy rate is of particular interest. We present estimated sample sizes to detect a 25% change in atrophy rate (effect size $\Delta = 0.25\mu$) for the MCI (pMCI and sMCI) and AD patient group in Table 2. We selected a power $(1 - \beta)$ of 0.8 ($z_{1-\beta} \approx 0.84$) and significance α of 5% ($z_{1-0.05/2} \approx 1.96$). Following recent communication in the neuroimaging community [11], discussing the importance of relating atrophy rates in dementia to normal atrophy during aging, we provide sample sizes that are both corrected and uncorrected for normal aging. We observed comparable corrected sample sizes for MCI for all evaluated methods, so that no significant difference can be assumed. Sample sizes for AD are smallest for 4D GC.

4 Conclusion

We presented a symmetric probabilistic extension of the boundary shift integral (PBSI) and employed it to classify 377 AD patients grouped by different clinical diagnoses. We observed an increased accuracy of up to 4.5% for the classification of stable versus progressive MCI groups, which shows the superiority of our model that incorporates prior spatial information. In the presented experiment we showed accuracies of over 70% for this clinically highly relevant prediction of MCI conversion. Relative volume changes and sample sizes of the proposed method seem reasonable and are comparable to published atrophy rates. Our model offers a spatially dependent weighting function γ that allows to incorporate prior information into the integral computation. While the more accurate determination of the ROI seems of less importance, we believe that incorporation of spatial information using γ might be the key to robust atrophy measurement. In future it will be very interesting to investigate the influence of parameters and to employ this approach for atrophy measurement on other brain structures.

References

1. Leung, K.K., Barnes, J., Ridgway, G.R., et al.: Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. In: *NeuroImage*, 51(4), pp. 1345–1359, 2010.
2. Leung, K.K., Ridgway, G.R., Ourselin, S., et al.: Consistent multi-time-point brain atrophy estimation from the boundary shift integral. In: *NeuroImage*, 59(4), pp. 3995–4005, 2012.
3. Chupin, M., Gerardin, E., Cuingnet, R., et al.: Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. In: *Hippocampus*, 19(6), pp. 579–587, 2009.
4. Wolz, R., Heckemann, R.A., Aljabar, P., et al.: Measurement of hippocampal atrophy using 4D graph-cut segmentation: Application to ADNI. In: *NeuroImage*, 52(1), pp. 109–118, 2010.
5. Freeborough, P.A., Fox, N.C.: The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. In: *IEEE TMI*, 16(5), pp. 623–629, 1997.
6. Van Leemput, K., Maes, F., Vandermeulen, D., et al.: Automated model-based tissue classification of MR images of the brain. In: *IEEE TMI*, 18(10), pp. 897–908, 1999.
7. Yushkevich, P.A., Avants, B.B., Das, S.R., et al.: Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: An illustration in ADNI 3 T MRI data. In: *NeuroImage*, 50(2), pp. 434–445, 2010.
8. Wolz, R., Aljabar, P., Hajnal, J.V., et al.: LEAP: Learning embeddings for atlas propagation. In: *NeuroImage*, 49(2), pp. 1316–1325, 2010.
9. Mueller, S.G., Weiner, M.W., Thal, L.J., et al.: The Alzheimer's Disease Neuroimaging Initiative. In: *Neuroimaging Clinics of North America*, 15(4), pp. 869–877, 2005.
10. Leung, K.K., Barnes, J., Modat, M., et al.: Brain MAPS: An automated, accurate and robust brain extraction technique using a template library. In: *NeuroImage*, 55(3), pp. 1091–1108, 2011.
11. Fox, N.C., Ridgway, G.R., Schott, J.M.: Algorithms, atrophy and Alzheimer's disease: Cautionary tales for clinical trials. In: *NeuroImage*, 57(1), pp. 15–18, 2011.