The predictive value of segmentation metrics on dosimetry in organs at risk of the brain.

Abstract

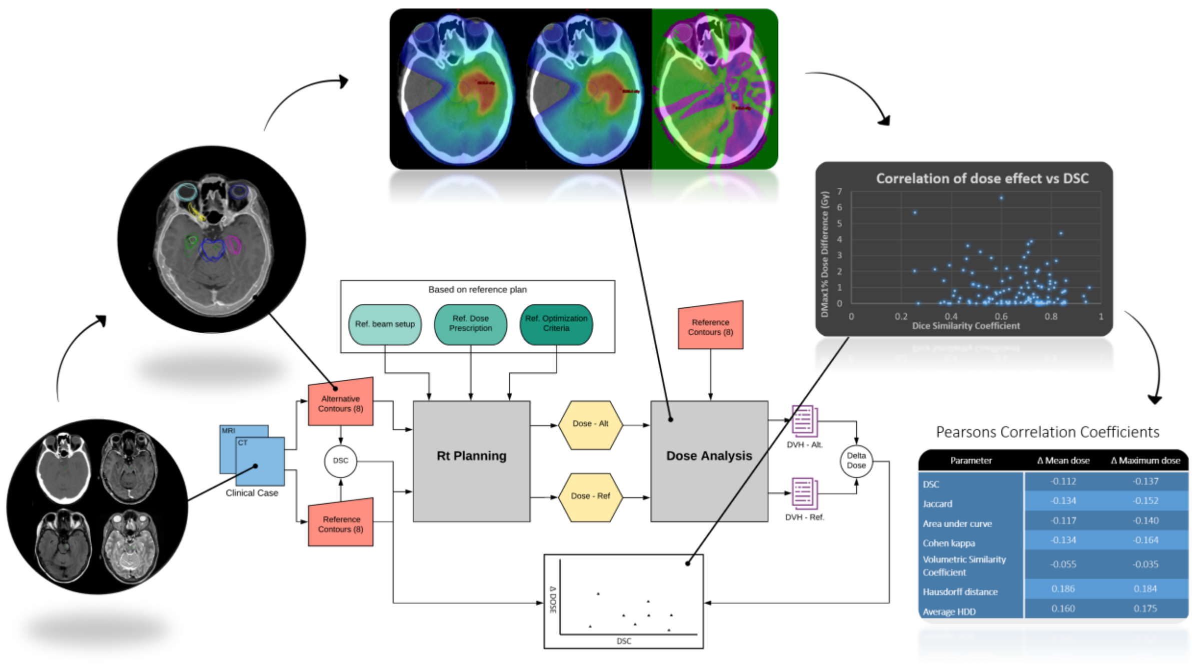
Background: Fully automatic medical image segmentation has been a long pursuit in radiotherapy (RT). Recent developments involving deep learning show promising results yielding consistent and time efficient contours. In order to train and validate these systems, several geometric based metrics, such as Dice Similarity Coefficient (DSC), Hausdorff, and other related metrics are currently the standard in automated medical image segmentation challenges. However, the relevance of these metrics in RT is questionable. The quality of automated segmentation results needs to reflect clinical relevant treatment outcomes, such as dosimetry and related tumor control and toxicity. In this study, we present results investigating the correlation between popular geometric segmentation metrics and dose parameter for Organs-At-Risk (OAR) in brain tumor patients, and investigate properties that might be predictive for dose changes in brain radiotherapy.

Methods: A retrospective database of glioblastoma multiforme patients (GBM) was stratified for planning difficulty where after 12 cases were selected and reference sets of OARs and radiation targets were defined. In order to assess the relation between segmentation quality -as measured by standard segmentation assessment metrics- and quality of RT plans, alternative contours for each OAR of the selected cases were obtained through three methods: (i) Manual contours by two additional human raters. (ii) Realistic manual manipulations of reference contours (two sets for each case). (iii) Through deep learning based segmentation results (one set for each case). On the reference structure set a reference plan was generated that was re-optimized for each corresponding alternative contour-set. The correlation between segmentation metrics, and dosimetric changes was obtained and analyzed for each OAR, by means of the mean dose and maximum dose to 1% of the volume (Dmax 1%).

Furthermore, we conducted specific experiments to investigate the dosimetric effect of alternative OAR contours with respect to the proximity to the target, particular shape and relative location to the target.

Results: We found a low correlation between the DSC, reflecting the alternative OAR contours, and dosimetric changes. The Pearson correlation coefficient between the mean OAR dose effect and the Dice was -0.11, -0.17 and -0.31 for the human-rater, manually manipulated, and the deep-learning based alternative contours, respectively. For Dmax 1%, we found a correlation of -0.08, -0.13, and -0.13 for the human-rater, manually manipulated, and the deep-learning based alternative contours, respectively. Similar findings were found for 22 other geometric similarity indices, suggesting a low correspondence between existing segmentation metrics and their possible dosimetric effect. Furthermore, we found that proximity to the target does not make contour variations more susceptible to the dose effect. However, the direction of the contour variation with respect to the relative location of the target seems to have a strong correlation with the dose effect.

Conclusions: This study shows a low correlation between segmentation metrics and dosimetric changes for OAR in brain tumor patients. Results suggest that the current metrics for image segmentation in RT, as well as deep learning systems employing such metrics, need to be revisited towards clinically oriented metrics that better reflect how segmentation quality affects dose distribution and related tumor control and toxicity.



Correlation of dose effect vs DSC
DMax1% Dose Difference (Gy)
Dice Similarity Coefficient
Based on reference plan
Ref. beam setup
Ref. Dose Prescription
Ref. Optimization Criteria
Reference Contours (8)
Alternative Contours (8)
MRI
CT
Clinical Case
DSC
Rt Planning
Dose - Alt
Dose - Ref
Dose Analysis
DVH - Alt.
DVH - Ref.
Delta Dose
Δ DOSE
DSC
Pearsons Correlation Coefficients
Parameter | Δ Mean dose | Δ Maximum dose
DSC | -0.112 | -0.137
Jaccard | -0.134 | -0.152
Area under curve | -0.117 | -0.140
Cohen kappa | -0.134 | -0.164
Volumetric Similarity Coefficient | -0.055 | -0.035
Hausdorff distance | 0.186 | 0.184
Average HDD | 0.160 | 0.175