Quantitative Tumor Segmentation for Evaluation of Extent of Glioblastoma Resection to Facilitate Multisite Clinical Trials[1,2]

James S. Cordova*[,†], Eduard Schreibmann[‡], Costas G. Hadjipanayis[§,¶], Ying Guo[#], Hui-Kuo G. Shu[‡,¶], Hyunsuk Shim*[,¶] and Chad A. Holder*

*Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA; [†]Medical Scientist Training Program, Emory University School of Medicine, Atlanta, GA; [‡]Department of Radiation Oncology, Emory University School of Medicine, Atlanta, GA; [§]Department of Neurosurgery, Emory University School of Medicine, Atlanta, GA; [¶]Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA; [#]Department of Biostatistics, Emory University School of Medicine, Atlanta, GA

Abstract

Standard-of-care therapy for glioblastomas, the most common and aggressive primary adult brain neoplasm, is maximal safe resection, followed by radiation and chemotherapy. Because maximizing resection may be beneficial for these patients, improving tumor extent of resection (EOR) with methods such as intraoperative 5-aminolevulinic acid fluorescence-guided surgery (FGS) is currently under evaluation. However, it is difficult to reproducibly judge EOR in these studies due to the lack of reliable tumor segmentation methods, especially for postoperative magnetic resonance imaging (MRI) scans. Therefore, a reliable, easily distributable segmentation method is needed to permit valid comparison, especially across multiple sites. We report a segmentation method that combines versatile region-of-interest blob generation with automated clustering methods. We applied this to glioblastoma cases undergoing FGS and matched controls to illustrate the method's reliability and accuracy. Agreement and interrater variability between segmentations were assessed using the concordance correlation coefficient, and spatial accuracy was determined using the Dice similarity index and mean Euclidean distance. Fuzzy C-means clustering with three classes was the best performing method, generating volumes with high agreement with manual contouring and high interrater agreement preoperatively and postoperatively. The proposed segmentation method allows tumor volume measurements of contrast-enhanced T_1-weighted images in the unbiased, reproducible fashion necessary for quantifying EOR in multicenter trials.

Translational Oncology (2014) 7, 40–47

Introduction

Glioblastoma is the most common and most aggressive primary brain neoplasm in adults. Current imaging evaluation for glioblastoma management relies heavily on the subjective analysis of T_1-weighted (T_1W) magnetic resonance images [1,2]. Simple unidimensional and bidimensional measurements of T_1W contrast-enhancing regions of the tumor are the crux of response criteria in clinical trials, although limitations of such methods have been reviewed previously in detail, particularly with regards to postsurgical tumor analysis [3–5]. For example, these linear methods of measurement are not well suited for evaluating curvilinear tumor remnants such as those along the edges

of a postoperative resection cavity. In addition, they may not accurately account for the presence of T_1-hyperintense blood products (methemoglobin) in and around the resection site, which can be confused with enhancing tumor tissue. Although not currently the standard of care, these morphologic nuances can be accounted for by a neuroradiologist with 3-dimensional (3D) volume–rendering software that allows for the manual tracing of images, or "contouring"; however, this process is time consuming, cost prohibitive, and suffers from limited reproducibility [6–8]. To overcome these limitations, many sophisticated algorithms have been developed for the automated segmentation of multiple images; however, few have achieved the simplicity, speed, accuracy, and limited user interaction required for routine clinical use [9–12]. Moreover, many of the software environments in which these techniques have been designed are not standardized for clinical use, compounding the challenge of their implementation in clinical trials [13]. Due to the inadequacies surrounding the manual and automated segmentation methods that are currently available, it would be desirable to develop a hybrid method of tumor segmentation that is adaptable to various clinically available tools for the reproducible segmentation of contrast-enhancing tumors in multicenter neurosurgical trials.

Coupling a flexible image sampling method, such as region-of-interest blob (ROI blob) generation, to an unsupervised statistical classification algorithm appears to exhibit the clinically desired balance for semiautomated tumor segmentation. Two well-established classification schemes that can be adapted for unsupervised image segmentation are Otsu's multilevel thresholding (Otsu) and Fuzzy C-means clustering (Fuzzy) [14–16]. The Otsu method uses discriminant statistical analysis for image histogram intensity thresholding, exhaustively determining intensity thresholds (between tissue classes) that minimize the intraclass variance among each class of voxels [17,18]. The division algorithm modifies each threshold by fitting the histogram with a number of probability curves and iteratively computing the variances and positions of each until curve overlap is minimized. The result is a 3D map with discrete classes containing voxels that exhibit similar signal intensities. When this method is used to divide a contrast-enhanced, T_1W ROI blob into three or four classes, the resultant map differentiates strongly and weakly enhancing regions from surrounding tissues.

The Fuzzy algorithm similarly computes cluster centroids and clusters voxels on the basis of intraclass/interclass signal intensity variance. Unlike Otsu, which classifies voxels into discrete clusters (hard clustering), Fuzzy treats each data element as a member of all clusters with an associated level of membership in each, which can be expressed as a continuous value (soft clustering) [19–21]. Therefore, Fuzzy classification results in a set of maps, each representing a class with a similar signal intensity range, where each voxel's value indicates the degree of its residence (expressed as a probability) in that class. Fuzzy has been investigated for tumor segmentation using multidimensional feature vectors for the last couple of decades with varying degrees of success [22–24]. However, to our knowledge, no investigation has been done to evaluate its performance coupled to ROI blob analysis.

In this study, we develop and validate a flexible semiautomated tumor segmentation tool for the assessment of preoperative and postoperative glioblastoma tumor burden on the basis of contrast-enhanced T_1W images. As proof of principle, we applied ROI blob/voxel classification methods (Otsu and Fuzzy) to patients with glioblastoma receiving fluorescence-guided surgery (FGS) with 5-aminolevulinic acid (5-ALA) and retrospectively matched control patients to: 1) illustrate the reliability and accuracy of the segmentation method and 2) corroborate the interim extent-of-resection

(EOR) findings of the ongoing phase II 5-ALA clinical trial [25–31]. The desired outcome is to develop a reliable and user-friendly segmentation method for the generation of T_1W contrast-enhancing tumor volumes that may be used in multicenter neurosurgical clinical trials in patients with glioblastoma.

Methods

Preoperative and Postoperative Imaging

Per the institutional review board–approved 5-ALA phase II clinical trial, high-resolution, 3D preoperative MR images, including 1 mm^3 T_1W magnetization–prepared rapid gradient–echo (MP-RAGE) images (repetition time/echo time = 1900/3.52, 256 × 256 matrix, flip angle = 9°, before and after IV administration of gadolinium-based contrast medium) generated 48 hours or less before surgery for 5-ALA patients were analyzed. Postoperative images consisted of T_1W MP-RAGE and/or low-resolution (1 × 1 × 5 mm, repetition time/echo time = 450-2000/9.2-20, 256 × 224-177 matrix, flip angle = 90°-130°) images acquired before and after the IV administration of gadolinium-based contrast medium postoperatively, within 24 hours of surgery. Control patients matched for glioblastoma tumor size and location were retrospectively selected from an Emory University Department of Radiation Oncology, Winship Cancer Institute (Atlanta, GA) database for case comparison. These preoperative and postoperative MR images exhibited similar characteristics in terms of the aforementioned pulse sequence parameters and the timing of the scans relative to surgery (24-48 hours postsurgery).

Image Analysis

The 510(k), Food and Drug Administration (FDA)-cleared medical imaging platform VelocityAI (Velocity Medical Solutions, Atlanta, GA) was used to outline tumors manually and generate ROI blobs using preoperative and postoperative T_1W scans with and without IV contrast. As per a European randomized multicenter FGS trial, complete tumor resection was defined as ≤0.175 cm^3 of residual contrast-enhancing tumor after surgery using volumetric assessment [26]. The conventional imaging definition of gross total resection (residual contrast-enhancing tumor < 1 cm^3) was also evaluated.

Computer-Assisted Manual Contouring

Preoperative and postoperative tumor volumes were manually delineated using VelocityAI by an experienced board-certified neuroradiologist (C.A.H.). Enhancing-tumor control points were initiated, tumor regions were interactively grown or contracted in two dimensions, and *post hoc* edits were made as needed. This manual method of contouring is commonplace in the generation of tumor volumes for both radiotherapy planning and validation of simulated segmentations [10,32,33].

Semiautomated Segmentation Method

Preoperative tumor ROI blobs for Otsu and Fuzzy processing were generated in VelocityAI by coarsely contouring around enhancing tumor on T_1W images, including the first and last images on which the tumor is visible, and every sixth axial slice in between and then automatically interpolating to a cohesive 3D structure (Figure 1, *A*
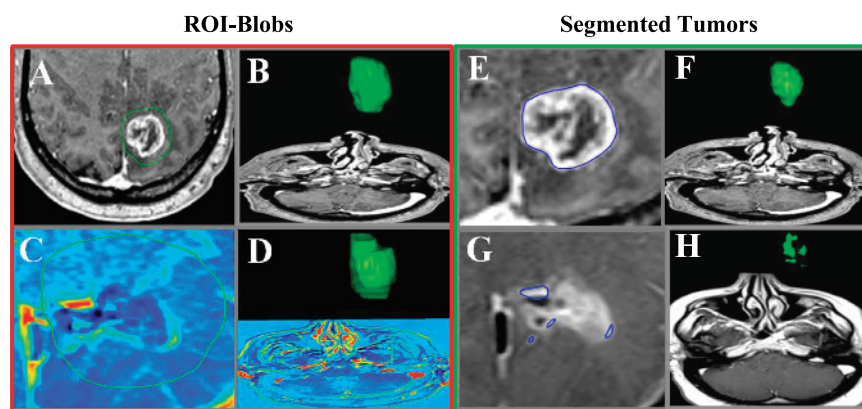
Figure 1. Preoperative (A and B) and postoperative (C and D) ROI blobs [2-dimensional (2D) and 3D] generated by coarse contouring. Subtraction image (C) accounts for blood product accumulation in resection cavity (dark region). Resultant 2D and 3D tumor segmentations for preoperative (E and F) and postoperative (G and H) contrast-enhanced T_1W images using Fuzzy3 algorithm.

and *B*). As few as two slices (at the superiormost and inferiormost poles of the tumor, respectively) and as many as every slice containing tumor can be contoured and interpolated to give a cohesive ROI blob; however, on the basis of practical clinical experience, we have found that contouring every sixth slice gives ROI blobs that adequately encompass even the most complex contrast-enhancing brain tumors within a reasonable time frame. This type of initialization can be done with any volume-rendering platform capable of exporting Digital Imaging and Communication in Medicine blobs, making implementation of the method simple at multiple sites.

ROI blobs and parent images were then exported and processed offline using C++-generated Otsu and Fuzzy algorithms. Algorithms used standard code from the Insight Segmentation and Registration Toolkit and were compiled as a command-line executable taking as input Digital Imaging and Communication in Medicine images (ROI blobs) and outputting segmentation masks in the same format. Algorithm preprocessing including fade correction for signal differences was due to magnetic field gradients in a bias-field estimation manner. No smoothing functions were used, ε was set at 0.01, and iterations were set at 30 as these were found to be optimal. For these algorithms, the number of tissue classes is a definable parameter, and any number of strata represented by a positive integer can be generated; however, stratification of voxels into more than five classes is of limited utility, as discrete tissue types (such as contrast-enhancing tumor rims) become erroneously subcategorized. Therefore, only maps stratified into three or four classes were generated with the naming convention as follows: classification of four classes using the Fuzzy and Otsu algorithms are named Fuzzy4 and Otsu4, respectively, and classification of three classes using the Fuzzy and Otsu algorithms are named Fuzzy3 and Otsu3, respectively. It should be noted that Otsu3 was abandoned in the preliminary analysis as resultant maps did not differentiate parenchyma from enhancing tissue and therefore is not reported herein. These maps were then imported into VelocityAI for display and evaluation.

To produce postresection tumor volumes, image difference maps (subtraction images) were generated by subtracting white matter–normalized, spatially coregistered precontrast T_1W images from postcontrast T_1W images. This was done to correct for the presence of T_1-hyperintense postoperative blood products and cavity deformation. Normalization of postoperative images included fade correction and precontrast image signal scaling with postcontrast/precontrast intensity ratio. These normalized subtraction images were then used to draw coarse ROI blobs and generate segmentation maps using both algorithms outlined above (Figure 2, *C* and *D*). This was done for all preoperative and postoperative images by two research specialists (nonradiologists), reader 1 (J.S.C.) and reader 2 (J.A.), independently and many days apart (>20 days).

Statistical Methods

Statistical analyses were performed with the Statistical Analysis System (version 9; SAS Institute, Cary, NC); all were two sided, and statistical significance was set at $P < .05$. The concordance correlation coefficient (CCC), a reproducibility index evaluating the agreement of two readings by measuring their combined variation from the line $y = x$, was used to measure the agreement of tumor volumes generated by readers 1 and 2 using Otsu4, Fuzzy4, and Fuzzy3 with those generated manually by the neuroradiologist [34–36]. Spatial agreement between each segmentation and the manual contour were evaluated using the Dice similarity index (Dice) and the mean Euclidean distance (MED)—volume overlap and surface distance metrics, respectively—for each pair, and means between readers were subjected to analysis of variance with Tukey's correction [37,38]. It is generally accepted that Dice values > 0.70 represent a significant overlap in structures, with values > 0.80 considered to be "good" and values > 0.90 considered to be "excellent." For surface distance, a value of 0 is an ideal MED; however, for practical purposes, a mean distance of ~1 mm is considered very good (~1 voxel width error in MP-RAGE images). Interrater agreement, in terms of CCC, was evaluated for the method that had the highest agreement with manual contouring as a metric of interrater variability. Wilcoxon signed rank test and χ^2/Fisher exact analysis were used to compare manually generated and algorithm-generated volume outcomes (i.e., residual tumor volume and EOR) for the 16 5-ALA: control pairs in the 37 case samples. Performance metrics (sensitivity, specificity, positive predictive value, and negative predictive value) were computed using a two-by-two contingency table for postresection tumor volume cutoffs of ≤ 1 cm^3 and ≤ 0.175 cm^3, with manual contouring by an experienced neuroradiologist as the "ground truth" for these 16 ALA:control pairs [39].

Results

Fuzzy3 Algorithm Shows Greatest Volume Agreement with Manual Contouring

A total of 37 cases were accrued for validation of the aforementioned segmentation methods, 16 of which had received FGS in the 5-ALA trial and 21 retrospective controls. Examples of preoperative and postoperative ROI blobs and segmented tumor volumes generated using Fuzzy3 are shown in Figure 1, *A* to *D* and *E* to *H*, respectively, and raw volume data for all cases and each method are included in Tables W1 and W2. The agreement for each method with manual contouring preoperatively and postoperatively is shown with CCCs, SEM, and 95% confidence intervals (CIs) in Figure 2. A complimentary evaluation of agreement using the method comparison analysis outlined by Bland and Altman can be found in Table W3 and Figures W1–W3 [40]. The extra five cases are included with 5-ALA cases for CCC, method comparison, and spatial analysis in Tables W1–W6, as well as Figures 2–4 and W1 to W3. Although differences in agreement between Fuzzy4 and Otsu4 were subtle preoperatively, greater differences became apparent postoperatively, with CCCs for Fuzzy4 approaching unity, whereas, those for the Otsu method remain below 0.84. This trend held true for both readers pre-

operatively and postoperatively, suggesting a reproducible finding attributable to the method. Moreover, preoperative and postoperative differences in interreader mean CCC between Otsu4 and Fuzzy4 were found to be statistically significant at $P = .027$ (Table W4) and $P = .002$ (Table W5), respectively.

On the basis of its near-perfect agreement with manual contouring both preoperatively and postoperatively, Fuzzy clustering was selected as the most promising segmentation method, and another simulation using only three classes was then investigated in an attempt to optimize the method. With the use of only three classes, Fuzzy3 segmentation volumes (Figure 2, row 3) exhibited higher agreement with manual contouring, in terms of CCC, than Fuzzy4 segmentations, both preoperatively and postoperatively. The differences in interrater mean CCCs were found to be statistically significant in the preoperative and postoperative settings as well ($P = .025$ in Table W4 and $P = .045$ in Table W5, respectively).

Fuzzy3 Algorithm Shows Greatest Spatial Agreement with Manual Contouring

To evaluate spatial overlap of algorithm-generated tumor segmentations *versus* manual contours, Dice and MED metrics were used
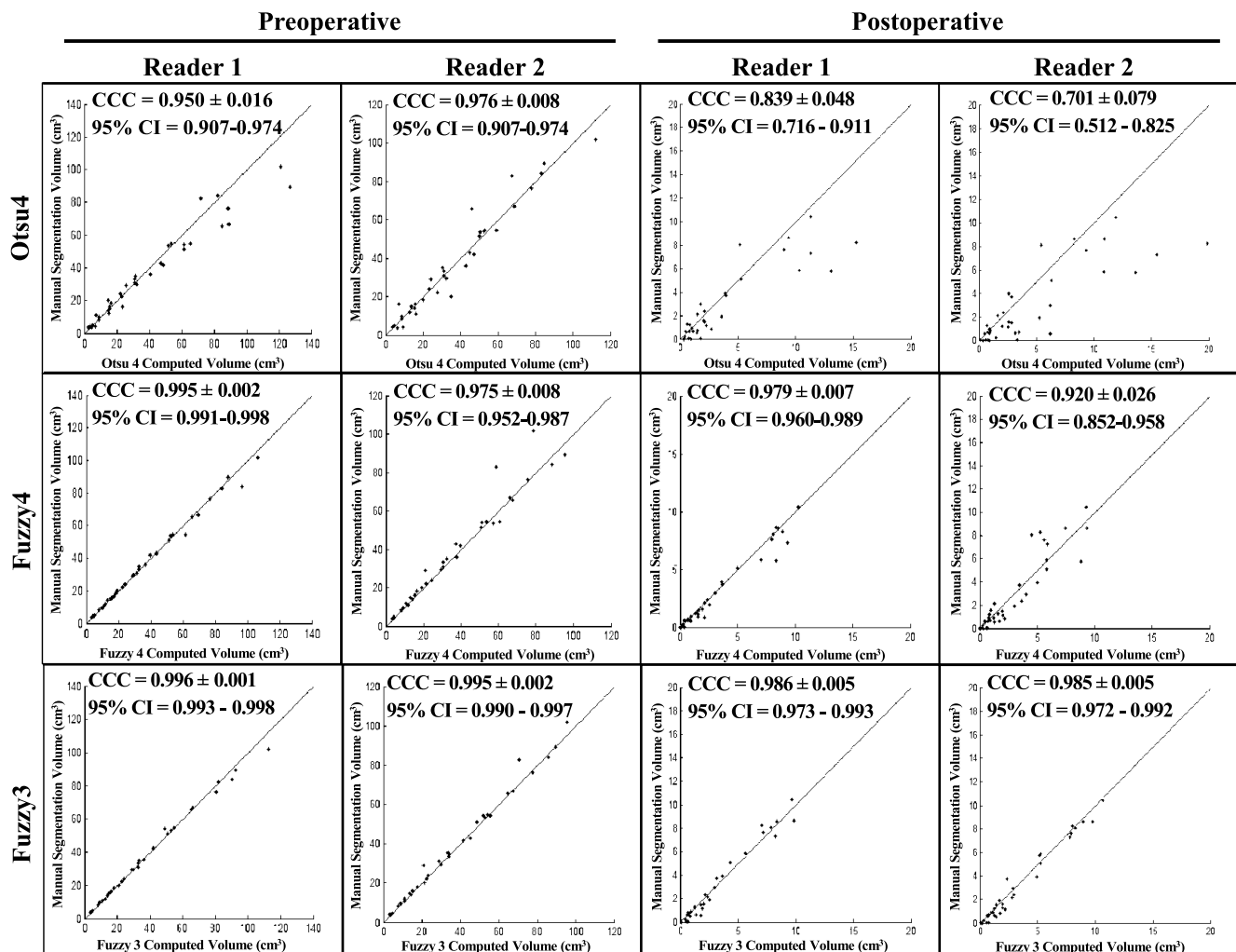

Figure 2. Preoperative and postoperative manual tumor contour volume *versus* semiautomated segmentation (Otsu4, Fuzzy4, or Fuzzy3 from top to bottom) by two separate readers. CCC, concordance correlation coefficient ± SEM; CI, confidence interval.
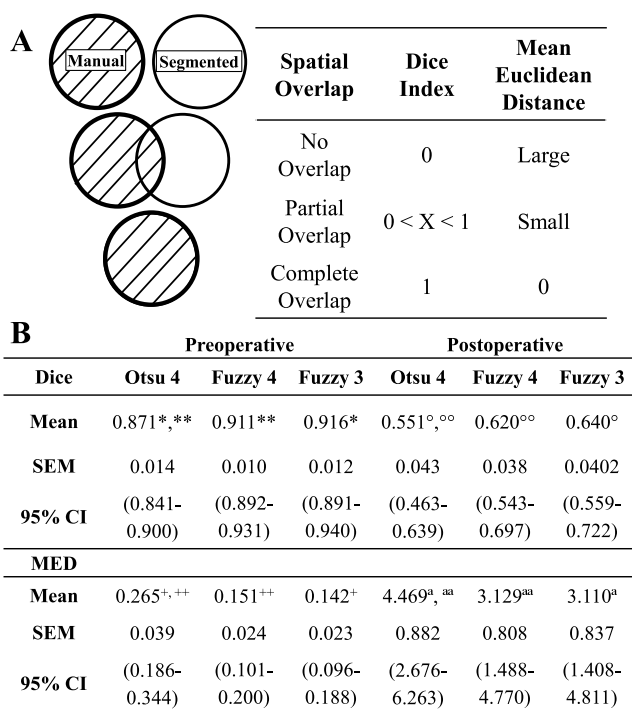
A

		Spatial Overlap	Dice Index	Mean Euclidean Distance
Manual	Segmented	No Overlap	0	Large
		Partial Overlap	0 < X < 1	Small
		Complete Overlap	1	0

B

Dice	Preoperative			Postoperative		
	Otsu 4	Fuzzy 4	Fuzzy 3	Otsu 4	Fuzzy 4	Fuzzy 3
Mean	0.871*,**	0.911**	0.916*	0.551°,°°	0.620°°	0.640°
SEM	0.014	0.010	0.012	0.043	0.038	0.0402
95% CI	(0.841-0.900)	(0.892-0.931)	(0.891-0.940)	(0.463-0.639)	(0.543-0.697)	(0.559-0.722)
MED						
Mean	0.265+,++	0.151++	0.142+	4.469ᵃ,ᵃᵃ	3.129ᵃᵃ	3.110ᵃ
SEM	0.039	0.024	0.023	0.882	0.808	0.837
95% CI	(0.186-0.344)	(0.101-0.200)	(0.096-0.188)	(2.676-6.263)	(1.488-4.770)	(1.408-4.811)

Figure 3. Two-dimensional illustration depicting impact of manual and segmented structure overlap on Dice and MED (A) along with mean preoperative and postoperative Dice and MED values with SEM and 95% confidence for Otsu and Fuzzy methods *versus* manual contouring (B). $*P < .05$, $**P < .05$, $+P < .05$, $++P < .05$, $°P < .05$, $°°P < .05$, $ᵃP < .05$, and $ᵃᵃP < .05$. CI, confidence interval.

(Figure 3*A*). Mean Dice and MED values with SEMs and 95% CIs for Otsu and Fuzzy methods *versus* manual contouring were computed for subjects preoperatively and postoperatively and reported in Figure 3*B*. Fuzzy4 again outperformed Otsu4, in terms of Dice and MED values, exhibiting a significantly higher mean Dice value and significantly lower MED preoperatively and postoperatively ($P < .05$). Subsequent Fuzzy3 simulations produced segmentations that also differed significantly in terms of Dice and MED from Otsu4; however, no significant difference in these values was found between segmentations produced using Fuzzy3 and Fuzzy4. Moreover, preoperative intrarater variability, as measured by the coefficient of variation for Dice and MED metrics, was found to be increased in Otsu4, exhibiting values of 0.156 and 1.35, respectively, relative to those for Fuzzy3 and Fuzzy4 that were similar (0.070/1.27 and 0.083/1.38, respectively).

Fuzzy3 Algorithm Exhibits Exceptionally High Interrater Agreement

Fuzzy3 outperformed both Otsu4 and Fuzzy4 in all measured parameters (CCC, Dice, and MED), although only nonstatistically significant trends were found *versus* Fuzzy4 in regards to spatial agreement with manual contouring. This, along with anecdotal evidence that Fuzzy4 requires more computation time than Fuzzy3 (as more centroids must be evaluated and more thresholds optimized), led to the selection of the Fuzzy3 method as the best candidate for interrater reproducibility analysis. The CCC, SEM, and 95% CIs between separate readers using Fuzzy3 for tumor volume generation are reported, along with their agreement plots, in Figure 4

for subjects preoperatively and postoperatively, as a measure of interrater agreement. The use of Fuzzy3 for tumor segmentation resulted in interrater CCCs for preoperative and postoperative tumor volumes that approached unity (0.990 and 0.983, respectively), with narrow CIs (0.981-0.995 and 0.981-0.995, respectively). This was particularly noteworthy postoperatively, as accurate postsurgical residual tumor measurements are typically difficult to determine, even using the most time-consuming manual segmentation methods, due to their complex morphology and the presence of T_1-hyperintense blood products.

Fuzzy3 Algorithm Performs Well per Diagnostic Performance Metrics

Chi-square comparison of Fuzzy3-generated postoperative tumor volumes *versus* manually contoured tumor volumes by an experienced neuroradiologist (ground truth) for 16 case-control pairs showed a sensitivity of 0.929 (95% CI = 0.661-0.998) and specificity of 1.0 (95% CI = 0.815-1.0) at a volume cutoff of ≤ 1.0 cm^3. Accuracy, positive predictive value, and negative predictive value at ≤ 1.0 cm^3 were 0.969, 1.0 (95% CI = 0.753-1.00), and 0.947 (95% CI = 0.740-0.999), respectively. At a postoperative tumor volume cutoff of ≤ 0.175 cm^3, sensitivity and specificity for Fuzzy3 were found to be 0.667 (95% CI = 0.223-0.957) and 1.0 (95% CI = 0.868-1.000), respectively, with accuracy, positive predictive value, and negative

CCC = 0.990 ± 0.003
95% CI = 0.981 - 0.995

Preoperative

CCC = 0.983 ± 0.006
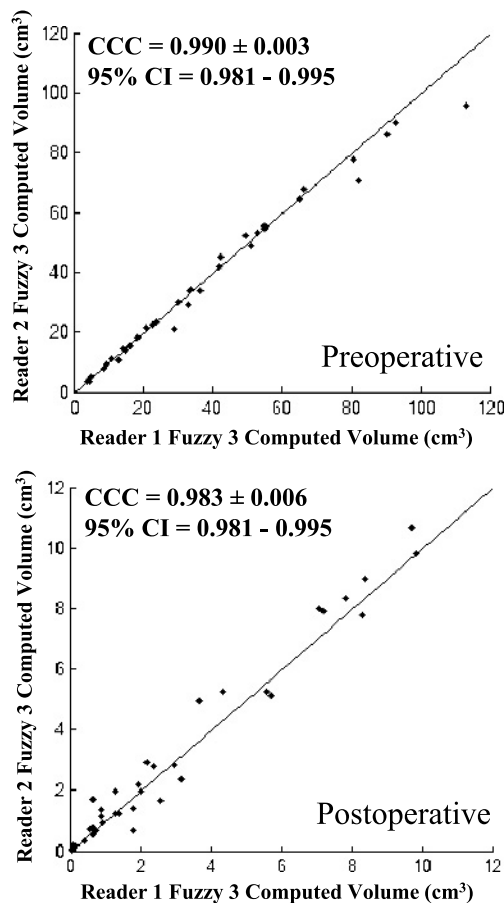95% CI = 0.981 - 0.995

Postoperative

Figure 4. Preoperative (top) and postoperative (bottom) tumor volumes were generated using Fuzzy3 by reader 1 *versus* reader 2. CCC, concordance correlation coefficient ± SEM; CI, confidence interval.

predictive value of 0.938, 1.0 (95% CI = 0.398-0.1.0), and 0.929 (95% CI = 0.765-0.991). The mean EOR, as computed using Fuzzy3 volumes, was found to be 94.5 ± 4.8% for the ALA patients and 86.5 ± 11.5% for the controls. Furthermore, no statistically significant difference between EOR determined using Fuzzy3 or manual contouring was found with P = .313 for ALA patients (manual = 95.0 ± 1.2%) and P = .115 for control patients (manual EOR = 86.5 ± 3.0%).

Discussion

In any clinical trial analyzing surgical resection, a major hindrance in image analysis must be addressed: the unbiased and reproducible measurement of preoperative and postoperative tumor volumes. We believe that Fuzzy coupled to ROI blob initialization offers a feasible solution to this hindrance. Fuzzy3 performs as well as other complex semiautomated and fully automated methods proposed in recent years. Fuzzy3's preoperative Dice values (mean = 0.92) lie well within the range of similar values generated using edge- and region-based segmentation algorithms (0.72-0.98), including content-based active contours, fuzzy connectedness, and fluid vector flow methods [41–45]. This holds true for comparison of Fuzzy3 to other classification/clustering methods—using neighborhood regularization, deformable models, and/or global constraints based on atlases—that exhibit Dice coefficients ranging from 0.40 to 0.90 [10]. These values include those generated by newly developed methods using conditional Markov fields, support vector machine classification, decision forests, and expectation-maximization + Markov random field methods [46–49]. Moreover, MED observed using Fuzzy3 (mean = 0.142 mm) preoperatively exhibited similar comparability to other segmentation studies using this metric (0.73-4.5 mm) [50–53]. Dice and MED values for postoperative segmentations barely failed to reach the generally known thresholds of reasonable segmentation (>0.70 and 1 mm, respectively) but still fell within the range of those reported in other studies. These lower scores are likely due to the large volume contribution of surrounding blood vessels to end segmentation. This is currently being addressed with postprocessing methods, including the use of shape identifiers to find and remove blood vessels in T_1W images.

Computer-assisted manual contouring decreases the interrater and intrarater variability of volume generation in multiple anatomic regions (i.e., cervix, lung, brain, pancreas, and liver); however, this requires a great deal of time and effort [8,13,54–56]. Sorensen et al. found that manual contouring of tumor volumes resulted in an average completion time of approximately 20 minutes per tumor, a time requirement that clearly limits its routine use [8]. Ongoing work shows that, when a group of cases (tumor volumes = 8.0-100.0 cm^3) is segmented by a research specialist using Fuzzy3 and a neuroradiologist using computer-assisted manual contouring, Fuzzy3 generates segmentations in significantly less time (186.7 ± 35.3 seconds vs 378.5 ± 49.9 seconds; P < .001). Furthermore, whereas the entirety of the manual contouring time consists of user interaction, this only accounts for an average of 86.4 ± 34.4 seconds (<50% of total time) using Fuzzy3, as the most time-intensive component of the Fuzzy3 segmentation is computation. Furthermore, the Fuzzy3 and manual segmentations were equally accurate, as no statistically significant difference was found in terms of the Dice means (Dice = 0.943 ± 0.018 and 0.945 ± 0.015, respectively; P = .741). Lastly, Fuzzy3 can be used for batch processing, allowing multiple tumors to be segmented sequentially and simultaneously—an advantageous quality when data from multiple centers are analyzed at a single site. The user (most likely an imaging technician) need not laboriously contour the intricate morphology of each tumor; he or she need only draw a simple coarse blob around the tumor for each patient and process multiple cases in a single batch. The algorithm will fill in the fine morphologic details of the tumor for the neuroradiologist to confirm afterward. The proposed semiautomated segmentation method delineates the intricacies of tumor morphology in a manner comparable to manual contouring with high fidelity but in less time and with less user interaction and, consequently, with less potential for user-introduced bias.

Semiautomated segmentation methods have been developed in an attempt to combine the high-level visual processing and specialized knowledge exhibited by humans with the objectivity of computers; however, most of these methods remain restricted to their individual development sites, having not gained widespread acceptance [24, 57–61]. This is likely due to an absence of segmentation algorithm/software transparency, the technical complexity of the algorithms, and the difficulty with which input data are generated. For example, artificial neural networks are capable of identifying very complex voxel-based dependencies within large 3D magnetic resonance imaging (MRI) data sets, offering clear advantages over less sophisticated methods [44]. However, implementing these processes is not trivial, as the algorithm must be trained, requiring a slow, tedious learning phase, and segmentation trouble shooting requires a proficiency in data abstraction processes, such as hierarchical self-organized mapping [9,62]. Conversely, ROI-guided semiautomated segmentation methods using algorithms similar to those used here have been shown to quantify enhancing tumor reproducibly in the presence of resection cavity collapse and nonneoplastic enhancing tissues in less time than manual contouring, lending credence to the standardization and use of similar methods in neurosurgical trials [63–65].

Precision in image registration is crucial to postoperative tumor segmentation, as evaluating the complex morphology of residual tumor requires the ability to differentiate residual enhancing tumor tissue from T_1-hyperintense blood products (methemoglobin), nonneoplastic enhancing structures, such as normal blood vessels, and parenchyma in the presence of an often deformed resection cavity. When the proposed algorithms are applied to ROI blobs that sample image difference maps (subtraction images), misregistration of the resection cavity and surrounding tissues may result in misclassification of voxels. For example, if voxels containing postsurgical blood products that are hyperintense on T_1W images are slightly misregistered, these voxels may subsequently be subtracted from voxels containing enhancing tumor, resulting in a lower residual tumor volume. As the current work used only rigid image registration algorithms for the fusion of precontrast and postcontrast images, the use of more sophisticated registration methods, including deformable registration algorithms, such as multiresolution modified basis spline or multimodality Demons, may prove to be advantageous. Due to the low error of VelocityAI's deformable registration algorithms, as shown by Kirby et al., a sample of cases with subcubic centimeter residual tumor volumes were segmented using Fuzzy3 after both rigid and B-spline deformable image registration to investigate the impact of registration on algorithm performance [66]. Although no significant difference was found, a trend in both MED and mean percentage of volume error for rigid and deformable segmentation (1.61 ± 0.73 mm/50.7 ± 35.3% and 0.89 ± 0.64 mm/47.6 ± 37.3%, respectively) suggests that deformable registration may increase the accuracy of postoperative tumor segmentation and lends credence to its use in future studies.

Although the Fuzzy3 method described herein exhibited high agreement and reproducibility for the systematic determination of

contrast-enhancing tumor volumes, it exhibited some weaknesses common to intensity histogram thresholding. As the threshold between classes/clusters is determined solely by the intensity of the voxels within the image histogram, other strongly enhancing, nonneoplastic tissues or high-intensity noise may be misclassified as contrast-enhancing tumor. Although this is a significant problem when using global thresholding techniques, it is less of an issue when using local thresholding, because generating versatile ROI blobs that exclude obviously nonneoplastic enhancing tissues (dura, blood vessels, and other examples) can mitigate segmentation errors. Furthermore, these regions may be accounted for by the classification algorithms or simple postprocessing procedures. The ROI selection and postprocessing techniques are more systematic, more reproducible, and more easily standardized than manual contouring and should be considered as important components of tumor segmentation using this methodology. In addition, in this study, these algorithms were applied to a population of MR images with variable acquisition parameters and spatial resolutions. Ideally, in future work, only high-resolution 3D images (i.e., MP-RAGE) would be acquired and analyzed, as this decreases misclassification due to partial volume effects. Furthermore, as deformable registration seems to be favorable, acquiring high-resolution images is important, as registration performance is inversely proportional to image spatial resolution [67].

The proposed method Fuzzy3 overcomes many of the perceived pitfalls of manual and other semiautomated segmentation methods by coupling a flexible shape of blob generation method with simple, local histogram thresholding/clustering. The proposed algorithm allows for significant control of segmentation parameters, including the selection of pertinent data for analysis (i.e., ROI blob) and the desired number of clusters to be classified, while maintaining an intuitive workflow. Moreover, Fuzzy3 is easily integrated with volume-rendering software that is currently being used for clinical purposes, such as radiation therapy planning, thereby expanding the potential for implementation across multiple sites. As such, Fuzzy3 should be considered for the generation of preoperative and postoperative tumor volumes for EOR analysis in multicenter trials for glioblastoma (or other tumors). Future directions concerning this method consist of addressing the aforementioned drawbacks of intensity histogram thresholding, further automating and standardizing the method, particularly in regard to ROI blob generation and postprocessing procedures, and further investigating the effect of deformable registration and spatial resolution on segmentation accuracy.

Acknowledgments

References

[1] Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U, et al. (2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* **352**, 987–996.

[2] Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJB, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, et al. (2009). Effects of radiotherapy with concomitant and adjuvant temozolomide *versus* radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* **10**, 459–466.

[3] van den Bent MJ, Vogelbaum MA, Wen PY, Macdonald DR, and Chang SM (2009). End point assessment in gliomas: novel treatments limit usefulness of classical Macdonald's criteria. *J Clin Oncol* **27**, 2905–2908.

[4] Macdonald DR, Cascino TL, Schold SC Jr, and Cairncross JG (1990). Response criteria for phase II studies of supratentorial malignant glioma. *J Clin Oncol* **8**, 1277–1280.

[5] Wen PY, Macdonald DR, Reardon DA, Cloughesy TF, Sorensen AG, Galanis E, Degroot J, Wick W, Gilbert MR, Lassman AB, et al. (2010). Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J Clin Oncol* **28**, 1963–1972.

[6] White DR, Houston AS, Sampson WF, and Wilkins GP (1999). Intra- and interoperator variations in region-of-interest drawing and their effect on the measurement of glomerular filtration rates. *Clin Nucl Med* **24**, 177–181.

[7] Tofts PS and Collins DJ (2011). Multicentre imaging measurements for oncology and in the brain. *Br J Radiol* **84**, S213–S226.

[8] Sorensen AG, Patel S, Harmath C, Bridges S, Synnott J, Sievers A, Yoon YH, Lee EJ, Yang MC, Lewis RF, et al. (2001). Comparison of diameter and perimeter methods for tumor volume calculation. *J Clin Oncol* **19**, 551–557.

[9] Gordillo N, Montseny E, and Sobrevilla P (2013). State of the art survey on MRI brain tumor segmentation. *Magn Reson Imaging* **31**, 1426–1438.

[10] Bauer S, Wiest R, Nolte LP, and Reyes M (2013). A survey of MRI-based medical image analysis for brain tumor studies. *Phys Med Biol* **58**, R97–R129.

[11] Gooya A, Pohl KM, Bilello M, Cirillo L, Biros G, Melhem ER, and Davatzikos C (2012). GLISTR: glioma image segmentation and registration. *IEEE Trans Med Imaging* **31**, 1941–1954.

[12] Gu YH, Kumar V, Hall LO, Goldgof DB, Li CY, Korn R, Bendtsen C, Velazquez ER, Dekker A, Aerts H, et al. (2013). Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. *Pattern Recognit* **46**, 692–702.

[13] Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D, et al. (2012). Radiomics: the process and the challenges. *Magn Reson Imaging* **30**, 1234–1248.

[14] Vermes E, Childs H, Carbone I, Barckow P, and Friedrich MG (2013). Autothreshold quantification of late gadolinium enhancement in patients with acute heart disease. *J Magn Reson Imaging* **37**, 382–390.

[15] Stoutjesdijk MJ, Zijp M, Boetes C, Karssemeijer N, Barentsz JO, and Huisman H (2012). Computer aided analysis of breast MRI enhancement kinetics using mean shift clustering and multifeature iterative region of interest selection. *J Magn Reson Imaging* **36**, 1104–1112.

[16] Liu HT, Sheu TW, and Chang HH (2013). Automatic segmentation of brain MR images using an adaptive balloon snake model with fuzzy classification. *Med Biol Eng Comput* **51**, 1091–1104.

[17] Bindu CH (2009). An improved medical image segmentation algorithm using Otsu method. *International Journal of Recent Trends in Engineering* **2**, 88–90.

[18] Otsu N (1979). Threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* **9**, 62–66.

[19] Ahmed MN, Yamany SM, Mohamed N, Farag AA, and Moriarty T (2002). A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Trans Med Imaging* **21**, 193–199.

[20] Pham DL and Prince JL (1999). Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Trans Med Imaging* **18**, 737–752.

[21] Bezdek JC, Hall LO, and Clarke LP (1993). Review of MR image segmentation techniques using pattern recognition. *Med Phys* **20**, 1033–1048.

[22] Phillips WE II, Phuphanich S, Velthuizen RP, and Silbiger ML (1995). Automatic magnetic resonance tissue characterization for three-dimensional magnetic resonance imaging of the brain. *J Neuroimaging* **5**, 171–177.

[23] Phillips WE II, Velthuizen RP, Phuphanich S, Hall LO, Clarke LP, and Silbiger ML (1995). Application of fuzzy c-means segmentation technique for tissue differentiation in MR images of a hemorrhagic glioblastoma multiforme. *Magn Reson Imaging* **13**, 277–290.

[24] Khotanlou H, Colliot O, Atif J, and Bloch I (2009). 3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models. *Fuzzy Sets and Systems* **160**, 1457–1473.

[25] Stummer W, Novotny A, Stepp H, Goetz C, Bise K, and Reulen HJ (2000). Fluorescence-guided resection of glioblastoma multiforme by using 5-aminolevulinic acid-induced porphyrins: a prospective study in 52 consecutive patients. *J Neurosurg* **93**, 1003–1013.

[26] Stummer W, Pichlmeier U, Meinel T, Wiestler OD, Zanella F, and Reulen HJ (2006). Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *Lancet Oncol* **7**, 392–401.

[27] Stummer W, Reulen HJ, Meinel T, Pichlmeier U, Schumacher W, Tonn JC, Rohde V, Oppel F, Turowski B, Woiciechowsky C, et al. (2008). Extent of resection and survival in glioblastoma multiforme: identification of and adjustment for bias. *Neurosurgery* **62**, 564–576; discussion 564-576.

[28] Stummer W, Stepp H, Möller G, Ehrhardt A, Leonhard M, and Reulen HJ (1998). Technical principles for protoporphyrin-IX-fluorescence guided microsurgical resection of malignant glioma tissue. *Acta Neurochir (Wien)* **140**, 995–1000.

[29] Stummer W, Stocker S, Novotny A, Heimann A, Sauer O, Kempski O, Plesnila N, Wietzorrek J, and Reulen HJ (1998). *In vitro* and *in vivo* porphyrin accumulation by C6 glioma cells after exposure to 5-aminolevulinic acid. *J Photochem Photobiol B* **45**, 160–169.

[30] Stummer W, Stocker S, Wagner S, Stepp H, Fritsch C, Goetz C, Goetz AE, Kiefmann R, and Reulen HJ (1998). Intraoperative detection of malignant gliomas by 5-aminolevulinic acid-induced porphyrin fluorescence. *Neurosurgery* **42**, 518–525; discussion 525-526.

[31] Stummer W, Tonn JC, Mehdorn HM, Nestler U, Franz K, Goetz C, Bink A, and Pichlmeier U (2011). Counterbalancing risks and gains from extended resections in malignant glioma surgery: a supplemental analysis from the randomized 5-aminolevulinic acid glioma resection study. *J Neurosurg* **114**, 613–623.

[32] Chang HH, Zhuang AH, Valentino DJ, and Chu WC (2009). Performance measure characterization for evaluating neuroimage segmentation algorithms. *Neuroimage* **47**, 122–135.

[33] Zhang YJ (1996). A survey on evaluation methods for image segmentation. *Pattern Recognit* **29**, 1335–1346.

[34] Lin LI (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255–268.

[35] Lin L, Hedayat AS, Sinha B, and Yang M (2002). Statistical methods in assessing agreement: models, issues, and tools. *J Am Stat Assoc* **97**, 257–270.

[36] Lin L, Hedayat AS, and Wu WT (2012). *Statistical Tools for Measuring Agreement*. Springer, New York, NY.

[37] Dice LR (1945). Measures of the amount of ecologic association between species. *Ecology* **26**, 297–302.

[38] Deza MM (2012). General definitions. In *Encyclopedia of Distances*. Springer New York, NY. pp. 3–43.

[39] Lalkhen AG and McCluskey A (2008). Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care, and Pain* **8**, 221–223.

[40] Bland JM and Altman DG (1999). Measuring agreement in method comparison studies. *Stat Methods Med Res* **8**, 135–160.

[41] Harati V, Khayati R, and Farzan A (2011). Fully automated tumor segmentation based on improved fuzzy connectedness algorithm in brain MR images. *Comput Biol Med* **41**, 483–492.

[42] Rexilius J, Hahn HK, Klein J, Lentschig MG, and Peitgen HO (2007). Multispectral brain tumor segmentation based on histogram model adaptation. *Medical Imaging 2007: Computer-Aided Diagnosis, Pts 1 and 2* **6514**, V5140 art. no. 65140V.

[43] Ho S, Bullitt E, and Gerig G (2002). Level-set evolution with region competition: automatic 3-D segmentation of brain tumors. In *Proceedings of the 16th International Conference on Pattern Recognition*. Vol I. R Kasturi, D Laurendeau, and C Suen (Eds). IEEE Computer Society, Los Alamitos, CA. pp. 532–535.

[44] Sachdeva J, Kumar V, Gupta I, Khandelwal N, and Ahuja CK (2012). A novel content-based active contour model for brain tumor segmentation. *Magn Reson Imaging* **30**, 694–715.

[45] Wang T, Cheng I, and Basu A (2009). Fluid vector flow and applications in brain tumor segmentation. *IEEE Trans Biomed Eng* **56**, 781–789.

[46] Bauer S, Nolte LP, and Reyes M (2011). Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. *Med Image Comput Comput Assist Interv* **6893**, 354–361, Miccai 2011, Pt Iii.

[47] Zikic D, Glocker B, Konukoglu E, Criminisi A, Demiralp C, Shotton J, Thomas OM, Das T, Jena R, and Price SJ (2012). Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. *Med Image Comput Comput Assist Interv* **15**, 369–376.

[48] Ruan S, Lebonvallet S, Merabet A, and Constans JM (2007). Tumor segmentation from a multispectral MRI images by using support vector machine classification. *2007 4th IEEE Int Symp Biomed Imaging: Macro to Nano* **1-3**, 1236–1239.

[49] Zhu Y, Young GS, Xue Z, Huang RY, You H, Setayesh K, Hatabu H, Cao F, and Wong ST (2012). Semi-automatic segmentation software for quantitative clinical brain glioblastoma evaluation. *Acad Radiol* **19**, 977–985.

[50] Weizman L, Ben Sira L, Joskowicz L, Constantini S, Precel R, Shofty B, and Ben Bashat D (2012). Automatic segmentation, internal classification, and follow-up of optic pathway gliomas in MRI. *Med Image Anal* **16**, 177–188.

[51] Deoni SC, Rutt BK, Parrent AG, and Peters TM (2007). Segmentation of thalamic nuclei using a modified k-means clustering algorithm and high-resolution quantitative magnetic resonance imaging at 1.5 T. *Neuroimage* **34**, 117–126.

[52] Chen A, Deeley MA, Niermann KJ, Moretti L, and Dawant BM (2010). Combining registration and active shape models for the automatic segmentation of the lymph node regions in head and neck CT images. *Med Phys* **37**, 6338–6346.

[53] Deeley MA, Chen A, Datteri R, Noble JH, Cmelak AJ, Donnelly EF, Malcolm AW, Moretti L, Jaboin J, Niermann K, et al. (2011). Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. *Phys Med Biol* **56**, 4557–4577.

[54] Pallud J, Taillandier L, Capelle L, Fontaine D, Peyre M, Ducray F, Duffau H, and Mandonnet E (2012). Quantitative morphological magnetic resonance imaging follow-up of low-grade glioma: a plea for systematic measurement of growth rates. *Neurosurgery* **71**, 729–739; discussion 739-740.

[55] Mayr NA, Taoka T, Yuh WT, Denning LM, Zhen WK, Paulino AC, Gaston RC, Sorosky JI, Meeks SL, Walker JL, et al. (2002). Method and timing of tumor volume measurement for outcome prediction in cervical cancer using magnetic resonance imaging. *Int J Radiat Oncol Biol Phys* **52**, 14–22.

[56] Grenacher L, Thorn M, Knaebel HP, Vetter M, Hassenpflug P, Kraus T, Meinzer HP, Büchler MW, Kauffmann GW, and Richter GM (2005). The role of 3-D imaging and computer-based postprocessing for surgery of the liver and pancreas. *Rofo* **177**, 1219–1226.

[57] Wang Y, Lin ZX, Cao JG, and Li MQ (2011). Automatic MRI brain tumor segmentation system based on localizing active contour models. In *Advanced Research on Information Science, Automation and Material System, Pts 1-6*. H Zhang, G Shen, and D Jin (Eds). Trans Tech Publications LTD, Zurich, Switzerland. pp. 1342–1346.

[58] Iftekharuddin KM, Zheng J, Islam MA, and Ogg RJ (2009). Fractal-based brain tumor detection in multimodal MRI. *Appl Math Comput* **207**, 23–41.

[59] Dou WB, Ruan S, Chen YP, Bloyet D, and Constans JM (2007). A framework of fuzzy information fusion for the segmentation of brain tumor tissues on MR images. *Image Vis Comput* **25**, 164–171.

[60] Lee CH, Schmidt M, Murtha A, Bistritz A, Sander M, and Greiner R (2005). Segmenting brain tumors with conditional random fields and support vector machines. In *Computer Vision for Biomedical Image Applications*. Y Liu, T Jiang, and CS Zhang (Eds). Springer, Berlin, Germany. pp. 469–478.

[61] Prastawa M, Bullitt E, Ho S, and Gerig G (2004). A brain tumor segmentation framework based on outlier detection. *Med Image Anal* **8**, 275–283.

[62] Lashkari AE (2012). A neural network-based method for brain abnormality detection in MR images using Zernike moments and geometric moments. *Int J Comput Appl* **4**, 1–8.

[63] Kanaly CW, Ding D, Mehta AI, Waller AF, Crocker I, Desjardins A, Reardon DA, Friedman AH, Bigner DD, and Sampson JH (2011). A novel method for volumetric MRI response assessment of enhancing brain tumors. *PLoS One* **6**, e16031.

[64] Tacher V, Lin M, Chao M, Gjesteby L, Bhagat N, Mahammedi A, Ardon R, Mory B, and Geschwind JF (2013). Semiautomatic volumetric tumor segmentation for hepatocellular carcinoma: comparison between C-arm cone beam computed tomography and MRI. *Acad Radiol* **20**, 446–452.

[65] Chow DS, Qi J, Guo X, Miloushev VZ, Iwamoto FM, Bruce JN, Lassman AB, Schwartz LH, Lignelli A, Zhao B, et al. (2013). Semiautomated volumetric measurement on postcontrast MR imaging for analysis of recurrent and residual disease in glioblastoma multiforme. *AJNR Am J Neuroradiol*. Available at: http://www.ajnr.org/content/early/2013/08/29/ajnr.A3724.full.pdf+html?sid=37e0fc78-f40f-4c54-a55c-943d777a6ed9.

[66] Kirby N, Chuang C, Ueda U, and Pouliot J (2013). The need for application-based adaptation of deformable image registration. *Med Phys* **40**, 011702.

[67] Hajnal J, Hawkes D, and Hill D (2001). In *Medical Imaging Registration*. J Hajnal, D Hawkes, and D Hill (Eds). CRC Press LLC, Boca Raton, FL. pp. 11–117.

Table W1. Preoperative Tumor Volumes Generated Using Manual, Otsu, and Fuzzy Methods.

Study ID	Contrast-Enhancing Tumor Volume (cm^3)						
	Manual	Reader 1 Otsu4	Reader 2 Otsu4	Reader 1 Fuzzy3	Reader 2 Fuzzy3	Reader 1 Fuzzy4	Reader 2 Fuzzy4
Case 1	3.59	2.41	6.10	3.58	3.51	3.547	3.547
Case 2	83.98	82.08	83.21	90.36	86.15	96.88	88.83
Case 3	22.36	23.23	27.69	22.47	22.64	22.56	21.14
Case 4	41.81	48.45	47.05	41.82	41.64	39.75	39.75
Case 5	8.38	9.20	8.72	8.49	8.01	8.25	8.32
Case 6	42.73	47.31	44.71	42.12	45.09	43.89	37.23
Case 7	11.94	15.24	12.77	12.72	10.85	12.05	10.45
Case 8	51.19	61.85	50.14	50.81	48.74	51.41	50.90
Case 9	4.15	6.59	3.68	4.38	3.34	4.02	3.61
Case 10	15.95	15.77	6.91	15.73	15.17	16.27	15.69
Case 11	4.80	4.87	4.65	4.87	4.88	5.07	4.16
Case 12	76.19	88.52	77.80	80.56	77.82	76.80	75.42
Case 13	66.75	89.27	68.63	66.30	67.62	69.69	65.81
Case 14	4.13	3.31	3.94	4.17	4.05	3.98	4.04
Case 15	53.34	51.98	50.44	52.94	52.9	53.09	57.10
Case 16	18.23	16.43	19.99	18.14	18.03	18.42	16.35
Case 17	13.97	14.54	15.16	14.00	14.49	13.45	14.07
Case 18	54.13	61.69	52.58	49.30	52.01	52.45	51.18
Case 19	10.92	7.24	16.30	10.85	11.10	11.02	11.60
Case 20	20.05	14.69	34.83	20.82	21.31	19.44	19.03
Case 21	30.94	31.03	31.11	32.75	28.95	31.46	30.65
Case 22	35.72	40.71	42.84	36.35	33.79	36.65	37.51
Case 23	54.41	65.54	59.19	54.88	55.74	61.82	60.76
Case 24	89.40	126.8	84.66	92.59	89.80	87.98	95.39
Case 25	3.92	4.19	9.30	3.60	3.67	3.66	3.39
Case 26	16.23	23.51	15.84	16.38	15.14	17.29	15.47
Case 27	29.66	32.50	32.64	29.98	30.03	29.54	29.11
Case 28	101.75	121.36	112.38	112.97	95.59	106.56	78.63
Case 29	65.59	84.86	45.88	65.10	64.63	65.51	67.64
Case 30	23.88	22.29	23.16	23.82	23.38	24.09	24.58
Case 31	33.40	30.77	31.17	33.11	33.89	33.05	30.63
Case 32	54.60	53.62	52.83	54.86	54.37	53.95	53.76
Case 33	29.22	25.97	24.23	28.9	20.90	28.79	21.04
Case 34	15.01	15.51	13.88	14.70	13.83	15.22	13.15
Case 35	9.67	8.98	8.51	9.32	9.13	9.81	9.01
Case 36	34.91	31.38	30.67	33.59	34.15	33.05	32.30
Case 37	82.58	72.00	67.66	81.93	70.54	84.29	58.78

Readers 1 and 2 are separate readers remote in time. Line demarcates the end of 32 cases in 5-ALA trial.

Table W2. Postoperative Tumor Volumes Generated Using Manual, Otsu, and Fuzzy Methods.

Study ID	Contrast-Enhancing Tumor Volume (cm³)						
	Manual	Reader 1 Otsu4	Reader 2 Otsu4	Reader 1 Fuzzy3	Reader 2 Fuzzy3	Reader 1 Fuzzy4	Reader 2 Fuzzy4
Case 1	0.22	0.41	1.46	0.43	0.30	0.28	0.29
Case 2	8.25	15.24	19.88	7.07	8.00	8.91	5.25
Case 3	1.49	2.11	2.84	1.77	1.42	1.62	1.93
Case 4	0.82	2.71	0.99	0.65	1.72	2.12	2.17
Case 5	0.62	0.47	0.90	0.62	0.69	0.39	0.75
Case 6	5.0	5.28	6.32	4.32	5.26	4.97	5.84
Case 7	0.08	0.97	0.90	0.05	0.05	0.03	0.00
Case 8	0.05	0.72	0.60	0.65	0.55	0.41	0.65
Case 9	0.01	0.28	0.33	0.11	0.19	0.10	0.31
Case 10	0.69	1.53	3.49	0.65	0.76	0.68	0.90
Case 11	0.04	1.80	3.33	0.06	0.03	0.13	0.66
Case 12	5.87	10.32	10.87	5.58	5.28	7.07	5.88
Case 13	3.94	3.93	2.59	3.66	4.95	3.60	5.01
Case 14	0.64	0.80	0.75	0.72	0.66	0.74	0.52
Case 15	0.52	1.44	6.21	0.86	1.38	0.41	1.28
Case 16	0.05	0.35	0.11	0.08	0.09	0.04	0.04
Case 17	1.28	0.61	0.70	1.26	1.97	1.44	1.58
Case 18	8.06	5.16	5.42	7.83	8.31	8.08	4.54
Case 19	1.20	0.83	2.03	1.28	1.21	1.30	0.84
Case 20	2.36	2.10	2.12	2.17	2.94	2.39	3.61
Case 21	1.58	2.07	2.59	2.02	1.95	1.98	0.96
Case 22	10.42	11.33	11.89	9.70	10.67	10.29	9.29
Case 23	8.60	9.37	10.89	8.35	8.98	8.55	9.31
Case 24	1.91	3.61	5.24	2.56	1.68	2.60	3.05
Case 25	0.64	1.11	3.18	1.38	1.21	0.84	1.71
Case 26	0.03	0.89	0.79	0.55	0.72	0.33	0.27
Case 27	2.12	1.53	1.62	2.36	2.79	2.15	1.29
Case 28	7.28	11.31	15.49	8.27	7.77	9.34	5.89
Case 29	5.77	13.0	13.64	5.70	5.14	8.33	8.80
Case 30	8.62	9.40	8.31	9.83	9.79	8.36	7.49
Case 31	2.95	1.80	6.21	2.94	2.83	3.09	4.05
Case 32	0.91	0.99	0.92	0.92	0.94	0.93	0.79
Case 33	3.73	3.95	2.84	3.16	2.35	3.70	3.46
Case 34	7.63	9.03	9.35	7.20	7.92	8.00	5.60
Case 35	0.91	0.99	0.95	0.88	1.15	1.58	1.23
Case 36	0.58	0.53	0.51	1.76	0.67	0.98	0.43
Case 37	1.14	2.29	2.54	1.93	2.17	1.56	2.01

Readers 1 and 2 are separate readers remote in time. Line demarcates the end of 32 cases in 5-ALA trial.
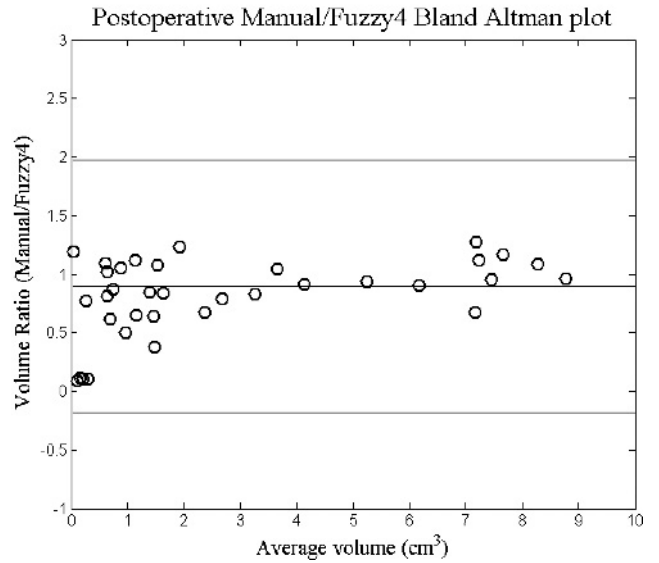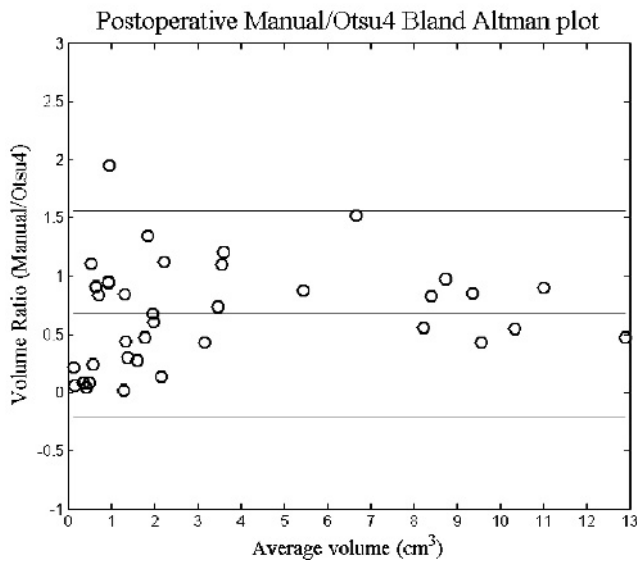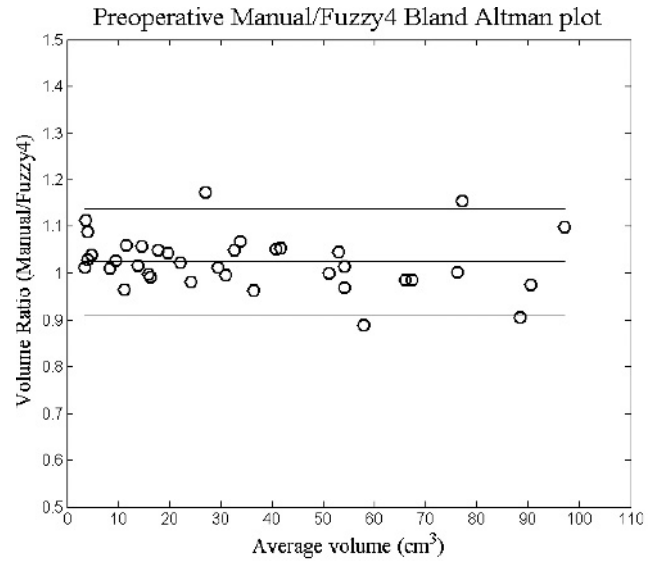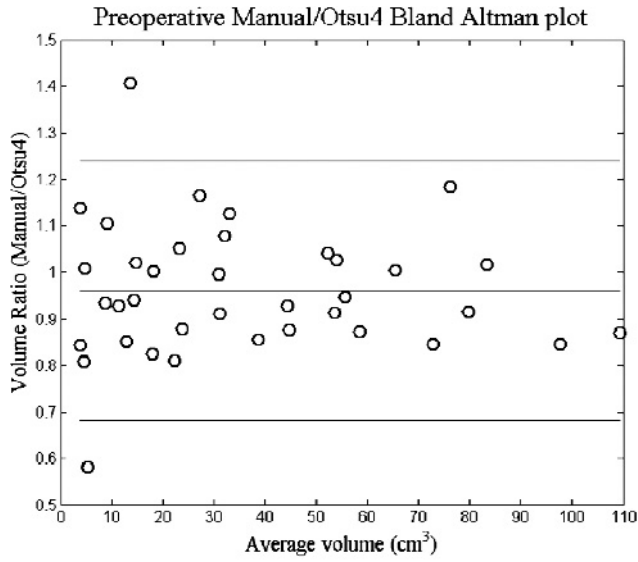
Figure W1. Preoperative and postoperative Bland-Altman plots of manual tumor contour volume *versus* mean semiautomated segmentation volume using Otsu4.

Figure W2. Preoperative and postoperative Bland-Altman plots of manual tumor contour volume *versus* mean semiautomated segmentation volume using Fuzzy4.
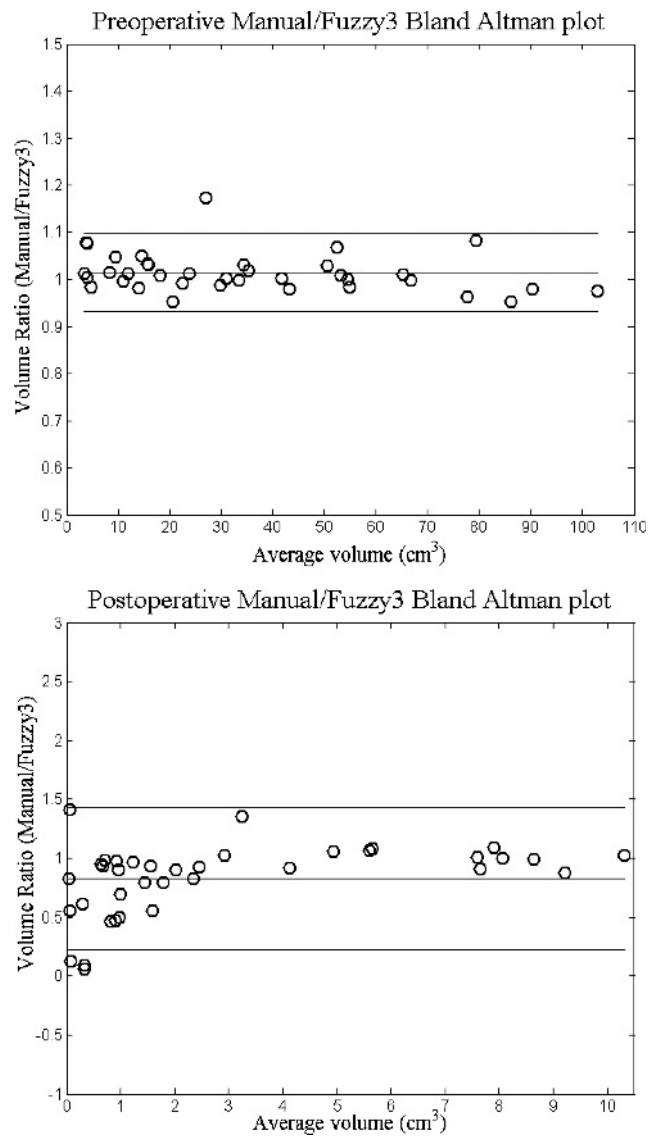
Figure W3. Preoperative and postoperative Bland-Altman plots of manual tumor contour volume *versus* mean semiautomated segmentation volume using Fuzzy3.

Table W3. Mean Ratios, Ratio SDs, Upper, and Lower 95% Limits of Agreement Preoperatively and Postoperatively for Bland-Altman Plots.

	Preoperative			Postoperative		
	Otsu4	Fuzzy4	Fuzzy3	Otsu4	Fuzzy4	Fuzzy3
Geometric mean ratio	0.961	1.024	1.014	0.678	0.894	0.828
SD	0.142	0.058	0.042	0.451	0.550	0.306
Upper 95% limit of agreement	1.239	1.137	1.097	1.562	1.972	1.428
Lower 95% limit of agreement	0.682	0.911	0.931	–0.206	–0.184	0.228

Table W4. Mean CCCs with SEM and 95% CIs for Preoperative Tumor Volumes Generated Using Automated Methods *versus* Manual Contouring by Two Separate Readers (Reader 1 and Reader 2).

	Otsu4	Fuzzy4	Fuzzy3
CCC	0.981[*,†]	0.993[*,‡]	0.998[†,‡]
SEM	0.006	0.002	0.001
95% CI	(0.963-0.990)	(0.988-0.998)	(0.997-0.999)

CCC, concordance correlation coefficient; CI, confidence interval.
[*]P = .027.
[†]P = .005.
[‡]P = .025.

Table W5. Mean CCCs with SEM and 95% CIs for Postoperative Tumor Volumes Generated Using Automated Methods *versus* Manual Contouring by Two Separate Readers (Reader 1 and Reader 2).

	Otsu4	Fuzzy4	Fuzzy3
CCC	0.780[*,†]	0.971[*,‡]	0.990[†,‡]
SEM	0.062	0.009	0.003
95% CI	(0.626-0.875)	(0.953-0.990)	(0.983-0.996)

CCC, concordance correlation coefficient; CI, confidence interval.
[*]P = .002.
[†]P = .001.
[‡]P = .045.

Table W6. CCCs with SEM and 95% CIs for Interreader Agreement of Preoperative and Postoperative Tumor Volumes Generated by Separate Readers (Reader 1 and Reader 2) Using Fuzzy3.

	Preoperative	Postoperative
CCC	0.990	0.983
SEM	0.003	0.006
95% CI	(0.981-0.995)	(0.981-0.995)

CCC, concordance correlation coefficient; CI, confidence interval.