Evaluating Glioma Growth Predictions as a Forward Ranking Problem

Karin A. van Garderen[1(✉)], Sebastian R. van der Voort[1], Maarten M. J. Wijnenga[1], Fatih Incekara[1,2], Georgios Kapsas[1], Renske Gahrmann[1], Ahmad Alafandi[1], Marion Smits[1], and Stefan Klein[1]

[1] Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands
k.vangarderen@erasmusmc.nl
[2] Department of Neurosurgery, Erasmus MC, Rotterdam, The Netherlands

Abstract. The problem of tumor growth prediction is challenging, but promising results have been achieved with both model-driven and statistical methods. In this work, we present a framework for the evaluation of growth predictions that focuses on the spatial infiltration patterns, and specifically evaluating a prediction of future growth. We propose to frame the problem as a ranking problem rather than a segmentation problem. Using the average precision as a metric, we can evaluate the results with segmentations while using the full spatiotemporal prediction. Furthermore, by applying a biophysical tumor growth model to 21 patient cases we compare two schemes for fitting and evaluating predictions. By carefully designing a scheme that separates the prediction from the observations used for fitting the model, we show that a better fit of model parameters does not guarantee a better predictive power.

Keywords: Glioma · Growth model · Validation · Magnetic resonance imaging · Brain

1 Introduction

As the diagnosis and delineation of glioma has improved with machine learning [4], researchers look towards the more challenging task of predicting the disease trajectory into the future [8,19]. However, the problem of tumor growth is challenging in many ways, not just by the lack of publicly available data. The variables of clinical importance, such as the speed of infiltration and proliferation, are unknown and the problem of estimating them from observations is ill-posed. Furthermore, the observations we do have are flawed as tumor cells are known to spread beyond the visible boundary on MR imaging [22].

Despite these challenges, biophysical growth models have shown promise in their ability to predict the spatial growth patterns for individual cases. They are model-driven and strongly rooted in a mechanistic understanding of tumor growth. Delineations of the tumor on MR imaging typically form the input for

A. Crimi and S. Bakas (Eds.): BrainLes 2021, LNCS 12962, pp. 100–111, 2022.
https://doi.org/10.1007/978-3-031-08999-2_8

individual model fitting, with follow-up imaging providing the gold standard of evaluation. Though other methods of evaluation exist, such as biopsy samples [10] or PET imaging [18], for most clinical cases consecutive delineations are the best approximation for a ground truth.

Due to the nature of the data, growth predictions are often framed as a segmentation problem. For example, by using an overlap metric such as the Dice Similarity Coefficient based on a sample in time [7,19]. Although this metric comes natural to the ground-truth data, it is less representative of the underlying problem. The main disadvantage of overlap-based metrics is that they treat all voxels equally, while some errors are more significant than others. Intuitively, we would want to assign more significance to false negative predictions at a large distance to the predicted tumor boundary as they represent a larger disagreement to the model and would likely require a large adjustment to predict correctly. This intuition is represented in metrics based on the segmentation boundary, such as the symmetric surface distance used in Konukoglu et al. [17]. But even a distance metric compares only to a single point in time, and using a boundary metric becomes less appropriate when the ground truth contains new disconnected lesions.

Another challenge in the evaluation of tumor growth predictions is the entanglement of model fit and prediction. All tumor growth models require an initial observation to fit model parameters. The goodness-of-fit is measured using the segmentation on this initial observation and the prediction is performed from the time of onset, through the initial observation towards the future [3]. The optimization of this inverse problem is an important topic for research, not in the least because the growth parameters can be of prognostic value by themselves [21], but often these methods are evaluated in simulated data. The clinical reality will not adhere to the strict assumptions made in the model, and therefore the predictive value of the model depends not only on the effectiveness of the model fitting but also on the correctness of the assumptions.

An ideal test of a prediction model would require a strict separation of model fitting and evaluation. However, in the problem of personalized tumor growth models this separation is not strictly possible because the initial condition used for the parameter fit is also part of the final tumor shape used for evaluation. Especially with models that simulate the full growth trajectory, there is a risk that model fit on the initial condition is strongly entangled with the prediction of growth. After all, if the shape of the initial lesion is not estimated correctly then this error will propagate to the estimation the future disease trajectory. This work explores the distinction between goodness-of-fit at the initial time-point, and predictive performance for future time-points by comparing two temporal evaluation schemes, one of which aims to strictly separate the initial condition from the predicted growth behavior.

In this work we propose the following contributions:

1. A novel framing of tumor growth as a ranking problem, with the Average Precision as the performance metric

2. The application of this evaluation framework on a biophysical tumor growth model and a dataset of 21 patient cases, to explore the relation between goodness-of-fit at the initial time-point, and predictive performance for future time-points.

2 Methods

2.1 Tumor Growth as a Ranking Problem

In this section, we propose that tumor growth prediction could be framed as a ranking problem, aimed at predicting the relative time-to-invasion of each voxel in the brain. Based on this perspective, we propose an evaluation metric for assessing the quality of the predictions (i.e., rankings) resulting from any growth model. This problem formulation is aimed at predicting infiltrative growth in a spatial sense, and simplifies the problem by disregarding the speed of growth and potential mass effect.

We assume that a growth model could produce a segmentation of the tumor $S(t)$ at any time $t > 0$. It may therefore assign to every location in the brain a time $T(x)$, which is the first time t when the tumor reaches that location. As we do not require an accurate estimation of the growth speed, we require only that the estimated $T(x)$ is a ranking of voxels in the brain, such that:

$$T(x_a) > T(x_b) \Leftrightarrow \exists t : x_a \notin S(t), x_b \in S(t). \tag{1}$$

The ranking can be evaluated by a sampling of the ground-truth segmentation S', by using the Average Precision (AP). The AP is defined as the area under the Precision-Recall (PR) curve:

$$AP = \Sigma_t (R(t) - R(t-1))P(t), \tag{2}$$

where $R(t)$ and $P(t)$ are the recall and precision at a threshold t on the time-to-invasion ranking T, leading to the predicted segmentation $S(t) = \{x : T(x) \leq t\}$, and comparing to the reference segmentation S':

$$P(t) = \frac{|S(t) \cap S'|}{|S(t)|}, R(t) = \frac{|S(t) \cap S'|}{|S'|}. \tag{3}$$

The AP metric weighs the precision scores are with the difference in recall, so that all tumor volume predictions $S(t)$ are taken into account from the tumor onset to the time when the recall is 1. This is when the ground-truth segmentation is completely encompassed by the prediction $S(t)$. An evaluation based on a single time t would represent a point on the PR curve. If we take a volume-based sample, where the estimated tumor volume equals the observed tumor volume, i.e. $|S(t)| = |S'|$, this is the time t where $R(t) = P(t)$.

Formulating the problem as a ranking and using the AP has a number of qualitative advantages. First, the ranking T has a direct local connection to the speed of the tumor boundary. If the ranking is smooth, the gradient of the

T represents the local movement of the visible tumor boundary. It automatically assigns a larger weight to certain parts of the prediction, depending on the assigned ranking T, regardless of any assumptions on the significance of distance in space or time. We might quantify the agreement between T and S locally by using the rank of the voxel $T(x)$ as a threshold on the PR curve. A local prediction $T(x)$ is in agreement with S' if it is part of the ground-truth segmentation ($x \in S'$) and can be included with high precision $P(T(x))$, or else if it falls outside S' but can be excluded with high recall $R(T(x))$. Figure 1 illustrates the computation of the AP metric and this local measure of disagreement.

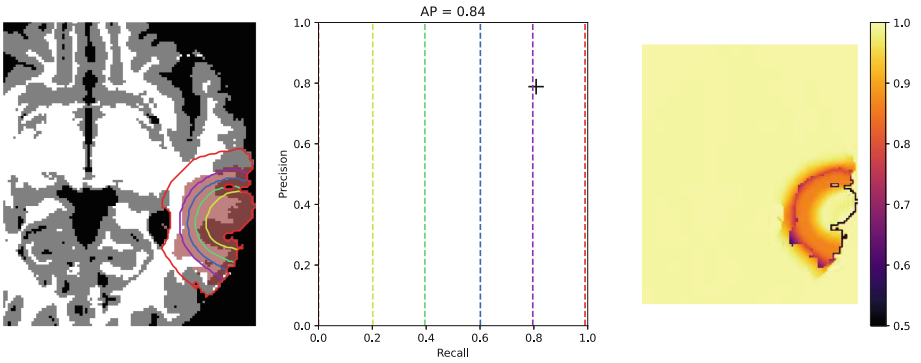

Fig. 1. Left: cross-section of tissue segmentation of a specific case with thresholds on the T map, generated by a tumor growth model, indicated as segmentation boundaries. The ground-truth segmentation S' is indicated by a red overlay. Middle: corresponding Precision-Recall curve with the same thresholds indicated. The sample with a corresponding volume is marked on the PR curve. Right: quantification of agreement by $R(T(x))$ outside S' and $P(T(x)))$ for voxels inside S'.

2.2 Example Growth Model

To illustrate the the proposed framework for evaluating tumor growth predictions, a traditional diffusion-proliferation model was used with anisotropic diffusion, informed by diffusion tensor imaging (DTI). This model is intended to illustrate the use of the evaluation framework, but it is not our aim to present a novel or improved growth model. The model is defined by a partial differential equation for the cell density c, which changes with each timestep dt according to:

$$\frac{dc}{dt} = \nabla(\mathbb{D}\nabla c) + \rho c(1-c), \tag{4}$$

$$\mathbb{D}\nabla c \cdot n_{\delta\Omega} = 0, \tag{5}$$

where ρ is the growth factor, $n_{\delta\Omega}$ is the normal vector at the boundary between the brain and CSF, and $\mathbb{D}$ is a tensor comprising an isotropic and anisotropic component:

$$\mathbb{D} = \kappa(x)\mathbb{I} + \tau F(x)\mathbb{T}(x), \tag{6}$$

where κ and τ are parameters to weigh the two components, $\mathbb{I}$ is the identity matrix, $F(x)$ is the local Fractional Anisotropy (FA) and $\mathbb{T}$ is the normalized diffusion tensor [11].

The isotropic diffusion depends on the local tissue type [14], as defined by a separate parameter κ_w and κ_g for voxels in the white matter ($\mathcal{W}$) and grey matter ($\mathcal{G}$) respectively:

$$\kappa(x) = \begin{cases} \kappa_w & x \in \mathcal{W} \\ \kappa_g & x \in \mathcal{G} \end{cases}$$

To go from a prediction of $c(t, x)$ to a time-to-invasion ranking $T(x)$, a threshold c_v is applied at each iteration such that $T(x) = \min_t c(t, x) > c_v$, where the visibility threshold is set as $c_v = 0.5$. The initial condition of the model is provided by an initial cell density $c(t = 0)$, which can be defined in two ways: 1) as a gaussian distribution centered at a location x_s and a standard deviation of 1mm; 2) based on a segmentation by setting the cell density at $c = c_v$ for voxels inside the segmentation [7].

The model was implemented in FEniCS [1] in a cubic mesh of 1mm isotropic cells, using a finite element approach and Crank-Nicolson approximations for the time stepping. It has four unknown parameters (ρ, τ, κ_w, κ_g) and, in case of the first approach for setting $c(t = 0)$, an initial location x_s. The method for fitting x_s is explained below.

Fit of Initial Point. A fit of the point x_s is essential for the model initialization from tumor onset, and its location depends on the model parameters. Konukoglu et al. [17] have shown that an eikonal approximation can effectively mimic the evolution of the visible tumor boundary. In this work, we use an eikonal approximation that assumes the visible tumor margin moves at a speed v of $v = 4\sqrt{\rho \mathrm{Tr}(\mathbb{D})}$, in order to estimate x_s for a given set of model parameters, by optimising the approximation of the initial tumor S_0 in terms of the Dice overlap at equal volume using Powell's method [20]. To be more robust to the optimization seed, considering that the optimization landscape may have mutliple local minima, the optimization was repeated for ten runs with different random seeds to increase the chance of finding the global optimum for x_s.

3 Experiments

3.1 Dataset

A retrospective dataset was selected from Erasmus MC of patients who a) were diagnosed with a low-grade glioma; b) were treated with surgical resection, but received no chemo- or radiotherapy; and c) had a DTI and 3D T1-weighted scan before resection, and two follow-up scans (before and after tumor progression). This resulted in data of 21 patients, after one dataset was excluded due to failed registration. Note that the time difference between the measurement of initial tumor and the two follow-up scans varied from a few months to several years.

3.2 Temporal Evaluation Schemes

In the typical timeline of fit and evaluation [14,17], described in Fig. 2 as the bidirectional scheme, the model is fitted on a tumor segmentation S_0 and then simulated from onset, through S_0, to the point of evaluation S_2. In other words, the prediction contains the behavior that it is fitted on.

We compare this method to a strictly forward evaluation scheme that separates the model fit from the prediction as much as possible. As described in Fig. 2 as the forward scheme, the parameters (in this case x_s) are fitted on an initial time-point S_0 and then used to make a prediction between two follow-up scans S_1 and S_2. By running the prediction from a segmentation S_1 instead of an initial location x_s, the potential error in fitting S_0 does not propagate to the evaluation, which is based purely on the growth behavior between S_1 and S_2 that is unknown when fitting the model.

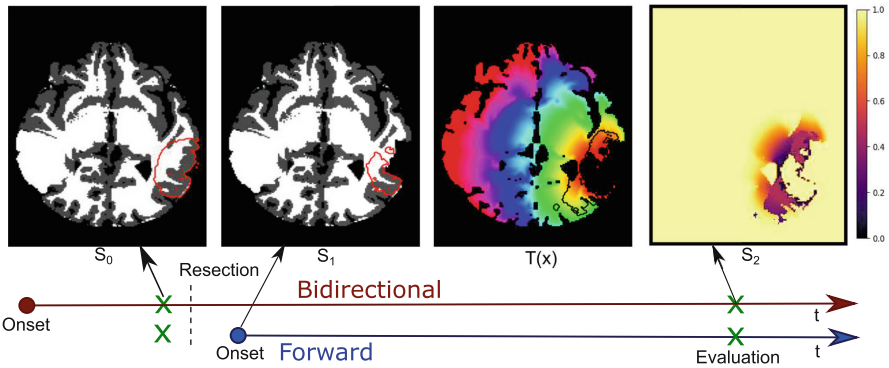

Fig. 2. Overview of two temporal evaluation schemes. Bidirectional: a growth model is fitted to the initial tumor and simulated from a seed point to generate a voxel ranking T. Forward: parameters are fitted to the initial tumor and then the model is initialized with a segmentation S_1 obtained after resection to generate the voxel ranking T. Images from left to right: example of tissue segmentation with S_0 outlined, tissue segmentation with resection cavity removed and S_1 outlined, example of final ranking T used for the evaluation with resection cavity and S_1 removed, quantification of agreement between $T(x)$ and S_2.

For our dataset, we need to consider the role of the tumor resection. In both schemes, the resection cavity as estimated by the aligment of the tissue at S_0 and S_1, is removed from the region of interest for evaluation. In the forward scheme, any voxels in the segmentation S_1 are also removed from the region of interest, leaving only the new growth visible in S_2 for evaluation. So where the bidirectional scheme evaluates predictive performance on the entirety of the remaining tumor, using S_0 only to initialize the location of onset, the forward scheme evaluates purely predictive performance based on the knowledge of S_1.

3.3 Data Preprocessing

Running a growth model from onset requires knowledge of the underlying healthy tissue. Removing pathology from an image is a research problem in itself, but commonly a registration approach with a healthy brain - often an atlas - is used [5,14,18]. In this study we used the contralateral side of the brain as a reference for healthy brain structure (similar to [6]). This is possible because in our dataset all lesions were strictly limited to one hemisphere. Using a registration of the T1-weighted image with its left-right mirrored version, all segmentations were transferred to the contralateral healthy side of the brain. To prevent unrealistic warping of the image due to image intensity changes in the tumor, while still capturing its mass effect, the b-spline registration was regularized with a bending energy penalty [16]. The weight of this penalty with the mutual information metric was tuned on a number of cases using visual inspection of the transformation.

The model input is a segmentation of the brain, separated into white matter ($\mathcal{W}$) and gray matter ($\mathcal{G}$), potentially an estimate of the local diffusion based on Diffusion Tensor Imaging (DTI), and a binary segmentation of the tumor. Segmentations of the brain and brain tissue were produced using HD-BET [13] and FSL FAST [23] respectively. For the pre-operative images, which did not include a T2W-FLAIR sequence, S_0 was segmented manually. Tumor segmentations S_1 and S_2 for consecutive images were produced using HD-GLIO [12,15] and corrected manually where necessary. Alignment with the space of S_0 was achieved with a b-spline registration, which was evaluated visually. Datasets were excluded if the registration did not produce a reasonable aligment.

As no registration or segmentation will be perfect, some inconsistencies remain that prevent a perfect prediction. To not punish the model unfairly, the voxels in S falling outside the brain were disregarded in the computation of the AP metric.

3.4 Parameters

As the variation of diffusive behavior within the brain is a defining factor for the tumor shape, and from a single observation it is impossible to estimate all parameters simultaneously, we kept the proliferation constant at $\rho = 0.01$ while using the parameters κ_w, κ_g and τ as parameters of interest. These parameters were not fitted but rather varied systematically, as listed in the legend of Fig. 3. For this range of seven growth model parameter settings, the AP performance was measured for goodness-of-fit on the baseline segmentation S_0 and predictive performance on S_2, according to the two evaluation schemes. The relation between goodness-of-fit and predictive performance was quantified using a patient-wise Spearman correlation across different growth model parameter settings. The mean of the patient-wise correlation coefficients was tested for a significant difference from zero using a one-sample t-test.

4 Results

Figure 4 shows two examples of the model input and results, in terms of the images used for tumor segmentation at the three timepoints, the segmentations and their mirrored counterparts and the results of a specfic model ($\kappa_w = 0.1$, $\kappa_g = 0.1$ and $\tau = 10$) using both the forward and bidirectional evaluation scheme. The local values of $R(T(x))$ and $P(T(x))$ indicate where the model results are most in disagreement with the ground-truth segmentation S_2.

Figure 3 shows a comparison of the goodness-of-fit, which is measured by the AP on the initial tumor segmentation S_0, and the final predictive performance on S_2.

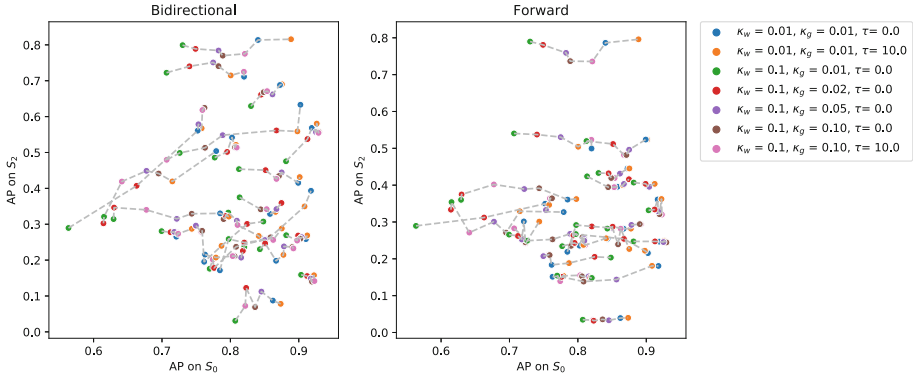

Fig. 3. Comparison of goodness-of-fit versus predictive performance for the two evaluation setups. Results for the same patient on different parameter sets are interconnected.

Comparing the performance between different growth model parameter settings, it is clear that goodness-of-fit is generally higher and more dependent on the model parameters than the predictive performance. From the growth model parameter settings, typically the best goodness-of-fit (AP on S_0) was achieved with low diffusion ($\kappa_w = 0.01$) while the worst fit was achieved when the difference in κ between white and gray matter was large ($\kappa_w = 0.1$, $\kappa_g = 0.01$ or $\kappa_g = 0.02$).

From the results of the bidirectional evaluation scheme, going from an intitial point through S_0 to S_2, it seems that there is a relation between the goodness-of-fit and the predictive performance. However, this relation disappears when using the forward evaluation scheme. These observations are confirmed by the mean patient-wise correlation coefficients, which were 0.24 ($\mathrm{p} = 0.06$) for the forward scheme and $-0.03(\mathrm{p} = 0.76)$ for the bidirectional scheme.

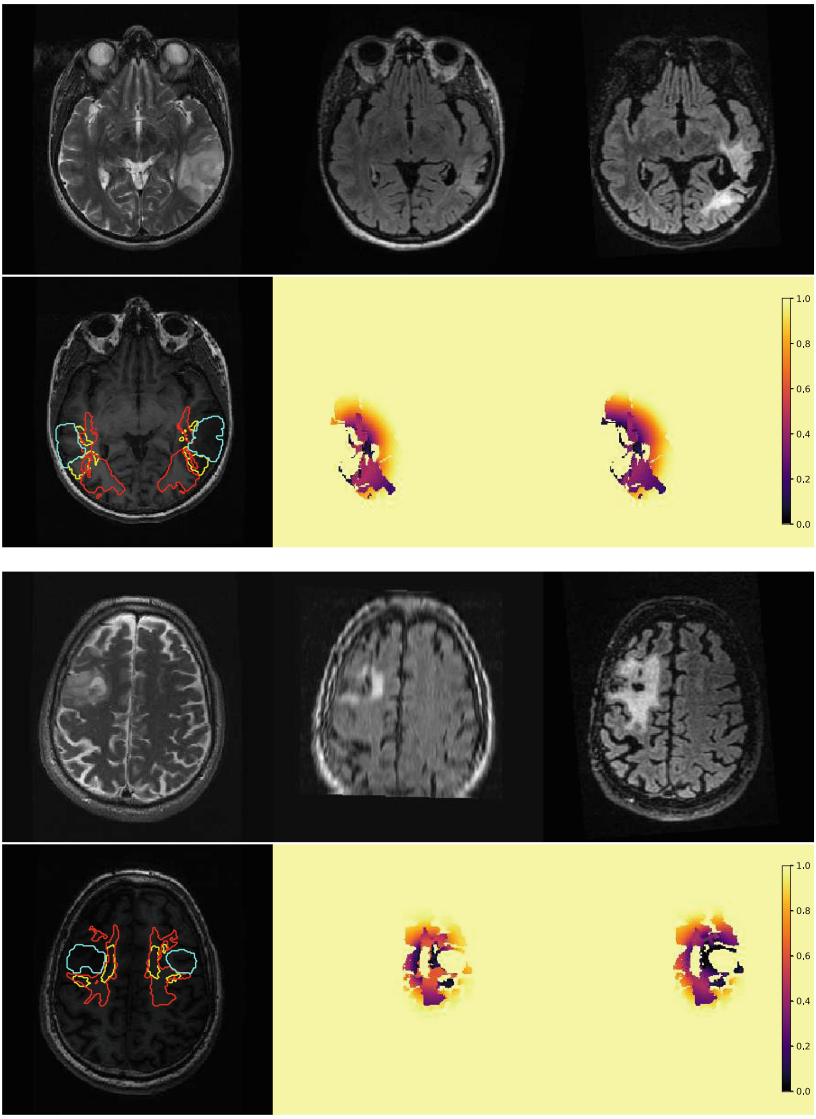

Fig. 4. Example of image processing results for two patients. Top row: T2W imaging showing the initial tumor (left) and T2W FLAIR images showing the tumor after surgery (middle) tumor and at recurrence (right). Bottom row, left: T1W imaging with boundary of resection cavity (cyan), S_1 (yellow) and S_2 (red). Both the original segmentations and the mirrored segmentations are shown. Bottom, middle: Visualization of the local quantification of agreement by $R(T(x))$ outside S_2 and $P(T(x)))$ for voxels inside S_2, for one parameter setting in the forward evaluation scheme. Bottom, right: same visualization for the bidirectional evaluation scheme, same parameter setting. (Color figure online)

5 Discussion

This work presents a formulation of the tumor growth predictions as a forward ranking problem, and describes the Average Precision metric for its evaluation. By formulating the problem in this way we can evaluate the full spatiotemporal results, even if the observations are only snapshots in the form of a segmentation. A further advantage is found in the direct link to local growth speed and quantification of the local model agreement. Though these advantages are only of a qualitative nature, and do not provide a direct benefit to the model itself, we believe it to be a useful step in the development and specifically evaluation of growth models. An important underlying assumption in this framework is that the time axis is not quantified, so the prediction does not provide information on the overall speed of growth or any potential mass effect. Predicting these factors is a highly relevant problem as well, but to predict both spatial distribution, mass effect and speed of growth would likely require at least multiple time-points for model fitting or additional clinical parameters. This is currently not feasible with the data available in clinical practice. For a model that does provide information on growth speed and mass effect, the AP metric could be combined with other metrics to separately evaluate the different factors of tumor growth.

The importance of problem formulation is further illustrated with the two temporal evaluation schemes. Specifically for personalized tumor growth models, which are fitted to an initial tumor shape, this work presents an alternative forward scheme that separates the goodness-of-fit from the evaluation of future predictions. In the forward scheme, the model is initiated with a segmentation instead of an initial point of onset, so that errors made in fitting the initial tumor do not propagate to the final prediction. The aim of this scheme is to evaluate the predictive value of the model and its parameters separately from the goodness-of-fit at the initial observation.

By comparing the bidirectional and forward evaluation schemes in a dataset of 21 patients, using a biophysical growth model, we show that the choice of evaluation greatly affects the relative performance of models. This is illustrated with different parameter setting of the same model, not with different models, but with the purpose of showing the difficulty of evaluating true predictive performance in general. In this case, for our specific model and parameter settings, the difference in performance between parameter settings can be attributed to a better fit of the initial situation, and not necessarily a prediction of unseen behavior. We must note, however, that often the goal in tumor growth modelling is to find the model that best fits the available data on a fundamental level, both initially and in the future, and overfitting is not an immediate concern with strongly model-driven research.

The dataset used in this research was a selection of patients that underwent surgical resection, but no radio- or chemotherapy. Although it is fair to assume that the diffusive behavior of the tumor is not affected during the surgery, so the model parameters would stay the same, the future growth pattern can be affected by the removal of tumor tissue. The decompression that occurs at resection also

complicates the registration of post-operative imaging, which led to the exclusion of one patient due to a failed registration. However, with surgical resection being the recommended treatment for most glioma patients, this is a complicating factor that is difficult to avoid in clinical datasets and in any application in clinical practice.

As new methods of tumor growth prediction are developed, and even fully data-driven models are emerging using machine learning, comparing model performance becomes increasingly relevant. For that purpose, the framing of the problem is essential. Between the actual mechanisms of tumor growth and the segmentation is a flawed observation on MR imaging, the rather difficult problem of segmentation and registration and an estimate of the time horizon. Those factors, combined with limited data and the fact that glioma are naturally unpredictable are a major reason why tumor growth models have relied heavily on simulations [9] and qualitative observations [2] for their validation. This work is a step towards the comparison and clinical evaluation of tumor growth predictions that fits their spatiotemporal nature, and allows for localized interpretation.

Acknowledgements. This work was supported by the Dutch Cancer Society (project number 11026, GLASS-NL) and the Dutch Organization for Scientific Research (NWO).

References

1. Alnaes, M.S., et al.: The FEniCS Project Version 1.5 3(100), 9–23 (2015)
2. Angeli, S., Emblem, K.E., Due-Tonnessen, P., Stylianopoulos, T.: Towards patient-specific modeling of brain tumor growth and formation of secondary nodes guided by DTI-MRI. NeuroImage Clin. **20**, 664–673 (2018)
3. Angelini, E., Clatz, O., Mandonnet, E., Konukoglu, E., Capelle, L., Duffau, H.: Glioma dynamics and computational models: a review of segmentation, registration, and in silico growth algorithms and their clinical applications. Curr. Med. Imaging Rev. **3**(4), 262–276 (2007)
4. Bakas, S., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv 124 (2018)
5. Bakas, S., et al.: GLISTRboost: combining multimodal MRI segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Handels, H. (eds.) BrainLes 2015. LNCS, vol. 9556, pp. 144–155. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30858-6_13
6. Clatz, O., et al.: Realistic simulation of the 3-D growth of brain tumors in MR images coupling diffusion with biomechanical deformation. IEEE Trans. Med. Imaging **24**(10), 1334–1346 (2005)
7. Elazab, A., et al.: Post-surgery glioma growth modeling from magnetic resonance images for patients with treatment. Sci. Rep. **7**(1), 1–13 (2017)
8. Elazab, A., et al.: GP-GAN: brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR Images. Neural Netw. **132**, 321–332 (2020)

9. Ezhov, I., et al.: Neural parameters estimation for brain tumor growth modeling. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 787–795. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_87
10. Gaw, N., et al.: Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric MRI. Sci. Rep. **9**(1), 1–9 (2019)
11. Gholami, A., Mang, A., Biros, G.: Mathematical Biology An inverse problem formulation for parameter estimation of a reaction-diffusion model of low grade gliomas. J. Math. Biol. **72**, 409–433 (2016)
12. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**, 203–211 (2020)
13. Isensee, F., et al.: Automated brain extraction of multisequence MRI using artificial neural networks. Hum. Brain Mapp. **40**(17), 4952–4964 (2019)
14. Jacobs, J., et al.: Improved model prediction of glioma growth utilizing tissue-specific boundary effects. Math. Biosci. **312**, 59–66 (2019)
15. Kickingereder, P., et al.: Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. Lancet Oncol. **20**(5), 728–740 (2019)
16. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging **29**(1), 196–205 (2010)
17. Konukoglu, E., et al.: Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic eikonal equations. IEEE Trans. Med. Imaging **29**(1), 77–95 (2010)
18. Lipkova, J., et al.: Personalized radiotherapy design for glioblastoma: integrating mathematical tumor models, multimodal scans, and Bayesian inference. IEEE Trans. Med. Imaging **38**(8), 1875–1884 (2019)
19. Petersen, J., et al.: Deep probabilistic modeling of glioma growth. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 806–814. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_89
20. Powell, M.J.D.: The BOBYQA algorithm for bound constrained optimization without derivatives. Technical report (2009)
21. Raman, F., Scribner, E., Saut, O., Wenger, C., Colin, T., Fathallah-Shaykh, H.M.: Computational Trials: unraveling motility phenotypes, progression patterns, and treatment options for glioblastoma multiforme. PLoS ONE **11**(1), e0146617 (2016)
22. Silbergeld, D.L., Chicoine, M.R.: Isolation and characterization of human malignant glioma cells from histologically normal brain. J. Neurosurg. **86**(3), 525–531 (1997)
23. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging **20**(1), 45–57 (2001)