Automatic segmentation of basal ganglia iron deposits from structural MRI

Andreas Glatz[1]
a.glatz@sms.ed.ac.uk
Maria C. Valdés Hernández[1]
Alexander J. Kiker[2]
Mark E. Bastin[1,4]
Susana Muñoz Maniega[1]
Natalie A. Royle[1]
Ian J. Deary[3]
Joanna M. Wardlaw[1]

[1] BRIC, Division of Clinical Neurosciences
University of Edinburgh, UK

[2] College of Medicine and Veterinary Medicine
University of Edinburgh, UK

[3] Department of Psychology
University of Edinburgh, UK

[4] Division of Medical and Radiological Sciences
University of Edinburgh, UK

Abstract

Brain iron deposits have recently been suggested as biomarkers for small brain vessel diseases. Here, we present a novel, automated method for segmenting brain iron deposits in the basal ganglia from structural MRI data. It is based on minimum-variance clustering of intensities from T_1- and T_2*-weighted volumes, and a supervised cluster selection algorithm. This method was evaluated with MR data from 24 subjects and compared with iron deposit masks segmented manually by an experienced rater. A median Jaccard similarity index of 0.64 between manual and automatically generated segmentation masks is promising and encourages further investigations to improve the computing speed and accuracy of the method.

1 Introduction

Iron is essential for many human cellular functions involving enzymes and prosthetic groups. In the brain, excess iron is usually stored in form of the soluble protein ferritin which provides protection against iron induced brain tissue damage and is used as an iron source during times of iron deficiency. Iron can also be found in form of the insoluble iron-complex hemosiderin, which is mainly associated with brain microbleeds and iron deposits in the basal ganglia. Histopathological investigations revealed that basal ganglia iron deposits (BGIDs) are closely related with small blood vessels and that they can show calcifications. Recently, a clinical study suggested brain microbleeds and BGIDs, which increase in prevalence and extent with age, as biomarkers for small vessel dysfunction[4].

MRI is commonly used to detect IDs in brain tissue. Water shielded paramagnetic molecules, such as hemosiderin, ferritin and deoxyhemoglobin, as well as calcium, dephase proton spins in their vicinity, which locally shortens the transverse relaxation time and causes hypointense regions on T_2- and T_2*-weighted magnitude images. T_1-weighted magnitude images can provide additional information about the chemical state of IDs[1]. A recent study

already showed that fused T_1- and T_2*-weighted magnitude images can be used to differentiate IDs from calcium because calcium appears hypointense in T_1-weighted magnitude images[6].

Currently, IDs in MRI data are segmented and rated manually by radiologists, which can be very time consuming and prone to a high degree of inter-observer variability. A recent study compared two semi-automated methods for segmenting IDs based on thresholding and minimum variance quantization as implemented in the software MCMxxxVI[7], which showed good intra- and inter-observer reliability. However, both methods have limitations regarding further automation and observer independence[6]. To address these issues, we have developed a new method for reliably and automatically identifying BGIDs from structural MRI data which we describe below.

2 Method

The automatic BGID segmentation method requires registered, bias corrected T_1- and T_2*-weighted volumes, binary masks of normal-appearing white matter, and basal ganglia structures, which are generated in a preprocessing step. In the subsequent segmentation step, potential BGID hypointensities are identified, collected and clustered. Individual clusters are scored and clusters with scores above an arbitrary threshold are selected to create BGID masks. The threshold can be chosen either manually or automatically.

2.1 Preprocessing

Firstly, the T_1- and T_2*-weighted volumes were affine registered to T_2-weighted volumes with FSL[1] FLIRT[2]. The bias field of the registered T_1- and T_2*-weighted volumes was removed with FSL FAST[10]. We generated masks of the normal-appearing white matter using MCMxxxVI as it has been validated on scans from older subjects[7]. Lastly, FSL FIRST was used to segment basal ganglia structures, caudate nucleus, globus pallidus and putamen, and to generate binary basal ganglia masks[3].

2.2 Hypointensities collection and clustering

Intensities of registered, bias-corrected T_1- and T_2*-weighted volumes which are potentially associated with BGIDs are identified in all slices and collected in a set $\mathscr{D}$. Firstly, ROI masks M_{ROI} are calculated by adding a margin to basal ganglia masks to compensate for possible segmentation errors. Intensities pairs $(I_{T1,reg,bc}, I_{T2*,reg,bc})$ within the ROI from registered, bias-corrected T_1- and T_2*-weighted volumes, and with an $I_{T2*,reg,bc}$ intensity below a cutoff intensity $I_{T2*,reg,bc,max}$ are added to the set $\mathscr{D}$. The cutoff intensity is defined as $I_{T2*,reg,bc,max} = m_{WM} - n\, s_{WM}$ with the cutoff factor $n \in \mathbb{R}^+$. White matter mean m_{WM} and standard deviation s_{WM} are estimated from $I_{T2*,reg,bc}$ intensities within the volume covered by the white matter mask. Finally, Ward's average-linkage agglomerative clustering method is used to partition the set $\mathscr{D}$ into k subsets by minimising the inter-cluster and maximising the intra-cluster variance[8].

[1] http://www.fmrib.ox.ac.uk/

2.3 Scoring and selection of clusters

After clustering, we assign a score S_i to each cluster according to the likelihood that they contain BGID intensities. A score:

$$S_i = f(I_{T2*,med,i}) = 1 - \frac{I_{T2*,med,i}}{I_{T2*,reg,bc,max}} \quad i = 1..k \tag{1}$$

is calculated for each of the k clusters after clustering with Ward's method. Here, $I_{T2*,med,i}$ is the median of $I_{T2*,reg,bc}$ intensities of each cluster. The scoring function f implements the believe that BGID intensities are on average darker than white or grey matter intensities.

BGID intensity clusters with a score above a cutoff S_{cutoff} are selected to generate BGID masks.The cutoff score can be chosen either manually by searching for the optimal score, or automatically based on a linear relationship with the factor $W_{\mathscr{D}}$ according to:

$$S_{cutoff} = \alpha + W_{\mathscr{D}}\beta = \alpha + \frac{I_{T2*,med}}{I_{T2*,max} - I_{T2*,min}}\beta \quad \alpha,\beta \in \mathbb{R} \tag{2}$$

with $I_{T2*,max}$, $I_{T2*,min}$ and $I_{T2*,med}$ as the maximum, minimum and median $I_{T2*,reg,bc}$ intensities of set $\mathscr{D}$, respectively. The parameters α and β can be estimated from the optimal cutoff score and corresponding width factor $W_{\mathscr{D}}$ of subjects in a training set.

3 Experimental results and discussion

The presented method was evaluated using T_1-, T_2*- and T_2-weighted volumes acquired from 24 generally healthy, community-dwelling older subjects from the Lothian Birth Cohort 1936[2], with imaging parameters described in [9]. We compared the BGID masks, obtained with manual and automatic cutoff score selection, with reference masks generated with a published, semi-automated method by an experienced rater[7]. The Jaccard index was used to measure the spatial, in-plane coincidence of generated and reference BGID masks[5].

Figure 1 shows segmentation results with $k = 16$ clusters and cutoff factor $n = 1.2$ for a representative subject. The Jaccard indices for manual and automatic cutoff score selection were 0.71 and 0.57, respectively. The automatic segmentation results were generated in a 5-fold cross-validation run. The manual segmentation result approximates the reference masks reasonably well whereas the automatic segmentation generally produces a slightly larger region than the reference masks. A Matlab[3] implementation of the cross-validation of the segmentation algorithm, using the pMatlab library[4] to distribute processing across six standard $2.66GHz$ processor cores, takes about $40min$ for all 24 subjects to complete.

Figure 2 shows how the number of clusters k influences the performance of the segmentation method at $n = 1.2$ for all subjects. The performance with manual cutoff score selection improves for $k = 16$ and stays about the same for increasing number of clusters. It seems that for $k = 8$ there is still a significant number of clusters which contain a mixture of BGIDs and grey or white matter intensities. Those clusters seem to be broken up for higher k values that enable a more fine-grained cluster selection. The performance with automatic cutoff score selection generally follows the same pattern as with manual cutoff score selection. The

[2] http://www.psy.ed.ac.uk/research/lbc/LBC1936.htm
[3] http://www.mathworks.com/
[4] http://www.ll.mit.edu/pMatlab/

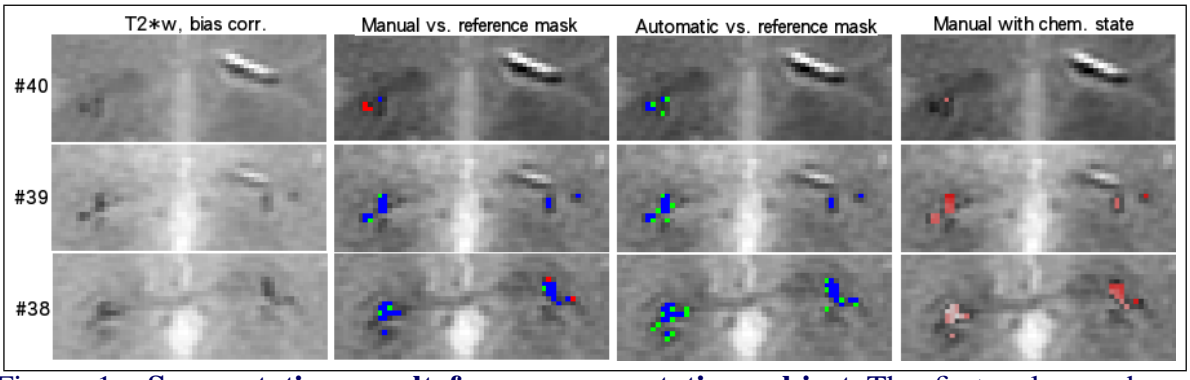

Figure 1: **Segmentation result for a representative subject** The first column shows cropped, bias corrected T_2*-weighted images with basal ganglia hypointensities from slice 38 to 40. Column two and three shows images from the first column with overlaid reference BGID masks, and either masks from the method with manual or automatic cutoff score selection, respectively. The colours in those two columns indicate the agreement between the BGID masks: bright (red, green) coloured voxels indicate that they appear in only one of the masks while dark (blue) voxels indicate a perfect match. The forth column shows the manual mask where higher colour saturation levels indicate brighter voxels in the T_1-weighted image, and possibly a different chemical state of the iron deposits[1].

linearity between $W_{\mathscr{D}}$ and S_{cutoff} (Equation 2) from all 24 subjects increases with increasing number of clusters k, with a linear correlation coefficient of 0.66, 0.84, 0.90, 0.91 for $k = 8, 16, 24, 32$, respectively. It seems that cluster granularity and the parameter linearity is sufficiently high at $k = 16$ for the automated method to work well.

For subjects with approximately the same numbers of BGIDs, grey and white matter intensities used for hierarchical clustering results in a cluster hierarchy with three main branches for BGIDs, grey and white matter intensity clusters, respectively. This suggests that BGID clusters seem to separate well from grey and white matter clusters, and the number of clusters could be decreased to $k = 3$ without any significant loss of accuracy. For subjects with small or large BGIDs, the BGID intensities either have too little or too much weight in the clustering process which causes an increased aggregation of BGID intensities with grey or white matter intensities. This seems to originate from the fact that Ward's method is generally biased towards forming equal sized clusters. Then an increase in the number of clusters is required to be able to separate the BGID clusters from the others.

Currently, this method is designed to segment IDs in the basal ganglia. Segmenting IDs in other brain regions may be more difficult due to the presence of blood vessels, areas filled with air or deposits of different minerals. All those features may cause hypointensities similar to IDs, which would have to be detected and filtered either before or after clustering.

4 Conclusions

We have developed and evaluated an automated segmentation method for BGIDs. The resulting BGID masks are promising and tend to be close to reference masks from an experienced rater. The next step will be to improve the accuracy and increase the computation speed of the method by evaluating other clustering and cluster selection algorithms.

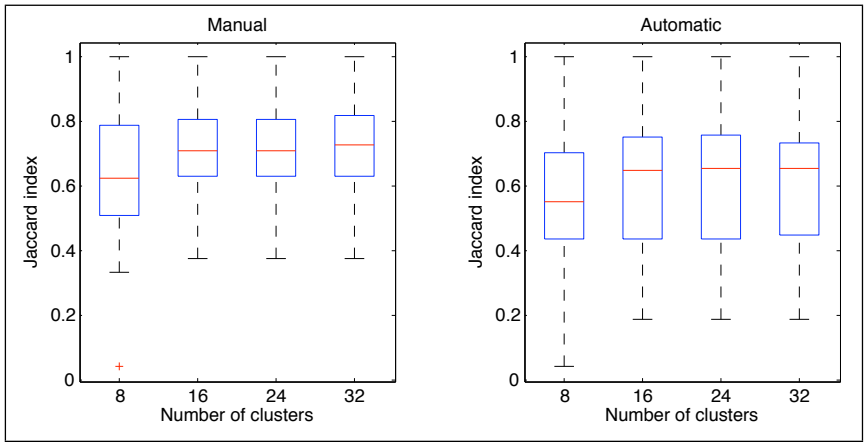

Figure 2: **Performance of the segmentation method for different number of clusters *k* with manual and automatic parameter selection** A boxplot shows the minimum, *25th*-, *50th*-, *75th*-percentile, and maximum of the distribution of Jaccard indices from each subject. The median Jaccard indices for each *k* are 0.62, 0.71, 0.71, 0.73 with manual parameter selection, and 0.55, 0.64, 0.64, 0.64 with automatic parameter selection.

References

[1] S.D. Brass et al. Magnetic resonance imaging of iron deposition in neurological disorders. *Topics in Magnetic Resonance Imaging*, 17(1):31, 2006.

[2] M. Jenkinson et al. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.

[3] B. Patenaude et al. A bayesian model of shape and appearance for subcortical brain. *Neuroimage*, in press.

[4] L. Penke et al. Brain iron deposits are associated with general cognitive ability and cognitive aging. *Neurobiol. Aging*, Jun 2010.

[5] D.W. Shattuck et al. Online resource for validation of brain segmentation methods. *NeuroImage*, 45(2):431–439, 2009.

[6] M.C. Valdés Hernández et al. Reliability of two techniques for assessing cerebral iron deposits from structural MRI. *Magnetic Resonance Imaging*, 2010.

[7] M.C. Valdés Hernández et al. New multispectral MRI data fusion technique for white matter lesion segmentation: method and comparison with thresholding in FLAIR images. *European Radiology*, 20:1684–1691, 2010.

[8] J.H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, pages 236–244, 1963.

[9] J.M. Wardlaw et al. Brain ageing, cognition in youth and old age, and vascular disease in the lothian birth cohort 1936: Rationale, design and methodology of the imaging protocol. *International Journal of Stroke*, in press.

[10] Y. Zhang et al. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *Medical Imaging, IEEE Transactions on*, 20(1):45–57, 2001.