Early Detection of Treatment Response for GBM Brain Tumor using ADC Map of DW-MRI

Jing Huo[1], Whitney Pope[1], Kazunori Okada[2], Jeffery Alger[3], Hyun Jung Kim[1], Yang Wang[1], Jonathan Goldin[1], and Matthew Brown[1]

[1] Department of Radiological Sciences, David Geffen School of Medicine, University of California in Los Angeles, {jhuo,wpope,gracekim,yangwang, jgoldin,mbrown}@mednet.ucla.edu
[2] Computer Science Department, San Francisco State University, kazokada@sfsu.edu
[3] Department of Neurology and Department of Radiological Sciences, David Geffen School of Medicine, University of California in Los Angeles, jralger@ucla.edu

Abstract. This preliminary study explores novel methods using diffusion weighted(DW) MR images as a biomarker to detect early GBM brain tumor response to treatment. Apparent diffusion coefficient(ADC) map, calculated from DW-MR images, can provide unique information of tumor response at cellular level. In this study, we investigate whether changes in ADC histograms between two scans, taken 5-7 weeks apart before and after treatment, could predict treatment effectiveness before lesion size changes are observed on later scans. The contribution of our work is to exploit quantitative pattern classification techniques for the prediction. For both pre- and post-treatment scans, we first compute the histogram from the ADC values covered within the tumor. Then we apply supervised learning on features extracted from the histogram for classification. We evaluated our approach with pool data of 86 patients with GBM under chemotherapy while 40 responded and 46 did not respond based on tumor size reduction. We compared Fisher's linear discriminant analysis, AdaBoost and random forests classifier using leave one out cross validation(LOOCV), resulting in the best accuracy of 67.44%.

1 Introduction

Glioblastoma multiforme(GBM) is the most common and aggressive type of primary brain tumor. There are many clinical trials underway to assess the ability of new drugs and strategies to treat glioblastoma and extend the duration of patient survival. The traditional way of assessing treatment response is to measure the size of the tumors after the treatment. However, efficacy can only be evaluated weeks or months after treatment.

Diffusion weighted MRI has tremendous potential for monitoring early changes in tumor cellularity that are thought to be reflective of treatment response [1]. It provides image contrast determined by microscopic motion of water molecules in the tissue. The mobility of water molecules is highly related to cell density within tumors.

Apparent diffusion coefficient(ADC) is the parameter to measure water molecule motions. In general, water movement inside cells is more restricted than outside. Thus, increasing cell density tends to lower ADC, whereas increased edema(more interstitial water) results in higher ADC values. Therefore, ADC values in treated brain tumors could not only theoretically increase due to cell kill(and thus reduced cell density), but also decrease due to inhibition of edema.

A number of related studies have investigated methods to overcome the overall complicated situation and separate the competing effects. In [2], they observe the shift of ADC histogram and conclude that the mean ADC value increases when tumor cells are killed. In [3], they calculate pixelwise the ADC value changes along with time and display it as a functional diffusion map for correlation with clinical response.

In this paper, we investigate statistical techniques and pattern classification methods to predict tumor responses using ADC map. We extract statistical features from the histogram of tumor ADC values, compare the feature differences between pre- and post-treatment scans, and compare three machine learning-based classification methods. By doing this, we explore the effectiveness of the machine learning approaches in this clinical context.

This paper is organized as follows. The next section describes the image analysis of the ADC map, histogram features and the three classifiers that we compared. In the following result section, we report the results of our comparative study for the different classifiers. The final section offers our discussion on the experimental results as well as our future work.

2 Method

2.1 Image Protocols and Image Analysis

ADC map is calculated from diffusion weighted(DW) images. DW images can be acquired with echo-planar pulse sequences plus DW gradients. The signal intensity of DW images is equal to the signal intensity on a T2-weighted(T2w) image decreased by an amount dependent on the rate of diffusion. [4]

$$SI = SI_0 * e^{-b*ADC} \tag{1}$$

with b being the diffusion sensitivity factor, ADC being the apparent diffusion coefficient, and SI_0 being the signal intensity when b=0 sec/mm^2. With b known, ADC maps are calculated from DW images by equation 1.

Three steps are as follows to get the tumor contour on ADC maps. First of all, radiologists contoured tumors on post-contrast T1-weighted(T1w) images using a semi-automated segmentation tool [5]. Next, tumor contours were mapped from T1w to ADC using rigid body transformation. The mapping was performed using DICOM header information, i.e. image position and image orientation, to compute transformation parameters. Finally, radiologists visually evaluated the contours on ADC images and manually corrected the tumor contours on ADC. An example of the mapping from T1w to ADC is shown in figure 1.
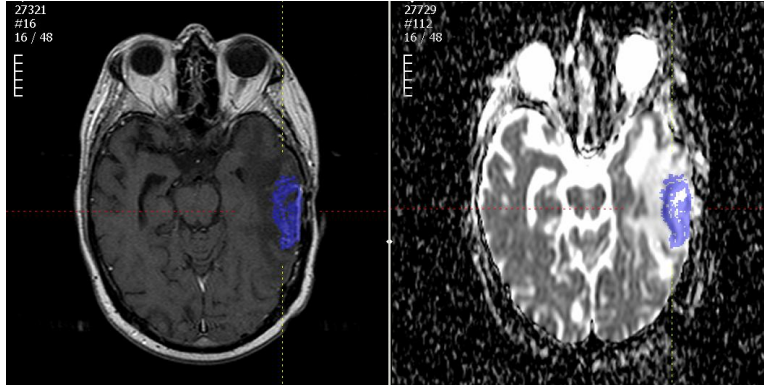
Fig. 1. Examples of the tumor region mapping from post-contrast T1w to ADC map: on the left is the post contrast T1w image with manually contoured tumor; on the right is the ADC map with mapped tumor.

Afterwards, the histogram of the ADC value within the tumor region was obtained. Figure 2 shows two examples of tumor ADC histograms for both pre- and post-treatment. The upper histogram shows the ADC value distribution before the drug treatment, while the lower one shows the ADC value distribution after the drug treatment. On the left is an example of non-responding tumors, while on the right is an example of responding tumors.

2.2 Feature Extraction and Classification

The difference of the features extracted from pre- and post-treatment histograms are used as the input to a tumor response classifier.

According to the clinical studies [2, 6, 1, 7, 3], the ADC value should change after treatment. In our data set, we observe the histograms for both responding and non-responding tumors. We find out that histograms change not only in mean, but also in shapes. Therefore, we bring in the idea of finding the patterns in ADC histogram changes by use of statistical classification methods.

The features we get from histograms are statistical features of the distribution of the ADC values within the tumor: mean, standard deviation, skewness, kurtosis, median, IQR(interquartile range), 25% percentile, and 75% percentile.

We obtain 8-dimensional feature vectors for both pre- and post-treatment tumor histograms. Afterwards, we calculate the difference between pre- and post-treatment tumor histogram by calculating both the absolute change and the change rate of the features. Therefore, we have 16-dimensional vector as the difference feature vector. Besides, we apply the earth mover's distance(EMD) [8, 9] as a metric to directly evaluate the distance between the histograms. The calculated EMD value is appended as the 17th element in the difference feature vector.
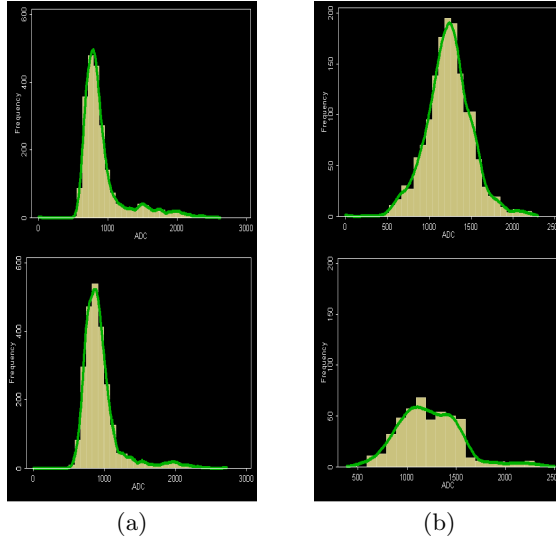
Fig. 2. Examples of histograms from two tumors and two time points of pre-(top row) and post-treatment(bottom row). (a):example of non-responding tumors. (b):example of responding tumors.

The 17-dimensional difference feature vector will be the input to the classifier. For classification, we investigate three classification techniques with different characteristics: fisher linear discriminant analysis, AdaBoost and random forests classifier. We will discuss the reasons why we choose these three classifiers in the results section.

Fisher's linear discriminant analysis(FLDA)[10] is a classification method that projects high-dimensional data onto a line, and perform classification in one dimensional space. The criterion for classification is to maximize the distance between the projected mean between classes and minimize the projected variance of each class. For our two-class case, the cost function and the solution are:

$$\max J(w) = \frac{|m_1 - m_2|^2}{s_1^2 + s_2^2} \quad with \quad solution: \quad w = S_w^{-1}(m_2 - m_1) \qquad (2)$$

where S_w stands for the within-class scatter matrix, m represents a mean, s^2 represents a variance, and the subscripts denote the two classes. As for classification criteria, assuming we have the projected class means well separated, we can choose the average of the two projected means as a threshold for classification.

$$y(x) = w^T x - 0.5 * w^T(m_1 + m_2) \qquad (3)$$

The AdaBoost algorithm, introduced by Freund and Schapire [11], is a boosting algorithm that can combine simple and moderately accurate classifiers into a final strong classifier to improve the final accuracy. It is iterative algorithm. In

each iteration, a weak classifier is selected to minimize the average training error. Afterwards, the weights on training samples are redistributed in such a way that the weight of accurately classified samples will be reduced while the weight of ill classified samples is raised. Therefore, AdaBoost "focuses in" on the informative or "difficult" ones [10]. The final classifier aggregates the selected weak classifier from each iteration, and the vote for each weak classifier is a function of its accuracy.

Random forests(RF) [12] is a classifier that combines many decision trees. Each tree depends on values of a random vector sampled independently and with equal distribution. Each tree casts a unit vote for the most popular case at input, and random forests outputs the class that is the mode of the classes output by individual trees. Breiman [13] suggests the generalization error for forests converges to a limit as the number of trees in the forest becomes large. The error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of features to split each node yields error rates that compare favorably to Adaboost but are more robust with respect to noise.

3 Results

3.1 Experimental Design

We included a total of 86 patients with GBM in our preliminary study. Tumors were diagnosed by board-certified radiologists as responding or non-responding to drugs based on the size change according to later scans. All the ones that present over 50% increase in volume is defined as non-responders, whereas the rest are defined as responders. The baseline scans and follow-up scans were 5-7 weeks apart. The DWI was performed in three or six orthogonal directions and diffusion weighting is b=1000 sec/mm^2. The axial plane resolution for DWI has 0.9375mm by 0.9375mm or 1.797mm by 1.797mm pixel size. The slice thickness for DWI is 3,5, or 7mm.

The statistical features were extracted from histograms of ADC values within the tumor region for both pre- and post-treatment scans. The difference between pre- and post-treatment features was calculated as the input to the classifiers. FLDA, AdaBoost, and RF tree classifiers were applied to the data, and results from the three classifiers were compared. We implemented FLDA in Matlab, while we used AdaBoost and RF classifier implemented in the open source data mining software Weka [14]. We validated the performance by LOOCV method.

3.2 Classification Performance

FLDA was evaluated with all permutations of 2-feature pairs for our ten-dimensional feature space. Among all the 136 combinations of feature pairs, the best classifier was with kurtosis ratio and 75 percentile difference, resulting in a correctly classified rate of 67.44%. Figure 3 shows the scatter plot of the data samples with the two features.
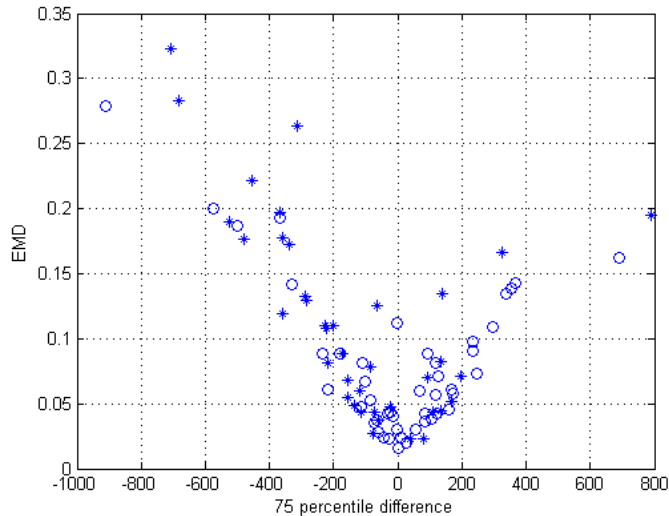
Fig. 3. The feature pair with the highest accuracy: ∗ denotes responding samples, while o denotes non-responding samples.

Table 1. Comparison between AdaBoost and random forests classifier

$Classifier$	$Sensitivity$	$Specificity$	$Accuracy$
AdaBoost	67.5%	63%	65.12%
Random forests	67.5%	63%	65.12%

Visual inspection on Figure 3 promptly reveals that our data is not linearly separable even in the feature space with the best classification rate. This indicates that non-linear classifiers may be more effective in solving our problem. For this reason, we also considered AdaBoost and RF classifiers. Our LOOCV experiments with the Weka implementation of AdaBoost resulted in 65.12% correct classification rate with 10 learning iterations in average. Not only the resulting accuracy is worse than the one for FLDA, but AdaBoost also selected the median difference, kurtosis ratio, 25 percentile difference, STD difference, skewness ratio, and kurtosis difference, which are different from the ones selected by the FLDA.

Next we evaluated the RF classifier as another non-linear classification approach. The report [15] suggests that RF classifier performs quite well, even in the presence of noise in training data, while AdaBoost is susceptible to the noise in training data in comparison with the bagging algorithm [16]. The results of
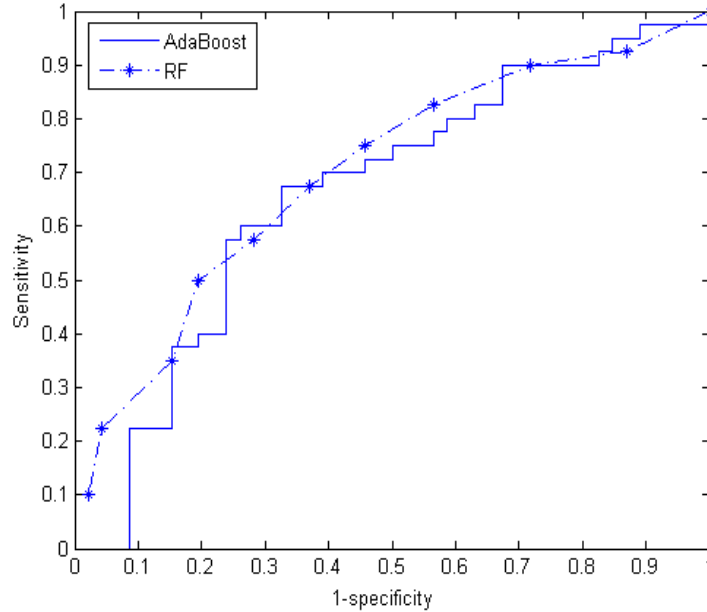
Fig. 4. The ROC curve drawn from Weka on leave-one-out cross validation, with green line for random forests classifier and red line for AdaBoost classifier.

our experiment with the Weka implementation show that the final random forest is composed of 10 trees, each of which is constructed considering five random features. The LOOCV accuracy of the resulting system was 65.12%, the same as AdaBoost classifier.

In table 1, the sensitivity, specificity, and accuracy drawn from Weka report for AdaBoost and RF classifier are compared. With the current dataset, AdaBoost and RF classifier report the same results, yet worse than FLDA.

4 Discussion

In our preliminary study, we exploited statistical pattern classification approaches towards early detection of treatment response using an ADC map.

Cell density and edema may be reflected in ADC values before size changes are apparent on standard MRI sequences. Therefore ADC holds promise as a biomarker, both in determining which tumors are more likely to respond to treatment, and to determine which tumors are actually responding. This will have major implications for clinical trials.

With our current dataset, we obtain comparable performance between all three classifiers tested. More future work will be to use mixture models to quan-

titize the tumor part and edema part since they show different patterns in histogram changes.

References

1. Ross, B.D., Moffat, B.A., Gebarski, S.S., Quint, D.J., Johnson, T.D., Junck, L., Robertson, P.L., Muraszko, K.M., Dong, Q., Meyer, C.R., Bland, P.H., McConville, P., Geng, H., Rehemtulla, A., Chenevert, T.L.: Evaluation of cancer therapy using diffusion magnetic resonance imaging. Molecular Cancer Therapeutics **2** (2003) 581–587
2. Chenevert, T.L., Stegman, L.D., Taylor, J.M.G., Robertson, P.L., Greenberg, H.S., Rehemtulla, A., etc., B.D.R.: Diffusion magnetic resonance imaging: an early surrogate marker of therapeutic efficacy in brain tumors. Journal of the National Cancer Institute **92**(24) (2000)
3. Moffat, B.A., Chenevert, T.L., Lawrence, T.S., Meyer, C.R., Johnson, T.D., Dong, Q., Tsien, C., Mukherji, S., Quint, D.J., Gebarski, S.S., Robertson, P.L., Junck, L.R., Rehemtulla, A., Ross, B.D.: The functional diffusion map: an imaging biomarker for the early prediction of cancer treatment outcome. Neoplasia **8(4)** (2006) 259–267
4. Pamela W. Schaefer, M.: Diffusion-weighted mr imaging of the brain. Radiology **217** (2000) 331–345
5. Otsu, N.: A threshold selection method from gray level histograms. IEEE Trans. Systems, Man and Cybernetics **9** (1979) 62–66
6. Barker, F.G., Prados, M.D., Chang, S.M., Gutin, P.H., Lamborn, K.R., Larson, D.A., Malec, M.K., McDermott, M.W., Sneed, P.K., Wara, W.M., Wilson, C.B.: Radiation response and survival time in patients with glioblastoma multiforme. J Neurosurg **84(3)** (1996) 442–448
7. Hamstra, D.A., Chenevert, T.L., Moffat, B.A., Johnson, T.D., Meyer, C.R., Mukherji, S., Quint, D.J., Gebarski, S.S., Fan, X., Tsien, C., Lawrence, T.S., Junck, L.R., Rehemtulla, A., Ross, B.D.: Evaluation of the functional diffusion map as an early biomarker of time-to-progression and overall survival in high-grade glioma. PNAS **102** (2005) 16759–16764
8. Y. Rubner, C.T., Guibas., L.J.: A metric for distributions with applications to image databases. (1998) 59–66
9. Ling, H., Okada, K.: An efficient earth mover's distance algorithm for robust histogram comparison. Volume 29(5). (2007) 840–863
10. Duda, R.O., Hart, P.E., Stork, D.H.: Pattern Classification. Wiley Interscience (2000)
11. Freund, Y., Schapire, R.E.: A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence **14** (1999) 771–780
12. Ho, T.K.: Random decision forest. In: Proceedings of the 3rd Internation Conference on Document Analysis and Recognition. (1995) 278–282
13. Breiman, L.: Random decision forest. Machine Learning **45** (2001) 5–32
14. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco (2005)
15. Pal, M.: Random forests for land cover classification. In: Geoscience and Remote Sensing Symposium, Proceedings of IEEE International. (2003) 3510– 3512
16. Breiman, L.: Bagging predictors. Machine Learning **24**(2) (1996) 123–140