SEGMA: an automatic SEGMentation Approach for human brain MRI using sliding window and random forests

SEGMA: an automatic SEGMentation Approach for human brain MRI using sliding window and random forests

Ahmed Serag[1*], Alastair G. Wilkinson[2], Emma J. Telford[1], Rozalia Pataky[1], Sarah A. Sparrow[1], Devasuda Anblagan[1], Gillian Macnaught[3], Scott Semple[3], James P. Boardman[1]

[1]MRC Centre for Reproductive Health, United Kingdom, [2]Royal Hospital for Sick Children, United Kingdom, [3]Centre for Cardiovascular Science, United Kingdom

Submitted to Journal:
Frontiers in Neuroinformatics

Article type:
Original Research Article

Manuscript ID:
228014

Received on:
31 Aug 2016

Revised on:
22 Nov 2016

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

A.S. designed and performed the experiments, and wrote the manuscript; A.S., J.P.B. and A.G.W. analysed output data; E.J.T., R.P. and S.A.S. recruited patients; G.M. and S.I.S. acquired imaging data. All authors approved the final submitted version, and agree to be accountable for its content.

Keywords

Brain, MRI, large-scale, Life-course, sliding window, Random forests, Classification, Tissue segmentation, Neonatal period, childhood and adolescence, adulthood

Abstract

Word count: 157

Quantitative volumes from brain magnetic resonance imaging (MRI) acquired across the life course may be useful for investigating long term effects of risk and resilience factors for brain development and healthy ageing, and for understanding early life determinants of adult brain structure. Therefore, there is an increasing need for automated segmentation tools that can be applied to images acquired at different life stages. We developed an automatic segmentation method for human brain MRI, where a sliding window approach and a multi-class random forest classifier were applied to high-dimensional feature vectors for accurate segmentation. The method performed well on brain MRI data acquired from 179 individuals, analysed in three age groups: newborns (38-42 weeks gestational age), children and adolescents (4-17 years) and adults (35-71 years). As the method can learn from partially labelled datasets, it can be used to segment large-scale datasets efficiently. It could also be applied to different populations and imaging modalities across the life course.

Funding statement

This work was supported by the Theirworld (http://www.theirworld.org), NHS Research Scotland, and NHS Lothian Research and Development. This work was undertaken in the MRC Centre for Reproductive Health which is funded by the MRC Centre grant MR/N022556/1.

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: Yes

Please provide the complete ethics statement for your manuscript. Note that the statement will be directly added to the manuscript file for peer-review, and should include the following information:

- Full name of the ethics committee that approved the study
- Consent procedure used for human participants or for animal owners
- Any additional considerations of the study in cases where vulnerable populations were involved, for example minors, persons with disabilities or endangered animal species

As per the Frontiers authors guidelines, you are required to use the following format for statements involving human subjects:
This study was carried out in accordance with the recommendations of 'name of guidelines, name of committee' with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the 'name of committee'.
For statements involving animal subjects, please use:
This study was carried out in accordance with the recommendations of 'name of guidelines, name of committee'. The protocol

was approved by the 'name of committee'.

If the study was exempt from one or more of the above requirements, please provide a statement with the reason for the exemption(s).
Ensure that your statement is phrased in a complete way, with clear and concise sentences.

Ethical approval was granted by the National Research Ethics Service (South East Scotland Research Ethics Committee) and NHS Research and Development.

Informed written parental consent was obtained.

Informed written parental consent was obtained.

SEGMA: an automatic SEGMentation Approach for human brain MRI using sliding window and random forests

A. Serag[1*], A.G. Wilkinson[2], E.J. Telford[1], R. Pataky[1], S.A. Sparrow[1], D. Anblagan[1,4], G. Macnaught[3], S.I. Semple[3,5], J.P. Boardman[1,4]

[1] MRC Centre for Reproductive Health, University of Edinburgh, Edinburgh, UK
[2] Department of Radiology, Royal Hospital for Sick Children, Edinburgh, UK
[3] Clinical Research Imaging Centre, University of Edinburgh, Edinburgh, UK
[4] Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK
[5] Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, UK

[*]**Correspondence:**
Ahmed Serag
Room W1.25 Queen's Medical Research Institute
47 Little France Crescent
Edinburgh EH16 4TJ
ahmed.serag@ed.ac.uk

Keywords: brain, MRI, large-scale, sliding window, classification

Abstract

Quantitative volumes from brain magnetic resonance imaging (MRI) acquired across the life course may be useful for investigating long term effects of risk and resilience factors for brain development and healthy ageing, and for understanding early life determinants of adult brain structure. Therefore, there is an increasing need for automated segmentation tools that can be applied to images acquired at different life stages. We developed an automatic segmentation method for human brain MRI, where a sliding window approach and a multi-class random forest classifier were applied to high-dimensional feature vectors for accurate segmentation. The method performed well on brain MRI data acquired from 179 individuals, analysed in three age groups: newborns (38-42 weeks gestational age), children and adolescents (4-17 years) and adults (35-71 years). As the method can learn from partially labelled datasets, it can be used to segment large-scale datasets efficiently. It could also be applied to different populations and imaging modalities across the life course.

1. Introduction

During early life, the brain undergoes significant morphological and functional changes, the integrity of which determines long-term neurological, cognitive and psychiatric functions (Tamnes et al., 2013). For instance, a wide range of problems including autism spectrum disorder, poor cognitive ageing, stroke and

neurodegenerative diseases of adulthood may have early life origins (Stoner et al., 2014, McGurn et al., 2008, Shenkin et al., 2009, Wardlaw et al., 2011, Hill et al., 2010). Improved understanding of cerebral structural changes across the life course may be useful for studying early life determinants and atypical trajectories that underlie these common problems.

Quantitative volumes from brain structural magnetic resonance imaging (MRI) acquired at different stages of life offer the possibility of new insight into cerebral phenotypes of disease, biomarkers for evaluating treatment protocols, and improved clinical decision-making and diagnosis. The literature presents a clear distinction between methods developed for different ages partly because the computational task is determined by properties of the acquired data and these are age-dependent (Cabezas et al., 2011, Despotovic et al., 2015, Isgum et al., 2015). For example, the infant brain presents challenges to automated segmentation algorithms developed for adult brain due to: wide variations in head size and shape in early life, rapid changes in tissue contrast associated with myelination, decreases in brain water, changes in tissue density, and relatively low contrast to noise ratio between grey matter (GM) and white matter (WM). Therefore, automated segmentation tools for modelling structure over years are limited, and this hampers research that would benefit from robust assessment of the newborn to the adult trajectory.

With regard to methodology, approaches for automatic segmentation of brain MRI can be classified into unsupervised (Gui et al., 2012, Leroy et al., 2011, Cai et al., 2007, Weglinski and Fabijanska, 2011) or supervised (Weisenfeld and Warfield, 2009, Wang et al., 2015, Cardoso et al., 2013, Cherel et al., 2015, Moeskops et al., 2015, Ashburner and Friston, 2005, Van Leemput et al., 2001, Fischl et al., 2002, Makropoulos et al., 2012, Serag et al., 2012b), (Kuklisova-Murgasova et al., 2011, Prastawa et al., 2005, Shi et al., 2010, Song et al., 2007, Altaye et al., 2008, Loh et al., 2015) approaches. Supervised approaches have proven to be very successful in medical image segmentation (Aljabar et al., 2009, Coupe et al., 2011, Lotjonen et al., 2010, Rousseau et al., 2011, Wang et al., 2013). However, as they rely on labelled training data (or atlases) to infer the labels of a test scan, most existing supervised approaches require a large number of training datasets to provide a reasonable level of accuracy and they usually carry a high computation cost due to their requirement of non-linear registrations between labelled data and the test scan (Iglesias and Sabuncu, 2015).

To address these challenges, here we describe a method for automatic brain segmentation of MR images, called **SEGMA** (**SEGM**entation **A**pproach). **SEGMA** differs from current supervised approaches in the following ways. First, SEGMA uses a sparsity-based technique for training data selection by selecting training data samples that are 'uniformly' distributed in the low-dimensional data space, and hence eliminates the need for target-specific training data (Serag et al., 2016). Second, SEGMA uses linear registration to provide an accurate segmentation (mainly to

ensure the same orientation and size for all subjects). This is useful because it reduces computation time compared with most supervised methods which require non-linear registrations between the training images and the target image. Finally, SEGMA uses a machine learning classification based on random forests (Breiman, 2001) where a class label for a given test voxel is determined based on its high-dimensional feature representation. In addition to incorporating more information into the feature set (compared with methods that use voxel intensity information only), we use a sliding window technique that moves over all positions in the test image and classifies all voxels inside the window at once, instead of assigning labels on a voxel by voxel basis. This technique has the advantage of speeding-up the classification process while minimising misclassifications compared with methods that use a global classifier (Vovk et al., 2011, Iglesias et al., 2011, Zikic et al., 2014). The feature extraction framework is illustrated in Fig. 1.

2. Materials and methods

2.1 Data and image acquisition

The study includes brain imaging data from 179 subjects, spanning the ages of 0–71 years, from three MRI datasets.

Dataset I. The first dataset contained MR images from 66 infants: 56 preterms (mean post-menstrual age [PMA] at birth 29.23 weeks, range 23.28–34.84 weeks) were acquired at term equivalent age (mean PMA 39.84 weeks, range 38.00–42.71 weeks), and 10 healthy infants born at full term (> 37 weeks' PMA). None of the infants had focal parenchymal cystic lesions. Participants of the newborns dataset were recruited to a larger study using MRI to study the effect of preterm birth on brain growth and long-term outcome. Ethical approval was granted by the National Research Ethics Service (South East Scotland Research Ethics Committee) and NHS Research and Development, and informed written parental consent was obtained.

A Siemens Magnetom Verio 3T MRI clinical scanner (Siemens Healthcare GmbH, Erlangen, Germany) and 12-channel phased-array head coil were used to acquire: [1] T1-weighted (T1w) 3D MPRAGE: TR = 1650 ms, TE = 2.43 ms, inversion time = 160 ms, flip angle = 9 degrees, acquisition plane = sagittal, voxel size = 1 × 1 × 1 mm^3, FOV = 256 mm, acquired matrix = 256 × 256, acceleration factor (iPAT) = 2; [2] T2-weighted [T2w] SPACE STIR: TR = 3800 ms, TE = 194 ms, flip angle = 120 degrees, acquisition plane = sagittal, voxel size = 0.9 × 0.9 × 0.9 mm^3, FOV = 220 mm, acquired matrix = 256 × 218. The image data used in this manuscript are available from the BRAINS repository (Job et al., 2016) (http://www.brainsimagebank.ac.uk).

Reference tissue segmentations for the dataset were generated using an Expectation-Maximization algorithm with tissue priors provided by the atlas from (Serag et al.,

2012a, Serag et al., 2012c). Ground truth accuracy of reference neonatal segmentations was evaluated by a radiologist experienced in neonatal brain MRI, who concluded that they were all plausible representations of anatomical classes. Quantitative evaluation of the reference segmentations was performed against manual segmentations from 9 subjects chosen at random. For each subject, three slices (those numbered as 25th percentile, median and 75th percentile of the slices containing brain tissue) were segmented. In order to remove bias towards any particular anatomical plane, three subjects were segmented in the axial plane, three in the coronal plane, and three in the sagittal plane. The quantitative analyses indicated high agreement for all tissues (mean Dice coefficient of 92%).

Dataset II. The second dataset contained T1w MRI scans and corresponding manual expert segmentation of 32 structures from 103 subjects (mean age 11.24 years, range 4.20-16.90 years) publicly available from the Child and Adolescent NeuroDevelopment Initiative (CANDI) at University of Massachusetts Medical School (Kennedy et al., 2012, Frazier et al., 2008) (http://www.nitrc.org/projects/candi_share). The data originates from four diagnostic groups: healthy controls (N = 29), schizophrenia spectrum (N = 20), Bipolar Disorder (N = 35), and Bipolar Disorder with psychosis (N = 19). The T1w images were acquired using a 1.5T Signa scanner (GE Medical Systems, Milwaukee, USA) with the following parameters: a three-dimensional inversion recovery-prepared spoiled gradient recalled echo coronal series, number of slices = 124, prep = 300 ms, TE = 1 min, flip angle = 25 degrees, FOV = 240 mm^2, slice thickness = 1.5 mm, acquisition matrix = 256 × 192, number of excitations=2.

Dataset III. The third dataset contained brain images and the corresponding manual expert segmentation of the whole brain into 32 structures from 18 healthy subjects including both adults and children; for the current study, we used only the adult data (N = 10, mean age 38, range 35-71 years). The dataset is publicly available from the Internet Brain Segmentation Repository (http://www.cma.mgh.harvard.edu/ibsr/) as IBSR v2.0(Rohlfing, 2012). The T1w images were acquired using the following parameters: scanner/scan parameters unspecified, acquisition plane = sagittal, number of slices = 128, FOV = 256 × 256 mm, voxel size = 0.8-1.0 × 0.8-1.0 × 1.5 mm^3.

2.2 Preprocessing

For brain extraction, we used the brain masks which are provided with each dataset; except dataset I which was brain extracted using ALFA (Serag et al., 2016). All images from all datasets were corrected for intensity inhomogeneity using the N4 method (Tustison et al., 2010).

2.3 Training data

The number of training examples often must be limited due to the costs associated with procuring, preparing and storing the training examples, and the computational costs associated with learning from them (Weiss and Provost, 2003). Therefore, we use in this work a sparsity-based technique to select a number of representative atlas images that capture population variability by determining a subset of *n*-dimensional samples that are 'uniformly' distributed in the low-dimensional data space (Serag et al., 2016). The technique works by first linearly registering (12 degrees of freedom) all images from each dataset to an appropriate common coordinate space, and image intensities are normalised using the method described by (Nyul and Udupa, 2000). For dataset I, the 40 weeks PMA template from the 4D atlas (Serag et al., 2012a) was used as the common space, which is the closest age-matched template to the mean age of the cohort, while datasets II and III were aligned to the common space defined by the International Consortium for Brain Mapping (ICBM) atlas(Mazziotta et al., 2001). Then, all N aligned images are considered as candidates for the subset of selected atlases. The closest image to the mean of the dataset is included as the first subset image. The consecutive images are selected sequentially, based on the distances to the images already assigned to the subset. Further details can be found in (Serag et al., 2016).

2.4 Features

We use machine learning to assign a label to all voxels in the test image, based on training a local classifier. Most existing methods for tissue classification only utilise information from voxel intensity, without considering other information. Here, in addition to voxel intensities, we incorporated various gradient-based features. Typically for each voxel v, a ten-dimensional feature vector $\mathbf{f}_v$ is extracted:

$$\mathbf{f}_v = \begin{bmatrix} I & I_x & I_y & I_z & r & \theta & \phi & I_{xx} & I_{yy} & I_{zz} \end{bmatrix}^T \qquad (1)$$

where I is the grey scale intensity value, I_x, I_y and I_z are the norms of the first order derivatives, and I_{xx}, I_{yy} and I_{zz} are the norms of the second order derivatives. The image derivatives are calculated through the filters $[-1 \quad 0 \quad 1]^T$ and $[-1 \quad 2 \quad -1]^T$. The gradient magnitude (r), azimuth angle (θ) and zenith angle (ϕ) are defined as follows:

$$r = \sqrt{I_x^2 + I_y^2 + I_z^2} \qquad (2)$$

$$\theta = \tan^{-1}\left(\frac{I_y}{I_x}\right) \qquad (3)$$

$$\phi = \cos^{-1}\left(\frac{I_z}{r}\right) \qquad (4)$$

where $r \in [0, \infty), \theta \in [0, 2\pi)$, and $\phi \in [0, \pi]$.

2.5 Random forests

In the last decade, random forests (RF) (Breiman, 2001) became a popular ensemble learning algorithm, as they achieve state-of-the-art performance in numerous medical applications (Chen et al., 2010, Geremia et al., 2011, Mitra et al., 2014, Pereira et al., 2016, Tustison et al., 2015, Zikic et al., 2014, Yi et al., 2009). A RF ensemble classifier consists of multiple decision trees. In order to grow these ensembles, often random vectors are generated that govern the growth of each tree in the ensemble. Typically, each tree is trained by combining "bagging" (Breiman, 1996) (where a random selection is made from the examples in the training set) and random selection of a subset of features (Ho, 1998), which construct a collection of decision trees exhibiting controlled variation.

A test sample is pushed down to every decision tree of the random forest. When the sample ends up in one leaf node, the label of the training sample of that node it is assigned to the test sample as tree decision. Then, the final predicted class for a test sample is obtained by combining, in a voting procedure, the predictions of all individual trees. More details on decision forests for computer vision and medical image analysis can be found in (Criminisi and Shotton, 2013).

2.6 Sliding-window based classification

A sliding window is used to move over all possible positions in the test image, and for each window, the voxels inside the window are classified into different tissues or structures. The vector in equation (1) represents the test sample for one voxel in a window, where the number of test samples is equal to the window size w. The training samples come from the voxels of the aligned atlas images that are located at the same location as the voxels belonging to the test window. This means that the number of training samples per window is equal to $k \times w$, where k is the number of training atlases and w is the window size, e.g. $5 \times 5 \times 5$, or $7 \times 7 \times 7$, etc.

A local RF classifier is then used to assign each voxel in the test image to a segmentation class. Figure 2 shows an example of classifying one test window. The SEGMA algorithm is summarised in Algorithm 1.

Algorithm 1. SEGMA algorithm

Set $\mathbf{f}_v$ to represent a feature vector for a voxel v
Set c_v to represent a segmentation class for a voxel v
Set k to represent the number of training data
Set w to represent the sliding window size
for each window W **do**
    Construct the training data matrix $\mathcal{T}_W^{Train} = \{\mathbf{f}_v^j | j = 1, \dots, k; v = 1, \dots, w\}$
    Train the RF_W classifier for window W using $\mathcal{T}_W^{Train}$
    Construct the test data matrix $\mathcal{T}_W^{Test} = \{\mathbf{f}_v | v = 1, \dots, w\}$
    Determine the labels c_v for all voxels inside the test window W by applying RF_W to $\mathcal{T}_W^{Test}$
end

2.6 Evaluation

A leave-one-out cross-validation procedure was performed for every dataset. Each subject from a dataset in turn was left out as a test sample and the remaining subjects were used as the training data where a subset of k atlases is selected. The comparison between automatic (A) and reference (M) segmentations was performed using the Dice coefficient (DC) which measures the extent of spatial overlap between two binary images, with range 0 (no overlap) to 1 (perfect agreement). The Dice values are expressed as a percentage and obtained using the following equation:

$$DC(A, M) = \frac{2|A \cap M|}{|A| + |M|} \times 100 \qquad (5)$$

2.7 Comparison against other methods

We compared SEGMA against commonly used segmentation methods: Majority Vote (MV) (Heckemann et al., 2006, Rohlfing et al., 2004), Simultaneous Truth And Performance Level Estimation (STAPLE) (Warfield et al., 2004). The registration scheme for these methods is based on non-linear image deformation (Rueckert et al., 1999, Modat et al., 2010).

To compare SEGMA against other RF segmentation methods, we implemented a global RF classifier, similar to (Iglesias et al., 2011, Zikic et al., 2014), and experimented training it using intensity and gradient-based features, and intensity feature only. Non-linear registration was used as above to map the training images to the test image coordinate space, and a RF classifier was trained using 100,000 randomly sampled voxels from each training image.

2.8 Statistical analyses

To test for differences between segmentation results, t-tests were used for normally distributed data, and Mann Whitney U was used to compare non-normal distributions (Shapiro-Wilk normality test was used). *P*-values < 0.05 were considered significant after controlling for Type I error using false discovery rate (FDR).

3 Results

To evaluate segmentation performance across the life course, SEGMA was applied to three publicly available datasets that provide MR brain images at different stages of the life course: neonatal period (38-42 weeks gestational age), childhood and adolescence (4-17 years), and adulthood (35-71 years). Figure 3 shows examples of brain segmentation results across the life course, and Fig. 4 shows the resulting Dice coefficient (i.e. the agreement between the automatic and reference segmentations).

3.1 Brain segmentation in neonatal period

We first applied the proposed segmentation method to a neonatal cohort (dataset I) consisting of 66 MR images and associated segmentation of the following tissues / structures: brainstem, cerebellum, cortex or GM, cerebrospinal fluid (CSF), deep GM and WM. Quantitative analyses (Fig. 4) indicated high accuracy for all tissues and structures with a mean Dice coefficient of 91%.

The highest accuracies obtained for brainstem, cerebellum, deep GM, and WM with mean Dice coefficient of 90–94%, while cortex and CSF had average Dice coefficients of 89% and 85%, respectively.

3.2 Brain segmentation in childhood and adolescence

To examine the performance of SEGMA in childhood and adolescence, we used 103 MR images from subjects aged 4-17 years (dataset II) with associated anatomical segmentation of 32 structures. Quantitative analyses (Fig. 4) indicated high accuracy for all tissues and structures with a mean Dice coefficient of 86%. Nine structures had an average Dice coefficient higher than 90%, 7 structures had an average Dice coefficient of 79–89%, and 2 structures had an average Dice coefficient of 51–67%.

3.3 Brains segmentation in adulthood

A dataset (dataset III) consisting of MR images and corresponding anatomical segmentation of 32 structures from 10 subjects (aged 38–71 years) was used to examine the performance of the segmentation algorithm in adulthood. Quantitative analyses (Fig. 4) indicated high accuracy of 83%. Seven structures had an average Dice coefficient higher than 90%, 9 structures had an average Dice coefficient of 75–89%, and 2 structures had an average Dice coefficient of 49–57%.

3.4 Comparison against other methods

SEGMA was compared with two commonly used segmentation methods [Majority Vote (MV) (Heckemann et al., 2006, Rohlfing et al., 2004), Simultaneous Truth And Performance Level Estimation (STAPLE) (Warfield et al., 2004)], and other RF-based segmentation methods. SEGMA improved overall segmentation accuracy compared with MV, STAPLE, global-RF-1 (trained using intensity and gradient features), and global- RF- 2 (trained using intensity feature only); Table 1 shows Dice coefficients averaged over all structures, generated by each segmentation method and applied to datasets I, II and III. (P<0.001; after FDR correction).

3.5 Reproducibility

As dataset I (neonatal period) included T1-weighted (T1w) and T2-weighetd (T2w) MR imaging, we used it to test the reproducibility of SEGMA across different MR modalities by segmenting the newborn brain using information from T1w and T2w data separately (Fig. 5). SEGMA provided consistent segmentation results across different structural MRI modalities of the newborn brain. There was no statistically significant difference between mean Dice scores estimated from the two groups (P = 0.8977).

3.6 Influence of parameters

We evaluated the influence of size of training data on segmentation accuracy, and found that increasing the size of the training data improves segmentation accuracy, evidenced by the increase in average Dice coefficient from 88% (7% training data) to 91% (30% training data) for neonates, and from 83% (5% training data) to 86% (20% training data) for children and adolescents. From our experiments, 5-10 training images were sufficient to yield accurate results.

Forest parameters such as tree depth and number of samples per leaf node were set according to pervious work (Wang et al., 2015, Zikic et al., 2014, Geremia et al., 2011), and in this work, we only evaluated the influence of number of trees on segmentation accuracy. The number of trees in the forest characterizes the generalization power. As the number of trees becomes large, segmentation accuracy increases, but training time increases and a threshold value is reached after which further improvement is not achieved. In this work, number of trees was set to 10.

With regard to window size, the smaller the window, the longer the classification time. Hence, window size needs to be chosen carefully as it provides a balance between accuracy and speed. Therefore, in this paper, we select the window size as $5 \times 5 \times 5$.

3.7 Relative importance of features

As partial volume effects in neonatal brain MRI present challenges for automatic segmentation methods, we evaluated the influence of each of the features on segmentation accuracy of the neonatal brain (dataset I). This was done by dropping one or a group of the ten features and running segmentation with the remaining features (features of the same type were dropped together). Therefore, an approximation of relative importance of each feature was obtained. Our experiments show that dropping the intensity feature significantly hinders the segmentation accuracy (Fig. 6a), whilst the accuracy is improved by incorporating gradient-based features. When all of the features are used, SEGMA yielded higher accuracy than each individual category ($P<0.001$; after FDR correction). Figure 6b also shows an example of the automatic neonatal cortical GM segmentation and how the dropping of each of the ten features affects the segmentation accuracy.

We then analysed the edge detection for various regions based on using all features (intensity combined with gradients) and grey scale intensity only. Figure 7 shows that gradient-based features improved edge detection for various regions of the adult and neonatal brain.

3.8 Computation time

One classification task on a 64-bit iMac® (Intel® Core i7 @ 3.5GHz x 4, 32 GB RAM) takes 5-7 minutes. The classification has benefited much from the sliding window strategy used. This is because instead of performing the classification in a voxel-wise manner, this is done for a batch of voxels at once. Assuming a window size of $5 \times 5 \times 5$, the classification time is decreased by 125 folds. In addition, multi-core processing or computer clusters could greatly enhance the speed; and then one brain classification could be performed in about (or less than) 1min.

4. Discussion

In this article, we present a new method for MRI brain segmentation (**SEGM**entation **Approach**, SEGMA). SEGMA was evaluated on three different datasets (span the ages 0–71 years) that provide different challenges to the brain segmentation task, and accurate results were obtained at all stages of development.

The method is trained using partially labelled datasets where a relatively small number of manually labelled images from the population under study are sufficient to provide accurate results. It is possible that training the method with a larger dataset might increase the segmentation accuracy. However, our goal was to design a methodology that can provide an acceptable, yet high accuracy result using a small number of training images (and thence a low computation cost).

The relatively lower performance for CSF could be caused by its bordering with GM (which is a complex shape). The boundary between GM and CSF is especially difficult to identify inside the sulci, where it is often poorly visible. In addition, the relatively lower performance for the children and adolescence, and adult datasets compared with the neonatal dataset could be attributable to scanner strength. Yet, the results obtained are comparable with those obtained using other methods tested on the same datasets (Rousseau et al., 2011, Zikic et al., 2014).

SEGMA uses a local RF classifier (trained by information from neighbouring voxels in the same window) to assign a label to each voxel, which makes it less susceptible to classification errors such as the partial volume misclassification on the CSF-GM and CSF-background boundaries (Cardoso et al., 2013, Isgum et al., 2015, Kuklisova-Murgasova et al., 2011, Moeskops et al., 2015). We chose to use random forests as the classification technique since they naturally handle multi-class classification problems and are accurate and fast (Chen et al., 2010, Criminisi and Shotton, 2013, Geremia et al., 2011). Also, the sliding window plays an important role in significantly speeding up the classification task (compared to voxel-wise approaches).

The method provides an accurate segmentation using only linear registration, which ensures the same orientation and size for all subjects. This is an advantage compared with most supervised methods, which require non-linear registrations between the training images and the test image which increases segmentation time to several hours thereby compromising clinical utility (Iglesias and Sabuncu, 2015). SEGMA also has the advantage of providing an accurate segmentation using a single modality (which is important as the available data might be limited to one modality), and features that characterise object appearance and shape (intensity and gradients). However, the method is flexible and new features can easily be added to the high-dimensional feature vector.

To conclude, we present a method for segmentation of human brain MRI that is robust and provides accurate and consistent results across different age groups and modalities. As SEGMA can learn from partially labelled datasets, it can be used to segment large-scale datasets efficiently. The idea of SEGMA is generic and could be applied to different populations and imaging modalities across the life course. SEGMA is available to the research community at http://brainsquare.org.

5. Author contributions

A.S. designed and performed the experiments, and wrote the manuscript; A.S., J.P.B. and A.G.W. analysed output data; E.J.T., R.P. and S.A.S. recruited patients; G.M. and S.I.S. acquired imaging data. All authors approved the final submitted version, and agreed to be accountable for its content.

6. Funding

This work was supported by the Theirworld (http://www.theirworld.org), NHS Research Scotland, and NHS Lothian Research and Development. This work was undertaken in the MRC Centre for Reproductive Health which is funded by the MRC Centre grant MR/N022556/1.

7. Acknowledgments

We are grateful to the families who consented to take part in the study and to the nursing and radiography staff at the Clinical Research Imaging Centre, University of Edinburgh (http://www.cric.ed.ac.uk) who participated in scanning the infants.

8. Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

9. References

ALJABAR, P., HECKEMANN, R. A., HAMMERS, A., HAJNAL, J. V. & RUECKERT, D. 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage,* 46**,** 726-38.

ALTAYE, M., HOLLAND, S. K., WILKE, M. & GASER, C. 2008. Infant brain probability templates for MRI segmentation and normalization. *Neuroimage,* 43**,** 721-30.

ASHBURNER, J. & FRISTON, K. J. 2005. Unified segmentation. *Neuroimage,* 26**,** 839-51.

BREIMAN, L. 1996. Bagging predictors. *Mach Learn,* 24**,** 123-140.

BREIMAN, L. 2001. Random Forests. *Mach Learn,* 45**,** 5-32.

CABEZAS, M., OLIVER, A., LLADO, X., FREIXENET, J. & CUADRA, M. B. 2011. A review of atlas-based segmentation for magnetic resonance brain images. *Comput Methods Programs Biomed,* 104**,** e158-77.

CAI, W. L., CHEN, S. C. & ZHANG, D. Q. 2007. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Pattern Recogn,* 40**,** 825-838.

CARDOSO, M. J., MELBOURNE, A., KENDALL, G. S., MODAT, M., ROBERTSON, N. J., MARLOW, N. & OURSELIN, S. 2013. AdaPT: An adaptive preterm segmentation algorithm for neonatal brain MRI. *Neuroimage,* 65**,** 97-108.

CHEN, H., XIAOQING, D. & CHI, F. 2010. Head Pose Estimation Based on Random Forests for Multiclass Classification. *20th International Conference on Pattern Recognition (ICPR)***,** 934-937.

CHEREL, M., BUDIN, F., PRASTAWA, M., GERIG, G., LEE, K., BUSS, C., LYALL, A., CONSING, K. Z. & STYNER, M. 2015. Automatic Tissue Segmentation of Neonate Brain MR Images with Subject-specific Atlases. *Proc SPIE Int Soc Opt Eng,* 9413.

COUPE, P., MANJON, J. V., FONOV, V., PRUESSNER, J., ROBLES, M. & COLLINS, D. L. 2011. Patch-based segmentation using expert priors:

Application to hippocampus and ventricle segmentation. *Neuroimage,* 54**,** 940-954.
CRIMINISI, A. & SHOTTON, J. 2013. Decision forests for computer vision and medical image analysis*, London; New York, Springer*.
DESPOTOVIC, I., GOOSSENS, B. & PHILIPS, W. 2015. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. *Comput Math Methods Med,* 2015**,** 450341.
DICE, L. R. 1945. Measures of the Amount of Ecologic Association Between Species. *Ecology,* 26**,** 297-302.
FISCHL, B., SALAT, D. H., BUSA, E., ALBERT, M., DIETERICH, M., HASELGROVE, C., VAN DER KOUWE, A., KILLIANY, R., KENNEDY, D., KLAVENESS, S., MONTILLO, A., MAKRIS, N., ROSEN, B. & DALE, A. M. 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron,* 33**,** 341-55.
FRAZIER, J. A., HODGE, S. M., BREEZE, J. L., GIULIANO, A. J., TERRY, J. E., MOORE, C. M., KENNEDY, D. N., LOPEZ-LARSON, M. P., CAVINESS, V. S., SEIDMAN, L. J., ZABLOTSKY, B. & MAKRIS, N. 2008. Diagnostic and sex effects on limbic volumes in early-onset bipolar disorder and schizophrenia. *Schizophr Bull,* 34**,** 37-46.
GEREMIA, E., CLATZ, O., MENZE, B. H., KONUKOGLU, E., CRIMINISI, A. & AYACHE, N. 2011. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *Neuroimage,* 57**,** 378-90.
GUI, L., LISOWSKI, R., FAUNDEZ, T., HUPPI, P. S., LAZEYRAS, F. & KOCHER, M. 2012. Morphology-driven automatic segmentation of MR images of the neonatal brain. *Med Image Anal,* 16**,** 1565-79.
HECKEMANN, R. A., HAJNAL, J. V., ALJABAR, P., RUECKERT, D. & HAMMERS, A. 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage,* 33**,** 115-26.
HILL, J., DIERKER, D., NEIL, J., INDER, T., KNUTSEN, A., HARWELL, J., COALSON, T. & VAN ESSEN, D. 2010. A surface-based analysis of hemispheric asymmetries and folding of cerebral cortex in term-born human infants. *J Neurosci,* 30**,** 2268-76.
HO, T. K. 1998. The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal,* 20**,** 832-844.
IGLESIAS, J. E., LIU, C. Y., THOMPSON, P. M. & TU, Z. 2011. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Trans Med Imaging,* 30**,** 1617-34.
IGLESIAS, J. E. & SABUNCU, M. R. 2015. Multi-atlas segmentation of biomedical images: A survey. *Med Image Anal,* 24**,** 205-219.
ISGUM, I., BENDERS, M. J., AVANTS, B., CARDOSO, M. J., COUNSELL, S. J., GOMEZ, E. F., GUI, L., HUPPI, P. S., KERSBERGEN, K. J., MAKROPOULOS, A., MELBOURNE, A., MOESKOPS, P., MOL, C. P., KUKLISOVA-MURGASOVA, M., RUECKERT, D., SCHNABEL, J. A., SRHOJ-EGEKHER, V., WU, J., WANG, S., DE VRIES, L. S. & VIERGEVER, M. A. 2015. Evaluation of automatic neonatal brain segmentation algorithms: the NeoBrainS12 challenge. *Med Image Anal,* 20**,** 135-51.
JOB, D. E., DICKIE, D. A., RODRIGUEZ, D., ROBSON, A., PERNET, C., BASTIN, M. E., BOARDMAN, J. P., MURRAY, A. D., AHEARN, T., WAITER, G. D., STAFF, R. T., DEARY, I. J., SHENKIN, S. D. &

WARDLAW, J. M. 2016. A brain imaging repository of normal structural MRI across the life course: Brain Images of Normal Subjects (BRAINS). *Neuroimage*.

KENNEDY, D. N., HASELGROVE, C., HODGE, S. M., RANE, P. S., MAKRIS, N. & FRAZIER, J. A. 2012. CANDIShare: a resource for pediatric neuroimaging data. *Neuroinform,* 10**,** 319-22.

KUKLISOVA-MURGASOVA, M., ALJABAR, P., SRINIVASAN, L., COUNSELL, S. J., DORIA, V., SERAG, A., GOUSIAS, I. S., BOARDMAN, J. P., RUTHERFORD, M. A., EDWARDS, A. D., HAJNAL, J. V. & RUECKERT, D. 2011. A dynamic 4D probabilistic atlas of the developing brain. *Neuroimage,* 54**,** 2750-63.

LEROY, F., MANGIN, J. F., ROUSSEAU, F., GLASEL, H., HERTZ-PANNIER, L., DUBOIS, J. & DEHAENE-LAMBERTZ, G. 2011. Atlas-free surface reconstruction of the cortical grey-white interface in infants. *PLoS One,* 6**,** e27128.

LOH, W. Y., CONNELLY, A., CHEONG, J. L., SPITTLE, A. J., CHEN, J., ADAMSON, C., AHMADZAI, Z. M., FAM, L. G., REES, S., LEE, K. J., DOYLE, L. W., ANDERSON, P. J. & THOMPSON, D. K. 2015. A New MRI-Based Pediatric Subcortical Segmentation Technique (PSST). *Neuroinform*.

LOTJONEN, J. M., WOLZ, R., KOIKKALAINEN, J. R., THURFJELL, L., WALDEMAR, G., SOININEN, H., RUECKERT, D. & ALZHEIMER'S DISEASE NEUROIMAGING, I. 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage,* 49**,** 2352-65.

MAKROPOULOS, A., LEDIG, C., ALJABAR, P., SERAG, A., HAJNAL, J. V., EDWARDS, A. D., COUNSELL, S. J. & RUECKERT, D. 2012. Automatic tissue and structural segmentation of neonatal brain MRI using expectation-maximization. *MICCAI Grand Challenge on Neonatal Brain Segmentation 2012 (NeoBrainS12)***,** 9-15.

MAZZIOTTA, J., TOGA, A., EVANS, A., FOX, P., LANCASTER, J., ZILLES, K., WOODS, R., PAUS, T., SIMPSON, G., PIKE, B., HOLMES, C., COLLINS, L., THOMPSON, P., MACDONALD, D., IACOBONI, M., SCHORMANN, T., AMUNTS, K., PALOMERO-GALLAGHER, N., GEYER, S., PARSONS, L., NARR, K., KABANI, N., LE GOUALHER, G., BOOMSMA, D., CANNON, T., KAWASHIMA, R. & MAZOYER, B. 2001. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci,* 356**,** 1293-322.

MCGURN, B., DEARY, I. J. & STARR, J. M. 2008. Childhood cognitive ability and risk of late-onset Alzheimer and vascular dementia. *Neurology,* 71**,** 1051-6.

MITRA, J., BOURGEAT, P., FRIPP, J., GHOSE, S., ROSE, S., SALVADO, O., CONNELLY, A., CAMPBELL, B., PALMER, S., SHARMA, G., CHRISTENSEN, S. & CAREY, L. 2014. Lesion segmentation from multimodal MRI using random forest following ischemic stroke. *Neuroimage,* 98**,** 324-335.

MODAT, M., RIDGWAY, G. R., TAYLOR, Z. A., LEHMANN, M., BARNES, J., HAWKES, D. J., FOX, N. C. & OURSELIN, S. 2010. Fast free-form deformation using graphics processing units. *Comput Methods Programs Biomed,* 98**,** 278-84.

MOESKOPS, P., BENDERS, M. J., CHIT, S. M., KERSBERGEN, K. J., GROENENDAAL, F., DE VRIES, L. S., VIERGEVER, M. A. & ISGUM, I.

2015. Automatic segmentation of MR brain images of preterm infants using supervised classification. *Neuroimage*.

NYUL, L. G. & UDUPA, J. K. 2000. Standardizing the MR image intensity scales: making MR intensities have tissue-specific meaning. *Proc SPIE Int Soc Opt Eng*, 496-504.

PEREIRA, S., PINTO, A., OLIVEIRA, J., MENDRIK, A. M., CORREIA, J. H. & SILVA, C. A. 2016. Automatic brain tissue segmentation in MR images using Random Forests and Conditional Random Fields. *J Neurosci Meth,* 270**,** 111-123.

PRASTAWA, M., GILMORE, J. H., LIN, W. & GERIG, G. 2005. Automatic segmentation of MR images of the developing newborn brain. *Med Image Anal,* 9**,** 457-66.

ROHLFING, T. 2012. Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable. *IEEE Trans Med Imaging,* 31**,** 153-163.

ROHLFING, T., BRANDT, R., MENZEL, R. & MAURER, C. R., JR. 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *Neuroimage,* 21**,** 1428-42.

ROUSSEAU, F., HABAS, P. A. & STUDHOLME, C. 2011. A Supervised Patch-Based Approach for Human Brain Labeling. *IEEE Trans Med Imaging,* 30**,** 1852-1862.

RUECKERT, D., SONODA, L. I., HAYES, C., HILL, D. L. G., LEACH, M. O. & HAWKES, D. J. 1999. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging,* 18**,** 712-721.

SERAG, A., ALJABAR, P., BALL, G., COUNSELL, S. J., BOARDMAN, J. P., RUTHERFORD, M. A., EDWARDS, A. D., HAJNAL, J. V. & RUECKERT, D. 2012a. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *Neuroimage,* 59**,** 2255-65.

SERAG, A., BLESA, M., MOORE, E. J., PATAKY, R., SPARROW, S., WILKINSON, A. G., MACNAUGHT, G., SEMPLE, S. I. & BOARDMAN, J. P. 2016. Accurate Learning with Few Atlases (ALFA): an algorithm for MRI neonatal brain extraction and comparison with 11 publicly available methods. *Sci Rep,* 6, 23470.

SERAG, A., GOUSIAS, I. S., MAKROPOULOS, A., ALJABAR, P., HAJNAL, J. V., BOARDMAN, J. P., COUNSELL, S. J. & RUECKERT, D. 2012b. Unsupervised Learning of Shape Complexity: Application to Brain Development. *MICCAI workshop on Spatio-temporal Image Analysis for Longitudinal and Time-Series Image Data.*

SERAG, A., KYRIAKOPOULOU, V., RUTHERFORD, M. A., EDWARDS, A. D., HAJNAL, J. V., ALJABAR, P., COUNSELL, S. J., BOARDMAN, J. P. & RUECKERT, D. 2012c. A Multi-channel 4D Probabilistic Atlas of the Developing Brain: Application to Fetuses and Neonates. *Ann. BMVA,* 2012**,** 1-14.

SHENKIN, S. D., BASTIN, M. E., MACGILLIVRAY, T. J., DEARY, I. J., STARR, J. M. & WARDLAW, J. M. 2009. Birth parameters are associated with late-life white matter integrity in community-dwelling older people. *Stroke,* 40**,** 1225-8.

SHI, F., YAP, P. T., FAN, Y., GILMORE, J. H., LIN, W. & SHEN, D. 2010. Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. *Neuroimage,* 51**,** 684-93.

SONG, Z., AWATE, S. P., LICHT, D. J. & GEE, J. C. 2007. Clinical neonatal brain MRI segmentation using adaptive nonparametric data models and intensity-based Markov priors. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)***,** 4791**,** 883-890.

STONER, R., CHOW, M. L., BOYLE, M. P., SUNKIN, S. M., MOUTON, P. R., ROY, S., WYNSHAW-BORIS, A., COLAMARINO, S. A., LEIN, E. S. & COURCHESNE, E. 2014. Patches of disorganization in the neocortex of children with autism. *N Engl J Med,* 370**,** 1209-19.

TAMNES, C. K., WALHOVD, K. B., DALE, A. M., OSTBY, Y., GRYDELAND, H., RICHARDSON, G., WESTLYE, L. T., RODDEY, J. C., HAGLER, D. J., JR., DUE-TONNESSEN, P., HOLLAND, D., FJELL, A. M. & ALZHEIMER'S DISEASE NEUROIMAGING, I. 2013. Brain development and aging: overlapping and unique patterns of change. *Neuroimage,* 68**,** 63-74.

TUSTISON, N. J., AVANTS, B. B., COOK, P. A., YUANJIE, Z., EGAN, A., YUSHKEVICH, P. A. & GEE, J. C. 2010. N4ITK: Improved N3 Bias Correction. *IEEE Trans Med Imaging,* 29**,** 1310-1320.

TUSTISON, N. J., SHRINIDHI, K. L., WINTERMARK, M., DURST, C. R., KANDEL, B. M., GEE, J. C., GROSSMAN, M. C. & AVANTS, B. B. 2015. Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with ANTsR. *Neuroinform,* 13**,** 209-225.

VAN LEEMPUT, K., MAES, F., VANDERMEULEN, D., COLCHESTER, A. & SUETENS, P. 2001. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Trans Med Imaging,* 20**,** 677-688.

VOVK, A., COX, R. W., STARE, J., SUPUT, D. & SAAD, Z. S. 2011. Segmentation priors from local image properties: without using bias field correction, location-based templates, or registration. *Neuroimage,* 55**,** 142-52.

WANG, H., SUH, J. W., DAS, S. R., PLUTA, J. B., CRAIGE, C. & YUSHKEVICH, P. A. 2013. Multi-Atlas Segmentation with Joint Label Fusion. *IEEE Trans Pattern Anal,* 35**,** 611-623.

WANG, L., GAO, Y., SHI, F., LI, G., GILMORE, J. H., LIN, W. & SHEN, D. 2015. LINKS: learning-based multi-source IntegratioN frameworK for Segmentation of infant brain images. *Neuroimage,* 108**,** 160-72.

WARDLAW, J. M., BASTIN, M. E., VALDES HERNANDEZ, M. C., MANIEGA, S. M., ROYLE, N. A., MORRIS, Z., CLAYDEN, J. D., SANDEMAN, E. M., EADIE, E., MURRAY, C., STARR, J. M. & DEARY, I. J. 2011. Brain aging, cognition in youth and old age and vascular disease in the Lothian Birth Cohort 1936: rationale, design and methodology of the imaging protocol. *Int J Stroke,* 6**,** 547-59.

WARFIELD, S. K., ZOU, K. H. & WELLS, W. M. 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging,* 23**,** 903-921.

WEGLINSKI, T. & FABIJANSKA, A. 2011. Brain tumor segmentation from MRI data sets using region growing approach. *Proceedings of VIIth International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH)***,** 185-188.

WEISENFELD, N. I. & WARFIELD, S. K. 2009. Automatic segmentation of newborn brain MRI. *Neuroimage,* 47**,** 564-72.
WEISS, G. M. & PROVOST, F. 2003. Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research,* 19**,** 315-354.
YI, Z., CRIMINISI, A., SHOTTON, J. & BLAKE, A. 2009. Discriminative, Semantic Segmentation of Brain Tissue in MR Images. *Medical Image Computing and Computer-Assisted Intervention (MICCAI),* 5762**,** 558-565.
ZIKIC, D., GLOCKER, B. & CRIMINISI, A. 2014. Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. *Med Image Anal,* 18**,** 1262-73.

10. Figures

Figure 1. Overview of the SEGMA feature extraction framework. The input test image is preprocessed for brain extraction and bias field correction, before computing gradients. Then, a sliding window is scanned across the input image at all positions where a feature vector for each voxel over the window is extracted using intensity and gradient information. The feature vectors are fed into a random forest classifier trained for structure / tissue classification.

Figure 2. An example of classifying one test window. The green square in the test image represents the test window. The green rectangle represents the extracted features from the test window (i.e. test samples). The red rectangle represents the extracted features from training data (i.e. training samples). The random forest classifier is trained using the training samples and the voxels inside the test window are classified into different classes based on test samples.

Figure 3. Examples of brain segmentation results across the life course (axial view) using SEGMA. The automated segmentation is based on T2-weighted scans for the neonatal period and T1-weighted scans for the rest of growth stages. The images are taken from single subjects at the shown ages, where neonatal period images come from dataset I; childhood and adolescence images come from dataset II; and adulthood images come from dataset III.

Figure 4. Bar plots of the Dice coefficient (with standard deviation as error bar) comparing segmentations derived from SEGMA with reference segmentations using (a) dataset I [neonatal period], (b) dataset II [childhood and adolescence], and (c) dataset III [adulthood].

Figure 5. Examples of SEGMA's output segmentation results (sagittal view) using T1-weighted (T1w) and T2-weighted (T2w) MR individually.

Figure 6. (a) Relative importance of each of the ten features, expressed as the segmentation accuracy, on removing the feature from the feature vector. The leftmost bar shows a baseline value – Dice coefficient, when all features are used. (b) An example of the automatic segmentation of cortical GM (coronal view), which shows how the dropping of each of the ten features affects the segmentation accuracy. The baseline segmentation is obtained by using all features.

Figure 7. Examples of edge detection for various regions (cortical grey matter, sub-cortical structures, brainstem and cerebellum) based on using all features (intensity combined with gradients) and intensity grey scale only, for a neonatal (dataset I) and an adult brain (dataset III).

11. Tables

Table 1. Dice coefficients averaged over all structures for datasets I, II and III. SEGMA is compared with MV, STAPLE, global-RF-1, and global-RF-2.

Dataset	SEGMA	global-RF-1	global-RF-2	MV	STAPLE
I	**90.68 %**	85.29 %	84.22 %	86.97 %	87.01 %
II	**86.05 %**	78.98 %	74.90 %	81.75 %	79.17 %
III	**82.56 %**	78.75 %	76.02 %	77.13 %	77.54 %

Figure 1.JPEG

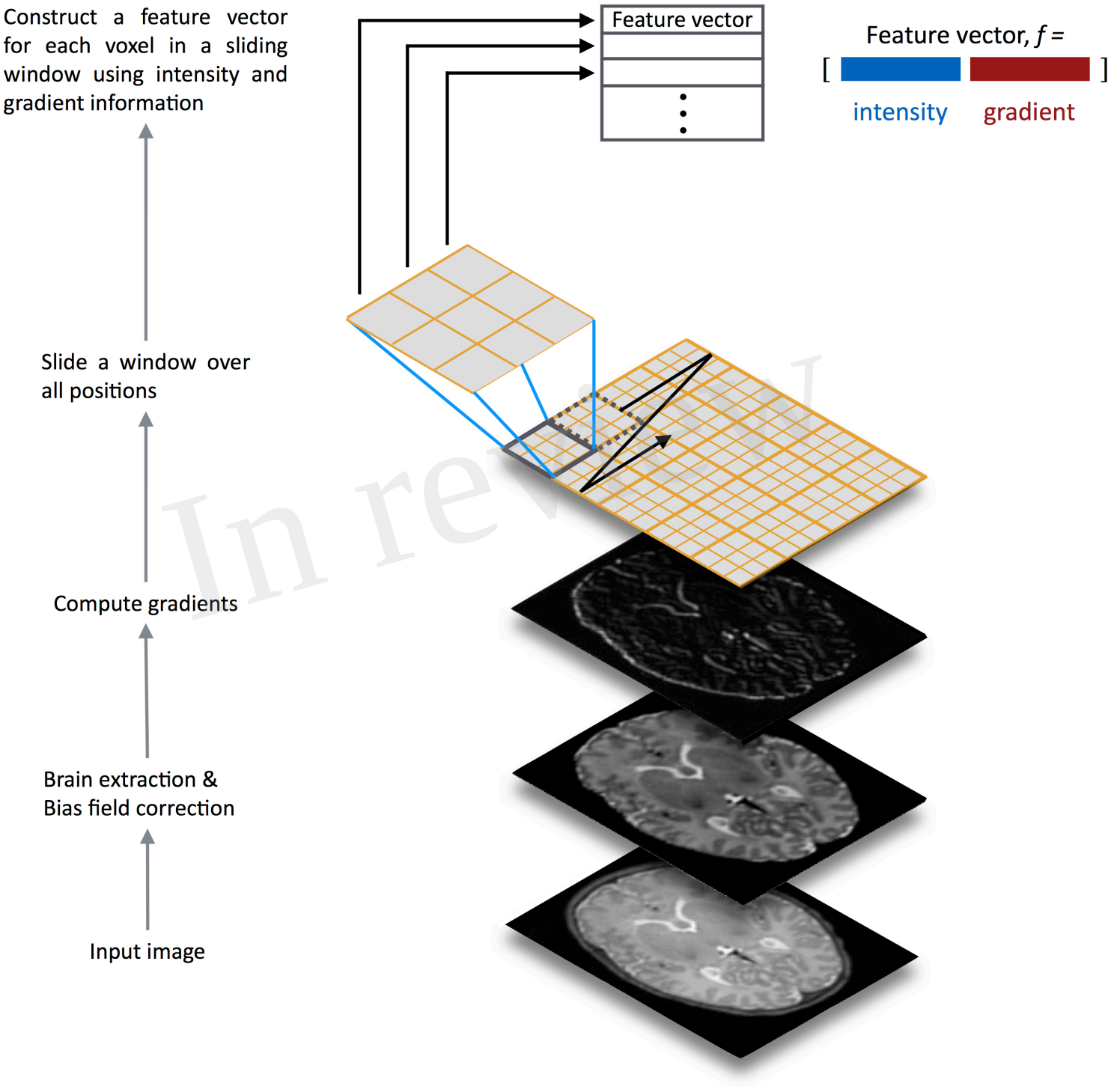

Figure 2.TIFF

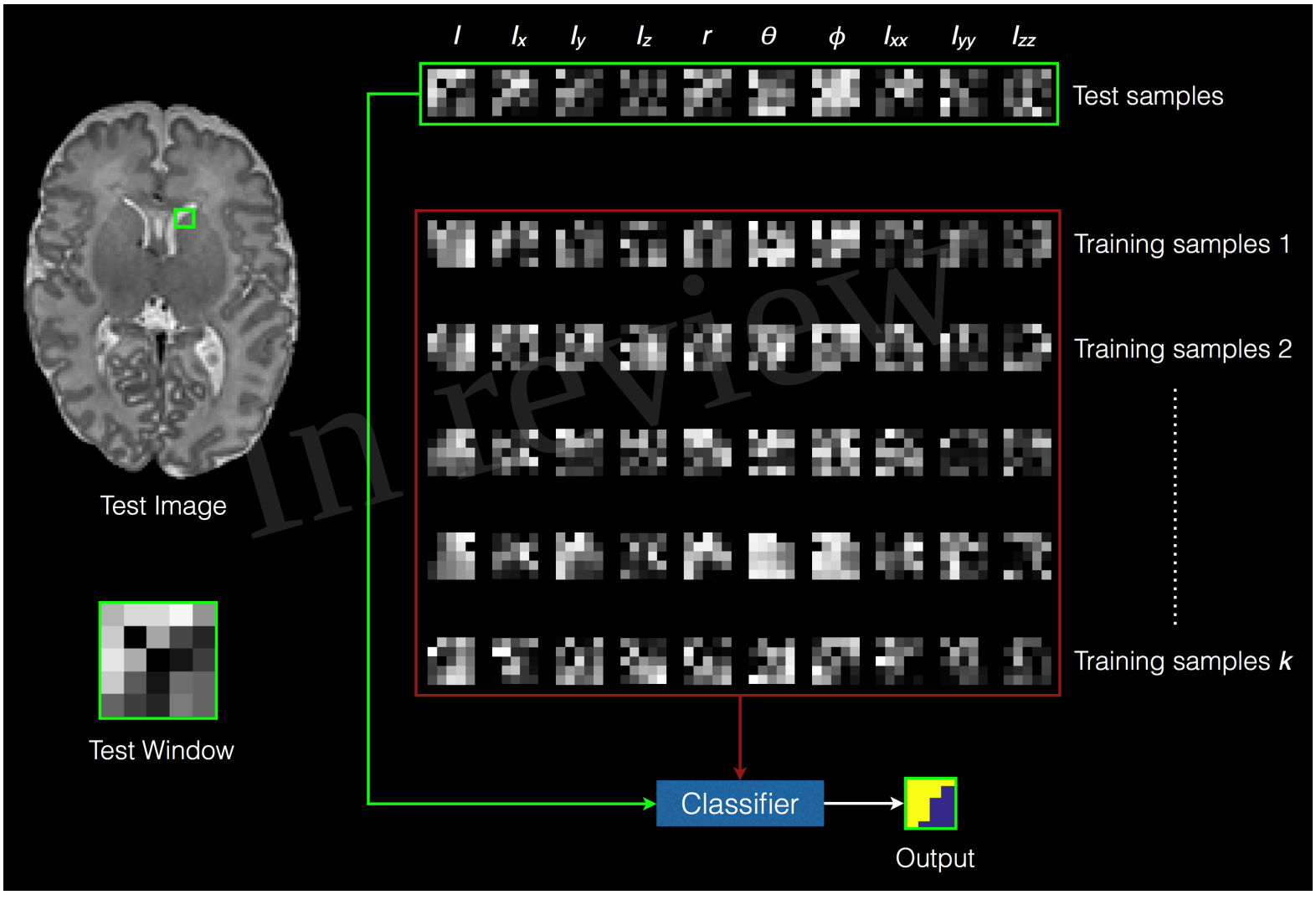

Figure 3.JPEG

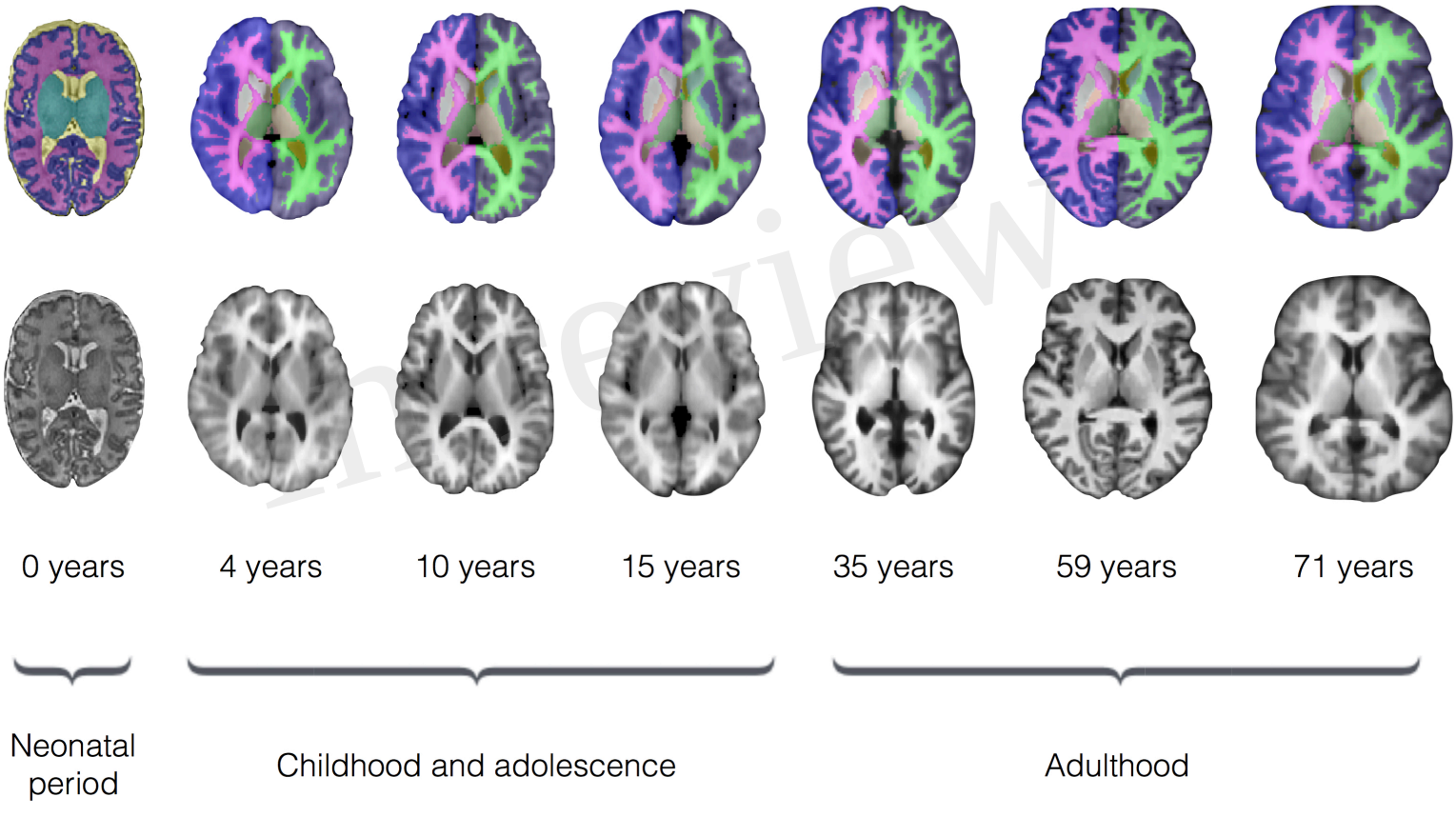

Figure 4.JPEG

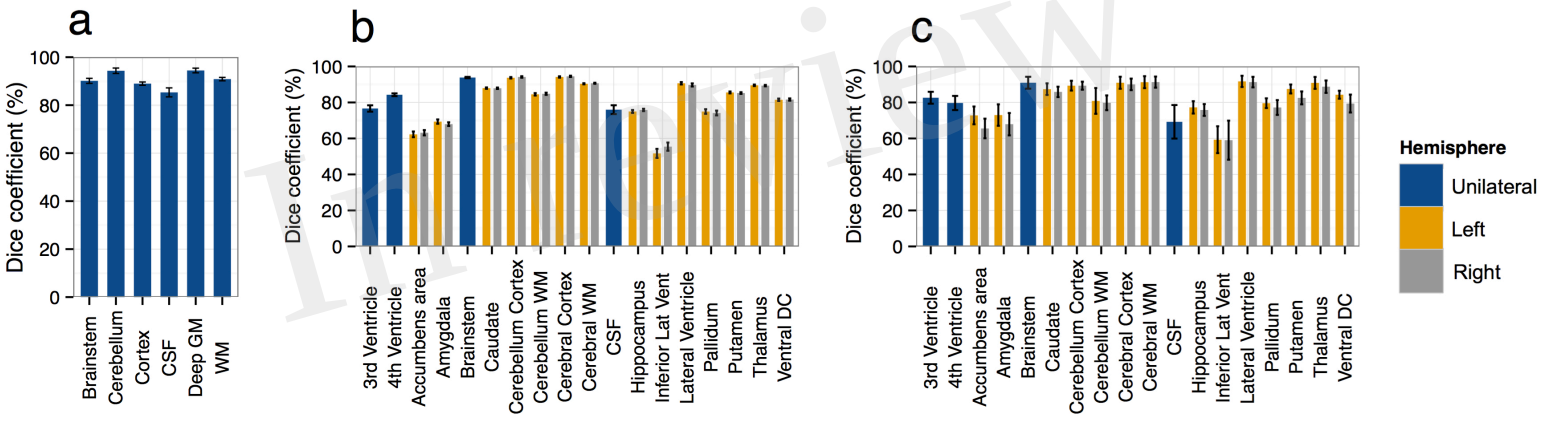

Figure 5.JPEG

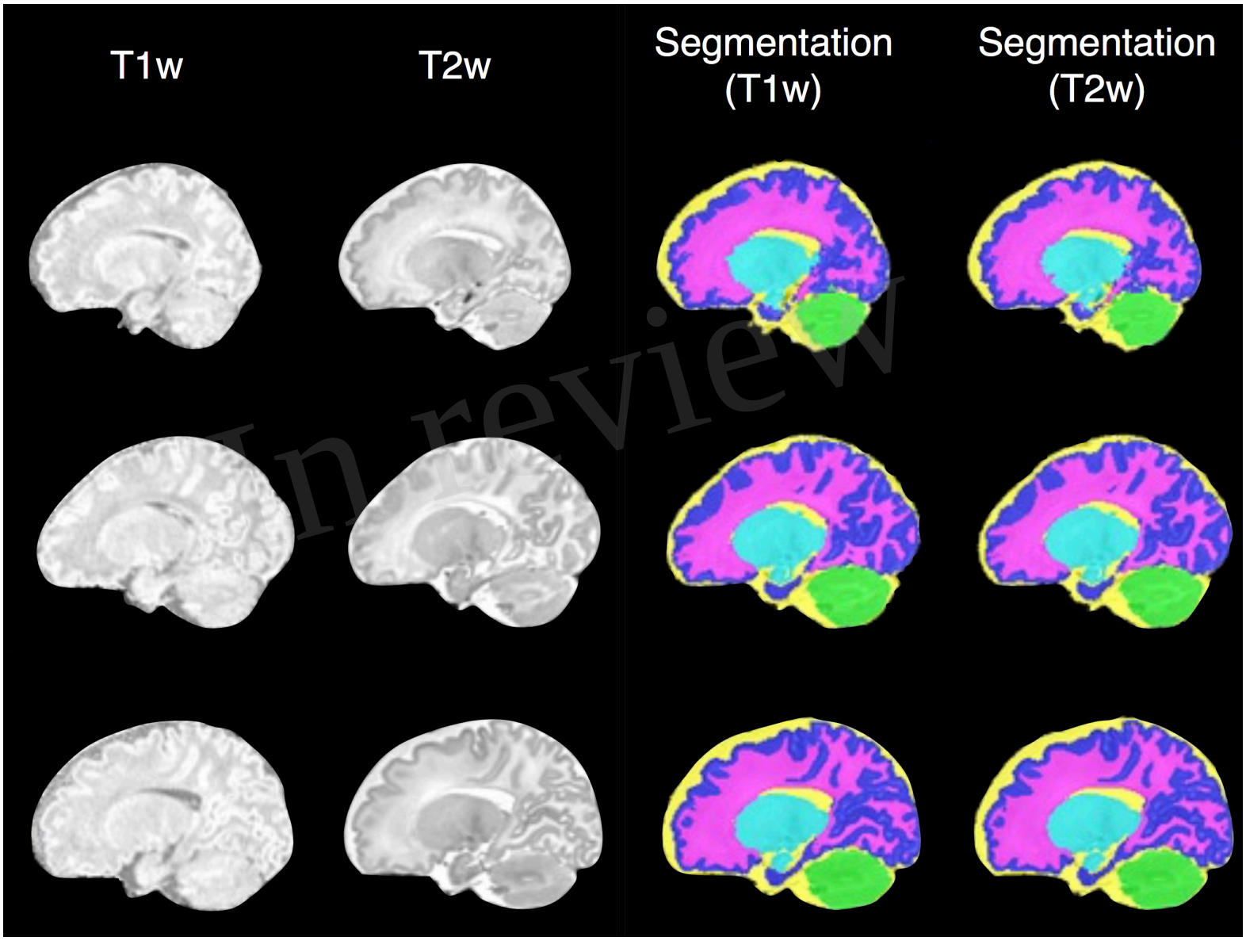

Figure 6.JPEG

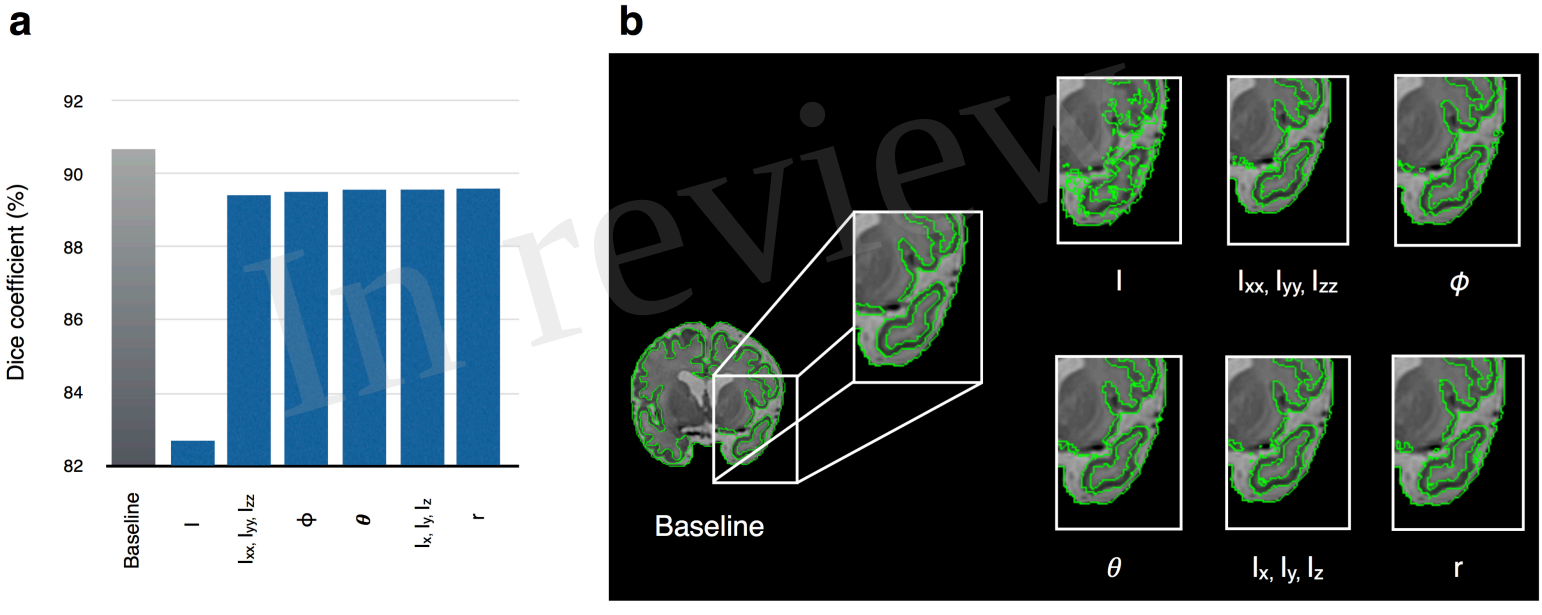

Figure 7.TIFF