CADrx for GBM Brain Tumors: Predicting Treatment Response from Changes in Diffusion-Weighted MRI

Jing Huo[1] MS, Matthew S Brown[1] PhD, and Kazunori Okada[2] PhD
[1]*Radiological Sciences, David Geffen School of Medicine at UCLA, USA*
[2]*Computer Science, San Francisco State University,USA*

ABSTRACT
The goal of this chapter is to describe a computer-aided therapeutic response assessment (CADrx) system for early prediction of drug treatment response for glioblastoma multiforme (GBM) brain tumors with diffusion weighted (DW) MR images. In conventional Macdonald assessment, tumor response is assessed nine weeks or more post-treatment. However, we will investigate the ability of DW-MRI to assess response earlier, at five weeks post treatment. The apparent diffusion coefficient (ADC) map, calculated from DW images, has been shown to reveal changes in the tumor's microenvironment preceding morphologic tumor changes. ADC values in treated brain tumors could theoretically both increase due to the cell kill (and thus reduce cell density) and decrease due to inhibition of edema. In this chapter, we investigate the effectiveness of features that quantify changes from pre- and post-treatment tumor ADC histograms to detect treatment response. There are three parts in this technique: First, tumor regions were segmented on T1w contrast enhanced images by Otsu's thresholding method, and mapped from T1w images onto ADC images by a 3D region of interest (ROI) mapping tool. Second, ADC histograms of the tumor region were extracted from both pre- and five weeks post-treatment scans, and fitted by a two-component Gaussian mixture model (GMM). The GMM features as well as standard histogram-based features were extracted. Finally, supervised machine learning techniques were applied for classification of responders or non-responders. The approach was evaluated with a dataset of 85 patients with GBM under chemotherapy, in which 39 responded and 46 did not, based on tumor volume reduction. We compared adaBoost, random forest and support vector machine classification algorithms, using ten-fold cross validation, resulting in the best accuracy of 69.41% and the corresponding area under the curve (Az) of 0.70.

1 INTRODUCTION
The general aim of this chapter is to discuss the application of machine learning techniques in treatment response monitoring using advanced MR imaging protocol in clinical trials. Computer-aided diagnosis (CADx) in GBM brain tumor is an active research area, and many promising MR methods have been developed for detecting and characterizing cancer, its treatments and adverse effects, e.g. T1-weighted MR, T2-weighted MR, MR spectroscopy, perfusion-weighted MR, and diffusion-weighted MR. In our study, we focused on T1-weighted and DW-MRI. We name our proposed system computer-aided therapeutic response assessment (CADrx).

We developed a computer-aided system to explore an early imaging biomarker using diffusion MR in a phase II clinical trial. Conventionally, tumor size change on T1w images is the only imaging biomarker that is accepted by the FDA as a surrogate endpoint of clinical outcome after chemotherapy and radiotherapy for phase III trials (2007). Diffusion MRI has been explored for early detection of GBM brain tumor treatment response prior to the tumor size changes.

Machine learning and statistical pattern recognition have potential for significant contributions to the biomedical community because they can improve the sensitivity and/or specificity of detection and diagnosis of disease, while at the same time increasing objectivity of the decision-making process (Sajda, 2006). The need for machine learning is perhaps greater than ever given the dramatic increase in medical data being collected, new detection, and diagnostic modalities being developed, as well as the complexity of the data types and importance of multimodal analysis. In all of these cases, machine learning can provide new tools for interpreting the high-dimensional and complex datasets with which the clinician is confronted (Sajda, 2006). In this chapter we explore three different classification methods: AdaBoost, random forest, and support vector machine in our CADrx system.

2 BACKGROUND

Glioblastoma multiforme (GBM) is the most common and most aggressive type of the primary brain tumors. GBM is an anaplastic, highly cellular tumor with poorly differentiated, round, or pleomorphic cells, occasional multinucleated cells, nuclear atypia, and anaplasia. The median survival time from the time of diagnosis without any treatment is 3 months, but with treatment, survival of 1-2 years is common. Although the prognosis of GBM is uniformly poor, treating patients in an attempt to improve the quality of life is worthwhile. GBM treatment consists of a combination of surgical resection, radiation therapy, and chemotherapy. Surgical resection is the mainstay of GBM treatment, and radiation therapy usually follows surgery. The U.S. Food and Drug Administration approved Avastin (bevacizumab) to treat patients with glioblastoma at progression after standard therapy. Bevacizumab (trade name Avastin, Genentech/Roche) is a humanized monoclonal antibody that recognizes and blocks vascular endothelial growth factor A (VEGF-A). VEGF-A is a chemical signal that stimulates the growth of new blood vessels (angiogenesis). Blood vessels grow uncontrollably in cancer, retinal proliferation of diabetes in the eye, and other diseases. Bevacizumab can block VEGF-A from creating new blood vessels. Bevacizumab was the first clinically available angiogenesis inhibitor in the United States.

The conventional way to assess treatment response in a phase II in vivo clinical trial is based on Macdonald criteria and evaluated on T1-weighted contrast enhanced (T1wCE) MR images. The Macdonald criteria define tumor response by use of tumor size change, steroids, and neurological function. There are four 'response' categories: Complete response (CR) is defined as disappearance of enhancing tumors, off steroids, and neurologically stable or improved. Partial response (PR) is >50% reduction in size of enhancing tumor, steroids stable or reduced, neurologically stable or improved. Progressive disease (PD) is >25% increase in size of enhancing tumor or any new tumor, or neurologically worse, and steroids stable or increased. Stable disease (SD) is all other situations. In our study, we used tumor volume to evaluate tumor sizes. More than 50% increase in volume is considered to be the PD based on the neuro-radiologist's recommendation (Huhn, 1999). However, efficacy can only be evaluated at least 8–10 weeks after treatment.

Diffusion weighted magnetic resonance imaging (DW-MRI) has the potential to work as a surrogate biomarker to reveal changes in the tumor microenvironment that precede morphologic tumor changes (Ross et al., 2003). DW-MRI depends on the microscopic mobility of water. This mobility, classically called Brownian motion, is due to thermal agitation and is highly influenced by the cellular environment of water. 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. Treatment response can be manifested as a change in tumor cellularity, which may precede tumor size changes. Thus, findings on DW-MRI could be an early sign of biologic changes (Padhani et al., 2009).

Apparent diffusion coefficient (ADC), derived from DW-MR images, is the quantitative measurement of water diffusion. ADC maps are calculated from DW-MR images using a two-point method as shown in the following equation:

$$ADC = -\ln[S(b)/S(0)]/b \quad (1)$$

with b being the diffusion sensitivity factor ranging between 700 and 1,000 s/mm^2, S(0) and S(b) being the image intensity when b = 0 and b = 700 – 1,000 s/mm^2. For DWI images, ADC maps are calculated from DW images by equation (1). For diffusion tensor imaging (DTI), ADC maps for each orientation are calculated and averaged as the final ADC map. Figure 1(b) shows an example of a derived brain ADC map.

Figure 1. (a) An example of the tumor segmented on a T1wCE image; (b) An example of the tumor ROI mapped from T1wCE to ADC map; (c) An example of the tumor ADC histogram fitted by two-component Gaussian mixtures.

Table 1. Summary of related methods in brain tumor segmentation. The type abbreviations are NC: Nasopharyngeal carcinoma; MNG: Meningiomas; MG - malignant gliomas; MS – multiple sclerosis.

Computer-aided detection and segmentation of GBM brain tumors is a challenging problem and in Table 1 we present a concise review of the prior art in automatic tumor segmentation. First of all, fuzzy clustering and knowledge-based analysis are popular methods explored by the early pioneers (Phillips et al., 1995; Clark et al., 1998; Fletcher-Heath et al., 2001). Fuzzy C-means (FCM) clustering is a widely used clustering method, since it does not require training data and is therefore operator independent. Phillip et al. (1995) was the first to apply FCM clustering to GBM brain tumor segmentation. He then correlated the segmentation with tumor histology, although he did not give a quantitative validation of the method. Clark et al. (1998) developed a system to segment the gadolinium enhanced GBM. This system was initialized by FCM and followed by knowledge based five-stage processing, evaluated with sixteen volumes from seven patients, and compared with the supervised k-nearest neighbor (kNN) method. The knowledge-based segmentation systems used the brain atlas as the prior information, and Fletcher-Heath et al. (2001) applied the system to the non-enhancing tumor. Secondly, Voxel-based classification method using statistical pattern classification techniques have been explored by others. Vinitski et al. (1999) developed a system using kNN to segment multiple sclerosis (MS) lesions and brain tumors from a limited number of patients. Prastawa et al. (2003) developed a system that models different brain tissues as Gaussian mixtures, added geometric constraints, and uses EM to estimate the model parameters and the segmentation result, with an atlas prior as initialization to EM. The system was extended for GBM tumor segmentation by adding the tumor and edema classes. Zhang et al. (2004) formulated the problem of GBM brain tumor segmentation into a one-class SVM. Corso et al. (2008) developed weighted aggregation (SWA) method based on graph shift algorithm and Bayesian inference for GBM brain tumor segmentation. Most of the studies above use multiple MRI sequences (T1w, T2w, proton density weighted, and Flair) for the automatic tumor and edema detection and segmentation. Liu et al. (2005) developed an interactive system adapting the fuzzy connectedness using multiple MRI sequences. Dube et al. (2007; 2008) used texture features and segmentation by the weighted aggregation (SWA) method for the GBM tumor segmentation on T1wCE images which is similar to part of our study. In our study, we developed a semi-automated method to segment tumors on T1wCE images; in addition, we mapped the tumor contours onto ADC images.

Recent studies that used DWI for GBM early prediction of treatment response are listed in Table 2. Ross et al reported that the ADC value increased significantly in effective therapeutic intervention in pre-clinical studies and presented two patients to support this hypothesis in a preliminary clinical study (Ross et al., 2003; Chenevert et al., 2000). Mardor et al. (2003) applied both low and high b-value and used mean ADC and diffusion index for treatment response evaluation. Moffat et al calculated voxel-by-voxel tumor ADC value changes over time and displayed it as a functional diffusion map for correlation with clinical response (Moffat et al., 2005; Hamstra et al., 2005). They reported that the number of voxels with increased ADC is related to treatment efficacy. Our previous work (Huo et al., 2009) showed promising results for using ADC histogram analysis, and we explored a more sophisticated classifier and designed experiments to show the advantages of two-component histogram modeling.

Table 2. Summary of related methods in GBM tumor treatment response using DWI.

3 MAIN FOCUS OF THE CHAPTER

3.1 Issues, Controversies, Problems

3.1.1 ADC measurement variation and reproducibility evaluation

Quantitative and qualitative response assessment is based on either treatment prediction using baseline ADC data or using ADC changes between baseline and follow-up(s). In the setting of a multi-center, multi-scanner therapeutic clinical trial, it is necessary to evaluate the reproducibility of ADC measurements in order to reliably use ADC values for response assessment.

In this chapter, a quality control (QC) tool is presented to evaluate between-scanner and between-visit variation of ADC measurement in the setting of a real-world multi-center drug clinical trial where multiple centers and multiple scanners are involved. Even following radiation and variable chemotherapy these patients still appear to have consistent normal brain WM ADC values by visual inspection. Thus, the tool was used to evaluate whether a variety of clinical MRI scanner models produce consistent measures of brain white matter ADC, and additionally, to evaluate the scan-rescan reproducibility in measured ADC at two visits.

3.1.2 Tumor definition on ADC maps

It is difficult for radiologists to directly delineate the tumor contours on derived ADC maps, because the GBM tumors are not well defined on these maps. Therefore, we segmented tumors on T1wCE images first, and then mapped the tumor contours onto the corresponding ADC maps.

First, all tumors were segmented on T1wCE images via a semi-automated method using the Otsu thresholding algorithm (Otsu, 1979) and seeded region growing (Adams and Bischof, 1994). Next, the tumor contours were mapped onto ADC maps from T1wCE images by a 3D ROI mapping tool based on the scanner geometry. The mapping technique only transformed voxels within the tumor ROI rather than the whole image volume; thus it was more computationally efficient than a registration technique. However, the mapping tool could not correct for patient motion; thus a board-certified radiologist was required to visually check the mapped results and perform manual corrections when necessary.

3.1.3 Modeling the two competing effects in the ADC change after treatment

The biggest challenge in using ADC as an early biomarker is that there are two competing effects in ADC changes after treatment. In general, water movement inside cells is more restricted than outside cells. Thus, increased 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. To specify the separate effects, we applied a two-component model to fit the tumor ADC histogram (Pope et al., 2009).

3.1.4 Developing a CADrx system in therapeutic response prediction

In previous studies using DWI as an early biomarker, either overall tumor ADC mean change or voxel-by-voxel ADC change is used the quantitative feature; In this chapter, we explore a larger feature set including histogram statistical features and metrics to directly compare histogram differences. Furthermore, we introduce machine learning techniques to perform feature selection and classification to separate responders and non-responders. In so doing, we developed a CADrx system for early therapeutic response prediction.

3.2 Solutions and Recommendations

3.2.1 ADC measurement variation and reproducibility evaluation

A QC method has been developed to evaluate the variation and reproducibility of ADC measurements using normal appearing brain white matter. A fixed size circular 2D ROI (r = 7 pixels) was manually drawn on the normal-appearing brain white matter above the ventricles and confirmed by a board-certified neuro-radiologist (W.P.), as shown in Figure 2. For each ROI, the ADC median and coefficient of variation (CV) were calculated. Median ADC value is used to compare across different scanner models. Median is used here instead of mean to minimize sensitivity to random noise. CV, defined as the ratio of standard deviation (STD) to the mean, is used to evaluate the dispersion of the whole ROI ADC measurements. CV was used instead of standard deviation (STD), because the STD of data should be understood in the context of the mean of the data. Figure 2 shows two images with different image quality, and thus different CVs and medians. Bilgili and Unal (2004) reported that varying ROI sizes in the brain WM did not yield statistically different ADC values.

Box-Cox transformation followed by Shapiro-Wilk normality test was used to meet normality assumptions in median and CV. For baseline inter-scanner variation analysis, differences in baseline ROI ADC median and CV were examined by the ANOVA test across different scanners, different magnetic fields, and different vendors. Between-visit ADC measurement agreement was examined by using an intra-class correlation coefficient (ICC), and differences between the two visits in ADC measurements were evaluated by using a paired t-test.

*Figure 2: Two example ADC maps with brain WM ROIs. The window level and width are set to 2800 and 1500*10^{-6}mm^2/s, respectively. The median and CV for each ROI is as follows: (a) median = 721*10^{-6}mm^2/s, CV = 0.11; (b) median = 884*10^{-6}mm^2/s, CV = 0.18.*

Figure 3 Box plots of the baseline brain WM ADC (units of 10^{-6}mm^2/s) median and CV: (a) Brain WM ROI median; (b) Brain WM ROI CV.

We included 52 patients (31 men and 21 women; age range 19-78 years old; mean age, 52 years old) imaged using 7 scanner models for the baseline ADC variation study. The Shapiro-Wilk normality test showed that the ROI ADC median (p=0.0014) and CV (p=0.0065) were not normally distributed. After log transformation of CV and inverse transformation of the median based on the Box-Cox model, data were normally distributed and thus eligible for the ANOVA test.

Figure 3 displays the box plots of the seven patient groups. The results showed that there was no significant difference in median ADC (p= 0.165) between any two of the seven scanner models. However there was a significant difference in CV (p= 0.0002). The multiple comparison test showed that the differences came from the following scanner pairs with significance levels: 1-7(p=0.033), 2-7(p=0.002), 3-7(p=0.00007), 5-7(p=0.009), 6-7(p=0.023).

For scan-rescan reproducibility analysis, we had 39 patients imaged on 5 scanners models with both baseline and follow-up data available. For each patient, we calculated the ADC changes for the ROI median and CV, and compared the difference across different scanners. The ANOVA test showed that there was no significant difference among the five scanners in terms of median change (p=0.62), and CV change (p=0.71).

When the 39 patients were combined, a paired t-test showed that there was no significant difference between baseline and follow-up ADC values in CV (p=0.44), but a significant difference in median (p=0.01). The intra-class correlation coefficient (ICC) consistency was 0.58 for the median, and 0.56 for CV. Three of the patients from Scanner 1 did not have consistent scanner parameter set up (number of diffusion sensitization directions) for baseline and follow-up scans. With these three patients excluded, the paired t-test between baseline and follow-up median ADC values showed a significance level of p=0.05 for the remaining 36 patients.

3.2.2 Semi-Automated Image Analysis on ADC Maps

All patients were scanned by both T1wCE MR images and DW-MR images. Since it is difficult to segment tumors accurately on derived ADC maps, we segmented tumors on T1wCE images first, and

then mapped the tumor contours onto the corresponding ADC maps.

Tumor segmentation on T1wCE MR images

All tumors were segmented on T1wCE images via a semi-automated method using the Otsu's thresholding algorithm (Otsu, 1979) and seeded region growing (Adams and Bischof, 1994). First, radiologists drew a line from inside of the tumor to the outside of the tumor on the approximate center slice of the tumor. Then intensity values along the line were collected to form a bimodal histogram, and the Ostu's thresholding method was used to find the optimal thresholding value. Afterwards, a 3D seeded region growing was applied to obtain the segmentation results in the whole volume. Radiologists reviewed the results and made manual corrections when necessary. Figure 1(a) shows an example of a segmented tumor on a T1wCE image. Threshold-based segmentation methods are a standard approach to calculation of tumor volume.

Tumor mapping from T1wCE images to ADC maps

It is difficult for radiologists to directly delineate the tumor contours on ADC maps, and the scanner-provided T1w images and the derived ADC maps are not inherently co-registered, because they have different slice thickness, different field of view (FOV), and different image resolutions. Therefore, a 3D ROI mapping tool was developed to map the tumor ROIs from T1wCE images onto ADC images based on the scanner geometry.

The mapping tool used an affine transformation with the parameters extracted from the DICOM header based on physical locations. Equation 2 shows the calculation of the 3D physical location of a given voxel. $\Delta i,j.k$ is the physical voxel size read from the tag "pixel spacing" and calculated from "slice location"; $X_{x,y,z}$, $Y_{x,y,z}$ is image orientation read from the tag "image orientation" which specifies the orientation of the image frame rows and columns, $Z_{x,y,z}$ is the z-direction orientation calculated from $X_{x,y,z}$, $Y_{x,y,z}$; $S_{x,y,z}$ is read from the tag "patient position" which specifies the physical location of the patient's anterior-left-upper corner; i, j, k are voxel index; and $P_{x,y,z}$ are the calculated physical location of the voxel in millimeters. The transformation matrices are calculated for both source and target ROI respectively. For each voxel in the source ROI, the physical location is first calculated, and then the inverse operation is performed to calculate the corresponding voxel coordinates of the target ROI. Finally, radiologists visually checked the contours on ADC maps and manually corrected the tumor contours on ADC when necessary. Figure 1(b) shows an example of the mapped tumor ROI on the ADC image.

Equation 2. The physical location calculation of a voxel (i,j,k).

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

Segmentation Performance

Figure 4 shows four examples of segmentation on T1wCE images and the mapped results on the derived ADC maps.

For quantitative evaluation of the tumor segmentation mapping results, we randomly selected 31 subjects' baseline data. The 31 tumors are from an ADC mapping database, 20 of which have different image resolutions between the T1wCE and ADC images in all three dimensions and 11 of which have exactly the same 3D image resolution in both modalities. We calculated the overlap ratio between the mapped ROI generated automatically by the tool and an ROI corrected by a neuro-radiologist. The overlap ratio (OR) is defined by Equation 3, where A and B are two tumor ROIs and size(.) is the number of voxels in that ROI. The results are shown in Table 3 with 20 out of 31 ROIs (64.5%) having an overlap ratio over 90%.

$$2 * size(A \cap B) / (size(A) + size(B)) \quad (3)$$

Table 3. Distribution of overlap ratios.

Figure 4. a)-d) and i)-l) show four examples of tumor segmentations on T1wCE images; e)-h) and m)-p) show the corresponding mapped tumor contours on ADC maps.

3.2.3 Feature Extraction and classification

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

Observations

Figure 5 shows examples of tumor ADC histograms for both pre-and post-treatment with responders and non-responders. 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 a volumetrically responding tumor, while on the right is an example of a non-responding tumor. From the figure, we observe that not only the location but also the shape of the responder's histogram changes after treatment. The two Gaussian mixture components change as well.

Figure 5. Examples of histograms from two tumors and two time points: (a), (c): example of a responding tumor for pre- and post-treatment respectively; (b), (d): example of a non-responding tumor for pre- and post-treatment respectively.

General histogram features

Different statistical features from tumor ADC histograms were extracted. According to clinical studies, the ADC value should change after treatment. In our data set, we observed that the histograms exhibit change not only in location, but also in shape. Therefore, we introduced the extraction of different ADC histograms features and explored changes in their pattern. The features are: mean, standard deviation, skewness, kurtosis, median, IQR (interquartile range), 25% percentile, and 75% percentile.

Features from GMM

Two-component Gaussian mixture modeling was applied to each tumor ADC histogram and the two-component features were extracted. 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 relating to tumor cellularity and edema. We assumed that the component with lower peak is influenced by tumor cellularity, and the component with higher peak by edema effects. We used a two component GMM to fit the ADC histogram for both baseline and follow-up scans and applied EM algorithm to estimate GMM parameters.

The features we obtained from the GMM-EM were 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 5 shows examples of tumor ADC histograms fitted by GMM with low ADC and high ADC curves overlaid.

Combining GMM features with the statistical features, we obtained 14-dimensional feature vectors for both pre- and post-treatment tumor histograms. Afterwards, we calculated the rate of change between the pre- and the post-treatment tumor histogram. Therefore, we had a 14-dimensional vector as the difference feature vector.

Earth Mover's Distance

Finally, we applied the earth mover's distance (EMD) (Rubner et al., 1998; Ling and Okada, 2007) as a metric to directly evaluate the distance between the pre- and post-treatment tumor ADC histograms.

Informally, if the histograms are interpreted as two different ways of piling up a certain amount of dirt over the region, 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 was appended as the 15th element in the difference feature vector. The calculated 15-dimensional vector was the input feature vector for classification.

Classification

In this study, we investigated three classification techniques with different characteristics: AdaBoost, random forests (RF) and support vector machine (SVM). We employed three classifiers to avoid biasing the results by selection of a single classification method. The reason we choose them is that the first two classifiers both include a feature selection mechanism. By applying these two classification techniques, we are seeking the best features that would separate responders from non-responders. SVM is reported to outperform several of the most frequently used machine learning techniques in structure–activity relationship (SAR) analysis. In this study, all classifiers were implemented in the open source data mining software Weka (Witten et al., 2005). Their performance was evaluated using 10-fold cross validation method.

Three experiments were performed. First, the conventional method of using mean ADC for treatment response classification was applied (Ross et al., 2003). Second, the AdaBoost, RF classifier, and SVM were applied to the difference feature vectors of general statistical histogram features without GMM features, and results from the three classifiers were compared. Finally, the three classifiers were applied using all statistical features including the GMM features, and the results were compared, and the results of accuracies from different classification techniques were compared with the conventional method of assessment by ADC mean changes by the test of proportion.

Classification Performance

Using the conventional method of mean ADC change (subjects with a mean ADC increase classified as responders and those with an ADC decrease as non-responders) (Ross et al., 2003; Chenevert et al., 2000), the accuracy is 29.4% (25/85), with a sensitivity of 17.95% and a specificity of 60.87% (see Table 4).

Table 4. Performance of the conventional mean ADC classification method.

The experiment with AdaBoost involved 10 learning iterations. The RF classifier was composed of 10 trees, each of which is constructed considering five random features. The SVM classifier used non-linear polynomial kernels and normalized all features.

The results for the experiment using only the general histogram features without GMM are shown in Table 5 with sensitivity, specificity, accuracy and area under the ROC curve (Az). The ROC curves are shown in Figure 6. The curve using conventional mean ADC was plotted by varying the threshold of the mean ADC change used for the classification, while the curve using the three ML techniques were plotted by varying the threshold on the probability assigned to the positive class.

As shown in Table 5, the accuracy is boosted from 29.4% to 63.53-67.06% by incorporating more features to describe histograms. There is a significant different between the conventional method using only the mean and the three ML methods using multiple histogram features ($p<0.001$). In Figure 7, the Az for the conventional method is below 0.5, which is worse than random guess, while the three ML methods improve the Az to over 0.6. Among the three ML methods, random forest shows slightly better performance but no statistically significant difference ($p>0.05$).

Table 5. Performance comparison among three classifiers without GMM features
Figure 6. ROC curve for three classifiers without GMM features.

With GMM features added, the three classifiers with the same parameter setups were applied to the data. The results are shown in Table 6 with sensitivity, specificity, accuracy and area under the curve (Az) of the ROC curve. The ROC curves are shown in Figure 7.

With GMM features added, adaBoost and RF classifiers showed overall improvement in accuracy, while SVM stays unchanged. Additionally, adaBoost and RF both showed decreased sensitivity and increased specificity, while SVM had unchanged sensitivity and specificity. But none of the three methods showed statistically significant difference.

Table 6. Performance comparison among three classifiers with GMM features.
Figure 7. ROC curve for three classifiers with GMM features.

5 FUTURE RESEARCH DIRECTIONS

5.1 ADC measurement quality control

For future studies, a larger subject population is required to increase statistical power to detect pairwise differences. We did not observe a significant difference in median ADC values across different scanner models. We performed the power analysis and achieved 52% power at the 0.05 level of alpha to detect differences in median ADC values among scanners. The power analysis test showed that we needed 13 patients per scanner to have enough statistical power to detect the difference in ROI median.

Furthermore, it is necessary to run the QC tool in every multi-center trial before the data analysis. Inconsistent scanner parameters, like DW gradient orientations, b-value, and image matrix size, may introduce bias to the study. However, in a real-world multi-center clinical trial, it is possible that scanner parameters are set differently among different vendors and sites. It is not possible to have exactly the same scanner parameters across all machines because different vendors have different optimal scanner setup. QC becomes extremely important in this context.

Lastly, conducting a controlled (phantom) study would be useful to identify the degree of between-scanner ADC variation. We were not able to address the degree of between-scanner ADC variation since we did not scan the same patient cohort on each scanner.

5.2 Tumor definition on ADC maps

The 3D ROI mapping tool is more computationally efficient compared to the co-registration techniques, but it cannot correct patient motion. Therefore, in our study, a board-certified radiologist's visually checked and edited all segmentation results as needed. In the future, a more sophisticated registration method with an image similarity measure may improve the accuracy of the tumor contours on ADC maps, and consequently improve the accuracy of the extracted features and the classifier performance.

5.3 CADrx system for assessing therapeutic response

A multi-category classification system should be explored as future work. One limitation of the study in this chaper is that we classified CR, PR and SD as responders for the ground truth to achieve a binary classification. GBM tumor physiology is complicated and thus SD and PR may have different patterns in terms of their ADC histogram change. The responder group could be further divided into two groups in the future.

In this chapter, Macdonald criteria at the eighth or tenth week after treatment was used for determining treatment response. In future work, time-to-progression and survival time would be a better endpoint to classify treatment response.

Besides DW, perfusion-weighted MR, MR spectroscopy, functional MR (fMRI) are also novel ways to evaluate tumor physiology, and they provide additional information besides DWI. The CADrx system can be easily extended to include those features, and to "intelligently" select the most informative features.

ADC values obtained on pre-operative MRI scans are reported to be of prognostic value in patients with glioblastoma (Pope et al., 2009; Yamasaki et al., 2009). The term "prognosis" refers to predicting the likely outcome of treatment. ADC, reported to be inversely proportional to tumor cellularity, is gaining interest in predicting GBM tumor prognosis. Our proposed framework now uses changes in DW-MRI for early prediction of treatment response; however, the framework with feature extraction and machine learning technique could be generalized to pre-treatment DW-MRI for prognostic assessment.

6 CONCLUSION

This chapter discussed a CADrx system using quantitative ADC histogram features and machine-learned classifiers in GBM treatment response assessment. ADC map quality was evaluated and image analysis was applied as the pre-processing step to the CADrx system. The CADrx system had better performance in treatment response assessment over conventional analysis using only a mean ADC value. This may have major implications for clinical trials, providing early treatment response assessment in GBM.

REFERENCES

Adams, R.; Bischof, L. (1994). Seeded region growing. *IEEE Trans. Syst. Man. Cybern. Int. 16*, 641-647.

Bilgili, Y; Unal B. (2004) Effect of Region of Interest on Interobserver Variance in Apparent Diffusion Coefficient Measures., *AJNR Am. J. Neuroradiol., 25,* 108 - 111.

Breiman, L. (2001). Random decision forest. *Mach. Learn. 45*, 5-32.

Bishop, C. (1995). *Neural Networks for Pattern Recognition*; Clarendon Press: Oxford, UK.

Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. (2001). Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput. And. Chem, 26*, 5–14.

Chenevert, T.L.; Stegman, L.D.; Taylor, J.M.; Robertson, P.L.; Greenberg, H.S.; Rehemtulla, A.; Ross, B.D. (2000). Diffusion magnetic resonance imaging: an early surrogate marker of therapeutic efficacy in brain tumors. *J. Natl. Cancer Inst., 92*, 2029-2036.

Clark, M.C.; Hall, L.O.; Goldgof, D.B.; Velthuizen, R.; Murtagh, R.; Silbiger, M.S. (1998). Automatic tumor segmentation using knowledge-based techniques. *IEEE Trans. Med. Imaging, 17*, 187–201.

Corso, J.J.; Sharon, E.; Dube, S.; El-Saden, S.; Sinha, U.; Yuille, A. (2008). Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification. *IEEE Trans. Med. Imaging, 27*, 629-640.

Duda, R.O.; Hart, P.E.; Stork, D.H. (2000). *Pattern classification*; Wiley Interscience: Malden, MA, USA.

Dube, S.; Corso, J.J.; Yuille, A.; Cloughesy, T.F.; El-Saden, S.; Sinha, U. (2008). Hierarchical Segmentation of Malignant Gliomas via Integrated Contextual Filter Response. *Proc. SPIE, 6914*, 69143Y.

Dube, S.; Corso, J.J.; Cloughesy, T.F. ; El-Saden, S.; Yuille, A.; Sinha, U. (2007). Automated MR image processing and analysis of malignant brain tumors: enabling technology for data mining. In *Data Mining Systems Analysis and Optimization in Biomedicine*; American Institute of Physics Proceedings: New York, NY, USA; Vol. 953, pp. 64-84.

Fletcher-Heath, L.M.; Hall, L.O.; Goldgof, D.B.; Murtagh, R.F. (2001). Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artif. Intell. Med., 21*, 43–63.

Freund, Y.; Schapire, R.E. (1999). A short introduction to boosting. *J. Jpn. Soc. For. Artif. Intell., 14*, 771-780

Giger, M.L. (2000). *Computer-aided diagnosis in medical imaging — A new era in image interpretation*; *Tech. Rep.*; World Markets Research Centre: London, UK, pp. 75-78.

Hamstra, D.A.; Chenevert, T.L.; Moffat, B.A.; Johnson, T.D.; Meyer, C.R.; Mukherji, S.K.; Quint, D.J.; Gebarski, S.S.; Fan, X.; Tsien, C.I.; Lawrence, T.S.; Junck, L.; Rehemtulla, A.; Ross, B.D. (2005).

Evaluation of the functional diffusion map as an early biomarker of time-to-progression and overall survival in high-grade glioma. *PNAS, 102*, 16759–16764.

Ho, S.; Bullitt, E.; Gerig, G. (2002). Level set evolution with region competition: Automatic 3-d segmentation of brain tumors. In *Proceedings of International Conference on Pattern Recognition*, Quebec, Canada, August, pp. 532–535.

Ho, T.K. (1995). Random decision forest. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition* Montreal, Canada, August, 1995; pp. 278-282.

Huo, J.; Kim, H.J.; Pope, W.B.; Okada, K.; Alger, J.R.; Wang, Y.; Goldin, J.G.; Brown, W.S. (2009). Histogram-based classification with Gaussian mixture modeling for GBM tumor treatment response using ADC map. *Proc. SPIE, 7260*, 72601Y.

Huo, J.; Alger, J.R.; Kim, H.J.; Pope, W.B.; Okada, K.; Goldin, J.G.; Brown, M.S. (2009). Between-scanner variation in normal white matter ADC in the setting of a multi-center clinical trial. *International Society Magnetic Resonance in medicine*, Haiwaii, USA, April 18-24.

Huhn, S.L.; Mohapatra, G.; Bollen, A.; Lamborn, K.; Prados, M.D.; Feuerstein, B.G. (1999). Chromosomal abnormalities in glioblastoma multiforme by comparative genomic hybridization: correlation with radiation treatment outcome. *Clin. Cancer Res. 5*, 1435-1443.

Kaus, M.; Warfield, S.; Nabavi, A.; Black, P.M.; Jolesz, F.A.; Kikinis, R. (2001). Automated segmentation of mr images of brain tumors. *Radiology, 218*, 586–591.

Lee, C.H.; Schmidt, M.; Murtha, A.; Bistritz, A.; Sander, J.; Greiner, R. (2005). Segmenting brain tumor with conditional random fields and support vector machines. In *Proceedings of Workshop on Computer Vision for Biomedical Image Applications at International Conference on Computer Vision*, Beijing, China, October, Vol. 3765, pp. 469-478.

Liu, J.; Udupa, J.; Odhner, D.; Hackney, D.; Moonis, G. (2005). A system for brain tumor volume estimation via mr imaging and fuzzy connectedness. *Comput. Med. Imaging Graph. 29*, 21–34.

Ling, H.; Okada, K. (2007). An efficient Earth mover's distance algorithm for robust histogram comparison. *IEEE Trans. Patt. Anal. Mach. Intell. 29*, 840-853.

Mardor, Y.; Pfeffer, R.; Spiegelmann, R.; Roth, Y.; Maier, S.E.; Nissim, O.; Berger, R.; Glicksman, A.; Baram, J.; Orenstein, A.; Cohen, J.S.; Tichler, T. (2003). Early detection of response to radiation therapy in patients with brain malignancies using conventional and high b-value diffusion-weighted magnetic resonance imaging. *J. Clin. Oncol. 21*, 1094-1100

Marzban, C. (2004). The ROC curve and the area under it as a performance measure. *Weather Forecast. 19*, 1106-1114.

Moffat, B.A.; Chenevert, T.L.; Meyer, C.R.; Mckeever, P.E.; Hall, D.E.; Hoff, B.A.; Johnson, T.D.; Rehemtulla, A.; Ross, B.D. (2005). The functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. *PANS, 102*, 5524-5529.

Niea, J.; Xue,.; Liu, T.; Young, G.S.; Setayesh, K.; Guo, L.; Wong, S.T.C. (2009). Automated brain tumor segmentation using spatial accuracy-weighted hidden Markov Random Field. *Comput. Med. Imaging Graph. 33*, 431–441.

Otsu N. (1979). A threshold selection method from gray level histograms. *IEEE Trans. Syst. Man. Cybern. 9*, 62-66.

Padhani, A.R.; Liu, G.; Mu-Koh, D.; Chenevert, T.L.; Thoeny, H.C.; Takahara, T.; Dzik-Jurasz, A.; Ross, B.D.; Cauteren, M.V.; Collins, D.; Hammoud, D.A.; Rustin, G.J.S.; Taouli, B.; Choyke, P.L. (2009). Diffusion-weighted magnetic imaging as a cancer biomarker: consensus and recommendations. *Neoplasia, 11*, 102–125.

Phillips, W.E.; Velthuizen, R.P.; Phupanich, S.; Hall, L.O.; Clarke, L.P.; Silbiger, M.L. (1995). Applications of fuzzy C-means segmentation technique for tissue differentiation in MR images of a hemorrhagic glioblastoma multiforme. *J. Magn. Reson. Imaging, 13*, 277–290.

Pope, W.B.; Kim, H.J.; Huo, J.; Alger, J.R.; Brown, W.S.; Gjertson, D.; Sai, V.; Young, J.R.; Tekchandani, L.; Cloughesy, T.; Mischel, P.S.; Lai, A.; Nghiemphu, P.; Rahmanuddin, S.; Goldin, J.G. (2009). Recurrent glioblastoma multiforme: ADC histogram analysis predicts response to bevacizumab treatment. *Radiology, 252*, 1-8.

Prastawa, M.; Bullitt, E.; Moon, N.; Leemput, K.V.; Gerig, G. (2003). Automatic brain tumor segmentation by subject specific modification of atlas priors. *Acad. Radiol.*, *10*, 1341–1348.
Rubner, Y.; Tomasi, C.; Guibas, L.J. (1998). A metric for distributions with applications to image databases. In *Proceedings of ICCV*, Bombay, India, January, pp. 59-66.
Ross, B.D.; Moffat, B.A.; Lawrence, T.S.; Mukherji, S.K.; 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. (2003). Evaluation of cancer therapy using diffusion magnetic resonance imaging. *Mol. Cancer Ther.*, *2*, 581-587.
Sajda, P. (2006). Machine learning for detection and diagnosis of disease. *Annu. Rev. Biomed. Eng.*, *8*, 537–65.
Vinitski, S.; Gonzalez, C.F.; Knobler, R.; Andrews, D.; Iwanaga, T.; Curtis, M. (1999). Fast tissue segmentation based on a 4D feature map in characterization of intracranial lesions fast tissue segmentation based on a 4D feature map in characterization of intracranial lesions. *J. Magn. Reson. Imaging*, *9*, 768–776.
Vapnik, V. (1979). *Estimation of Dependencies Based on Empirical Data*; Nauka: Moscow, Russia.
Witten, I.H.; Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: San Francisco, CA, USA.
Yamasaki, F.; Sugiyama, K.; Ohtaki, M.; Takeshima, Y.; Abed, N.; Akiyamad, Y.; Takabad, J.; Amatyac, V.J.; Saitoa, T.; Kajiwaraa, Y.; Hanayaa, R.; Kurisua, K. (2009). Glioblastoma treated with postoperative radio-chemotherapy: Prognostic value of apparent diffusion coefficient at MR imaging. *Eur.J. Aiol*., (in press).
Zhu, Y.; Yan, H. (1997). Computerized tumor boundary detection using a hopfield neural network. *LEEE Trans. Med. Imaging*, *16*, 55–67.
Zhang, J.; Ma, K.; Er, M.H.; Chong, V. (2004). Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine. In *Proceedings of International Workshop on Advanced Image Technology*, Singapore, January, pp. 207–211.
US Food and Drug Administration, (2007) Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. *Federal Register*, *72*, No. 94.
http://en.wikipedia.org/wiki/Support_vector_machine (accessed November 10, 2009).

Key Terms

Keyword: CADrx, GBM, DWI, ADC, QC, GMM-EM, EMD, adaBoost, SVM, Random Forest

CADrx : computer-aided therapeutic response system
GBM: glioblastoma multiforme
DWI: diffusion weighted imaging
ADC: apparent diffusion coefficient
QC: quality control
GMM: Gaussian mixture model
EM: expectation-maximization
EMD: earth mover distance
SVM: support vector machine