Unlimited Post-Treatment Glioma MR Images via Conditional Diffusion

Elijah Renner ELIJAHCRENNER@GMAIL.COM
Thetford Academy
304 Academy Rd, Thetford Center, VT 05075

Advisor: Dr. Alaa Youssef AYOUSSEF@STANFORD.EDU
Stanford Center for Artificial Intelligence in Medicine and Imaging
1701 Page Mill Rd, Palo Alto, CA 94304

Abstract

The scarcity of labeled post-treatment glioma MR images limits effective automatic segmentation of key features in brain MR images. Addressing this issue, GliomaGen is introduced, an anatomically informed generative diffusion model that uses a modified Med-DDPM structure to create high-quality MR images from anatomical masks. GliomaGen takes four modalities and six segmentation labels, including a new head area, as input. The developed GliomaGen pipeline augments existing masks to expand the BraTS 2024 Post-Treatment Glioma dataset by 2124 masks, which are later used to synthesize the largest BraTS 2024 Adult Post-Treatment Glioma derivative synthetic dataset ($N = 2124$). Evaluations of GliomaGen with quantitative metrics MS-SSIM, FID, and KID show high fidelity, particularly for t1c (FID: 55.2028 $\pm$ 3.7446) and t2w (FID: 54.9974 $\pm$ 3.2271) modalities. Segmentation tests with nnU-Net show hybrid training matches real-data performance, but inconsistencies and noise in generated volumes prevented state-of-the-art segmentation from being achieved. These findings show the potential of conditional diffusion models to address data constraints in the BraTS 2024 Adult Post-Treatment Glioma context, and also prompt further iteration on the GliomaGen pipeline.

Keywords: glioma, generative models, diffusion, segmentation, mask

1 Introduction

Review of AI in neuroradiology In recent years, AI tools have demonstrated remarkable capabilities in neuroradiology. Havaei et al. (2017), Lin et al. (2023), and Ferreira et al. (2024) demonstrate that convolutional neural networks excel in brain tumor segmentation tasks like the BraTS Challenge and that the state of the art has rapidly improved in the past decade.

Data scarcity The performance of neural networks hinges on the availability of high-quality and diverse training data Dee; et al.. In neuroradiology, annotated datasets are particularly scarce due to privacy concerns and the high cost of expert labeling The. This is evident in the BraTS 2024 Adult Glioma dataset, where gliomas present unique challenges for segmentation. Gliomas often blend seamlessly with surrounding brain tissue, lacking clear boundaries and infiltrating adjacent regions, which complicates manual and automated annotation Song et al. (2024). While the size of the BraTS Challenge has steadily increased,

helping advance the state of the art in detailed[1] glioma segmentation, its size ($N \approx 1350$) still limits model performance Sage Bionetworks.

Related work Synthetic data generation is a promising emerging solution to data scarcity. Since Ho et al. published Ho et al. (2020) in 2020, denoising diffusion probabilistic models (DDPMs, or simply diffusion models for brevity) have gained attention for their superior performance when generating synthetic clinical imagery compared to generative adversarial networks (GANs). Müller-Franzes et al. Müller-Franzes et al. (2023) showed that a latent diffusion model[2] (LDM) outperformed StyleGAN-3 Karras et al. (2021), cGAN Krause and et al. (2021), and ProGAN Karras et al. (2018) on fundus, histological, and chest X-ray generation tasks, respectively. Other studies Bluethgen et al. (2024); Lüpke et al. (2024); Pinaya et al. (2022); Akbar et al. (2024) publish similar results where diffusion models outperform GANs. Although these diffusion models achieve high-fidelity generations, they lack precise anatomical conditioning mechanisms. Bluethgen et al. (2024), for example, uses text-based anatomical conditioning, which adds subjectivity during inference, while Lüpke et al. (2024) does not incorporate any anatomical conditioning, Pinaya et al. (2022) incorporates only numerical conditioning through variables like age, gender, ventricular volume, etc., and Akbar et al. (2024) does not use any conditioning. Towards precise anatomical conditioning, novel frameworks like Med-DDPM Dorjsembe et al. (2024), SegGuidedDiff Konz et al. (2024), and Mei et al. (2024) allow a diffusion model to be conditioned on a labeled feature mask[3].

Despite achieving exceptional results, these methods have yet to release a large annotated synthetic dataset or leverage the diversity of emerging multi-class datasets like BraTS 2024 Adult Glioma (enhancing tissue, non-enhancing tumor core, surrounding non-enhancing FLAIR hyperintensity, and resection cavity).

Contribution This paper presents GliomaGen, an anatomically-conditioned diffusion model tailored to BraTS 2024 Adult Post-Treatment Glioma de Verdier et al. (2024) capable of generating high-fidelity brain MR images. The code and training procedures are shared publicly for the reproduction of this study. The modularity of the GliomaGen repository eli allows it to be adapted to other BraTS domains beyond post-treatment glioma. Following training, GliomaGen is used to synthesize a synthetic BraTS 2024 Adult Post-Treatment Glioma derivative ($N = 2124$), which is the largest labeled dataset of its kind to date. A high-quality subset will be released publicly on HuggingFace in the coming weeks, along with the weights and training code for GliomaGen. Finally, training a nnU-Net on a high-quality subset of the synthetic dataset demonstrated that while current methods can produce detailed structures, using synthetic data in isolation to train downstream models may yield poor segmentation performance.

1. I.e., with more than 1 feature to segment
2. A diffusion model that utilizes a latent space
3. Note that these are *not* LDMs since they make use of the image space such that anatomical precision isn't lost in an arbitrary latent space

2 Data Preparation

The BraTS 2024 Adult Post-Treatment Dataset de Verdier et al. (2024) is collected via Synapse. The dataset contains $N = 1350$ samples, and 50 are withdrawn into a validation set. Each training example has 4 corresponding modalities (t1c, t1n, t2f, and t2w) and a mask annotated by expert radiologists with the labels background (0), non-enhancing tumor core (1), surrounding non-enhancing FLAIR hyperintensity (2), enhancing tumor (3), and resection cavity (4). To align the dataset with the diffusion pipeline, a label in the anatomical masks for 'head' is added by first incrementing each non-background label by 1. Then, Otsu thresholding is used via scikit-image **?** to detect the head region, which is assigned label 1. With the head mask, the resulting anatomical masks have 6 channels (see Figure 1). Adding the head region expedites the model's learning of how to generate the head region. All samples were reviewed for artifacts introduced by the head mask addition procedure, and 67 samples with holes in the head mask were removed, making the final dataset size $(\text{train}, \text{val}) = (1235, 48)$.

Next, the samples are all linearly normalized to $[0, 1]$ for consistency, since the original BraTS volumes have varying intensity ranges. Finally, volumes are padded and cropped: $(182, 218, 182) \rightarrow (192, 192, 144)$. The new dimensions ensure compatibility with the 3D U-Net implemented in the diffusion routine since dimensions that are multiples of 16 ensure non-integer dimensions don't arise in downsampling.

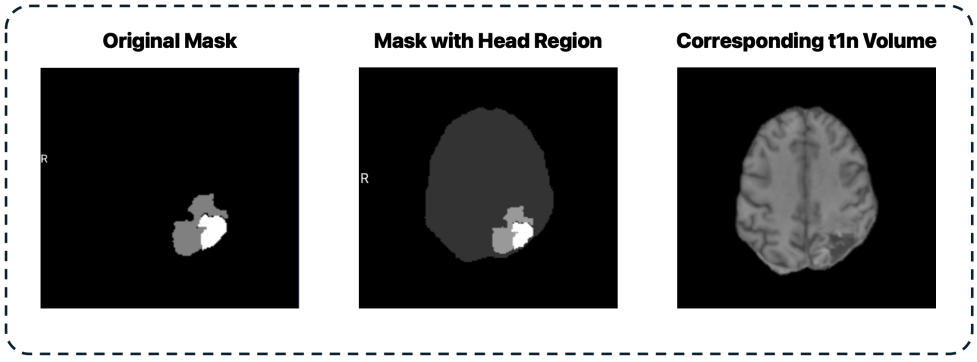

Figure 1: Anatomical mask before and after head addition alongside its corresponding t1n volume

3 Diffusion Pipeline

Med-DDPM Med-DDPM Dorjsembe et al. (2024) is the architecture used for GliomaGen. Whereas in the vanilla DDPM Ho et al. (2020), the noisy input x_t is the only input into the denoising U-Net, Med-DDPM concatenates the labeled anatomical mask input c with the noisy image x_t, yielding the input

$$\tilde{x}_t^{(10,w,d,h)} = x_t^{(4,w,d,h)} \oplus c^{(6,w,d,h)}$$

This channel-wise concatenation allows the denoiser to be informed about anatomical structures and label placement, enabling highly-controlled generation. To process $\tilde{x}$, the authors

adjust the U-Net's input channels to equal the sum of channels in $\tilde{x}$ and c (1 modality + 2 labels in the case of Med-DDPM's single-modality synthesis).

Med-DDPM (see Figures 2 and 3) uses the original forward diffusion process q outlined by Ho et al. (2020), which adds small amounts of random (Gaussian) noise $\epsilon \sim \mathcal{N}(0, 1)$ according to a variance schedule $\bar{\alpha}_t$ at each timestep t to an image x_0 for a total number of timesteps T. In GliomaGen, $T = 1000$ is used instead of Med-DDPM's 250 steps to gain more precise generations. They define the noisy sample x_t as

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon$$

for timesteps $1 \leq t \leq T$. Med-DDPM's cosine noise schedule

$$\bar{\alpha}_t = \frac{f(t)}{f(0)}, f(t) = \cos^2\left(\frac{t/T + s}{1 + s} \cdot \frac{\pi}{2}\right)$$

is used to control noise addition where s is a small offset parameter to prevent extreme values near $t = 0$. The schedule smoothly introduces noise throughout the forward process, ensuring a continuous loss of structure rather than abrupt degradation.

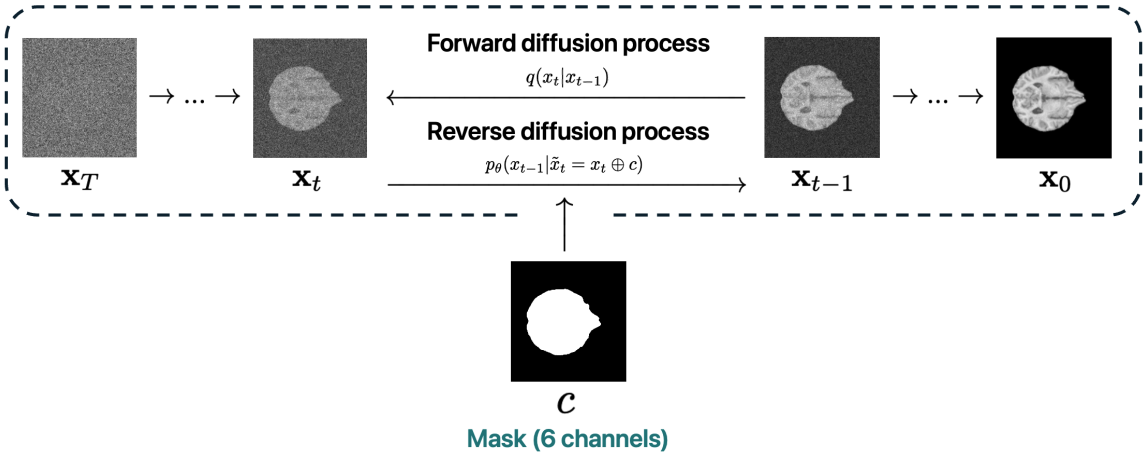

Figure 2: Diffusion processes with Med-DDPM concatenation rule adapted to the modified BraTS 2024 Adult Post-Treatment Glioma dataset with the head class.

To recover images from a noisy state, the same reverse process p_θ containing a 3D U-Net architecture is used.

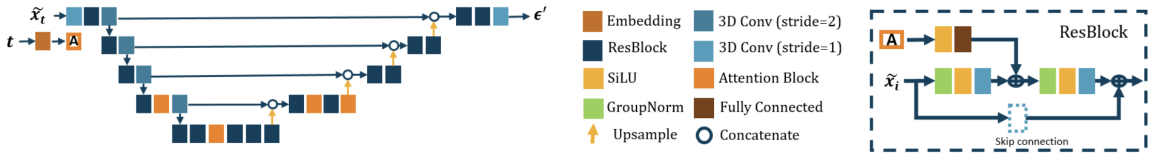

Figure 3: U-Net noise predictor ϵ_θ architecture from Med-DDPM Dorjsembe et al. (2024)

However, for the context of the modified BraTS 2024 Adult Post-Treatment Glioma dataset containing the head label,

$$n_{channels} = (4 \text{ modalities} + 6 \text{ anatomical mask labels}) = 10$$

; therefore, we modify the U-Net's head to support the 10-channel input $\tilde{x}_t^{(10,w,d,h)}$ accordingly. The denoising process is then written as

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\tilde{x}_t, t)\right) + \sigma_t z$$

where $z \sim \mathcal{N}(0, \mathbf{I})$, $\sigma_t = \sqrt{\beta_t}$, $\beta_t \in (0, 1)$, and ϵ_θ is the trained noise predictor U-Net model.[4] This step is repeated iteratively until x_0, the final denoised image, is obtained.

The $L1$ loss function

$$L1 = \frac{1}{n}\sum_{i=1}^{n}|\epsilon_i - \epsilon_i'|$$

is used to measure the distance between the original noise added in the forward process ϵ_i with the noise predicted by the denoising U-Net ϵ_i' where $\epsilon' = \epsilon_\theta(\tilde{x}_t, t)$ and $n = w \cdot h \cdot d$ is the total number of pixels in the volume. Then, the Med-DDPM algorithms are as described in Figure 4, modified to handle BraTS 2024 Adult Post-Treatment Glioma's 10 channel inputs of size $(w, d, h) = (192, 192, 144)$.

a Training for GliomaGen

Require: $c^{(6,192,192,144)}$

repeat
  $x_0^{(4,192,192,144)} \leftarrow q\left(x_0^{(4,192,192,144)}\right)$
  $t \leftarrow \text{Uniform}(\{1, \ldots, T\})$
  $\epsilon \leftarrow \mathcal{N}(0, \mathbf{I})$
  $x_t^{(4,192,192,144)} \leftarrow \sqrt{\bar{\alpha}_t}\, x_0^{(4,192,192,144)} + \sqrt{1-\bar{\alpha}_t}\,\epsilon$
  $\tilde{x}_t^{(10,192,192,144)} \leftarrow x_t^{(4,192,192,144)} \oplus c^{(6,192,192,144)}$
  Take gradient descent step on $\nabla_\theta L1\left(\epsilon, \epsilon_\theta\left(\tilde{x}_t^{(10,192,192,144)}, t\right)\right)$
until convergence

b Sampling for GliomaGen

Require: $c^{(6,192,192,144)}$

Sample $x_T^{(4,192,192,144)} \leftarrow \mathcal{N}(0, \mathbf{I})$
Initialize with $c^{(6,192,192,144)}$
for $t = T$ **down to** 1 **do**
  if $t > 1$ **then**
    $z \leftarrow \mathcal{N}(0, \mathbf{I})$
  else
    $z \leftarrow 0$
  end if
  $\tilde{x}_t^{(10,192,192,144)} \leftarrow x_t^{(4,192,192,144)} \oplus c^{(6,192,192,144)}$
  $\epsilon' \leftarrow \epsilon_\theta\left(\tilde{x}_t^{(10,192,192,144)}, t\right)$
  $x_{t-1}^{(4,192,192,144)} \leftarrow \frac{1}{\sqrt{\alpha_t}}\left(x_t^{(4,192,192,144)} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\,\epsilon'\right) + \sigma_t\, z$
end for
return $x_0^{(4,192,192,144)}$

Figure 4: Modified training and sampling algorithms for GliomaGen that process conditioned inputs with dimensions $c^{(6,192,192,144)} \oplus x_0^{(4,192,192,144)} \rightarrow \tilde{x}_t^{(10,192,192,144)}$.

The entire diffusion pipeline is implemented in PyTorch PyTorch (2014).

4. See Appendix B for the intuition behind this formulation.

4 Experiments

4.1 Diffusion Model Training

First, the modified BraTS dataset of 1235 training samples and 48 validation samples is loaded. To grant the U-Net a greater ability to learn features, it is initialized with 64 channels. The number of timesteps is increased to $T = 1000$ for more gradual noise removal, improving fidelity. The model was trained for 245000 iterations, with the learning rate schedule in Table 1.

Step Range	**Learning Rate**
0–65,000	1×10^{-5}
65,001–75,000	7.5×10^{-6}
75,001–85,000	5×10^{-6}
85,001–100,000	2.5×10^{-6}
100,001–245,000	1×10^{-6}

Table 1: Learning rate schedule over training steps.

Gradient accumulation is used to mimic a higher batch size not achievable on the A100 GPU used for training due to memory constraints. At 170000 iterations, the U-Net encoder is frozen to prioritize learning fine details in the encoder while retaining learned high-level features. This also increased $\frac{\#\text{ iterations}}{\text{sec}}$ from $\frac{34}{60} \rightarrow \frac{38}{60}$, an 11% increase in training efficiency, by reducing the number of parameters to compute gradients for at each step.

4.2 Dataset Synthesis

Generating additional synthetic data requires masks, so a procedure was designed to augment the original $N = 1235$ masks. For each original mask, a flip along the sagittal plane is performed with 50% probability. Then, a transformation limited to ± 1 voxels in each dimension and scaling by a factor $s \sim \mathcal{U}(0.98, 1.02)$ is applied. If the head region is no longer within the bounds, up to 50 attempts are made to find a valid transformation; if none is found, the original mask is retained. Note that for retained masks that did not have a flip, no augmented mask will be produced for the corresponding original mask. Hence, the procedure may generate fewer augmentations than masks in the original dataset.

Using the mask-wise procedure described in Algorithm 1, a set of 2124 anatomical masks was generated by creating 2 augmentations per original mask. 346 masks were discarded after multiple augmentation attempts failed. Dataset synthesis was then carried out by feeding each synthetic anatomical mask through GliomaGen. After 48 hours of inference, the complete synthetic dataset of size $N = 2124$ was constructed.

Algorithm 1 Anatomical Mask Augmentation

Require: Original mask $M \in \mathbb{Z}^{D \times H \times W}$, number of augmentations K

for $i = 1$ to K **do**
  Random Flip: With probability 0.5, set $M' \leftarrow \texttt{flip}(M)$; otherwise $M' \leftarrow M$.
  Compute Head Region: Compute centroid $\mathbf{c} = (c_z, c_y, c_x)$ from label 1 voxels in M'.
  repeat
    Sample scale $s \sim \mathcal{U}(0.98, 1.02)$ and translations $(\Delta z, \Delta y, \Delta x) \in \{-1, 0, 1\}$.
    Define the affine matrix $A = \frac{1}{s} I_{3\times 3}$ and offset $\mathbf{o} = \mathbf{c} - s\,\mathbf{c} + (\Delta z, \Delta y, \Delta x)$.
    Compute the transformed mask $M'' \leftarrow \texttt{affine\_transform}(M', A, \mathbf{o})$ ▷ Applies the affine transformation defined by matrix A and offset $\mathbf{o}$ to M' (i.e., scales and translates the mask).
  until M'' has a head region (1) within the image bounds
  if $1 \in M''$ **then**
    Save M'' as the augmented mask.
  else
    Discard and revert to M'.
  end if
end for

Figure 5: Procedure for augmenting one mask K times

4.3 Segmentation Benchmarking

In order to determine the validity of synthetic data for use in downstream tasks like segmentation, several identical nnU-Net[5] Isensee et al. (2018) models were trained on different datasets:

1. Only real data ($N = 1235$).
2. Original data ($N = 1235$) + all synthetic samples ($N = 2124$).
3. Original data ($N = 1235$) + half of synthetic samples ($N = 1062$).
4. Original data ($N = 1235$) + high-quality subset ($N = 598$) of synthetic data.
5. Only the high-quality subset ($N = 598$) of synthetic data.
6. Half of original data ($N = 617$) + high-quality subset ($N = 598$) of synthetic data.

High-quality subset curation The high-quality subset of generated images was curated by selecting images without noise or artifacts from the first 1062 images to be generated. GliomaGen occasionally produces images with significant noise, which are omitted from the high-quality subset. Images that do not align with their input masks (e.g., if a tumor is absent) are also omitted.

5. nnU-Net is chosen due to its modularity. It was easily adaptable to the context of BraTS 2024 Adult Post-Treatment Glioma.

The nnU-Net auto-configuration tool was used, and the resulting 1000-epoch training plan consisted of a U-Net with 6 stages of 3D convolutions with a kernel size of $[3, 3, 3]$. Each nnU-Net configuration had 4 input channels, each corresponding to an MRI modality (t1c, t1n, t2w, t2f). During training, a batch size of 2 was used. An SGD optimizer with an initial learning rate $\alpha = 0.01$, momentum of 0.99, and weight decay of 3×10^{-5} was used for learning. A combined dice and cross-entropy loss is applied to decoder layers at multiple stages to prevent vanishing gradients. A comprehensive, low-level technical report of nnU-Net is provided in the nnU-Net paper Isensee et al. (2018).

5 Results

5.1 GliomaGen

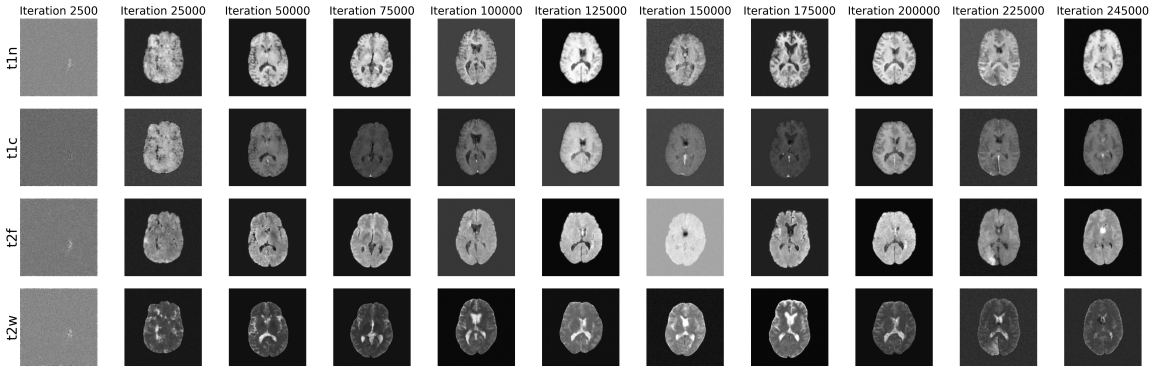

Figure 6: Sample generated modalities over 245000 training iterations, showing the model's learning of anatomical features over time.

The diffusion model was trained for 127 hours. Over 245000 iterations, the model learned to generate anatomical structures. Figure 6 shows the quality of generated volumes over iterations. Note that as training progresses, GliomaGen gains a better understanding of how to produce internal structures.

Modality*	**MS-SSIM (↑)**	**FID (↓)**	**KID (↓)**
t1n[a]	0.7005 ± 0.2585	58.4627 ± 3.8681	0.0305 ± 0.0011
t1c[b]	0.7647 ± 0.2106	55.2028 ± 3.7446	0.0293 ± 0.0019
t2w[c]	0.6513 ± 0.2881	54.9974 ± 3.2271	0.0291 ± 0.0010
t2f[d]	0.7842 ± 0.1551	70.4296 ± 4.1727	0.0370 ± 0.0018
Mean ± SD†	0.7252 ± 0.06	59.7731 ± 6.5	0.0315 ± 0.003

Table 2: Quantitative results for GliomaGen's multi-modality MRI synthesis. Higher MS-SSIM (↑) indicates better structural similarity; lower FID (↓) and KID (↓) suggest better perceptual quality. Values are reported as mean ± standard deviation across samples.

*Modality abbreviations: [a]T1-weighted without contrast, [b]T1-weighted with contrast, [c]T2-weighted, [d]T2-FLAIR. †Mean ± standard deviation calculated across modalities (not samples).

GliomaGen was fed the 48 validation masks as conditioning to generate 48 corresponding MRI volumes. The predicted volumes were quantitatively compared against the ground truth using multi-scale structural similarity index (MS-SSIM) Wang et al. (2003), Fréchet inception distance (FID), and KID metrics. All 3 metrics quantify how close the model is to real MRI scans. MS-SSIM measures the structural similarity between generated and real images using local luminance, contrast, and structure, while Fréchet inception distance uses an Inception Szegedy et al. (2014) network to extract features. The Fréchet distance[6] between the extracted feature distributions and the real images is then taken, yielding FID, which has been shown to correlate with human assessment Xu et al. (2018). Kernel inception distance (KID) is similar to FID but uses a kernel-based method for measuring the similarity between generated and real volumes. The comprehensive results are visible in Table 2.

GliomaGen achieved the best FID (54.9974 ± 3.2271) and KID (0.0291 ± 0.0010) for generating t2w modalities, but struggled significantly to produce high-fidelity modalities like t2f, where a FID of 70.4296 ± 4.1727 and KID of 0.0370 ± 0.0018 were measured. The overall mean FID of 59.7731 ± 3.7531 is within an acceptable range but indicates room for further work.

MS-SSIM scores of 0.7842 ± 0.1551 and 0.7647 ± 0.2106 for the t2f and t1c modalities, respectively, indicate strong structural similarity was retained in generated volumes. However, for t2w and t1n modalities, MS-SSIM scores of 0.6513 ± 0.2881 and 0.7005 ± 0.2585, respectively, reveal GliomaGen may falter in generating accurate structures across all modalities. Hence, while GliomaGen can produce realistic structures for certain modalities, further refinements to the training routine are needed to enhance its capability to produce modality-specific features and more anatomical detail. The variance across MS-SSIM values also indicates that generated images are inconsistent in fidelity, which also suggests further refinement is needed.

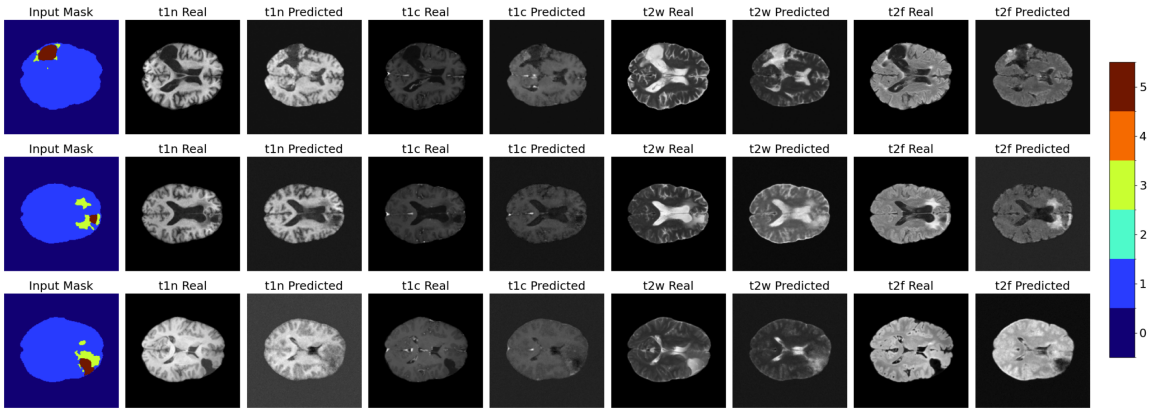

Figure 7: Generated MRI modalities for corresponding validation input anatomical masks aside ground truth. The labels for the input mask are background (0), head (1), non-enhancing tumor core (2), surrounding non-enhancing FLAIR hyperintensity (3), enhancing tumor (4), and resection cavity (5).

6. A measure of similarity between curves

Figure 7 shows sample generations given unseen anatomical masks. Note that GliomaGen is capable of producing detailed structures, but nonetheless may produce noisy volumes, like in row three of Figure 7. The generated modalities in row three of Figure 7 show visible noise and blurriness, unlike the samples in rows two and three. This presents an opportunity for GliomaGen to be further iterated on such that noisy volumes are less frequently generated.

5.2 Segmentation Benchmarking

Model	Kappa	Accuracy	F1 (NETC)	F1 (FLAIR)	F1 (Enhancing)	F1 (Resection)	Avg Dice	Pearson's r
Real + 1062 Synthetic	0.7958	0.8969	0.7291	0.9403	0.8915	0.9407	0.8116	0.9919
Real + 598 Synthetic (Subset)	0.7913	0.8945	0.7352	0.9394	0.8898	0.9383	0.8085	0.9938
Real Only	0.8071	0.9033	0.7199	0.9447	0.8951	0.9424	0.8167	0.9923
Real + 2124 Synthetic	0.7931	0.8953	0.6998	0.9408	0.8858	0.9363	0.7840	0.9913
Synthetic Subset Only (598)	0.4159	0.6367	0.4457	0.7759	0.7636	0.7024	0.5592	0.9295
Half of Real (617) + (598) Synthetic (Subset)	0.7942	0.8965	0.6962	0.9429	0.8865	0.9316	0.8042	0.9922

Table 3: Segmentation performance of nnU-Net models trained on different data configurations.

The segmentation performance across different training data configurations is recorded in Table 3. Models combining real and synthetic data scored comparably to real-data-only training. Across all models trained on real and synthetic data, Cohen's Kappa values fell in the range of 0.7913 to 0.7957, with accuracy exceeding 0.89. Notably, the model trained using only synthetic data exhibited substantially lower performance (Kappa=0.4159, Accuracy=0.6367), highlighting current limitations when using only synthetic data. Still, note that the nnU-Net trained on half of the real data with 598 high-quality synthetic samples retained performance on par with the model trained using only real data. And, for some metrics, it even surpassed the model trained using the same synthetic subset with all real data in overall accuracy (0.8965) and F1 (FLAIR) (0.9429), demonstrating that the amount of real training data can be significantly reduced by using synthetic data.

Model	**Avg HD95 (mm)**	**Avg ASSD (mm)**
Real + 1062 Synthetic	3.78	1.71
Real + 598 Synthetic (Subset)	4.00	1.83
Real Only	3.35	1.43
Real + 2124 Synthetic	4.01	1.70
Synthetic Subset Only (598)	12.37	5.98
Half of Real (617) + (598) Synthetic (Subset)	4.99	1.75

Table 4: Boundary quality metrics (HD95 and ASSD) across nnU-Net segmentation models trained on different datasets. HD95 (95th percentile Hausdorff Distance) measures worst-case boundary error, while ASSD (Average Symmetric Surface Distance) quantifies average boundary deviation. Lower values indicate better segmentation accuracy.

Boundary quality metrics (Table 4) further elucidate that synthetic-only training may be ineffective, given that training only on the synthetic subset yielded HD95 and ASSD values three to four times worse than those of real and synthetic-data models. Nonetheless, hybrid models maintained HD95 < 4.01 mm and ASSD < 1.83 mm, with the model trained on real data and 1062 synthetic samples achieving the comparable boundary preservation

(HD95=3.78 mm, ASSD=1.71 mm) to real data only. Volumetric correlations remained strong (Pearson's r $>$ 0.99) for hybrid models, while the synthetic subset-only model showed reduced, correlation (r=0.9295). Finally, aligning with the results in Table 3, the model trained using half of the real data and the high-quality synthetic dataset achieved an average ASSD (1.75) on par with the entire real and synthetic dataset subset, while a jump in HD95 (4.99) was observed compared to the model trained on real data and the synthetic subset.

6 Discussion

Implications The open-source presented pipeline for training GliomaGen and generating synthetic datasets via mask augmentation provides a replicable foundation for use in other medical domains beyond brain MR images. The to-be-released subset of the synthetic BraTS dataset will serve as the baseline for further downstream improvements to the GliomaGen framework. Segmentation benchmarking revealed that synthetic data has the potential to substantially shrink dataset size while retaining performance. This suggests that synthetic data generated by the pipeline can be leveraged in resource-limited settings to improve model performance.

Limitations This exploratory study demonstrates both the effectiveness and pitfalls of applying current diffusion-based methods to complex datasets like BraTS 2024 Adult Post-Treatment Glioma. Quantitative results indicate that certain MRI modalities, particularly T2-FLAIR, exhibit elevated noise levels and inconsistent anatomical detail. These inconsistencies may be harmful to segmentation performance, and require further evaluation. Testing also revealed that synthetic-only segmentation models lag behind those trained on a combination of real and synthetic data. Additionally, the testing and revision of methods proposed in this study were constrained by the intense computational costs associated with training diffusion models, suggesting that optimizing efficiency will be important before they're broadly adopted.

7 Conclusion and Future Directions

This work introduced GliomaGen, an anatomically conditioned diffusion model for BraTS 2024 Adult Post-Treatment Glioma. The GliomaGen pipeline will be publicly available on GitHub, and its modularity makes it a valuable pipeline for synthetic data generation in other domains beyond post-treatment glioma. A large BraTS 2024 Adult Post-Treatment Glioma derivative dataset was synthesized, and will be hosted on HuggingFace in the coming weeks. GliomaGen achieved high-fidelity generation for certain modalities, with optimistic quantitative metrics (FID, KID, MS-SSIM), demonstrating the potential for diffusion models in the context of BraTS 2024. The segmentation benchmarking results showed that GliomaGen synthetic data can be used to reduce dataset size while retaining high-quality segmentation performance. Future work on the GliomaGen pipeline will involve refining model architectures and training routines to further reduce inconsistency in generations. Feedback from clinical experts will also be sought to inform future iterations of GliomaGen. And, eventually, the procedure will be used to generate synthetic data in other imaging domains.

8 Acknowledgements

I would like to sincerely thank a few people who were fundamental to this research:

- Dr. Alaa Youssef for her selfless mentorship and guidance throughout the entire project.
- Dr. Jeremiah Brown for insightful preliminary discussions about medicine and project feasibility.
- Neuroradiologist Dr. Ryan Cusic for his advice and mentorship.
- Vast.ai Ren for generous cloud computing resources for model training.

References

Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions — sn computer science. https://link.springer.com/article/10.1007/s42979-021-00815-1. (Accessed on 09/13/2024).

Vast. URL https://vast.ai/. [Online; accessed 2025-02-08].

The effects of data balancing approaches: A case study - sciencedirect. https://www.sciencedirect.com/science/article/pii/S1568494622009024. (Accessed on 09/12/2024).

elijahrenner/gliomagen. https://github.com/elijahrenner/gliomagen. (Accessed on 09/11/2024).

Muhammad Usman Akbar, Måns Larsson, and Anders Eklund. Brain tumor segmentation using synthetic mr images – a comparison of gans and diffusion models, 2024. URL https://arxiv.org/abs/2306.02986.

Christian Bluethgen, Pierre Chambon, Jean-Benoit Delbrouck, Rogier van der Sluijs, Małgorzata Połacin, Juan Manuel Zambrano Chaves, Tanishq Mathew Abraham, Shivanshu Purohit, Curtis P. Langlotz, and Akshay S. Chaudhari. A vision–language foundation model for the generation of realistic chest x-ray images. *Nature Biomedical Engineering*, Aug 2024. ISSN 2157-846X. doi: 10.1038/s41551-024-01246-y. URL https://doi.org/10.1038/s41551-024-01246-y.

Maria Correia de Verdier et al. The 2024 brain tumor segmentation (brats) challenge: Glioma segmentation on post-treatment mri, 2024. URL https://arxiv.org/abs/2405.18368.

Zolnamar Dorjsembe, Hsing-Kuo Pao, Sodtavilan Odonchimed, and Furen Xiao. Conditional diffusion models for semantic 3d brain mri synthesis. *IEEE Journal of Biomedical and Health Informatics*, 28(7):4084–4093, July 2024. ISSN 2168-2208. doi: 10.1109/jbhi.2024.3385504. URL http://dx.doi.org/10.1109/JBHI.2024.3385504.

Willemink et al. Preparing medical imaging data for machine learning — radiology. https://pubs.rsna.org/doi/10.1148/radiol.2020192224. (Accessed on 10/22/2024).

André Ferreira, Naida Solak, Jianning Li, Philipp Dammann, Jens Kleesiek, Victor Alves, and Jan Egger. How we won brats 2023 adult glioma challenge? just faking it! enhanced synthetic data augmentation and model ensemble for brain tumour segmentation, 2024. URL https://arxiv.org/abs/2402.17317.

Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35:18–31, January 2017. ISSN 1361-8415. doi: 10.1016/j.media.2016.05.004. URL http://dx.doi.org/10.1016/j.media.2016.05.004.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020. URL https://arxiv.org/abs/2006.11239.

Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, and Klaus H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018. URL https://arxiv.org/abs/1809.10486.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation, 2018. URL https://arxiv.org/abs/1710.10196.

Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks, 2021. URL https://arxiv.org/abs/2106.12423.

Nicholas Konz, Yuwen Chen, Haoyu Dong, and Maciej A. Mazurowski. Anatomically-controllable medical image generation with segmentation-guided diffusion models, 2024. URL https://arxiv.org/abs/2402.05210.

Jeremias Krause and et al. Deep learning detects genetic alterations in cancer histology generated by adversarial networks - pubmed, May 2021. URL https://pubmed.ncbi.nlm.nih.gov/33565124/. [Online; accessed 2025-01-05].

Jianwei Lin, Jiatai Lin, Cheng Lu, Hao Chen, Huan Lin, Bingchao Zhao, Z. Shi, Bingjiang Qiu, Xipeng Pan, Zeyan Xu, Biao Huang, Changhong Liang, Guoqiang Han, Zaiyi Liu, and Chu Han. Ckd-transbts: Clinical knowledge-driven hybrid transformer with modality-correlated cross-attention for brain tumor segmentation. *IEEE Transactions on Medical Imaging*, PP:1–1, 02 2023. doi: 10.1109/TMI.2023.3250474.

Sven Lüpke, Yousef Yeganeh, Ehsan Adeli, Nassir Navab, and Azade Farshad. Physics-informed latent diffusion for multimodal brain mri synthesis, 2024. URL https://arxiv.org/abs/2409.13532.

Siyuan Mei, Fuxin Fan, Fabian Wagner, Mareike Thies, Mingxuan Gu, Yuliang Sun, and Andreas Maier. Segmentation-guided knee radiograph generation using conditional diffusion models, 2024.

Gustav Müller-Franzes, Jan Moritz Niehues, Firas Khader, Soroosh Tayebi Arasteh, Christoph Haarburger, Christiane Kuhl, Tianci Wang, Tianyu Han, Teresa Nolte, Sven Nebelung, Jakob Nikolas Kather, and Daniel Truhn. A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. *Scientific Reports*, 13(1), July 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-39278-0. URL http://dx.doi.org/10.1038/s41598-023-39278-0.

Walter H. L. Pinaya, Petru-Daniel Tudosiu, Jessica Dafflon, Pedro F da Costa, Virginia Fernandez, Parashkev Nachev, Sebastien Ourselin, and M. Jorge Cardoso. Brain imaging generation with latent diffusion models, 2022. URL https://arxiv.org/abs/2209.07162.

PyTorch. scikit-image: image processing in python [peerj], 10 2014. URL https://peerj.com/articles/453/. [Online; accessed 2025-02-08].

info@sagebase.org Sage Bionetworks. Brain tumor segmentation (brats) challenges - syn53708126 - wiki. URL https://www.synapse.org/Synapse:syn53708126/wiki/626320. [Online; accessed 2025-01-05].

Baoqin Song, Xiu Wang, Lijing Qin, Shehbaz Hussain, and Wanjun Liang. Brain gliomas: Diagnostic and therapeutic issues and the prospects of drug-targeted nano-delivery technology. *Pharmacological Research*, 206:107308, 2024. ISSN 1043-6618. doi: https://doi.org/10.1016/j.phrs.2024.107308. URL https://www.sciencedirect.com/science/article/pii/S1043661824002536.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014. URL https://arxiv.org/abs/1409.4842.

Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, 2003. doi: 10.1109/ACSSC.2003.1292216.

Qiantong Xu, Gao Huang, Yang Yuan, Chuan Guo, Yu Sun, Felix Wu, and Kilian Weinberger. An empirical study on evaluation metrics of generative adversarial networks, 2018. URL https://arxiv.org/abs/1806.07755.

Appendix A. Intuition

Reverse diffusion process formula

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\tilde{x}_t, t) \right) + \sigma_t z$$

The term $x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\tilde{x}_t, t)$ removes the predicted noise from the current noisy image, which is then scaled by the factor $\frac{1}{\sqrt{\alpha_t}}$ to rescale the denoised image for consistency in the reverse process. Finally, $\sigma_t z$ is added as randomness to maintain diversity in generated images. $\sigma_t = \sqrt{\beta_t}$ where β_t is the stepwise noise variance derived from $\bar{\alpha}_t$, the total noise up to step t. β_t then functions as a scaling factor of additional noise added at each step.