Automated Lesion Discrimination and Outlining

D. S. Wack[1,2], M. Dwyer[1,3], S. Hussein[1], C. Caiola[1], P. Hojczyk[1], and R. Zivadinov[1,3]

[1]Buffalo Neuroimaging Analysis Center, State University of New York at Buffalo, Buffalo, New York, United States, [2]Depts. of Nuclear Medicine and Neurology, State University of New York at Buffalo, Buffalo, New York, United States, [3]Dept. of Neurology, State University of New York at Buffalo, Buffalo, New York, United States

INTRODUCTION: We describe a novel method for detection of regional lesion abnormalities in the brain. While the focus of this study was on lesion detection in multiple sclerosis (MS), the method is immediately transferable to any lesion abnormalities in a variety of systemic disorders of the central nervous system (CNS). The application of an Automated Lesion Detection and Outliner (ALDO) method, can have a major impact for research use. ALDO is a two step process. The first and central step uses Stochastic Discrimination as a pattern recognition method to classify voxels on a point by point basis. The second step uses the results of the first step and the original scan to outline detected lesions.

Stochastic Discrimination (SD) is a method of pattern recognition originally developed by E. M. Kleinberg. The method has been shown to perform quite well on many standardized test sets. In a comparison of 23 other methods, on 7 problem sets, Kleinberg's implementation of SD performed the best on five tests, and fifth on the other two. An independently created implementation of Kleinberg's method has been developed for this study since Kleinberg's implementation is not publicly available. Our new method had very similar rankings on the test sets. Our SD method works by randomly creating weak models that assign to each voxel a number between -.1 and .1. A negative value would indicate the model classifies a voxel as a non-lesion; a positive value indicates the weak model classifies a voxel as from a lesion. An individual weak model's performance will be poor and close to that of classification by chance, since the weak models are randomly created The SD method collects those randomly created weak models that perform a little better than chance. The collection of classifiers is averaged together to give a classification for each voxel. The distribution of the combined weak models forms both a positive and negative peak representing the two classifications. As the central limit theorem predicts the individual peaks tend towards normal, and the peaks' variance decrease as the number of weak models increase. Figure 1 shows this distribution using a log scale; using a linear scale it would not be possible to track the small percentage of voxels near zero that remain difficult to classify. Differences in classifying the regions by experts, especially at the edge of a lesion, make a perfect separation difficult. SD can handle any n-class problem by calculating each class versus the rest of the classes. The final classification would be determined by the class comparison with the highest value. SD methods are resistant to overtraining, and performance on test sets typically continue to increase even after the performance on the training set has leveled off.

Figure 1. The log distributions of lesion and non-lesion voxels.

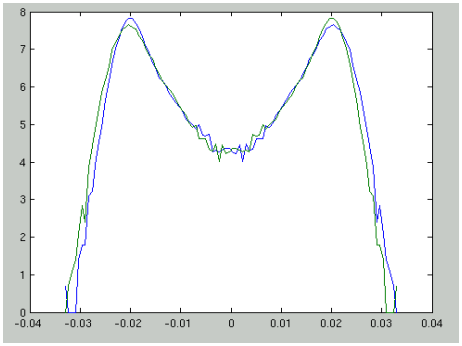
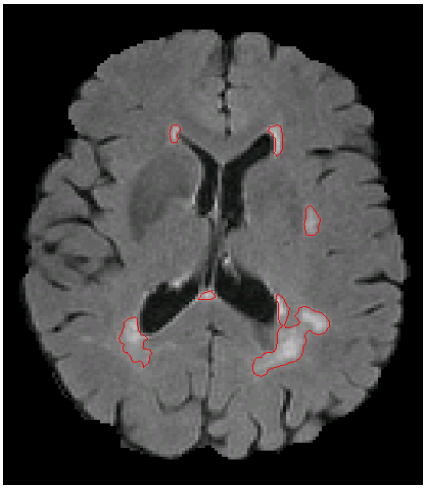


Figure 2. The ALDO ROIs drawn on test MRI



METHODS: The scans are first pre-processed using FMRIB's Brain Extraction Tool (BET) in order to eliminate all non-brain voxels. Only brain voxels above the 60th percentile in intensity for a given transaxial plane are considered for further analysis. The result of the first stage (pattern recognition) is an image with the same orientation as the original, such that higher intensity voxels correspond to voxels from the original image that is more likely to be representative of a lesion. The second stage of the process aims to both maximize the results of the first stage, and create a final product useful to the experts who need to review and possibly modify the results. To this end, the second stage searches the results from the first step and forms clusters meeting a given minimum value for the maximum of the cluster, and extends the cluster according to a lower threshold. These clusters, are then contoured using an algorithm that shares similarity with the Java Image Manipulation (JIM) software. The final result of the second stage is a set of outlines that can be further examined and edited in JIM (see Figure 2). Hence, this allows the expert operator to use an automated lesion detection method, while keeping fully editable options readily available.

ALDO was performed on 40 MS scans obtained from 4 different scanners and two different strengths (1.5T and 3T). The lesions ROIs were marked by one of the center's experts. The lesion masks classify each voxel as either from or not from a lesion. This pattern recognition set was very large both in terms of the size and number of the training vectors. A training vector for our SD method consisted of the classification together with a 40x40x3 voxel neighborhood (a 4800 element vector) and used a total of 400,000 training vectors. The resulting weak models were tested on a set of 39 scans, from 4 different scanners (9-10 scans per scanner, representing both 1.5 and 3T scanners). After approximately 2 weeks of training there were 84,000 weak models generated for the training set. Additionally the times needed to correct the results of our automated method were measured for 87 scans. This series of scans were processed using an earlier version of the method, which differed slightly in how weak models were formed and calculated a voxel's intensity percentile in reference to the entire brain, rather than by slice.

RESULTS: The median Kappa value for the 39 pairs of masks was .61 which indicates a substantial agreement between the hand drawn ROIs and those created from the automated method. The minimum, median, and maximum lesion volumes were 2.8, 10.7, and 44.9 CC, respectively as determined from the hand drawn ROIs from the expert. A Spearman's correlation relating total lesion volume was calculated for the 39 pairs of scans, (Rho = .95 , p =9.78 x 10^{-14}). The performance was also measured on a lesion by lesion basis for lesions above .45 cm^2. The automated method identified 93.26% percent of the lesions .45 cm^2 or greater from the gold standard. Furthermore the gold standard lesions identified 94.36% of the lesions .45 cm^2 or greater from automated method, which is an indication of the methods control on creating false positive lesions. A direct false positive measure is not reported for the lesion detection, since one lesion from one mask may correspond to several smaller lesion masks from the other mask. There was a strong correlation between the true size of lesion and the found Kappa value (Spearman Rho = .57, p = .0001). The unedited outlined lesions found from the automated method are displayed for one slice in Figure 2. Using an earlier version of ALDO, a mean of 14.1 minutes (N=87) was needed to correct the ROIs. Forty minutes is considered the center's average time to draw the ROIs without ALDO.

CONCLUSIONS: The performance of the system provides a method that can be regularly used in a research setting to provide lesion identification. A notable design feature is creating JIM format files facilitating inspection and correction to be done later edited by a human expert. This hybrid approach offers the human expert an advanced starting point for marking lesions. In comparing our results it is important to consider the lower median lesion volume of the scan set, and that training and test sets were performed over 4 scanners. Overall classification performance was directly related to underlying lesion size, as has been observed previously in the literature. We expect our method to continue to improve in performance as new training points are added, and longer training periods are permitted.