# SkepticalNet: A Safety-Aware Interactive Segmentation Framework for Neuro-Oncology

Efstathios Kyriazis[1(✉)], Grigorios Kalliatakis[1], Sotirios Bisdas[2], Manolis Tsiknakis[1], and Kostas Marias[1]

[1] Foundation for Research and Technology - Hellas (FORTH)
[2] University College London
{stkyriazis, gkalliatak, tsiknaki, kmarias}@ics.forth.gr
sotirios.bisdas@nhs.net

**Abstract.** Precise tumor segmentation is fundamental to neuro-oncological planning and longitudinal monitoring. Existing interactive foundation models exhibit strong generalization capabilities, theoretically enabling the segmentation of diverse brain lesion types and sub-regions. However, they remain ill-suited for this domain due to their reliance on single-modality inputs and "trigger-happy" behavior—often hallucinating structures assuming that every user interaction targets a valid lesion. In this work, we propose SkepticalNet, a native 3D, multi-modal framework based on nnInteractive, trained with a *skeptical* prior. By integrating a novel background interaction sampling protocol during training, we enable our model to distinguish between valid and erroneous user guidance. We evaluate this approach on a diverse multi-cohort BraTS dataset encompassing adult/pediatric gliomas and meningiomas. Our method significantly outperforms the state-of-the-art nnInteractive foundation model, achieving a global mean Dice of 85.9% (vs. 76.3%) while reducing the annotation burden. Crucially, we demonstrate that the model learns a context-aware safety profile, effectively suppressing activations triggered on healthy tissue while preserving the ability to robustly correct imprecise clicks near tumor boundaries. Inference code and weights are available at https://github.com/stathisky-repo/SkepticalNet.

**Keywords:** Interactive Segmentation · Brain Tumor Segmentation · Safety-Aware AI · Hallucination Suppression · Multi-modal Learning

## 1 Introduction

Accurate and robust segmentation of brain lesions is a prerequisite for radiotherapy planning and surgical navigation [3]. While fully automated methods based on Convolutional Neural Networks (CNNs) and Transformers have achieved expert-level performance on standard benchmarks [9,4], they often lack the flexibility to generalize across diverse clinical scenarios. A model trained on adult glioma typically fails in pediatric or post-treatment cases due to domain shifts in anatomy and texture. This necessitates a unified tool that bridges these domain gaps effectively, obviating the need for ensembles of specialist models.

In the broader landscape of medical image segmentation, the field is shifting towards interactive foundation models. Conventional SAM adaptations [10,2] and native 3D frameworks such as SAM-Med3D [13] and nnInteractive [5] demonstrate impressive versatility across diverse anatomies. However, these generalist models face critical limitations when applied to neuro-oncology. First, standard SAM adaptations rely on 2D slice-wise processing, which ignores 3D volumetric context and impedes annotation efficiency. Second, existing 3D models typically process single-channel inputs, discarding the complementary multi-modal information (T1, T1c, T2, FLAIR) essential for distinguishing complex tumor subregions [6]. Most critically, their core generalization mechanism, the "segment anything" prior, creates a safety paradox. Designed to extract valid structures from any prompt, these models are inherently agnostic to the biological nature of the target. Consequently, when applied to pathology segmentation, they exhibit "trigger-happy" behavior: an erroneous click on healthy tissue forces the model to segment local anatomical features. While this agnostic capability is a strength in open-world settings, it manifests as dangerous hallucination (an unconstrained segmentation response) in clinical workflows strictly limited to pathology, compromising *safety* (the reliable suppression of out-of-context prompts).

**Contributions.** In this work, we present SkepticalNet, a unified, safety-aware interactive generalist framework specifically adapted for neuro-oncology. Building upon nnInteractive, our contributions are four-fold:

1. **Native Multi-Modal 3D Architecture:** We introduce a native volumetric architecture that leverages full multi-modal MRI inputs, enabling precise segmentation of distinct tumor sub-regions (e.g., edema, necrosis, enhancing tumor) as defined by user intent.
2. **Safety-Aware Training:** We propose a novel training strategy that utilizes background interaction sampling to enforce context-awareness. By teaching the model to validate user prompts against local anatomy via a zero-output constraint—effectively learning a *skeptical* prior—we enable the suppression of activations triggered on healthy tissue, while preserving the intrinsic capability to correct imprecise interactions near tumor boundaries.
3. **Topology-Aware Interaction:** We introduce a stochastic pathfinding algorithm that respects the irregular topology of infiltrative brain lesions, providing a more realistic training signal than standard linear scribble simulations.
4. **Extensive Validation:** We demonstrate that our method outperforms the state-of-the-art nnInteractive foundation framework in both native and fine-tuned settings across four diverse BraTS benchmarks (Pre- and Post-Treatment Adult Glioma, Meningioma, and Pediatric Glioma), achieving superior performance with an improved safety profile.

## 2 Methodology

### 2.1 Network Architecture and Multi-Modal Adaptation

Our model builds upon the nnInteractive framework [5], which utilizes a 3D residual U-Net backbone. While the native implementation processes an 8-channel

input (1 image + 7 interaction guidance maps), we modify the first convolutional layer to accept an 11-channel input comprising four MRI modalities (T1, T1c, T2, FLAIR) and the original interaction channels. To leverage the generalized representations of the pretrained model, we initialize this expanded layer by replicating the weights of the original image channel across the four new modality channels, scaling each by a factor of 1/4. This normalization preserves the expected magnitude of initial feature activations, ensuring training stability while treating all modalities as equally weighted inputs during the initial phase.

### 2.2 Context-Aware User Simulation and Safety-Aware Training

Unlike standard fully automated segmentation, where the output space is fixed, our interactive model must adapt to varying user intents. In neuro-oncology, lesions manifest as complex hierarchies of *pathology-specific compartments*—ranging from atomic sub-regions (e.g., edema, necrosis, resection cavities) to composite structures (e.g., whole tumor, tumor core). To support this flexibility across diverse datasets, we formulate the training objective as a binary segmentation task where the target class $c_{target}$ dynamically samples either an atomic or a composite sub-region. This ensures all possible classes per-sample are supervised throughout training, seamlessly resolving user intent. To mitigate class imbalance, we explicitly bias this sampling toward minority classes.

Standard interactive training protocols typically simulate user prompts solely on error regions (where prediction $\neq$ ground truth). While effective for refining segmentation boundaries, this approach biases the model towards foreground-only supervision, depriving it of explicit signals from healthy tissue. To mitigate this and improve clinical reliability, we propose a hybrid sampling strategy where each training iteration is stochastically assigned to either **Foreground Supervision** ($p = 0.75$) or **Background Supervision** ($p = 0.25$). The variable $p$ denotes the branching probabilities governing our stochastic training tree (Fig. 1). All such ratios were experimentally optimized via grid search on the validation set to maximize segmentation accuracy and robustness to invalid prompts.

**Foreground Supervision.** In this mode, the model focuses on segmenting a valid lesion sub-region (Fig. 1, Left). For the initial step (where no prediction exists), a positive prompt (point, bounding box, scribble, or lasso) is simulated on the ground truth mask. In subsequent iterations, corrective interactions are generated based on the current error map:

- **Error Region Selection:** A False Positive (FP) or False Negative (FN) component is sampled with a probability proportional to its volume. This prioritizes significant morphological errors while maintaining the diversity of the training signal.
- **Interaction Placement:** Once a component is selected, an interaction type is chosen and placed according to the spatial priors defined in nnInteractive.
- **Negative Reinforcement:** Distinct from standard error-correction protocols, we introduce a probability ($p = 0.10$) of placing a negative prompt

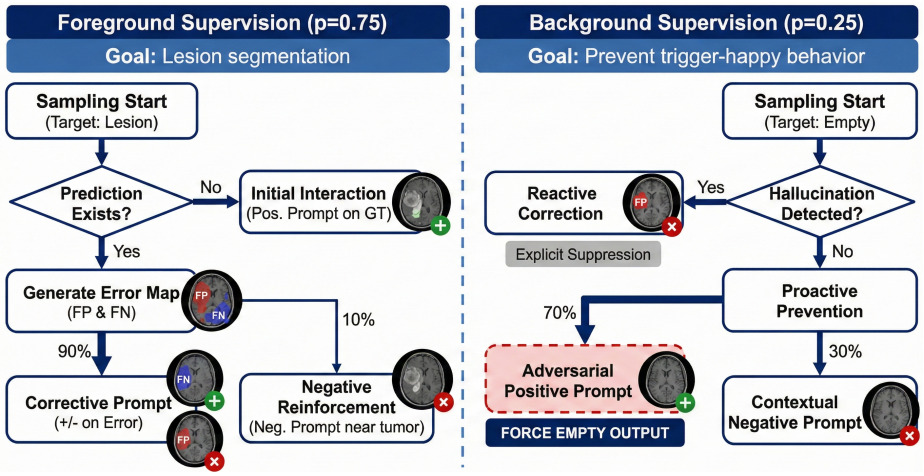


**Fig. 1.** Overview of the proposed hybrid Safety-Aware interactive training protocol.

within the proximal background adjacent to the lesion. This simulates a radiologist explicitly excluding confounding anatomy (e.g., vessels or resection cavities) to reinforce boundary constraints, even in the absence of a current prediction error.

**Background Supervision.** To prevent the "trigger-happy" behavior common in interactive models, we introduce a Safety-Aware training branch (Fig. 1, Right). In this mode, the target output is explicitly defined as an *empty mask*, enforcing a zero-output constraint. Our sampling strategy is conditioned on the current prediction state:

- **Scenario A: Proactive Prevention.** For clean predictions (no false positives), we simulate interactions to reinforce hallucination robustness. Sampling is restricted to regions $> 4$ voxels from the tumor boundary. Assuming 1 mm isotropic resolution, this 4 mm margin aligns with standard 2–5 mm radiotherapy uncertainty buffers used to account for microscopic extension [11], defining a clinically unambiguous healthy zone.
  1. **Adversarial Positive Prompts ($p = 0.70$):** A positive prompt is placed on healthy tissue. The model is penalized for any non-zero output, effectively learning a *skeptical* prior that prioritizes local image features over user guidance when the two conflict (e.g., accidental clicks).
  2. **Contextual Negative Prompts ($p = 0.30$):** A negative prompt is placed on healthy tissue to reinforce the association of negative constraints with non-target anatomy.
- **Scenario B: Reactive Correction.** If the model fails to suppress an input (generating a hallucination), a negative prompt is placed directly on the false positive region. We constrain these corrective clicks to be strictly contained

within healthy tissue ($GT = 0$) to avoid conflicting supervision near the lesion boundary. This acts as a second line of defense, teaching the model to reliably suppress explicit hallucinations based on user feedback.

This dual-mode strategy ensures that accidental or noisy interactions do not result in confusing false positives, establishing a robust clinical safety profile.

### 2.3 Topology-Aware Scribble Generation

Brain lesions frequently exhibit irregular, non-convex morphologies (e.g., C-shaped resection cavities) where standard line scribbles are prone to inadvertently traversing healthy tissue. This introduces erroneous signals that can compromise training. To mitigate this, we replace the default linear interpolation of nnInteractive with a **Topology-Constrained Stochastic Walk (TCSW)** algorithm. Given a start and a target point within the foreground of a single 2D slice, TCSW generates a path strictly constrained to the lesion mask by iteratively selecting the 8-connected neighbor that minimizes the geodesic distance to the target. To simulate human motor variance, uniform stochastic noise is added to the geodesic distance map prior to each pathfinding step, enforcing different minimal connecting paths. This ensures scribbles remain topologically valid yet realistically imperfect.

## 3 Experimental Setup

### 3.1 Datasets and Preprocessing

We evaluate our method on four distinct benchmarks from the BraTS challenge [1,12,8,7], covering a diverse range of clinical scenarios: Pre-Treatment Glioma ($N = 1251$), Post-Treatment Glioma ($N = 1621$), Meningioma ($N = 1000$), and Pediatric Glioma ($N = 99$). Note that $N$ refers to the number of MRI acquisitions (timepoints), and a single patient may possess multiple longitudinal scans.

**Data Splitting and Preprocessing.** To ensure rigorous evaluation, each dataset was partitioned into training (70%), validation (10%), and test (20%) sets. We enforced a strict patient-wise split, ensuring that all longitudinal timepoints from a single patient are assigned to the same partition to prevent data leakage. Additional stratification was employed to enforce the proportional representation of rare tumor sub-classes across all data subsets. Preprocessing was limited to standard z-score normalization of intensity values per modality.

**Balanced Sampling Strategy.** Given the significant imbalance in dataset sizes (e.g., 99 Pediatric vs. 1621 Post-treatment samples), standard uniform sampling would cause the model to underfit minority domains. To mitigate this, we employ a *Balanced Dataloader* which independently draws each sample from one of the four datasets with equal probability (25%), ensuring fully stochastic batch compositions without fixed allocations per domain. To prevent overfitting on frequently re-sampled minority cases, we apply aggressive spatial and intensity augmentations (e.g., elastic deformations, gamma, and Gaussian noise), ensuring robust input diversity.

### 3.2 Training Protocol

The model was trained for a total of 500 epochs using the Adam optimizer with an initial learning rate of $1 \times 10^{-4}$. To allow the input weights (initialized via the replication strategy described in Sec. 2.1) to adapt to the specific contrast properties of the multi-modal data without destabilizing the pretrained backbone, we employed a linear warmup phase for the first 25 epochs. Subsequently, a Cosine Annealing scheduler was applied to gradually decay the learning rate to a minimum of $1 \times 10^{-6}$. All experiments were conducted on a single NVIDIA A100 GPU (40GB VRAM) using a batch size of 2 samples per iteration.

## 4 Experiments and Results

### 4.1 Quantitative Performance

We evaluate segmentation performance on the held-out test set across all tumor sub-classes, utilizing a simulated interaction policy with a maximum budget of 8 clicks (interactions) per target. We benchmark against two baselines: the official pre-trained **nnInteractive (Native)** model (representing off-the-shelf generalist performance) and a single-modality fine-tuned variant (**FT-T1c**) trained on our BraTS dataset. The latter serves as a controlled baseline to isolate gains attributable strictly to multi-modal integration. All reported improvements over baselines are statistically significant (Wilcoxon signed-rank test, $p < 0.01$).

**Segmentation Accuracy (Table 1, Top).** SkepticalNet yields the highest accuracy across all datasets, achieving a global mean Dice of 85.9% compared to 76.3% (Native) and 81.9% (FT-T1c). The performance advantage is most distinct in the Post-Treatment cohort, where our multi-modal approach outperforms the FT-T1c baseline by +4.9 points (82.7% vs. 77.8%) and the Native baseline by +12.2 points (82.7% vs. 70.5%). This confirms that single-modality models are effectively blind to specific tumor sub-regions that lack distinct contrast signatures. The global median Dice (90.6%) notably exceeds the mean, indicating robust performance across the majority of cases rather than just a few outliers. In the Meningioma cohort, our method effectively matches human inter-rater reliability with a median Dice of 95.4%.

**Interaction Efficiency (Table 1, Bottom).** Beyond raw accuracy, SkepticalNet noticeably reduces the annotation burden. We report the Success Rate (SR)—the percentage of cases that reach a high-precision target (Dice $\geq 0.90$) within the budget—and the average clicks required for those successful cases. Our model achieves a global SR of 55.3%, nearly double that of the Native baseline (30.3%) and substantially higher than the FT-T1c model (39.6%) while requiring fewer clicks (2.0 vs. 2.6 vs. 2.3). In Post-Treatment cases, the baselines fail to reach the precision threshold in more than 75% of cases, whereas our method maintains a viable workflow with an SR of 42.2%. Furthermore, in the challenging Pediatric cohort, our method reduces the click burden by over 50% (1.8 vs. 4.0) and simultaneously improves the success rate (32.2% vs. 21.1%).

**Table 1.** Quantitative Performance. **Top:** Segmentation Accuracy (Dice %). We report Mean (Median). **Bottom:** Interaction Efficiency for high-precision tasks (Dice $\geq 0.90$). We report Average Clicks (Success Rate %). Best results in **bold**.

| Method | Glioma-Pre | Glioma-Post | Men. | Ped. | Global Avg |
|---|---|---|---|---|---|
| *Segmentation Accuracy: Mean (Median) Dice %* | | | | | |
| Native | 80.5 (85.9) | 70.5 (75.5) | 83.6 (92.0) | 73.9 (81.4) | 76.3 (82.5) |
| FT-T1c | 84.7 (89.6) | 77.8 (81.6) | 88.0 (93.6) | 77.2 (84.8) | 81.9 (86.8) |
| **Ours** | **87.9 (92.2)** | **82.7 (88.3)** | **90.7 (95.4)** | **80.1 (86.8)** | **85.9 (90.6)** |
| *Efficiency: Clicks (Success Rate %) for Target Dice $\geq$ 0.90* | | | | | |
| Native | 2.9 (35.0%) | 3.3 (14.3%) | 2.0 (60.4%) | 4.2 (20.0%) | 2.6 (30.3%) |
| FT-T1c | 2.4 (50.0%) | 2.9 (21.4%) | 1.7 (67.9%) | 4.0 (21.1%) | 2.3 (39.6%) |
| **Ours** | **1.9 (63.5%)** | **2.6 (42.2%)** | **1.5 (76.0%)** | **1.8 (32.2%)** | **2.0 (55.3%)** |

### 4.2 Safety Analysis: Skeptical Suppression

To isolate the impact of our *skeptical* training protocol, we benchmark against a **Standard Baseline** (an identical 4-modal architecture trained without background supervision) by simulating positive interactions on healthy tissue. We consider two spatial regimes: (1) **Distal** ($> 4$ voxels from tumor), an unambiguous non-target zone where positive prompts are clearly invalid; and (2) **Proximal** ($\leq 4$ voxels), a spatially ambiguous zone where the model must distinguish between noise and imprecise targeting. We employ two metrics to quantify safety: the **Activation Rate (AR)**, defined as the frequency of non-empty mask generation; and the **Target Precision** ($P_{target} = |P \cap G|/|P|$), measuring the precision of the predicted non-empty mask $P$ relative to the ground truth tumor $G$, to distinguish true hallucinations from snaps to valid pathology.

**Distal Regime: Global Suppression.** As shown in Table 2, the baseline exhibits "trigger-happy" behavior, generating activations in 98.9% of distal interactions. The moderate precision ($P_{target} = 60.8\%$) indicates that these predictions often drift toward distant pathologies, but are spatially unconstrained, resulting in substantial hallucinations. SkepticalNet drastically reduces this risk, achieving a Global AR of just 1.5%. This confirms that the model successfully learns a zero-output prior for unsupported signals on healthy tissue.

**Proximal Regime: Intent Recognition.** In the proximal zone, the behavior of the two models diverges. While the baseline indiscriminately segments local pathological structures in 100% of cases, SkepticalNet exhibits sophisticated context-awareness by differentiating between prompt types:

- **Dense Prompts (Bbox, Lasso):** The model identifies these as invalid prompts. Since their region encloses no tumor, the model suppresses the output (AR $\approx 2.6\%$), prioritizing visual evidence over erroneous guidance.
- **Sparse Prompts (Point, Scribble):** The model interprets these as imprecise targeting attempts. It retains a high activation rate ($\approx 70\%$) and achieves a high Target Precision ($P_{target} \approx 89\%$), indicating robust boundary correction (snapping) to the valid pathology.

**Table 2.** Safety analysis in Distal ($> 4$ voxels) and Proximal ($\leq 4$) regimes. We report global metrics alongside interaction-specific breakdowns (Sparse vs. Dense).

| **Method** | **Distal (Safety)** | | **Proximal (Robustness)** | |
|---|---|---|---|---|
| | $AR \downarrow$ | $P_{target} \uparrow$ | $AR \downarrow$ | $P_{target} \uparrow$ |
| Baseline (Global) | 98.9% | 60.8% | 100.0% | 85.6% |
| **Ours (Global)** | 1.5% | — | 36.3% | 86.5% |
| *Sparse (Point/Scribble)* | 1.9% | — | 70.1% | **88.6**% |
| *Dense (Bbox/Lasso)* | **1.1**% | — | **2.6%** | 84.5% |

This behavior aligns with clinical intuition: the model enforces strict feature validation for dense prompts (rejecting empty regions), while retaining the flexibility to correct "near-miss" clicks (sparse prompts) to the adjacent pathology.

## 5 Discussion

In this work, we addressed the limitations of interactive foundation models in neuro-oncology, confirming that "one size fits all" approaches struggle with the specific safety and radiological challenges of brain tumor segmentation.

**Safety through Context-Awareness.** A critical barrier to deploying interactive AI is the "trust gap". Standard models, driven by foreground-only sampling, exhibit a bias to segment *something*—even on healthy tissue—violating the clinical "Do No Harm" principle. Our proposed *skeptical* training strategy drastically reduces this risk (AR $< 2\%$) by teaching the model a zero-output prior for background interactions. This ensures that the network acts as a reliable partner that rejects accidental or erroneous guidance rather than a tool that requires constant supervision against hallucinations.

**The Necessity of Multi-Modal Context.** Comparisons with single-modality baselines highlight a fundamental physical constraint: geometry alone cannot compensate for missing contrast. While T1c-based models can identify enhancing cores, they inherently lack the signal information to unambiguously separate non-enhancing regions (e.g., edema or resection cavities) from healthy tissue. SkepticalNet leverages complementary signals (e.g., FLAIR/T2) to resolve these ambiguities, resulting in substantial gains in complex cohorts (+4.9% Dice on Post-Treatment). This confirms that for comprehensive tumor volumetry, multimodal integration is not a luxury, but a requirement.

**Efficiency and Survivorship Bias.** A superficial analysis might suggest that the Native baseline is relatively efficient ($\approx$2.6 clicks) for Dice $\geq 0.90$. However, this metric suffers from massive *survivorship bias*. In this setting, the baseline fails in $\approx$70% of cases, effectively filtering out complex lesions and skewing efficiency calculations toward simple targets. In contrast, our model doubles the Success Rate (55.3%) while averaging fewer clicks (2.0). This indicates that our method is not only faster but fundamentally more robust, successfully resolving the complex pediatric and post-treatment cases where generalist models fail.

**Limitations and Future Work.** While 4-sequence mpMRI is the standard-of-care, our fixed 11-channel input restricts missing-modality resilience. Additionally, validation is currently limited to pre-processed BraTS datasets. Future work must evaluate generalizability to unstandardized artifacts, tiny lesions (e.g., brain metastases), and extra-cranial applications. Finally, user studies are needed to validate our framework's real-world impact on radiologist contouring time.

## 6 Conclusion

We presented SkepticalNet, a safety-aware, multi-modal interactive segmentation framework specifically tailored for neuro-oncology. By enforcing *skeptical* behavior via a novel background interaction protocol, we achieved superior accuracy and efficiency compared to generalist baselines. Our findings demonstrate that domain-specific adaptation is essential to mitigate hallucinations in foundation models, providing a robust solution to accelerate clinical tumor volumetry.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Baid, U., Ghodasara, S., Mohan, S., et al.: The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification (2021), https://arxiv.org/abs/2107.02314
2. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., Sun, H., He, J., Zhang, S., Zhu, M., Qiao, Y.: SAM-Med2D (2023), https://arxiv.org/abs/2308.16184
3. Despotović, I., Goossens, B., Philips, W.: Mri segmentation of the human brain: Challenges, methods, and applications. Computational and Mathematical Methods in Medicine **2015**(1), 450341 (2015). https://doi.org/https://doi.org/10.1155/2015/450341, https://onlinelibrary.wiley.com/doi/abs/10.1155/2015/450341
4. Ferreira, A., Solak, N., Li, J., Dammann, P., Kleesiek, J., Alves, V., Egger, J.: How we won BraTS 2023 adult glioma challenge? just faking it! enhanced synthetic data augmentation and model ensemble for brain tumour segmentation (2024), https://arxiv.org/abs/2402.17317
5. Isensee, F., Rokuss, M., Krämer, L., Dinkelacker, S., Ravindran, A., Stritzke, F., Hamm, B., Wald, T., Langenberg, M., Ulrich, C., Deissler, J., Floca, R., Maier-Hein, K.: nnInteractive: Redefining 3d promptable segmentation (2025), https://arxiv.org/abs/2503.08373
6. Işın, A., Direkoğlu, C., Şah, M.: Review of mri-based brain tumor image segmentation using deep learning methods. Procedia Computer Science **102**, 317–324 (2016). https://doi.org/https://doi.org/10.1016/j.procs.2016.09.407, https://www.sciencedirect.com/science/article/pii/S187705091632587X, 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria
7. Kazerooni, A.F., Khalili, N., Liu, X., et al.: The brain tumor segmentation (BraTS) challenge 2023: Focus on pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs) (2024), https://arxiv.org/abs/2305.17033

8. LaBella, D., Adewole, M., Alonso-Basanta, M., et al.: The ASNR-MICCAI brain tumor segmentation (BraTS) challenge 2023: Intracranial meningioma (2023)
9. Luu, H.M., Park, S.H.: Extending nn-UNet for brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 173–186. Springer International Publishing, Cham (2022)
10. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024)
11. Niyazi, M., Andratschke, N., Bendszus, M., Chalmers, A., Erridge, S., Galldiks, N., Lagerwaard, F., Navarria, P., Munck af Rosenschold, P., Ricardi, U., Bent, M., Weller, M., Belka, C., Minniti, G.: Estro-eano guideline on target delineation and radiotherapy details for glioblastoma. Radiotherapy and Oncology **184**, 109663 (04 2023). https://doi.org/10.1016/j.radonc.2023.109663
12. de Verdier, M.C., Saluja, R., Gagnon, L., et al.: The 2024 brain tumor segmentation (BraTS) challenge: Glioma segmentation on post-treatment MRI (2024), https://arxiv.org/abs/2405.18368
13. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., Fu, B., Zhang, S., He, J., Qiao, Y.: SAM-Med3D: Towards general-purpose segmentation models for volumetric medical images (2024), https://arxiv.org/abs/2310.15161