European Journal of Radiology 95 (2017) 75–81

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

Comparison of perioperative automated versus manual two-dimensional tumor analysis in glioblastoma patients

Frauke Kellner-Weldon[a,*], Christoph Stippich[b], Roland Wiest[a], Vera Lehmann[a], Raphael Meier[c], Jürgen Beck[d], Philippe Schucht[d], Andreas Raabe[d], Mauricio Reyes[c], Andrea Bink[b]

[a] Support Center for Advanced Neuroimaging – Institute for Diagnostic and Interventional Neuroradiology, University Hospital Inselspital and University of Bern, Bern, Switzerland
[b] Department of Radiology, Division of Diagnostic and Interventional Neuroradiology, University Hospital, Basel, Switzerland
[c] Institute of Surgical Technology and Biomechanics, University of Bern, Bern, Switzerland,
[d] Department of Neurosurgery, University Hospital Inselspital and University of Bern, Bern, Switzerland

ARTICLE INFO

Keywords:
Computer assisted reading
automated data analysis
glioblastoma
machine learning
MRI

ABSTRACT

Objectives: Current recommendations for the measurement of tumor size in glioblastoma continue to employ manually measured 2D product diameters of enhancing tumor. To overcome the rater dependent variability, this study aimed to evaluate the potential of automated 2D tumor analysis (ATA) compared to highly experienced rater teams in the workup of pre- and postoperative image interpretation in a routine clinical setting.

Materials and methods: From 92 patients with newly diagnosed GB and performed surgery, manual rating of the sum product diameter (SPD) of enhancing tumor on magnetic resonance imaging (MRI) contrast enhanced T1w was compared to automated machine learning-based tumor analysis using FLAIR, T1w, T2w and contrast enhanced T1w.

Results: Preoperative correlation of SPD between two rater teams (1 and 2) was $r = 0.921$ ($p < 0.0001$). Difference among the rater teams and ATA ($p = 0.567$) was not statistically significant. Correlation between team 1 vs. automated tumor analysis and team 2 vs. automated tumor analysis was $r = 0.922$ and $r = 0.897$, respectively ($p < 0.0001$ for both). For postoperative evaluation interrater agreement between team 1 and 2 was moderate (Kappa 0.53). Manual consensus classified 46 patients as completely resected enhancing tumor. Automated tumor analysis agreed in 13/46 (28%) due to overestimation caused by hemorrhage and choroid plexus enhancement.

Conclusions: Automated 2D measurements can be promisingly translated into clinical trials in the preoperative evaluation. Immediate postoperative SPD evaluation for extent of resection is mainly influenced by postoperative blood depositions and poses challenges for human raters and ATA alike.

1. Introduction

Glioblastoma is the most common primary brain tumor in adults [1]. Complete resection of this tumor entity is currently not possible, due to its infiltrating growth [2]. Still, the largest possible extent of resection is the primary goal of surgery as it has been shown to improve overall survival [3,4]. Therefore, an accurate preoperative evaluation of the tumor is needed prior to surgery. Current recommendations for the measurement of tumor size (Response-Assessment Neuro-Oncology (RANO) working group) [5], continue to employ two-dimensional (2D) product diameters of enhancing tumor on MRI. To overcome the rater dependent variability and the time consuming measurement [6–8], automated delineation methods that segment tumor as a three-dimensional (3D) volume have been developed [9,10].

Automated and manual tumor subcompartment delineation has shown comparable performance in terms of prognosis and correlation with Visually AcceSAble Rembrandt Images (VASARI) features [11].

The aim of this study was to evaluate the performance of automated tumor analysis (ATA) to identify complete vs. incomplete resections in comparison with human image interpretations and 2-D diameter measurements. Complete resection (CRET) was defined as absence of any contrast-enhancing tumor volume on ceT1w imaging after surgical

Abbreviations: GB, glioblastoma; MRI, magnetic resonance imaging; FLAIR, fluid attenuation inversion recovery; SPD, sum product of diameter; ATA, automated tumor analysis; ceT1w, contrast enhanced T1weighted

* Corresponding author at: Institute for Diagnostic and Interventional Neuroradiology, Inselspital, Bern University Hospital, Freiburgstrasse 4, 3010, Bern, Switzerland.
E-mail address: Frauke.Kellner-Weldon@insel.ch (F. Kellner-Weldon).

http://dx.doi.org/10.1016/j.ejrad.2017.07.028
Received 20 March 2017; Received in revised form 21 July 2017; Accepted 31 July 2017

procedure.

Although 3-D measures are increasingly recognized as alternative surrogate markers for tumor progression, 2-D measures are still recommended for the routine clinical follow-up. The goal of this study was to evaluate i) the agreement between automatic and manual estimates of 2-D tumor measures in preoperative images and ii) if automated tools can detect the amount of resection if 2-D measures are applied.

2. Materials and methods

2.1. Study population

We retrospectively identified 92 patients who were diagnosed with a histologically proven GBM and who underwent resective surgery and immediate postoperative MRI after resection. All patients were examined between 2009 and 2013. Inclusion criteria were newly diagnosed, untreated and histologically confirmed GB (WHO IV), performed surgery, Karnofsky performance ≥70 and age > 18 y. Exclusion criteria were prior other malignancy, biopsy performed previously on the GB, postoperative MRI later than 72 h. The study was approved by the Local Research Ethics Commission. All patients provided written informed consent.

2.2. Automated image processing

We used the automatic brain tumor analysis software BraTumIA, which has been clinically evaluated for longitudinal tumor volumetry in previous studies. Within the framework of this study, we were interested in the automatic detection of enhancing tumor, which is part of the more extended volume analysis that is offered by the software. A detailed description of the software and its potential applications has been published previously [12,13]. In short, BraTumIA is a supervised machine learning based software that relies on expert annotated training data to learn the relationships between imaging features and tissue classification. It relies on multisequence MRI (T1weighted (w), contrast enhanced (ce) T1w, T2w, fluid attenuated inversion recovery (FLAIR)) to perform automated tumor analysis. It performs co-registration of the multisequence images, skull-stripping (i.e. brain extraction), and tissue classification. Beyond volumetric measurements, BraTumIA also provides measures of (2D) tumor diameters.

As human reference, four board-certified neuroradiologists with a mean of 13 years (range, 7–17y) of experience in neuroradiology, working in two different university hospitals, rated the imaging data in two teams (team 1 and team 2, each consisting of two raters from both hospitals) in a consensus fashion.

2.3. Manual annotations and quality control

Manual annotations were performed on contrast enhanced T1weighted images (ceT1w) after checking for confounding blood products on T1w images. As annotations we used sum of the products of diameters (SPD) measures, and presence/absence of residual tumor. Tumor SPD measures (in mm^2) were acquired according to the RANO recommendations, using a hospital picture archiving system (Sectra IDS7, Linköping, Sweden) by each team. Similarly, each team rated the presence/absence of residual tumor. In case of disagreements, a final joined consensus reading was achieved by team 3, which consisted of one rater from each team 1 and 2.

For automated analysis, BraTumIA was trained on an independent dataset of 54 pre- and post-operative cases, as described in [12]. Upon automated analysis of the segmented pre- and postoperative MRI data sets, two readers performed quality control of the results to see if (i) the data were inconclusively processed due to motion artefacts or (ii) if the skull stripping process had failed. If any of these conditions was found, the MRI dataset was removed.

2.4. Statistical methods

Statistical analysis was performed with IBM SPSS Statistics 21.ink and the R software package. Agreement on presence/absence of residual tumor between teams 1 and 2 was assessed using the Cohen's kappa statistic. A Kruskal-Wallis test was used to assess multiple differences between manual and automated preoperative mean values of tumor diameters. Paired differences were analyzed using a Wilcoxon signed-rank test. Correlation of results for mean SPD-values between raters and automated analysis was analyzed using Pearson's r. A value of $p < 0.05$ was considered statistically significant.

3. Results

92 patients (47 male, 45 female; mean 62 y, range 60–80 y) met the inclusion criteria. Time between pre- and postoperative MRI was 7.5 days, (range 1–55, median 5.5 days) and between operation and postoperative MRI 1.15 days (range 0–2 days). All in-house examinations (pre-/postoperative) (n = 81/92) were performed on MR scanners of the same vendor (Siemens Medical Solutions, Erlangen, Germany- 1.5 T Avanto (n = 30/49), 1.5 T Aera (n = 12/12), 3.0 T Verio (n = 21/20) and 3.0 Trio (n = 18/11)). Of the remaining 11 preoperative MR images six were performed on Siemens scanners [1.5 T Avanto (n = 3), 1.5 T unknown scanner type (n = 2), 1.0 T unknown scanner type (n = 1)], three on Philips scanners [1.5 T Intera (n = 1), 1.5 T unknown scanner type (n = 1), 3.0 T unknown scanner type (n = 1)], and two on GE scanners [1.0 T unknown scanner type (n = 1), and 3.0 T Discovery MR750 (n = 1)]. The sequence parameters of the T1w, T2w, FLAIR and contrast-enhanced T1 w (ceT1) images are shown in Table 1. In-house patients received 0.1 mmol/kg gadolinium based contrast agent (Gd-DTPA). Data on contrast agent and dose for non in-house patients was not available. For an overview of processed patient data see Fig. 1.

3.1. Manual annotation of 2D diameters

Correlation (Pearson's R-value) of SPD values measured between

Table 1
Pre- and postoperative MRI sequence parameters included for automated analysis (n = 60 patients).

preoperative	postoperative	Sequence T1 non-enhanced	Sequence T1 contrast-enhanced	Sequence T2 weighted	Sequence FLAIR
7	7	*T1 5* mm[a]	*T1C 5* mm	*T2 5* mm	FLAIR 4–5 mm
2	1	*T1 5* mm	*T1C 5* mm	T2 3D 1 mm	FLAIR 4–5 mm
25	28	*T1 5* mm	T1C 3D 1 mm	*T2 5* mm	FLAIR 4–5 mm
11	12	*T1 5* mm	T1C 3D 1 mm	T2 3D 1 mm	FLAIR 4–5 mm
2	0	*T1 5* mm	T1C 3D 1 mm	*T2 5* mm	FLAIR 3D 1 mm
2	1	T1 3D 1 mm	T1C 3D 1 mm	*T2 5* mm	FLAIR 4–5 mm
11	10	T1 3D 1 mm	T1C 3D 1 mm	T2 3D 1 mm	FLAIR 4–5 mm
0	1	T1 3D 1 mm	*T1C 5* mm	T2 3D 1 mm	FLAIR 4–5 mm

[a] cursive: spin echo sequences; all others gradient echo sequences.