A Learnable Prior Improves Inverse Tumor Growth Modeling

Jonas Weidner, Ivan Ezhov, Michal Balcerak, Marie-Christin Metz, Sergey Litvinov, Sebastian Kaltenbach, Leonhard Feiner, Laurin Lux, Florian Kofler, Jana Lipkova, Jonas Latz, Daniel Rueckert, Bjoern Menze, Benedikt Wiestler

November 7, 2024

Abstract

Biophysical modeling, particularly involving partial differential equations (PDEs), offers significant potential for tailoring disease treatment protocols to individual patients. However, the inverse problem-solving aspect of these models presents a substantial challenge, either due to the high computational requirements of model-based approaches or the limited robustness of deep learning (DL) methods. We propose a novel framework that leverages the unique strengths of both approaches in a synergistic manner. Our method incorporates a DL ensemble for initial parameter estimation, facilitating efficient downstream evolutionary sampling initialized with this DL-based prior. We showcase the effectiveness of integrating a rapid deep-learning algorithm with a high-precision evolution strategy in estimating brain tumor cell concentrations from magnetic resonance images. The DL-Prior plays a pivotal role, significantly constraining the effective sampling-parameter space. This reduction results in a fivefold convergence acceleration and a Dice-score of 95%.

J. Weidner, I. Ezhov, L. Feiner und L. Lux are with the Dept. of Computer Science and the Center for Translational Cancer Research, TranslaTUM, Technical University of Munich, Germany (e-mail: j.weidner@tum.de, ivan.ezhov@tum.de, leo.feiner@tum.de, laurin.lux@tum.de).
B. Wiestler and M. Metz are with the Dept. of Neuroradiology, School of Medicine and Health, Technical University of Munich, Germany (e-mail: b.wiestler@tum.de, marie.metz@tum.de)
S. Litvinov and S. Kaltenbach are with the School of Engineering and Applied Sciences, Harvard University, MA, USA (e-mail: slitvinov@seas.harvard.edu, skaltenbach@seas.harvard.edu)
Michal Balcerak and Björn Menze are with the Dept. of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland (e-mail: michal.balcerak@uzh.ch, bjoern.menze@uzh.ch)
F. Kofler is with Helmholtz AI, Helmholtz Zentrum Munich, Germany and with the Center for Translational Cancer Research, TranslaTUM, Technical University of Munich (e-mail: florian.kofler@helmholtz-munich.de)
J. Lipkova is with the Department of Pathology, University of California Irvine, CA, USA (e-mail: jlipkova@hs.uci.edu)
J. Latz is with the Department of Mathematics, University of Manchester, UK (e-mail: jonas.latz@manchester.ac.uk)
D. Rueckert is with the Department of Computing, Biomedical Image Analysis Group, Imperial College London, UK and with the Klinikum rechts der Isar, Technical University Munich, Germany (e-mail: daniel.rueckert@tum.de)
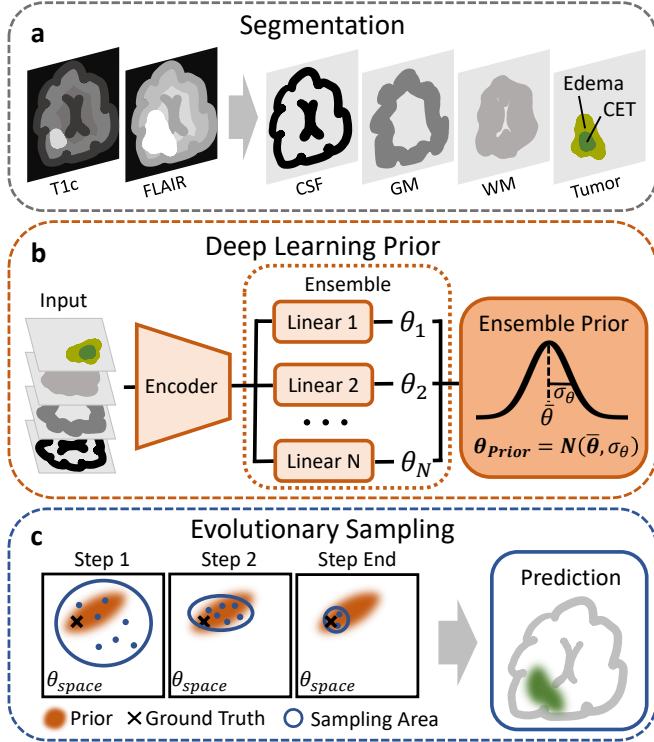Bjoern Menze and Benedikt Wiestler share senior authorship.

Figure 1: Fitting a tumor model with a combination of DL and evolution strategy. **(a)** T1c and FLAIR MR images are segmented into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). The respective tumor segmentations are combined into one channel. **(b)** DL provides a fast estimation of the tumor model parameters $\theta = \{x, y, z, \mu_D, \mu_\rho\}$. A Gaussian prior is constructed based on an encoder followed by an ensemble of linear layers. **(c)** CMA-ES incorporates this DL-Prior and a Dice-score-based likelihood to further optimize the tumor model parameters.

1 Introduction

Glioblastoma is the most common primary malignant brain tumor with a median survival time of only 14-17 months [1] and a yearly incidence rate of 3.53 per 100,000 [2]. Despite recent advances in understanding the biology of this disease and prolonged survival with modern combinations of radio- and chemotherapy [3], there is still a tremendous need for improving the treatment of these patients. A central challenge is the tendency of glioblastoma cells to diffusely spread from the primary tumor into the surrounding brain.

Radiotherapy represents a central pillar in brain tumor treatment. The goal of radiation therapy for brain tumors is striking a delicate balance between delivering an effective dose of radiation to the tumor and minimizing damage to surrounding healthy brain tissue. Current glioblastoma radiation therapy planning predominantly relies on MR and CT imaging, with the clinical target volume defined as a uniform 15 mm margin around the resection cavity plus T1 contrast-enhancing tumor [4]. While this approach aims to address the diffuse spread, it lacks true patient individualization due to its uniformity. Imaging-based individualization of radiotherapy promises to meet the need for more effective therapy.

Radiotherapy planning is limited to the visible boundaries of the tumor on imaging modalities, which is known to only show the "tip of the iceberg" [4]. Physical modeling has been explored to predict tumor cell concentrations based on growth patterns to overcome this limitation and uncover the full extent of the tumor. These models aim to incorporate tumor pathophysiology through the spatial and temporal dynamics of the tumor. In the literature, various models are used, ranging from agent-based cellular automaton over multi-scale models to grid-based continuous models, which are applied in this paper [5–9].

Traditionally, various sampling-based approaches are used to solve the inverse tumor problem, i.e., to estimate the growth model parameters best describing the observed tumor. Simply put, these methods run many forward tumor growth simulations [10], called samples, in an optimized way to converge toward the most likely set of parameters or towards an entire posterior probability distribution of the parameters [11]. For instance, Bayesian inference executed by Markov-chain Monte Carlo can lead to high-quality model fits [12]. With typical forward solvers using the Lattice Boltzmann methods and multi-scope adapted grid methods, a runtime of about 1-3 min per forward run can be achieved for simple tumor models [11, 12]. However, multiple hundred up to a few thousand samples must be simulated for a reasonable inference result. Thus, although sampling strategies are very precise and, in principle, attractive for clinical use given their convergence robustness, they have a strongly growing complexity with the number of parameters relevant to advanced models. Therefore, these methods are prohibitively costly to calculate, running multiple days per patient [12], forming a significant hurdle for their routine clinical use.

Recently, neural networks have been employed to estimate parameters for brain tumor models [13–16] or tumor growth volumes directly [17]. Convolutional neural networks (CNNs) are designed to analyze images and identify patterns, making them well-suited for analyzing medical images and estimating growth model parameters. Applying CNNs for brain tumor modeling involves

2

training the network to recognize specific features within the MR images associated with tumor growth and progression. After training, these networks can solve the inverse problem of predicting tumor model parameters from an imaging observation to drive a deterministic physics simulation. Despite the fast inference time of less than 5 minutes, the main drawback of a pure DL approach lies in its inflexibility towards modal adaptations and the lack of error guarantees, which can lead to incorrect parameter estimation [18]. This lack of robustness represents a significant challenge for their clinical translation.

Recently, a data-driven model regularized by soft physics-informed penalties demonstrated the ability to predict tumor recurrence better than the current standard procedure [19]. The lack of hard physical constraints within the model presents a limitation, as it may lead to biophysically implausible local minima, which could be addressed by incorporating a physically constrained prior.

The successful integration of deep learning with traditional methods for diverse inverse problems across various fields [20–22] motivated us to explore this approach for brain tumors.

Contributions

As outlined above, both approaches to solving the inverse problem have individual drawbacks, namely computational cost and lacking robustness. Here, we incorporate the deep learning (DL) approach as a prior for downstream evolutionary sampling. Our methodology combines the speed of DL inference with the reliability and precision of sampling strategies. In more detail,

- We introduce an ensemble of CNNs to predict tumor growth model parameters in patient space, reducing the mean squared error to simulation ground truth by 80% compared to existing methods.

- We reduce the effective parameter space for subsequent precise evolutionary sampling, based on the DL prediction, effectively accelerating convergence time by a factor of five without sacrificing precision.

2 Methods

2.1 Framework Overview

Our goal is to predict the tumor concentration for each voxel, which is not visible in conventional imaging modalities, based on physically constrained modeling (Figure 1). As input to our model, contrast-enhanced T1 (T1c)

and FLAIR MRI scans are used, highlighting higher and lower tumor concentrations, respectively (Figure 1 a). In a preprocessing step, the tumor regions are segmented in both scans into the subregions contrast-enhancing tumor (CET) and edema, using BraTS Toolkit [23]. The tissue segmentation is generated based on the T1c scan using the S3 software [12].

The tumor and tissue segmentations are fed into a neural network to predict the physical tumor model parameters (Figure 1 b). The premise is that the network learns the growth dynamics based on the underlying tissue. Therefore, an overlay of both tumor segmentations is fed into the network as a single channel(edema: 0.33, CET:0.66), whereas the other input channels contain the patient-specific tissue information: white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) to facilitate patient-specific modeling. The tissue information in each channel is continuous and can be interpreted as the ratio of the given tissue type per voxel. The network output consists of scalar parameters that determine the physical process of tumor development. From those parameters, it is possible to simulate the growth process of the tumor within the patient's brain.

We discuss the network architecture in section 2.5. The DL-Prior is included in subsequent evolutionary sampling, which is explained in section 2.6. Evolutionary sampling is a computational technique that leverages multiple forward simulations to converge toward the optimal parameters governing tumor behavior. The DL-Prior plays a crucial role in guiding the exploration of this parameter space.

2.2 Tumor Growth Model

The tumor progression can be distinguished into two main parts: growth and diffusion. From those two phenomena, the tumor concentration c can be modeled by the following partial differential equation:

$$\frac{\partial c}{\partial t} = \nabla \cdot (D \nabla c) + \rho c (1 - c) \tag{1}$$

The first term describes the diffusion process with diffusion coefficient D, and the second term describes the growth process over time t. In the case of the applied Fisher-Kolmogorov equation [24], a logistic growth process with a growth parameter ρ is assumed.

2.3 Tumor Simulation

For the simulation, three types of brain tissue are distinguished: WM, GM and CSF. For each voxel i the probability p of all tissues adds up to one: $p_{WM} + p_{GM} + p_{CSF} = 1$.

3

The diffusion coefficient is assumed to differ for each tissue type [12]. The ratio of diffusion within WM and GM is fixed to $D_{GM} = \frac{1}{10}D_{WM}$, whereas no diffusion is allowed inside CSF ($D_{CSF} = 0$), as the tumor physiologically cannot grow there.

Additional parameters are the tumor origin $\{x, y, z\}$ and the total time of tumor growth T. We select a single tumor origin as the starting point of the simulation. This simulated origin cannot necessarily be directly associated with the actual tumor-originating cell in medical cases. Thus, the original set of parameters for the forward solver is $\theta_{orig} = \{x, y, z, \rho, D_{WM}, T\}$. As only one time-point is realistically available for radiotherapy planning, θ_{orig} is ambiguous: for example, a faster-growing tumor over a shorter time and a slower-growing tumor over a longer time are indistinguishable based on a single time point. Therefore we introduce end-time independent parameters $\mu_D = \sqrt{D_{WM}T}$ [cm] and $\mu_\rho = \sqrt{\rho T}$ [unitless] leading to the reduced parameter set: $\theta = \{x, y, z, \mu_D, \mu_\rho\}$. The 3D tumor cell density c is a function of these parameters. For simulation, the GliomaSolver [12] software is used.

2.4 Data

For the network training, 30,000 simulations are created using 140 real patient tissues from the BraTS dataset [25] to simulate synthetic tumors. By explicitly incorporating tissues from BraTS patients, whose anatomy is already altered by tumors, we introduce pathological anatomy variations that the network is likely to encounter in real patient data. This approach aims to make the model invariant to tumor-induced anatomical differences and typical tissue deformations, thereby enhancing its ability to generalize to diverse patient data. The dataset is split into 28000 training samples and 2000 validation samples.

For the additional test set for extensive validation of all methods, including costly additional sampling, 180 simulations are conducted in another 140 BraTS patient anatomies with random parameters.

For validation, we apply all methods to preoperative imaging data from nine real glioblastoma patients to compare their ability to fit preoperative tumor segmentations instead of the comparison to synthetic ground truth.

2.5 Deep Learning Network

2.5.1 Network Training

The network encoder (Figure 1 b) is based on a residual neural network (ResNet) [26]. In contrast to former experiments [14], the network input consists of four channels with dimensions $128 \times 128 \times 128$, concatenating the three patient-specific tissue masks and the tumor segmentation map. All other parameters and conditions are maintained as in the prior experiments [14]. The network outputs are the five model parameters θ optimizing an L_2 loss. The simulation parameters are sampled based on clinical experience in the same way as introduced in the Learn-Morph-Infer (LMI) method [14].

For the original model, the input tumor is generated by random thresholding of the synthetic ground truth tumor at values $c_{Edema}^{th} \in [0.05, 0.5]$ and $c_{CET}^{th} \in [0.5, 0.85]$. All voxels containing ground truth tumor concentrations above these thresholds are considered edema ($c_{Edema}^{th} \leq c < c_{CET}^{th}$) or CET ($c \geq c_{CET}^{th}$) segmentations for training.

The training was conducted on an NVIDIA Quadro RTX 8000 with 50GB RAM. We trained for 85 hours over 50 epochs.

2.5.2 Network Ensemble

An ensemble of ten networks is introduced to increase the model robustness towards multiple edema and CET thresholds. The last layer of the network ensemble is fine-tuned for a certain threshold ($c_{Edema}^{th}, c_{CET}^{th}$) in the ground truth training data. The premise is that the ensemble mean is more stable toward single network failure, and the ensemble standard deviation can quantify the uncertainty of the prediction. Technically, we fixed all weights except the last linear layer and retrained it on different thresholds. A Gaussian prior is generated from this equally weighted ensemble (see section 2.6.1).

2.6 Evolutionary Sampling

Introducing a prior drastically reduces the effective parameter space for the subsequent sampling strategy. The following sections describe the integration of this prior into a covariance matrix adaptation evolutionary strategy (CMA-ES) [27, 28].

2.6.1 Prior

A Gaussian distribution around the DL-ensemble estimated parameters is defined as prior. The prior distribution is based on the ensemble mean ($\bar{\theta}_{dl}^i$) and its standard deviation (σ_{dl}^i) for each parameter $i \in \theta$. It is assumed that the individual parameters follow independent Gaussian distributions.

$$P(\theta|\bar{\theta}_{dl}, \sigma_{dl}) \sim \prod_i \frac{1}{\sigma_{dl}^i \sqrt{2\pi}} \exp\left(-\frac{(\theta^i - \bar{\theta}_{dl}^i)^2}{2(k\sigma_{dl}^i)^2} \right) \quad (2)$$

The scaling parameter k is introduced to relax or tighten the variance of the prior based on empirical findings. It is found that the model sampling performs better if we relax the prior contribution compared to the likelihood. This is done by increasing k to a value of 10. In a strict Bayesian formulation the prior does not depend on the data. As patient data is used for DL inference, the prior does depend on the data, so technically, we apply empirical Bayes [29].

2.6.2 Likelihood

The likelihood describes the probability of observing the MRI tumor segmentations (data $\mathcal{D} = \{Y_{\text{CET}}, Y_{\text{Edema}}\}$) given a set of tumor model parameters θ. It is composed of the likelihoods of edema and CET where $c^{th} = \{c^{th}_{CET}, c^{th}_{Edema}\}$.

$$P(\mathcal{D}|\theta, c^{th}) = P(Y_{\text{CET}}|\theta, c^{th}_{CET}) \cdot P(Y_{\text{Edema}}|\theta, c^{th}_{Edema}) \tag{3}$$

The respective probability of observing the segmentations with a given set of parameters is assumed to be described by the Dice-score (DSC)[1] [30] between the segmentation and the proposed concentration above the threshold c^{th}_{Edema}, and equivalently for T1c c^{th}_{CET}.

$$P(Y_{\text{Edema}}|\theta, c^{th}_{Edema}) = \text{DSC}(Y_{\text{Edema}}, c(\theta) > c^{th}_{Edema}) \tag{4}$$

2.6.3 Posterior

The posterior is calculated with the DSC-based likelihood and the DL-based prior.

$$P(\theta|\mathcal{D}) \sim P(\mathcal{D}|\theta, c^{th}) \cdot P(\theta|\bar{\theta}_{dl}, \sigma_{dl}) \tag{5}$$

The maximum a posteriori probability (MAP) estimate provides the most likely tumor model parameters $\hat{\theta}_{\text{MAP}}$ under the described assumptions.

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} P(\theta, c^{th}|\mathcal{D}) \tag{6}$$

$$= \arg\max_{\theta} \int_{c^{th}} P(\theta, c^{th}|\mathcal{D}) dc^{th} \tag{7}$$

$$\overset{(5)}{=} \arg\max_{\theta} \int_{c^{th}} P(\mathcal{D}|\theta, c^{th}) dc^{th} \cdot P(\theta|\bar{\theta}_{dl}, \sigma_{dl}) \tag{8}$$

$$\overset{*}{\approx} \arg\max_{\theta} P(\mathcal{D}|\theta, \overline{c^{th}}) \cdot P(\theta|\bar{\theta}_{dl}, \sigma_{dl}) \tag{9}$$

---

[1]Approximating the likelihood function with the Dice-score is mathematically inaccurate. For example, a vanishing Dice-score is not equivalent to a zero probability of observing the data. Nevertheless, the real likelihood function is unknown, and modeling it would imply further assumptions [12] that we want to avoid.

We applied the midpoint approximation (*) over the likelihood for the average thresholds $\overline{c^{th}} = \{0.25, 0.675\}$, as the threshold values are sampled over a uniform distribution, and the difference in likelihood is small on a relative scale (about max 5%) for various thresholds. As there is no analytical solution for $\hat{\theta}_{\text{MAP}}$, the optimization is carried out by CMA-ES as described in the following.

2.6.4 Sampling Algorithm

The novelty of our approach lies in integrating a DL-Prior into sampling strategies. This is realized by the state-of-the-art covariance matrix adaptation evolution strategy algorithm [27] (Figure 1 c).

The CMA-ES algorithm is employed due to the inherently ill-posed, non-convex, and non-differentiable nature of the inverse tumor growth problem [31]. Furthermore, CMA-ES exhibits robust performance on unsteady problems frequently encountered at tissue interfaces. For instance, a tumor might not grow at all if its origin lies in a voxel where CSF is dominant, whereas it may proliferate unimpeded in an adjacent white matter voxel.

The idea of the CMA-ES is based on sampling within a multivariate Gaussian distribution in parameter space. In each step, samples are evaluated according to their fitness. In our case, the fitness function is represented by the estimated posterior $P(\theta|\mathcal{D})$, describing the probability of the parameter set θ given the measured data $\mathcal{D}$ (Eq. 5). The fittest samples are selected, and the sampling mean and the covariance matrix of the multivariate Gaussian are adapted accordingly [32]. Eventually, the sampling distribution converges to the MAP $\hat{\theta}_{\text{MAP}}$ (Eq. 9). We sampled for a maximum of 600 simulations.

To ensure an equitable comparison of runtime between sampling methods with and without a prior, we implement a specific initialization protocol for the absence of a prior, the naive sampling. The tumor origin is initialized at the center of mass of the FLAIR segmentation. The diffusion coefficient μ_D and μ_{rho} are initialized at its lower limit, given that the forward simulation runtime increases linearly with μ_D. This initialization results in a sampling strategy that converges towards optimal parameters while commencing with a small tumor and subsequently increasing in size. As smaller tumors facilitate expedited simulation time, this initialization improves the entire sampling runtime. For this comparative analysis, a subset of 13 patients was randomly selected for cost reasons. We sampled until a convergence up to 99 % of the maximum posterior up to a total of 2000 samples for the case without prior and 600 samples for the case with DL-Prior. All simulations were conducted on an AMD EPYC 7313 16-core CPU. We published our

code on GitHub[2].

2.7 Evaluation

The predicted parameters can be compared to the ground truth as an intermediate evaluation. Here, the distance to the ground truth origin and the deviation from the ground truth growth and diffusion parameters are of interest. Note, however, that due to the ill-posedness of the problem, several combinations of parameters can describe a given tumor observation equally well. In view of the potential clinical application for individualized radiotherapy, the accurate estimation of tumor cell concentration is more important.

Thus, different metrics are applied to evaluate the similarity between predictions and ground truth concentration. The mean absolute error (MAE) and mean squared error (MSE) provide information about the pixel-wise differences between the proposed tumor and the ground truth. The problem with those metrics lies in the averaging over all pixels. For example, a tumor where every voxel concentration differs a bit can have the same error as a tumor where some areas are utterly distinct while some regions fit very well. Therefore, the commonly used Dice-score [30] is also applied. The Dice-score measures the overlap of two binary segmentations. A tumor density threshold must be applied as the solver's output is continuous.

For synthetic data, where ground truth tumor concentration is available, any clinically relevant threshold can be selected to binarize the ground truth and the proposed tumor density map. As a benchmark, the atlas-specific network LMI is integrated into our synthetic workflow [14].

To ensure statistical robustness, we have highlighted the significant results by marking them with an asterisk (*). We set the threshold for significance to a p-value of 1%. If not stated otherwise, we compared the different methods to our proposed method, "DL-Prior + Sampling". As we cannot assume Gaussian distributions and as we analyze paired data, we used the Wilcoxon signed-rank test.

3 Results

In this section, we present the evaluation of various methodologies to predict tumor cell distribution. We compare the following methods:

- **LMI**, which is a DL method requiring atlas registration [14].

- **DL**, which is the in-patient-space DL approach with a single network.

- **DL Ensemble**, which is the mean of an ensemble of networks, finetuned for different Edema and CET thresholds.

- **Naive Sampling**, which is the CMA-ES.

- **DL-Prior + Sampling**, which is the subsequent evolutionary sampling incorporating a DL-Prior based on the DL ensemble.

First, we qualitatively compare the proposed tumor concentrations, highlighted in Figure 2. Subsequently, we delve into a detailed comparison of estimated parameters. This is illustrated in Table 1, where deviations from ground truth parameters are analyzed. The following evaluation of tumor concentrations, as detailed in Figure 4 and Table 2, provides the most relevant metric of different methodologies. Additionally, we compare the runtime efficiency of these methods in Figure 5, evaluating the convergence times of each approach. Finally, we evaluate our method with nine real patients (Figure 6, Table 4).

3.1 Deep Learning Methods

In the first results part, we focus on the DL methods described in Figure 1 b.

3.1.1 Qualitative Results

In Figure 2, a qualitative comparison of the different approaches is visualized. It is visible that the LMI tumor prediction (b) does not align perfectly with the underlying tissue. This discrepancy is caused by the registration process from the atlas space into the patient space. Furthermore, the LMI prediction overestimates the tumor size. Our in-patient space method (c) slightly underestimates the tumor size, whereas the ensemble mean (d) overestimates it.

3.1.2 Estimated Parameter Comparison

The parameter estimation is compared to the simulated ground truth (Table 1). All parameter estimations are Gaussian distributed, as expected[3]. Therefore, only the mean deviation and the RMSD are reported.

---

[2]github.com/jonasw247/a-learnable-prior-improves-inverse-tumor-growth-modeling

[3]The distribution of origin-deviation is not Gaussian itself, but it is composed of three Gaussian distributions (x,y,z).
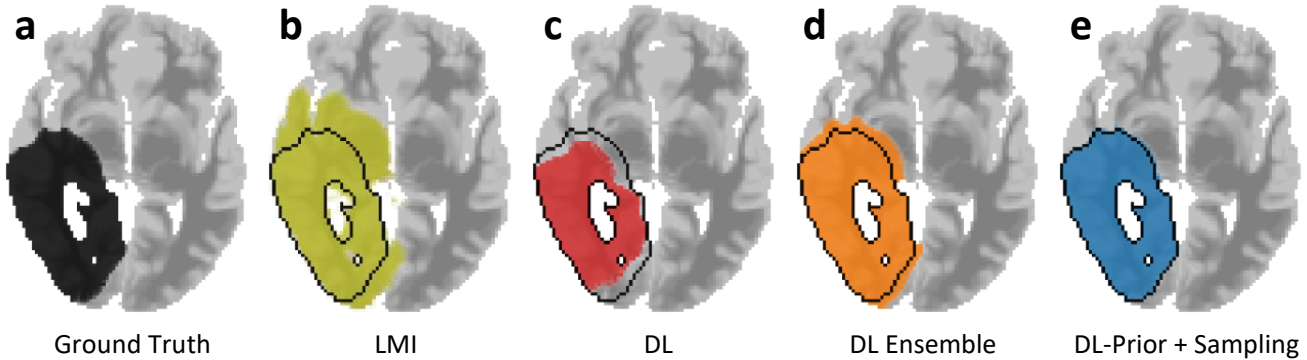
Figure 2: Example tumors created by different models are compared. The ground truth tumor is shown in black (a). Its outline is displayed in the other images as well. The atlas-based LMI approach (b), our DL-Prior (c), the ensemble mean (d), and the combination of the DL-Prior and sampling (e) are compared.
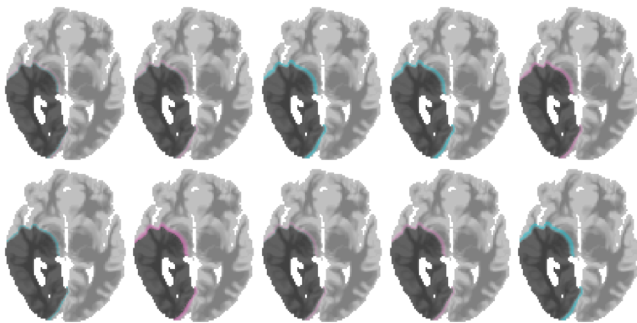

Figure 3: Example simulations with parameters created by the ensemble of the networks. The difference (cyan, positive; pink, negative) to the ensemble mean (gray, Figure 2 d) is shown.

The estimation of the tumor origin is generally precise, with an average deviation of 3.7 mm to the ground truth for the single network (3.6 mm for the ensemble mean). This is twice as good as the LMI approach, with a mean displacement of 8.7 mm. As a baseline, the edema segmentation's center of mass is included as a rough estimation. The CET center of mass mean displacement is 12.7 mm away from the ground truth, demonstrating the superiority of our method.

For μ_D, the time-independent diffusion parameter, a three-fold improvement in root means square deviation (RMSD) is found for DL over LMI. An average overestimation (16.5 mm) can be found for LMI, whereas our approach underestimates μ_D slightly (-2.4 mm). Regarding the unitless, time-independent growth parameter μ_ρ, a two-fold improvement (in RMSD) is found, whereas no systematic deviation is detected.

The DL ensemble (orange), finetuned with different edema and CET thresholds, performs similarly to our original (DL, non-ensemble) model (red) on parameter level.

3.1.3 Estimated Tumor Concentration

Due to the variations in the employed methodologies, exclusive reliance on parameter estimation is inadequate. A comprehensive evaluation of the resulting tumor concentration is imperative, especially considering the intended use for individualized radiotherapy. Therefore, we compare the MAE, the MSE, and the Dice-score for different threshold values.

The errors of the presented methods are shown in Table 2. It is found that our proposed in-patient-space DL ensemble and single networks (DL) perform significantly better than the atlas-registration-based LMI.

Additionally, we compare the Dice-score between thresholded ground truth and predicted tumor cell distributions for a clinically more relevant measure. Figure 4 shows a general decrease in the Dice-score for increasing thresholds independent of the applied method. This decrease is intrinsic to the Dice, which is biased towards larger volumes.

Compared to the atlas-based DL approach (LMI), there is a significant improvement in the Dice-score (Figure 4) for all of the shown methods. The single network already reaches Dice-scores of about 90% in the clinically relevant range of low tumor concentrations. This result is slightly, but not significantly, outperformed by the ensemble mean. Still, the ensemble provides an additional variance utilized by the subsequent sampling approach in the form of a prior.

Table 1: The tumor model parameter mean displacement (MD) and the root-means-squared deviation (RMSD) from the simulation ground truth are shown for different methods. MD is a criterion for the systematic deviation, the shift between the means. RMSD is a measure of the spread of differences in distribution and, thus, determines the quality of the fit. The distance from the edema and CET center of mass to the ground truth tumor origin is included. A significant improvement in origin estimation of DL-Prior + sampling is found over the core center of mass. For μ_D and μ_ρ, subsequent sampling results in worse parameter results. This highlights that the problem is ill-posed, meaning that finding a better fit to the observed tumor data can result in deviations or instability in the estimated parameter values (see also Table 2)

Metric	Parameter	DL-Prior + Sampling	DL	DL Ensemble	LMI	Edema Center of Mass	CET Center of Mass
	Origin [mm]	**3.51 ± 0.19**	3.72 ± 0.20	3.57± 0.19	8.72 ± 0.27 *	15.87 ± 0.73 *	12.66 ± 0.56 *
MD	μ_D [mm]	-0.30± 0.08	**-0.24± 0.06**	-0.25± 0.06	1.66± 0.15 *		
	μ_ρ	**0.01 ± 0.02**	-0.02 ± 0.02	0.05 ± 0.02	0.01 ± 0.06		
	Origin [mm]	4.38 ± 0.24	4.56 ± 0.24	**4.37 ± 0.24**	10.30 ± 0.31 *	18.64 ± 0.88 *	14.9± 0.96 *
RMSD	μ_D [mm]	1.05 ± 0.10	0.84 ± 0.06	**0.82 ± 0.06**	2.56 ± 0.13 *		
	μ_ρ	0.40 ± 0.06	0.33 ± 0.03	**0.33 ± 0.03**	0.74 ± 0.07		

Table 2: The mean absolute (MAE) and mean squared error (MSE) are compared between the predicted tumor concentration and the ground truth for different algorithms. A significant improvement of DL-Prior + sampling over the second-best method is demonstrated.

	MAE [10^{-4}]	MSE [10^{-5}]	Dice 0.5
DL-Prior + Sampling	**16 ± 2**	**30 ± 7**	**0.93 ± 0.01**
DL	24 ± 2 *	68 ± 10 *	0.88 ± 0.01 *
DL Ensemble	27 ± 3 *	72 ± 11 *	0.89 ± 0.01*
LMI	79 ±6 *	401± 40 *	0.75 ± 0.01 *

3.2 Subsequent Evolutionary Sampling

In the second part of the results, we focus on subsequent sampling methods in combination with the DL-Prior, as described in Figure 1 c.

3.2.1 Qualitative Results

The combined method of DL-Prior and CMA-ES sampling (Figure 1 e) closely aligns with the ground truth (Figure 1 a). These qualitative findings are validated with quantitative results in the following sections.

3.2.2 Estimated Parameter Comparison

Similar results to DL and DL ensemble are found for subsequent evolutionary sampling (Table 1). Limited information, introduced by stochastic thresholding of test data, might lead to a performance threshold that is impossible to overcome theoretically. For the combined version of sampling and prior, a slightly worse result is found for μ_D and μ_ρ, whereas the origin deviation is comparable to the ensemble mean and significantly better than the center of mass.

Even though parameter estimation does not improve with the network ensemble and subsequent evolutionary sampling, the rationale for training the ensemble is founded on the anticipation of achieving more stable predictions and providing variance estimation for the DL-Prior to optimally initialize the subsequent sampling. Further, the ill-posedness of the problem allows multiple parameter combinations to result in similar tumors. The following section examines this premise by comparing the resulting tumor concentrations.
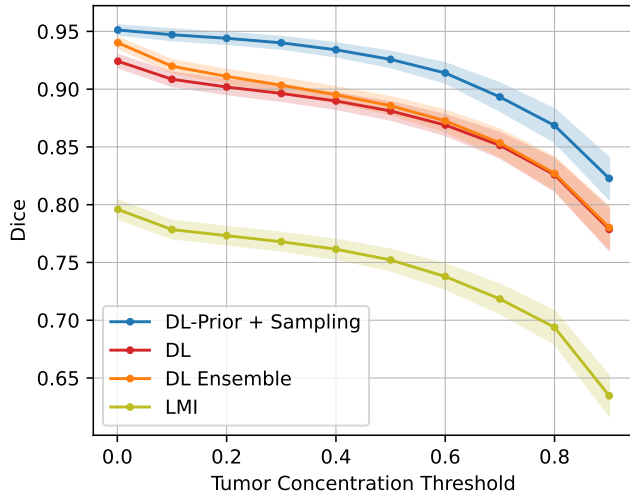

Figure 4: Mean Dice-score results between prediction and ground truth at different threshold concentrations. The single network approach is compared to the ensemble mean, LMI and the final combined version with CMA-ES sampling + DL-Prior.

3.2.3 Estimated Tumor Concentration

Subsequent sampling decreases the error to the ground truth of the estimated tumor concentration as shown in Table 2. This highlights the synergistic combination of sampling with a DL-Prior. Further, the combined version of CMA-ES sampling and DL-Prior achieves a significantly better Dice-score than the DL-Prior alone over the whole range of tumor concentrations (Figure 4). Within the clinically relevant range, up to 40% tumor concentration, a Dice overlap of 95% is found, further demonstrating the superiority of combining DL with traditional sampling.

3.2.4 Runtime comparison

In Figure 5, we compare the convergence times across various methods for a subset of samples. Convergence time for the evolutionary sampling method is defined as the duration required to achieve 99% of the maximum posterior probability. All calculations were performed on a single CPU to maintain consistency in comparison. The convergence time associated with evolutionary sampling using a DL-Prior is not directly comparable to the inference time of DL alone, as the latter is predominantly determined by the duration of a single forward simulation run, while the duration of the DL inference is negligible. When comparing against naive evolutionary sampling, integrating a DL-Prior into our optimization strategy demonstrates an 80% improvement in convergence time.

To guarantee the integrity of our comparative analysis regarding time, we have included performance metrics for the specific subset of samples utilized in the temporal comparison, as detailed in Table 3. Sampling enhanced with a DL-Prior demonstrates improved results in terms of the Dice. However, the observed differences between these methods fall within the range of the method-specific error margins. These observations suggest that both approaches converge towards a similar minimum, legitimizing the time comparison of Figure 5.
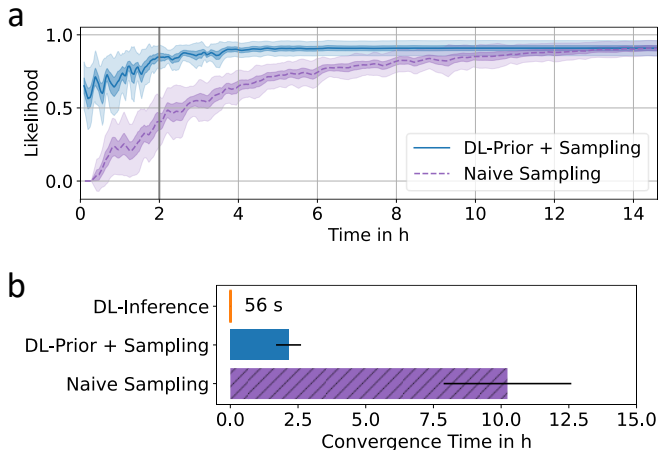

Figure 5: (a) The mean likelihood over time for Naive Sampling and DL-Prior + Sampling is shown. Shaded areas represent standard error over different patients. DL-Prior + Sampling consistently yields higher likelihoods, with a significant difference around the clinically relevant 2-hour mark, resulting in a likelihood of 0.85 ± 0.03, while Naive Sampling only reached 0.41 ± 0.05. This translates to a p-value of 1.2×10^{-5} and an effect size of 2.0. (b) The mean convergence time of our DL-inference (orange) and our combination of classical sampling with DL-Prior (blue) is compared to sampling without prior (purple, dashed). The paired t-test results in an effect size of 1.0 and a p-value of 0.003

Table 3: Tumor concentration comparison between DL-Prior with subsequent sampling and sampling without a prior for a subset of patients. The Dice-score at different thresholds (0.01, 0.1, 0.5) is given, and no substantial difference was found. Simulation time differences are shown in Figure 5.

Metric	DL-Prior + Sampling	Naive Sampling
Dice at 0.01	**0.94** $\pm$ 0.02	0.90 $\pm$ 0.02
Dice at 0.1	**0.94** $\pm$ 0.03	0.92 $\pm$ 0.03
Dice at 0.5	0.93 $\pm$ 0.02	**0.95** $\pm$ 0.03

4   Validation on Patient Data

To demonstrate that simulation advancements can translate into clinical application, we tested our method on nine patients. An example is shown in Figure 6. As already found in Figure 2, the LMI simulation is not limited to white and gray matter regions and, therefore, grows into the CSF region due to registration inaccuracies. DL and DL ensemble already roughly estimate the tumor location, while overestimating the volume, while DL-Prior + sampling and naive sampling estimate the segmentation well. The faster convergence becomes evident, considering the likelihood over time.

The quantitative ability of different methods to fit real tumor segmentations is compared in Table 4. The Dice-score is compared for CET and edema regions. It is clearly visible that sampling improves the simulation fit dramatically. The differences in mean Dice-score between naive sampling and prior included sampling are small, but those results do not allow for statistical statements. Therefore, we qualitatively analyze the results case-specific in Figure 7. Regarding the runtime, a clear advantage is seen for DL-Prior + sampling.

Table 4: For nine real patients, the Dice-score to the tumor segmentation is compared for different algorithms, thresholded at the $\overline{c^{th}} = \{0.25, 0.675\}$. Additionally, the convergence time is reported for each method. It becomes apparent that naive sampling and sampling with a prior reach similar Dice-score with significant time differences in runtime.

	Dice CET	Dice Edema	Runtime
DL-Prior + Sampling	0.68 ± 0.03	0.68 ± 0.02	(4.9 ± 1.2) h
Naive Sampling	0.70 ± 0.03	0.70 ± 0.02	(16.3 ± 1.7) h*
DL	0.39 ± 0.06*	0.44 ± 0.04*	< 1 min*
DL Ensemble	0.44 ± 0.08*	0.47 ± 0.05*	< 1 min*
LMI	0.34 ± 0.09*	0.55 ± 0.04*	< 1 min*

Our method generally demonstrates strong performance, as illustrated in Figure 5, which also shows significant runtime improvements. However, there are occasional instances where it encounters challenges. We analyze various example scans in Figure 7 to assess the types of tumors that our method accurately characterizes and to identify specific scenarios where our approach falls short, highlighting potential for future research. For the small tumor (a), the likelihood plot indicates a strong advantage for including the DL-Prior convergence, which is reached much faster than with naive sampling. The naive sampling approaches the optimal prior region only after exploring a larger space, while the diffusion and growth rate stay close to the initial prior. For the larger tumor (b), the prior estimate of the origin is close to the final
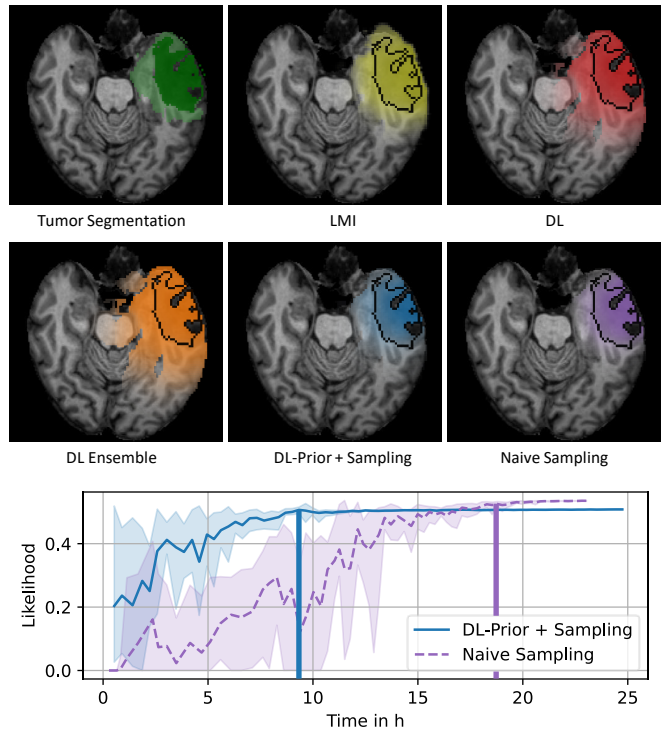

Figure 6: Real patient example with different inversion strategies. The first image shows the ground truth tumor segmentation before surgery. Light green shows the edema region, while dark green represents the CET, which is included in the following images as a black outline. The mean likelihood development over time comparing DL-Prior sampling to naive sampling is plotted. The margin shows the minimum and maximum values for each CMA-ES epoch, and the vertical lines depict the time points of convergence. The likelihood for the other methods is 0.39 for LMI, 0.23 for DL, and 0.13 for DL ensemble.

estimate of the naive process. The prior sampling early advantage diminishes over time, as the DL-Prior restricts the diffusion rate to decrease further, which seems beneficial for this tumor. However, for the example with large mass effect (c), all methods struggle to fit the tumor, which is expected, as the mass effect is not part of the physical tumor model. Due to the insufficient tumor model, the DL-Prior becomes obsolete, and no improvement in likelihood is achieved. Neither method fits the scans well in this case.

In summary, the introduction of a DL-Prior helps the sampling strategy to converge faster to the optimal solution (Figure 5, 7 a). In some cases, the underlying tumor model is clearly oversimplified (7 b), and in a few other

10

cases, the prior can lead the sampling in the wrong direction (7 c). The last problem could be overcome by relaxing the prior constraint over time.

5 Discussion

We developed a novel technique combining DL with CMA-ES sampling to estimate brain tumor density based on magnetic resonance images, synergistically exploiting the unique advantages of each approach.

A new DL-based method is introduced, operating solely in patient image space compared to previous atlas-based methods [14] by explicitly incorporating patient anatomy into the network. With our method, a better parameter estimation is found (Table 1). Additionally, a better Dice overlap to the ground truth concentration is observed (Figure 4). We assume that the abundance of registration errors between the patient and the atlas space causes this improvement.

No significant improvement in parameter estimation is achieved by combining multiple fine-tuned networks into an ensemble, whereas a significant improvement in Dice overlap is found compared to the single network. This can be attributed to the fact that the network ensemble provides higher stability in prediction [33].

The primary strength of the proposed method lies in the combination of DL and CMA-ES sampling. Within the confines of our simulation, we observed that the combined method provides a faster and more accurate representation of tumor dynamics compared to individual approaches (Figure 4).

In synthetic data, the combined optimization time was improved by a factor of five compared to the optimization without prior (Figure 5) while the performance is increased to 95% Dice compared to 80% reached by a previous DL approach [14] (Figure 4).

For real data, the advantage of sampling over pure DL methods becomes apparent. The DL-Prior reduced the convergence time by a factor of three with similar results.

Regarding brain tumor modeling in general, despite being enclosed inside the skull, it is highly debatable to which degree the complex biological system of brain tumors can be modeled using deterministic models, to which degree it results from chaotic behavior, and what role intratumoral heterogeneity and evolution play. Despite this, modeling-based approaches open a unique perspective on the diffuse spread of tumor cells, which might be sufficient for radiotherapy improvement.

Regarding the validity of our findings, the model's limitations are primarily centered around its simplicity, which, while beneficial for understanding the proposed combination of two modeling worlds, may not accurately represent the underlying complex biological systems. The model's exclusion of necrosis and mass effect factors is a significant limitation. Additionally, the isotropic diffusion, particularly the distinction between white and gray matter, is somewhat arbitrary, potentially oversimplifying these complex dynamics. Furthermore, the main findings of the method are confined within the narrow boundaries of these simulations, where tumor behavior is strictly deterministic. Therefore, extended real data validation with advanced tumor models is critical to validate our approach.

Regarding clinical implications, we demonstrate a method to combine general, flexible sampling methods with specialized DL. Changing the DL model, e.g., with respect to higher tumor model complexity, can be time-consuming since the network must be retrained. However, this should not be an issue in clinical use, as the model will not be updated frequently.

6 Outlook

For an impact on radiotherapy planning, more sophisticated tumor growth models are needed to explain the data. The presented combination of DL and classical sampling enables more complex models and models with more parameters to be solved in clinically relevant time frames. Such models are crucial for several well-studied tumor properties like mass effect [34, 35], nutrient flow [36], and necrosis [37]. The mass effect takes the tumor-induced tissue shift into account, whereas nutrient flow and necrosis formation affect the growth dynamic of the tumor based on complex interaction with itself and surrounding tissue.

It has been shown that tumors grow along fiber tracts, explaining the so-called butterfly tumors [38, 39]. Additional imaging modalities like diffusion tensor imaging contain information about fiber tracts that can be incorporated in the tumor model [40–42]. Therefore, the shown framework can be adapted and finetuned in multiple ways. For example, by incorporating multiple simulation time points into the CMA-ES selection process or by relaxing the prior weight over time.

Finally, an evaluation of large-scale clinical data is needed. As the availability of longitudinal data is highly limited due to the rapid surgery after detection for most patients, other validation metrics are considered. Prediction of tumor recurrence represents one of the most promising approaches. Besides the influence of treatment procedures like surgery and radiotherapy, it was shown that a certain degree of recurrence prediction is possible [43–45]. Showing the ability to predict recurrence with a physically constrained model would provide
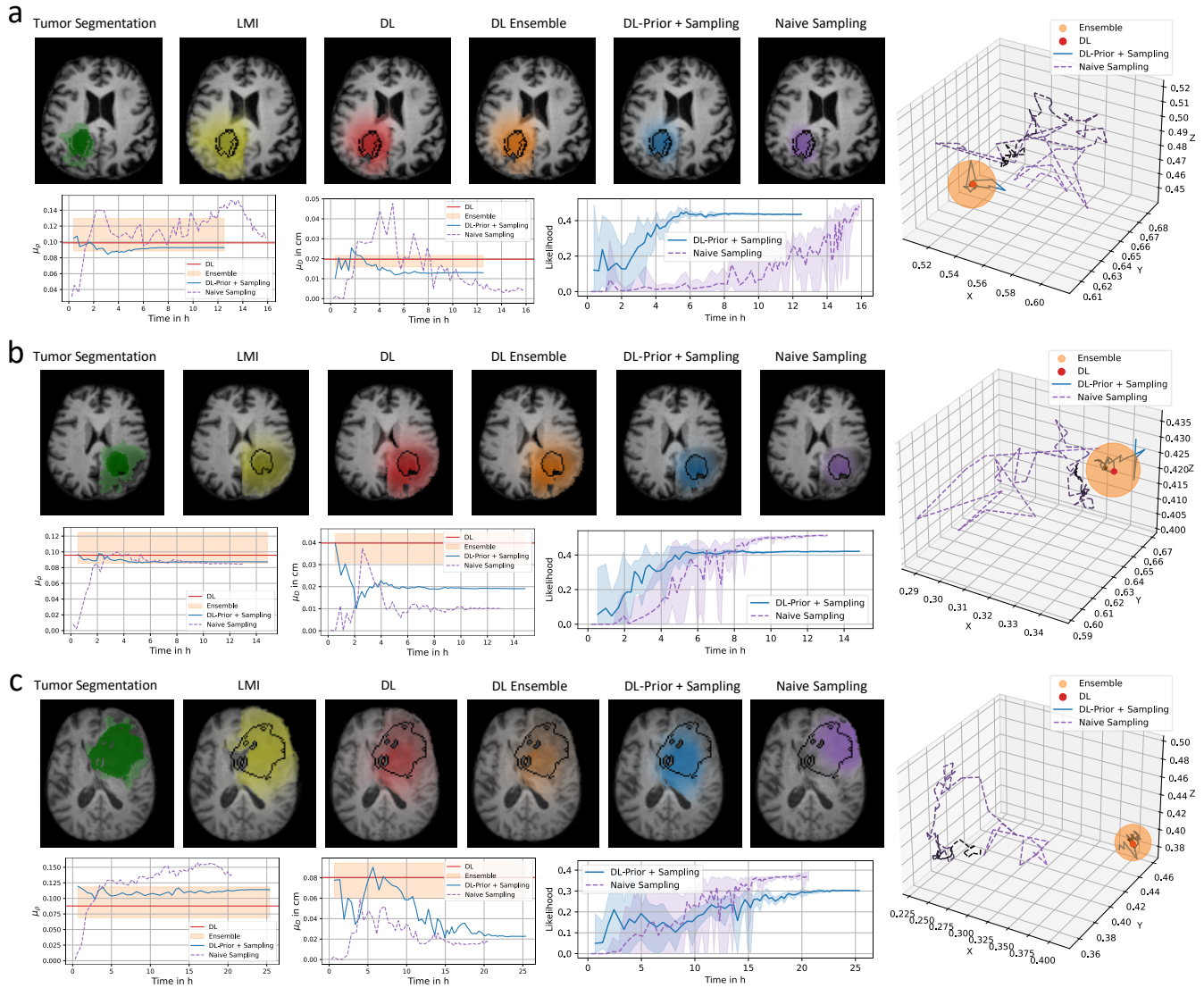
Figure 7: Evaluation of our method focusing on multiple scans of varying sizes and mass effects, particularly analyzing challenging cases. The final tumor concentration predictions are shown in the first row. Below, the time developments of μ_D, μ_ρ, the likelihood, and the tumor origin are shown. The DL and ensemble results are shown in red and orange. It becomes apparent that for small tumors without mass effect (a), the DL-Prior estimation provides a solid estimate leading to faster convergence. Large mass effects (c) can result in a failure of the network, constraining the sampling process into a suboptimal parameter set, which becomes apparent concerning the origin in this case.

an excellent argument for proceeding with clinical trials testing model-based radiotherapy planning.

7 Conclusion

This study introduces a novel framework that synergistically combines deep learning with evolutionary sampling to enhance brain tumor modeling. By leveraging a DL ensemble for initial parameter estimation and incorporating it into a high-precision evolution strategy, we achieve a fivefold acceleration in convergence time and a Dice-score of 95%. This approach significantly constrains the sampling parameter space, offering a powerful and efficient solution for individualized radiotherapy planning with the potential to substantially improve patient outcomes.

References

[1] M. J. van den Bent, M. Geurts, P. J. French, M. Smits, D. Capper, J. E. Bromberg, and S. M. Chang, "Primary brain tumours in adults," *The Lancet*, vol. 402, no. 10412, pp. 1564–1579, 2023.

[2] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2013–2017," *Neuro-oncology*, vol. 22, no. Supplement_1, pp. iv1–iv96, 2020.

[3] M. Weller, M. van den Bent, M. Preusser, E. Le Rhun, J. C. Tonn, G. Minniti, M. Bendszus, C. Balana, O. Chinot, L. Dirven *et al.*, "EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood," *Nature reviews Clinical oncology*, vol. 18, no. 3, pp. 170–186, 2021.

[4] M. Niyazi, N. Andratschke, M. Bendszus, A. J. Chalmers, S. C. Erridge, N. Galldiks, F. J. Lagerwaard, P. Navarria, P. M. af Rosenschöld, U. Ricardi *et al.*, "ESTRO-EANO guideline on target delineation and radiotherapy details for glioblastoma," *Radiotherapy and Oncology*, vol. 184, p. 109663, 2023.

[5] S. Subramanian, A. Ghafouri, K. M. Scheufele, N. Himthani, C. Davatzikos, and G. Biros, "Ensemble inversion for brain tumor growth models with mass effect," *IEEE Transactions on Medical Imaging*, vol. 42, no. 4, pp. 982–995, 2022.

[6] A. C. S. Jørgensen, C. S. Hill, M. Sturrock, W. Tang, S. R. Karamched, D. Gorup, M. F. Lythgoe, S. Parrinello, S. Marguerat, and V. Shahrezaei, "Data-driven spatio-temporal modelling of glioblastoma," *Royal Society Open Science*, vol. 10, no. 3, p. 221444, 2023.

[7] K. Schaller, "'go or grow': the key to the emergence of invasion in tumour progression?" *Mathematical medicine and biology: a journal of the IMA*, vol. 29, no. 1, pp. 49–65, 2012.

[8] J. Moreira and A. Deutsch, "Cellular automaton models of tumor development: a critical review," *Advances in Complex Systems*, vol. 5, no. 02n03, pp. 247–267, 2002.

[9] E. Lima, J. Oden, B. Wohlmuth, A. Shahmoradi, D. Hormuth II, T. Yankeelov, L. Scarabosio, and T. Horger, "Selection and validation of predictive models of radiation effects on tumor growth based on noninvasive imaging data," *Computer methods in applied mechanics and engineering*, vol. 327, pp. 277–305, 2017.

[10] M. Pabisz, J. Muñoz-Matute, and M. Paszyński, "Augmenting MRI scan data with real-time predictions of glioblastoma brain tumor evolution using exponential time integrators," *arXiv preprint arXiv:2402.02273*, 2024.

[11] M. Lê, H. Delingette, J. Kalpathy-Cramer, E. R. Gerstner, T. Batchelor, J. Unkelbach, and N. Ayache, "Bayesian personalization of brain tumor growth model," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part II 18*. Springer, 2015, pp. 424–432.

[12] J. Lipkova, P. Angelikopoulos, S. Wu, E. Alberts, B. Wiestler, C. Diehl, C. Preibisch, T. Pyka, S. E. Combs, P. Hadjidoukas *et al.*, "Personalized radiotherapy design for glioblastoma: integrating mathematical tumor models, multimodal scans, and bayesian inference," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1875–1884, 2019.

[13] S. Pati, V. Sharma, H. Aslam, S. P. Thakur, H. Akbari, A. Mang, S. Subramanian, G. Biros, C. Davatzikos, and S. Bakas, "Estimating glioblastoma biophysical growth parameters using deep learning regression," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6*. Springer, 2021, pp. 157–167.

[14] I. Ezhov, K. Scibilia, K. Franitza, F. Steinbauer, S. Shit, L. Zimmer, J. Lipkova, F. Kofler, J. C. Paetzold, L. Canalini *et al.*, "Learn-Morph-Infer: a new way of solving the inverse problem for brain tumor modeling," *Medical Image Analysis*, vol. 83, p. 102672, 2023.

[15] R. Z. Zhang, I. Ezhov, M. Balcerak, A. Zhu, B. Wiestler, B. Menze, and J. Lowengrub, "Personalized predictions of glioblastoma infiltration: Mathematical models, physics-informed neural networks and multimodal scans," *ArXiv*, 2023.

[16] A. Zhu, "Accelerating parameter inference in diffusion-reaction models of glioblastoma using

physics-informed neural networks," *SIAM Undergraduate Resources Online*, 2022.

[17] J. Petersen, F. Isensee, G. Köhler, P. F. Jäger, D. Zimmerer, U. Neuberger, W. Wick, J. Debus, S. Heiland, M. Bendszus *et al.*, "Continuous-time deep glioma growth models," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*. Springer, 2021, pp. 83–92.

[18] T. G. Grossmann, U. J. Komorowska, J. Latz, and C.-B. Schönlieb, "Can physics-informed neural networks beat the finite element method?" *arXiv preprint arXiv:2302.04107*, 2023.

[19] M. Balcerak, I. Ezhov, P. Karnakov, S. Litvinov, P. Koumoutsakos, J. Weidner, R. Z. Zhang, J. S. Lowengrub, B. Wiestler, and B. Menze, "Individualizing glioma radiotherapy planning by optimization of a data and physics informed discrete loss," *arXiv preprint arXiv:2312.05063*, 2023.

[20] Z. Xiong, Y. Gao, Y. Liu, A. Fazlollahi, P. Nestor, F. Liu, and H. Sun, "Quantitative susceptibility mapping through model-based deep image prior (MoDIP)," *NeuroImage*, vol. 291, p. 120583, 2024.

[21] Q. Zhao, D. B. Lindell, and G. Wetzstein, "Learning to solve pde-constrained inverse problems with graph networks," *arXiv preprint arXiv:2206.00711*, 2022.

[22] J. Jin, E. Ollivier, R. Touret, M. McKinley, K. Sabra, and J. Romberg, "PETAL: Physics emulation through averaged linearizations for solving inverse problems," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[23] F. Kofler, C. Berger, D. Waldmannstetter, J. Lipkova, I. Ezhov, G. Tetteh, J. Kirschke, C. Zimmer, B. Wiestler, and B. H. Menze, "Brats toolkit: translating brats brain tumor segmentation algorithms into clinical and scientific practice," *Frontiers in neuroscience*, p. 125, 2020.

[24] H. L. Harpold, E. C. Alvord Jr, and K. R. Swanson, "The evolution of mathematical modeling of glioma proliferation and invasion," *Journal of Neuropathology & Experimental Neurology*, vol. 66, no. 1, pp. 1–9, 2007.

[25] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz,

J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[27] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," *Evolutionary computation*, vol. 9, no. 2, pp. 159–195, 2001.

[28] N. Hansen, S. D. Müller, and P. Koumoutsakos, "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)," *Evolutionary computation*, vol. 11, no. 1, pp. 1–18, 2003.

[29] G. Casella, "An introduction to empirical bayes data analysis," *The American Statistician*, vol. 39, no. 2, pp. 83–87, 1985.

[30] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[31] A. Gholami, A. Mang, and G. Biros, "An inverse problem formulation for parameter estimation of a reaction–diffusion model of low grade gliomas," *Journal of mathematical biology*, vol. 72, pp. 409–433, 2016.

[32] A. Auger and N. Hansen, "Tutorial CMA-ES: evolution strategies and covariance matrix adaptation," in *Proceedings of the 14th annual conference companion on Genetic and evolutionary computation*, 2012, pp. 827–848.

[33] C. Ju, A. Bibaut, and M. van der Laan, "The relative performance of ensemble methods with deep convolutional neural networks for image classification," *Journal of Applied Statistics*, vol. 45, no. 15, pp. 2800–2818, 2018.

[34] J. Lipková, B. Menze, B. Wiestler, P. Koumoutsakos, and J. S. Lowengrub, "Modelling glioma progression, mass effect and intracranial pressure in patient anatomy," *Journal of the Royal Society Interface*, vol. 19, no. 188, p. 20210922, 2022.

[35] S. Subramanian, K. Scheufele, N. Himthani, and G. Biros, "Multiatlas calibration of biophysical brain tumor growth models with mass effect,"

in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23.* Springer, 2020, pp. 551–560.

[36] H. B. Frieboes, X. Zheng, C.-H. Sun, B. Tromberg, R. Gatenby, and V. Cristini, "An integrated computational/experimental model of tumor invasion," *Cancer research*, vol. 66, no. 3, pp. 1597–1604, 2006.

[37] X. Zheng, S. Wise, and V. Cristini, "Nonlinear simulation of tumor necrosis, neo-vascularization and tissue invasion via an adaptive finite-element/level-set method," *Bulletin of mathematical biology*, vol. 67, pp. 211–259, 2005.

[38] M. Finneran, D. Marotta, D. Altenburger, and E. Nardone, "Long-term survival in a patient with butterfly glioblastoma: A case report. cureus 12 (2), e6914," 2020.

[39] J. Siddiqui and A. S. Krishnan, "Butterfly glioma," *N Engl J Med*, vol. 378, no. 281, pp. 20 182 934 238 610–1056, 2018.

[40] L. Hathout, V. Patel, and P. Wen, "A 3-dimensional DTI MRI-based model of GBM growth and response to radiation therapy," *International journal of oncology*, vol. 49, no. 3, pp. 1081–1087, 2016.

[41] K. Painter and T. Hillen, "Mathematical modelling of glioma growth: the use of diffusion tensor imaging (DTI) data to predict the anisotropic pathways of cancer invasion," *Journal of theoretical biology*, vol. 323, pp. 25–39, 2013.

[42] K. A. van Garderen, S. R. van der Voort, M. M. Wijnenga, F. Incekara, A. Alafandi, G. Kapsas, R. Gahrmann, J. W. Schouten, H. J. Dubbink, A. J. Vincent *et al.*, "Evaluating the predictive value of glioma growth models for low-grade glioma after tumor resection," *IEEE Transactions on Medical Imaging*, 2023.

[43] M.-C. Metz, M. Molina-Romero, J. Lipkova, J. Gempt, F. Liesche-Starnecker, P. Eichinger, L. Grundl, B. Menze, S. E. Combs, C. Zimmer *et al.*, "Predicting glioblastoma recurrence from preoperative MR scans using fractional-anisotropy maps with free-water suppression," *Cancers*, vol. 12, no. 3, p. 728, 2020.

[44] M. Lundemann, P. Munck af Rosenschöld, A. Muhic, V. A. Larsen, H. S. Poulsen, S.-A. Engelholm, F. L. Andersen, A. Kjær, H. B. Larsson, I. Law *et al.*, "Feasibility of multi-parametric PET and MRI for prediction of tumour recurrence in patients with glioblastoma," *European journal of nuclear medicine and molecular imaging*, vol. 46, pp. 603–613, 2019.

[45] M.-C. Metz, I. Ezhov, J. C. Peeken, J. A. Buchner, J. Lipkova, F. Kofler, D. Waldmannstetter, C. Delbridge, C. Diehl, D. Bernhardt *et al.*, "Toward image-based personalization of glioblastoma therapy: A clinical and biological validation study of a novel, deep learning-driven tumor growth model," *Neuro-Oncology Advances*, vol. 6, no. 1, p. vdad171, 2024.