Automated White Matter Hyperintensity Detection Using A False Discovery Rate Method

H. Zheng[1], V. Zagorodnov[2], and M. W. Chee[3]

[1]Cognitive Neuroscience Lab, Duke-NUS Graduate Medical School, Singapore, [2]Nanyang Technological University, Singapore, [3]Cognitive Neuroscience Lab, Duke-NUS Graduate Medical School, Singapore, Singapore

Introduction

We describe a novel fully-automatic, white matter hyper-intensity (WMH) detection algorithm based on 2D fluid-attenuated inversion-recovery (FLAIR) images. Higher WMH burden correlates with lower levels of cognitive function. Hypertension and diabetes mellitus are known to contribute to WMH burden. A more detailed quantification of the relationship between WMH burden and additional risk factors is desirable to fully characterize cognition in healthy elderly. Achieving this goal requires obtaining measurements from hundreds to thousands of subjects in order to yield trustworthy results. This effort would greatly benefit from automated WMH detection. We were further motivated to develop a novel method when we could not tailor a well-tested method that works well with 1.5 T data (1) but not with 3T data, possibly on account of differences in tissue contrast and the high intensity rim at the cortical surface that occasionally occurs with 3T images. Previous automated segmentation schemes can be classified as clustering-based (2, 3), where WMH is modeled as a separate tissue class, and outlier-based (4), where WMH is treated as an outlier to the main brain tissue classes. As clustering is often performed in iterative fashion, it is sensitive to initialization and can be unreliable when the WMH class is too small (small WMH burden). Treating WMH as outliers involves setting a threshold at constant k times the standard deviation of the null distribution. The results of this method are sensitive to the choice of the constant k (4). Here, we determined whether thresholding based on a fixed False Discovery Rate (FDR) could yield results that better correspond with manual segmentation as compared to thresholding based on significance level.

Methods

We evaluated twenty 2D FLAIR image sets (TR 10000 ms TI 2500ms TE 96 ms, 0.43*0.43*6.5mm voxels) containing images with roughly equivalent numbers of low, medium and high WMH burdens. These images were obtained from relatively healthy elderly volunteers ranging from 55-80 years of age. Prior to automatic evaluation, WMH were manually segmented by a trained observer to provide ground truth information. In the automated pipeline, skull removal was first carried out on the each individual's corresponding T1-weighted MPRAGE image. The resultant mask was applied to the FLAIR image. We further used the T1-weighted image to estimate a coarse WM mask that was used to exclude non-WM tissues in the FLAIR image. Inter-slice intensity differences were corrected by adjusting intensity distributions of each slice to a common mode. The histogram of the resultant volume has an approximately Gaussian shape with a long right tail. Due to difficulty of estimating parameters of the WMH distribution, the threshold was chosen to achieve a fixed FDR i.e. the expected ratio of falsely detected voxels to the total volume of detected WMH. Additionally, we excluded WMH clusters containing fewer than 5 contiguous voxels and high signal intensity voxels lining the lateral ventricles in the absence of adjacent confluent WMH lesions.

Results

We compared the FDR method with optimal thresholding and significance-based thresholding (4). The optimal threshold for each brain was found using an exhaustive search of thresholds that yielded the highest Jaccard similarity index (SI) with manually segmented images. SI was defined as (2*True Positive / (2*True Positive + False Positive + False Negative) and can range from 0 to 1. We set the significance-based threshold at k=2.39, as it as empirically found to provide the best average performance for this dataset. The results are summarized in Table 1.

	SI decline of our approach compared to optimal SI	SI decline of significance-based thresholding (k=2.39) compared to optimal SI	Mean (standard deviation) of SI using FDR
Severe WMH (n=8)	0.002 (0.36%)	0.104 (16.2%)	0.66 (0.10)
Moderate WMH (n=6)	0.003 (0.33%)	0.056 (10.2%)	0.57 (0.17)
Mild WMH (n=6)	0.035 (9.38%)	0.113 (30.7%)	0.32 (0.15)
All	0.012 (3.06%)	0.092 (15.0%)	0.53 (0.19)

Table 1. Comparison of threshold selection approaches

Compared to optimum thresholding, the FDR method resulted in negligible decline in SI for images with severe and moderate WMH burdens. For mild cases in which there is disagreement even among human operators, SI declined an average of 10%. In contrast, the decline in SI with significance-based thresholding (k=2.39) was more substantial and ranged from 10% to 30%. Thus, the chosen value of k=2.39 works well for some images only. On the other hand, FDR based thresholding adaptively varied k between 2.0 to 2.44 depending on the image. The benefit over clustering approaches is better reliability in cases with small WMH burden. The achieved similarity indices ranged from 0.32 to 0.66, with higher values corresponding to images with high WMH burden. Compared with significance-based detection, our study shows better similarity to manual segmentation over a large range of WMH burdens (from 1.2 cm^3 to 45cm^3).

Conclusion

FDR-based threshold selection does not require estimation of the parameters of the WMH distribution and was demonstrated to be more universal than thresholding based on significance level. While using a single FDR threshold, simplified pre- and post-processing routines, and a single FLAIR image, might theoretically limit WMH detection accuracy, our results are comparable to those of other automatic pipelines using multi-spectral (T2/PD/FLAIR) images. In the future we plan to extend our algorithm to support multi-spectral data, adaptive thresholding, and more sophisticated smoothness constraints (e.g. MRF priors).

References

1. Jack, C. R. J., O'Brien, P. C., Rettman, D. W., Shiung, M. M., Xu, Y., Muthupillai, R., Manduca, A., Avula, R. & Erickson, B. J. (2001) *Journal of Magnetic Resonance Imaging* **14,** 668-676.
2. Zijdenbos, A. P., Forghani, R. & Evans, A. C. (2002) *IEEE Transactions on Medical Imaging* **21,** 1280-1291.
3. Kikinis, R., Guttmann, C. R. G., Metcalf, D., Wells I. I. I., W. M., Ettinger, G. J., Weiner, H. L. & Jolesz, F. A. (1999) *Journal of Magnetic Resonance Imaging* **9,** 519-530.
4. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A. & Suetens, P. (2001) *IEEE Transactions on Medical Imaging* **20,** 677-688.