# Histogram-Based Classification with Gaussian Mixture Modeling for GBM Tumor Treatment Response using ADC Map

Jing Huo[1], Hyun J. Kim[1], Whitney B. Pope[1], Kazunori Okada[2], Jeffery R. Alger[1,3], Yang Wang[1], Jonathan G. Goldin[1], Matthew S. 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 Department of Computer Science, College of Science and Engineering, 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 study applied the Gaussian Mixture Model (GMM) to tumor apparent diffusion coefficient (ADC) histogram to evaluate glioblastoma multiforme (GBM) tumor treatment response using diffusion weighted (DW) MR images. ADC mapping, calculated from DW images, has been shown to reveal changes in the tumor's microenvironment preceding morphologic tumor changes. In this study, we investigated effectiveness of features that represents changes from pre- and post-treatment tumor ADC histograms to detect treatment response. This work mainly contributes to model the ADC histogram as the composition of two components, fitted by GMM with expectation maximization (EM) algorithm. For both pre- and post-treatment scans taken 5-7 weeks apart, we obtained the tumor ADC histogram, calculated the two-component features, as well as the other standard histogram-based features, and applied supervised learning for classification. We evaluated our approach with data of 85 patients with GBM under chemotherapy, in which 33 responded and 52 did not respond based on tumor size reduction. We compared AdaBoost and random forests classification algorithms, using ten-fold cross validation, resulting in the best accuracy of 69.41%.

**Keywords:** GBM, ADC, histogram, GMM-EM, AdaBoost, Random Forest.

## 1. INTRODUCTION

Traditional way to assess treatment response of Glioblastoma multiforme (GBM) brain tumor is tumor size change before and after chemotherapy or radiotherapy. Apparent diffusion coefficient (ADC) map, calculated from diffusion weighted (DW) MR images, has the potential to work as surrogate biomarker to reveal changes in the tumor microenvironment that precede morphologic tumor changes[1].

DW MRI can provide information related to microscopic environment at cellular level. Because water diffusion is strongly affected by molecular viscosity and membrane permeability between intra- and extracellular compartments, DW MRI can be used to characterize highly cellular regions of tumors versus acellular regions. The problem of treatment response detection can be manifested as a change in cellularity within the tumor over time, which is much earlier and more sensitive before the tumor size change is visible[1].

Apparent diffusion coefficient (ADC) is the parameter to measure water molecule motions at cellular level. A number of studies have been going on to explore quantitative results to correlate the ADC changes and the treatment response outcomes. Ross et al. reported ADC value increase significantly for the effective therapeutic intervention in pre-clinical studies and presented two patients to support this hypothesis in preliminary clinical study[1,2]. Moffat et al. calculated voxel-wise the tumor ADC value changes along with time and displayed it as a functional diffusion map for

correlation with clinical response[3]. In our previous work, we applied quantitative ADC histogram analysis to baseline and follow-up contrast-enhanced tumors and exploited pattern classification techniques to evaluate treatment response[4].

The tumor ADC change after treatment is complicated and affected by two main factors. In general, water movement inside cells is more restricted than outside. Thus, increasing cell density tends to lower ADC values, whereas increased edema (more interstitial water) results in higher ADC values. Therefore, theoretically, ADC values in treated brain tumors could not only increase due to the cell kill (and thus reduced cell density), but also decrease due to inhibition of edema. None of the listed studies above have specified the separate effects.

In this paper, we investigated explicit quantitative ways to deal with competing effects of cellularity and edema for early treatment response of GBM brain tumors. Pope et al.[5] applied Gaussian Mixture Model (GMM) fitting to baseline tumor ADC histograms in his study, and used the lower peak mean value for tumor recurrence prediction. In this paper, we applied the same method to tumor ADC histograms on both baseline and follow-up ADC maps. We made the assumption that there are two heteroscedastic components within the tumor. Then we fitted the tumor ADC histogram with GMM[6] and calculated the features of the two components with expectation-maximization (EM) algorithm[6] as well as the original tumor ADC histogram features. In this way, we explored the effectiveness of the machine learning approaches in this clinical context.

## 2. METHODOLOGY

### 2.1 Image Protocols and Image Analysis

All ADC maps were calculated from DW images with the same in-house software using a two-point method as shown in the following equation:

$$ADC = -\ln[S(b)/S(0)]/b \tag{1}$$

with b being the diffusion sensitivity factor ranging between 700 and 1000 s/mm$^2$, $S(0)$ and $S(b)$ being the image intensity when b=0 and b=700 or 1000 s/mm$^2$. For DWI trace images, we calculated ADC maps from DW images by equation (1). For DTI, we calculate ADC for each orientation and average them as the ultimate ADC map.

Three steps were followed to get tumors contoured on ADC maps: First, radiologists contoured tumors on post-contrast T1-weighted (T1w) images using a semi-automated segmentation tool[7]. The tool is initialized by a user-defined line starting from the inside of the tumor and ending at the outside of the tumor. Intensity histogram on the line is used to find the optimal threshold value by use of Otsu's thresholding method. Afterwards, seeded region growing is applied to segment the tumor. Next, tumor contours were mapped from T1w to ADC using a 3D rigid body transformation. For each voxel in the source tumor region of interest (ROI), a physical location is first calculated, and then the coordinates in the target images are calculated accordingly. The transformation matrix to calculate the physical locations is formed for source and target images respectively by the information stored in DICOM header, i.e. pixel size, slice thickness, image position and image orientation. The physical location is calculated as follows:

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = \begin{bmatrix} X_x\Delta_i & Y_x\Delta_j & Z_x\Delta_k & S_x \\ X_y\Delta_i & Y_y\Delta_j & Z_y\Delta_k & S_y \\ X_z\Delta_i & Y_z\Delta_j & Z_z\Delta_k & S_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i \\ j \\ k \\ 1 \end{bmatrix} \tag{2}$$

With $\Delta_{i,j,k}$ as the voxel size read from the tag "pixel spacing" and "slice thickness"; $X_{x,y,z}, Y_{x,y,z}$ as image orientation read from the tag "image orientation" which specifies the orientation of the image frame rows and columns, $Z_{x,y,z}$ as the z-direction orientation calculated from $X_{x,y,z}, Y_{x,y,z}$; $S_{x,y,z}$ as read from the tag "patient position" which specifies the anterior-left-upper corner; i, j, k as voxel index; and $P_{x,y,z}$ as the calculated physical location of the voxel in millimeters. Finally, radiologists visually evaluated the contours on ADC images and manually corrected the tumor contours on ADC. Afterwards, the histogram of the ADC value within the tumor region was obtained. Figure 1(A) shows an example of the mapping from the T1w image to the ADC map.

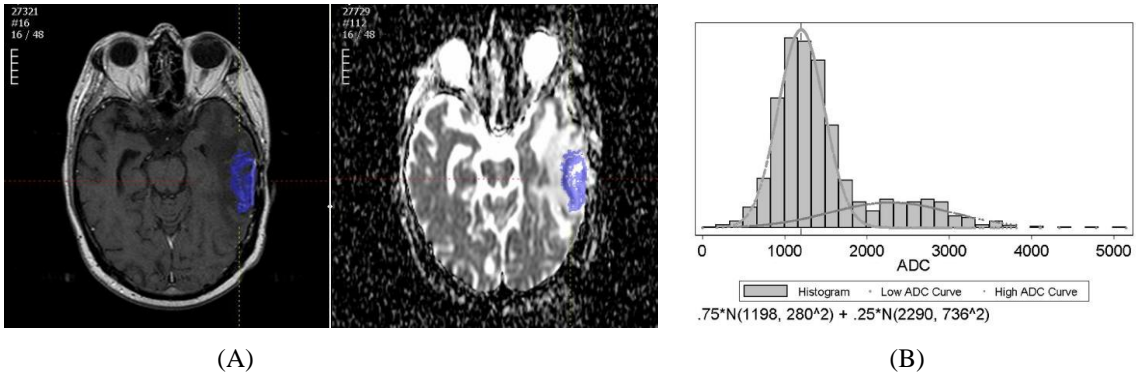(A)                                              (B)

Figure 1(A). An example of the tumor region mapping from post-contrast T1w to ADC map: on the left is the post contrast T1w image with the tumor contour; on the right is the ADC map with mapped tumor; (B) An example of the tumor ADC histogram fitted by two-component Gaussian mixtures.

Afterwards, the histogram of the ADC value within the tumor region was obtained. Figure 2 shows examples of tumor ADC histograms for both pre-and post-treatment. 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 responding tumors, while on the right is an example of non-responding tumors.
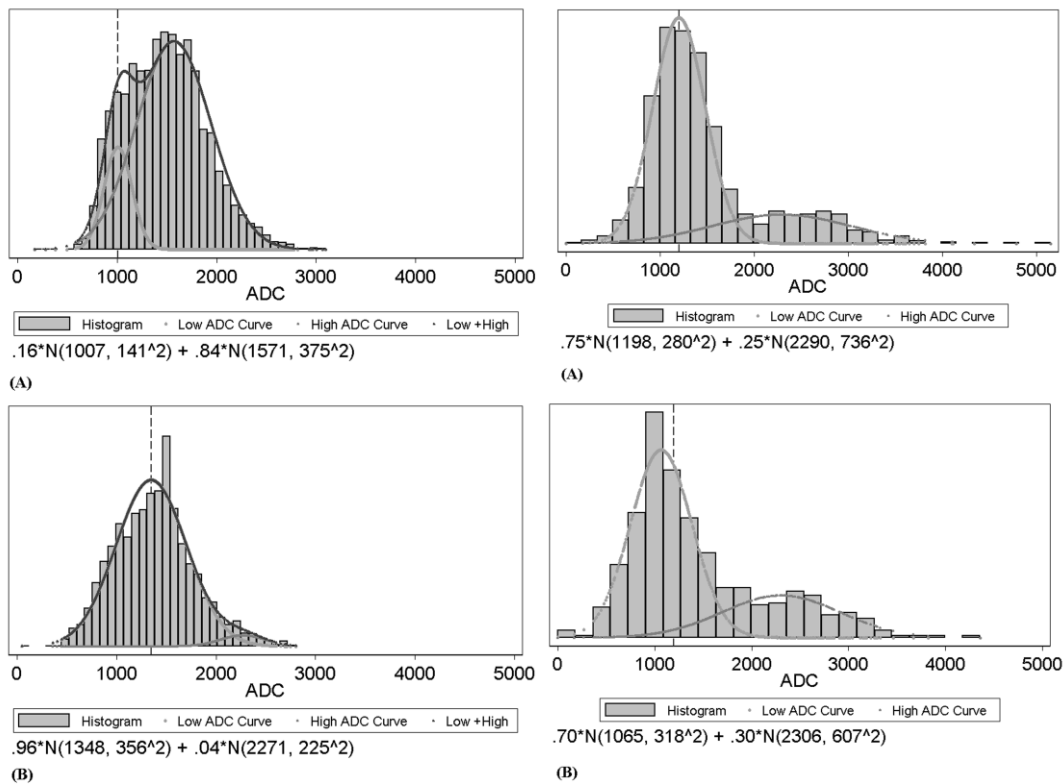


Figure 2. Examples of histograms from two tumors and two time points of pre-(top row) and post-treatment (bottom row). (A):example of responding tumors. (B): example of non-responding tumors.

## 2.2 Feature Extraction and Classification

The absolute change of the features extracted from pre- and post-treatment tumor ADC histograms were used as the input to a tumor response classifier.

In general, water movement inside cells is more restricted than outside. Thus, increasing cell density tends to lower ADC values, whereas increased edema (more interstitial water) results in higher ADC values. Due to the competing effects of tumor cell density and edema, we made the assumption that the obtained tumor ADC histogram was composed of two components: cellularity and edema. We assumed that the component with lower peak accounts for the tumor cellularity effect, while the component with higher peak accounts for the edema effects. We used a two component GMM as shown in equation (3) to fit the ADC histogram for both baseline and follow-up scans and applied EM algorithm to estimate GMM parameters.

$$f(x) = \sum_{i=1}^{2} \alpha_i G_i(x), \text{ with } G_i(x) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \tag{3}$$

The EM algorithm can be used to compute the parameters of a parametric mixture model distribution: the weight of the components $\alpha_i$, the Gaussian parameters $\mu_i$, and $\sigma_i$. It is an iterative algorithm with two steps: an expectation step (E-step) and a maximization step (M-step). In the E-step, with the current parameter estimates of the mixture components, the algorithm calculates the expectation values for the membership variables of all data points. In the (m+1) iteration, the expectation is:

$$p_{ni}^{(m+1)} = \frac{\alpha_i^{(m)} G_i^{(m)}(x)}{\sum_{i=1}^{2} \alpha_i^{(m)} G_i^{(m)}(x)} \tag{4}$$

In M-step, the algorithm maximizes the likelihood function and updates the corresponding parameters. The following solutions can be developed:

$$\mu_i^{(m+1)} = \frac{\sum_{n=1}^{N} y_n p_{ni}^{(m+1)}}{\sum_{n=1}^{N} p_{ni}^{(m+1)}} \quad (\sigma_i^{(m+1)})^2 = \frac{\sum_{n=1}^{N} (y_n - \mu_i^{(m+1)})^2 p_{ni}^{(m+1)}}{\sum_{n=1}^{N} p_{ni}^{(m+1)}} \quad \alpha_i = \frac{1}{n} \sum_{n=1}^{N} p_{ni}^{(m+1)} \tag{5}$$

The features we obtained from the GMM-EM are named as lower peak mean (LPM), lower peak variance (LPV), lower peak proportion (LPP), higher peak mean (HPM), higher peak variance (HPV) and higher peak proportion (HPP). Figure 1(B) shows an example of tumor ADC histogram fitted by GMM. The histograms shown in Figure 2 are fitted by the two-component GMM.

In our previous study[4], we explored the general statistical features from ADC histograms: mean, standard deviation, skewness, kurtosis, median, IQR (interquartile range), 25th percentile, and 75th percentile. We appended these features because they showed potential in early detection of treatment responders or non-responders.

We obtain 14-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 the absolute change of the features. Therefore, we have 14-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 pre- and post-treatment tumor ADC histograms. Intuitively, if the histograms are interpreted as two different ways of piling up a certain amount of dirt over the region D, the EMD is the minimum cost of turning one pile into the other; where the cost is assumed to be amount of dirt moved times the distance by which is moved. The calculated EMD value is appended as the 15th element in the difference feature vector.

The 15-dimensional difference feature vector will be the input to the classifiers. For classification, we investigated two classification techniques with different characteristics: AdaBoost and random forests (RF) classifier.

The reason we chose them was that they both have feature selection mechanism. By applying these two classification techniques, we are seeking for the best features that would separate responders from non-responders which is very meaningful in clinical practice.

The AdaBoost algorithm, introduced by Freund and Schapire[10], is an iterative algorithm that can boost weak classifiers into a strong classifier and improve the final classification accuracy. In each iteration, a feature is working as a weak classifier and the best feature 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 on the most "difficult" ones[6]. The final classifier aggregates the selected weak classifier from each iteration, and the weight for each weak classifier depends on its error rate. However, AdaBoost can be sensitive to noise and may introduce the overfitting problem.

Random forests (RF)[11] 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[12] 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

A total of 85 patients with GBM were included 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% decrease in volume is defined as responders, whereas the rest are defined as non-responders. The baseline scans and follow-up scans were 5-7 weeks apart. The T1 images have slice thickness from 1mm to 5mm, while the DWI images have a slice thickness of 5 or 6 mm. The axial plane resolution for T1w has 0.9375mm by 0.9375mm pixel size, while DW images have the same or 1.797mm by 1.797mm pixel size.

The six parameters from fitted GMM model and eight statistical features from ADC histograms were obtained within the tumor region for both pre- and post-treatment scans. The absolute difference between pre- and post-treatment features as well as EMD between pre- and post-treatment ADC histograms were calculated as the input to the classifiers. AdaBoost and RF tree classifiers were applied to the difference feature vectors, and results from the two classifiers were compared. AdaBoost and RF classifier were implemented in the open source data mining software Weka[13]. The performance was validated by 10-fold cross validation method.

## 3.2 Classification Performance

The experiment with AdaBoost resulted in 68.24% correct classification rate with 10 learning iterations in average. The selected features are LPM, HPV, 25th percentile, kurtosis, LPP, IQR.

The experiment with RF classifier showed that the final random forest is composed of 10 trees, each of which is constructed considering five random features. The 10-fold cross validation accuracy of the resulting system was 69.41%, moderately better than the AdaBoost classifier. In table 1, the sensitivity, specificity, and accuracy drawn from Weka report for AdaBoost and RF classifier are compared. In figure 3, ROC curves for AdaBoost and RF classifier are compared. With the current dataset, RF classifier works slightly better than AdaBoost classifier.

Compared to our previous exploratory study[4], the overall accuracy is increased from 67.44% to 69.41%, which showed improvement of the classification system using ADC histogram as biomarker to evaluate treatment response. For RF classifier, the performance has been improved from 65.12% to 69.41% with mixture model features included. As for AdaBoost classifier, the selected features included three mixture model features out of the total 15 features. These provided strong proof to support the assumption that the ADC histogram is a mixture of two components: cellularity and edema.

| Classifier | Sensitivity | Specificity | Accuracy |
|---|---|---|---|
| AdaBoost | 61.54% | 71.19% | 68.24% |
| Random forest | 62.96% | 72.41% | 69.41%, |

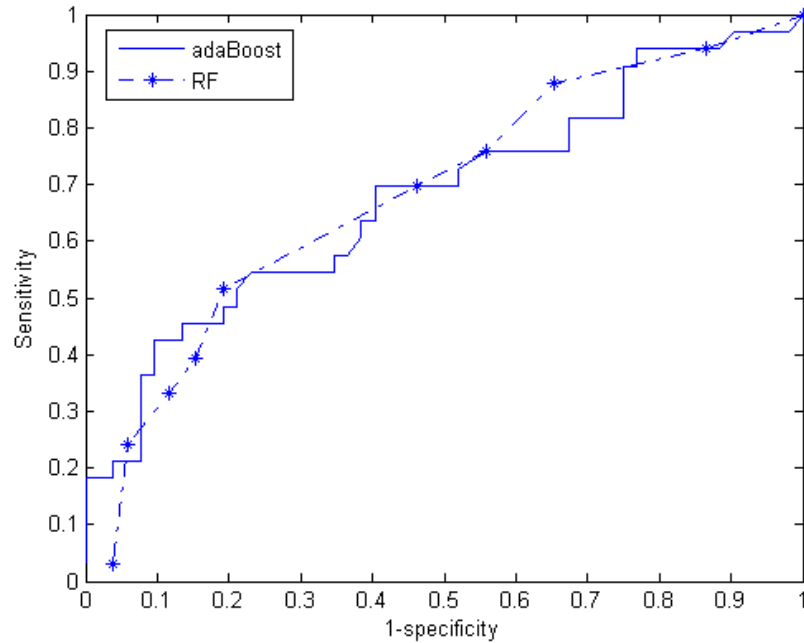Table 1 Comparison between AdaBoost and random forests classifier



Figure 3 The ROC curve drawn from Weka on ten-fold cross validation.

## 4. DISCUSSIONS

In our preliminary study, we applied GMM to model the ADC histogram to interpret the competing effects of cellular density and edema in tumor ADC change. Our system with the proposed new features resulted improved performance to the previous systems found in literature[4] in using ADC maps to early detect the treatment response.

With our dataset, we see potentials of using ADC map 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. For future work, a larger set of subjects will definitely be needed to have strong conclusions.

## REFERENCE

[1] Ross, B.D., Moffat, B.A., Lawrence, T.S., Mukherji, S.K., Gebarski, S.S., Quint, D.J., Johnson, T.D., Junck, L. and 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(6), 581-587 (2003).
[2] Chenevert, T.L., Stegman, L.D., Taylor, J.M., Robertson, P.L., Greenberg, H.S., Rehemtulla, A. and Ross, B.D., "Diffusion magnetic resonance imaging: an early surrogate marker of therapeutic efficacy in brain tumors", Journal of the National Cancer Institute 92(24), 2029-36 (2000).

[3] Moffat, B.A., Chenevert, T.L., Meyer, C.R., Mckeever, P.E., Hall, D.E., Hoff, B.A., Johnson, T.D., Rehemtulla, A. and Ross, B.D., "The Functional Diffusion Map: An Imaging Biomarker for the Early Prediction of Cancer Treatment Outcome", Neoplasia 8(4), 259–267 (2006).

[4] Huo, J., Pope, W.B., Okada, K., Brown, M.S., Kim, H.J., Alger, J.R., Wang, Y and Goldin, J.G., "Early Detection of Treatment Response for GBM Brain Tumor using ADC Map of DW-MRI", Proc MICCAI 2008 workshop on Computational Diffusion MRI, 183-190 (2008).

[5] Pope, W.B., Kim, H.J., Huo, J, Alger, J.R., Brown, M.S., Gjertson, D., Sai, V., Lai, A., Nghiemphu, P., Cloughesy, T. and Goldin, J.G., "ADC histogram analysis predicts response to bevacizumab treatment in GBM patients". Accepted to Radiology on Jan $12^{th}$ (2009).

[6] Duda, R.O., Hart, P.E. and Stork, D.H., [Pattern classification], Wiley Interscience, (2000).

[7] Otsu, N., "A threshold selection method from gray level histograms", IEEE transactions on Systems, Man and Cybernetics 9(1), 62-66 (1979).

[8] Rubner, Y., Tomasi, C. and Guibas, L.J., "A metric for distributions with applications to image databases", Proceedings of ICCV, 59-66 (1998).

[9] Ling, H. and Okada, K., "An Efficient Earth Mover's Distance Algorithm for Robust Histogram Comparison", IEEE transactions on Pattern Analysis and Machine Intelligence 29(5), 840-853 (2007).

[10] Freund, Y. and Schapire, R.E., "A short introduction to boosting", Journal of Japanese Society for Artificial Intelligence 14, 771-780 (1999).

[11] Ho, T.K., "Random decision forest", Proc of the 3rd International Conference on Document Analysis and Recognition, 278-282 (1995).

[12] Breiman, L., "Random decision forest", Machine Learning 45, 5-32 (2001).

[13] Witten, I.H. and Frank, E., [Data Mining: Practical machine learning tools and techniques], Morgan Kaufmann, San Francisco, (2005).