Toward a minimum viable MRI protocol for glioma segmentation in resource-constrained settings

Izak S. Pretorius[1, 3] [0009-0003-0200-0602], Thandiwe Zhanje[2,3] [0009-0008-4530-037X], Udunna Anazodo[3,4] [0000-0001-8864-035X],Aondonna Iorumbur[3,5] [0009-0006-5621-0981], Confidence Raymond[3,4,5,6] [0000-0003-3927-9697], and Willem P.E. Boonzaier[1,3] [0009-0007-7216-6543]

[1] Department of Medical Physics, University of the Free State, Bloemfontein, South Africa
[2] Division of Biomedical Engineering, Department of Human Biology, University of Cape Town, Cape Town, South Africa
[3] SPRINT AI Training for medical image translation (SPARK) Program, Montreal, Canada
[4] Montreal Neurological Institute, McGill University, Montreal, Canada
[5] Medical Artificial Intelligence Laboratory, Lagos Nigeria
[6] Department of Biomedical Engineering, McGill University, Montreal Canada

Stefanpretorius6@gmail.com

Abstract. Multiparametric MRI is the clinical standard for glioma assessment, yet complete four-sequence protocols are rarely guaranteed in Sub-Saharan African settings, where conventionally trained segmentation models degrade sharply when sequences are missing. We investigate the minimum set of MRI sequences required for reliable automated glioma segmentation under these constraints. Using the BraTS-Africa 2023 dataset and an nnU-Net v2 residual-encoder backbone, we train a baseline and five probabilistic modality-dropout variants and evaluate all configurations across nine modality combinations. Independent per-modality dropout ($p = 0.25$) was the most robust. It reduced boundary error by ~45% relative to the baseline under full input and retained a mean Dice of 0.83 across all combinations, where the baseline collapsed. Guided by these results, we identify two sequences (T1Gd + T2-FLAIR) as a minimum viable protocol for automated segmentation. A dedicated reduced-modality model and the dropout model retained mean Dice of 0.901 and 0.894, respectively, on this two-sequence input, versus 0.560 for the baseline. A single dropout-trained model thus delivers near-specialist accuracy on a reduced imaging protocol while remaining resilient to missing sequences for glioma segmentation in resource-constrained clinics.

Keywords: Glioma segmentation, MRI, Missing modality, Modality dropout, nnU-Net, BraTS-Africa, Resource-constrained imaging

1 Introduction

Gliomas are the most common primary malignant brain tumours and are associated with substantial morbidity and mortality worldwide. While mortality has decreased in high-income settings due to improved diagnostics and treatment, the same cannot be

seen in low- middle-income countries (LMICs), such as in sub-Saharan Africa (SSA) where mortality has increased by approximately 25%[1].

Despite limited availability in resource-constrained settings, magnetic resonance imaging (MRI) remains the clinical standard for glioma evaluation [2]. Multiparamet-ric MRI (mpMRI) protocols for glioma imaging typically consist of T1-weighted (T1), post-gadolinium T1-weighted (T1Gd), T2-weighted (T2), and T2-fluid-attenuated inversion recovery (T2-FLAIR) sequences, which provide complementary information on tumour anatomy and biology [3].

Accurate delineation of glioma subregions on MRI is essential for treatment planning, surgical decision-making, and follow-up but can be severely labour-intensive in resource-constrained settings. Recent advances in deep learning have substantially improved automated glioma segmentation performance and provided a robust and consistent method for delineation. Benchmark initiatives such as the Brain Tumour Segmentation (BraTS) challenges have enabled the development of high-performing models for delineating enhancing tumour (ET), non-enhancing tumour core (NETC), and surrounding FLAIR hyperintensity (SNFH) from mpMRI data. Although much progress has been made, specifically democratizing these challenges to the African context with the BraTS-Africa challenge and open access dataset, most winning models still utilized classic encoder-decoder networks, relying heavily on multi-channel inputs from mpMRI to process complementary soft-tissue details simultane-ously [4].

While this assumption of complete mpMRI datasets holds in curated research datasets, in resource-limited clinical environments, mpMRI is often unrealistic due to protocol variability, limited scanner availability [5], restricted access to gadolinium-based contrast agents, shortened scan times, and incomplete data archiving practices [2, 6]. Consequently, when deployed in clinical environments where these ideal, complete inputs cannot be guaranteed, these rigid architectural constraints frequently lead to degradation in segmentation performance [6]. Recent work on missing-modality learning has explored approaches such as modality dropout, feature fusion, and modality imputation [7]. To bridge this gap, we adopt modality dropout for gli-oma segmentation specifically in resource-constrained clinical environments.

2 Previous work

A wide range of deep learning architectures and training strategies have been benchmarked for automated brain tumour segmentation. A meta-analysis by van Kempen et al. (2021) reported a pooled Dice Similarity Coefficient (DSC) of 0.84 (95% CI: 0.82–0.86) across 10 studies, highlighting the promise of these methods for both low-grade and high-grade glioma segmentation [8].

Modality dropout - the process of randomly masking specific imaging channels during training - is the foundational technique used to simulate incomplete data streams and has been used to mitigate the vulnerability of automated segmentation systems to missing structural sequences at inference. Originally introduced as the "ModDrop" training strategy by Neverova et al. (2016) in the context of adaptive multi-modal fusion, this technique was designed to learn cross-modality correlations while strictly

prohibiting false co-adaptations between diverse data representations [9]. In the context of brain tumour segmentation, frameworks like Multimodal Masked Autoencoder (M3AE) and Semantic-guided Masked Mutual Learning (SMML) have used modality dropout which has further improved segmentation ro-bustness and performance [10, 11].

3 Methodology

3.1 Dataset Description

The BraTS-Africa[1] 2023 dataset provides 146 mpMRI cases of patients diagnosed with central nervous system neoplasms acquired across 6 imaging centres in Nigeria, representing a population from Sub-Saharan Africa (SSA). Of the 146 cases, 95 are presumed adult diffuse glioma [12]. Each case contains all four co-registered, skull-stripped MRI sequence volumes (T1, T1Gd, T2 and T2-FLAIR) at 1mm³ isotropic resolution acquired at 1.5T using 5 different scanners from 4 different vendors. Each institution used a different imaging protocol creating variability expected from standard clinical acquisitions. Expert annotations delineate three tumour substructures: ET, NETC, and SNFH.

3.2 Deep Learning Framework

nnU-Net version 2, a self-configuring framework for biomedical image segmentation, was used for all preprocessing, training, and inference staging [13]. All models employed the 3D full-resolution configuration with a medium residual encoder architecture. For all other hyperparameters, the nnU-Netv2 defaults were used. All models were trained on a single NVIDIA RTX 5080 (16 GB) GPU.

Inference hardware and runtime. All inference was performed on a single NVIDIA GeForce RTX 5080 (16 GB) GPU, with an AMD Ryzen 9 9950X (16 cores / 32 threads) available for preprocessing and CPU-fallback benchmarking. The baseline model (383 M parameters) required ~1.7 GB peak GPU memory at inference. A single forward pass averaged 0.041 s on the GPU (24.5 it/s) versus 6.90 s on CPU (0.14 it/s), a ~169× speedup, confirming that the models run in near real-time on a single consumer GPU and remain usable, if substantially slower, on CPU-only hardware where no accelerator is available.

3.3 Model Evaluation and Statistical Analysis

Models were evaluated using the 95th percentile Hausdorff distance (HD95) and DSC achieved in modality-dropped-out inference experiments summarised in Table 1. Statistical significance testing was performed using a Wilcoxon signed-rank test, with statistical significance indicated when the p-value is smaller than 0.05. The model was

[1] https://www.cancerimagingarchive.net/collection/brats-africa/

trained using the nnU-Net framework's 5-fold cross-validation split, utilising only fold 0 for training, with its corresponding validation subset serving as the validation dataset.

3.4 Experimental Phase Design

The study comprised two phases intended to reflect the modality-limited scenarios encountered in Sub-Saharan African clinics. Phase 1 focused on developing a model to handle missing modalities, generalising across the modality combinations (Table 1), while Phase 2 examined the minimum imaging requirements for accurate glioma segmentation.

3.5 Phase 1: Probabilistic Modality Dropout Training

During training, a modality dropout strategy was employed as a form of data augmentation. The dropout occurred in two stages: a dropout gate decided whether dropout should occur, and the modality dropout function removed available MRI sequences, as shown in Figure 1. A baseline model was trained as a control (experiment 1), consisting of the default nnU-Net architecture containing no dropout. The other 8 experiments described in Table 1 were chosen to represent cases where a single modality was dropped (experiments 2-5), cases where local clinical protocols and restrictions motivated decision-making (experiment 6 &7), including the most clinically used modalities ie. T1Gd and T2-FLAIR and a case where no contrast agent is available and finally, experiments 8 and 9 explore extreme variants where the most clinical used modalities are explored as single modality imaging protocol. Five experimental variants of modality dropout, as summarized in Table 2, were studied for each of the 9 experimental in Table 1. Dropped modalities were replaced with images containing all zero-pixel values and dimensions matching the other modalities.

Table 1. Inference experiment modality combinations

Experiment	T1	T2	T1Gd	T2-FLAIR
1	✓	✓	✓	✓
2	✓		✓	✓
3	✓	✓		✓
4		✓	✓	✓
5	✓	✓	✓	
6			✓	✓
7		✓		✓
8			✓	
9				✓

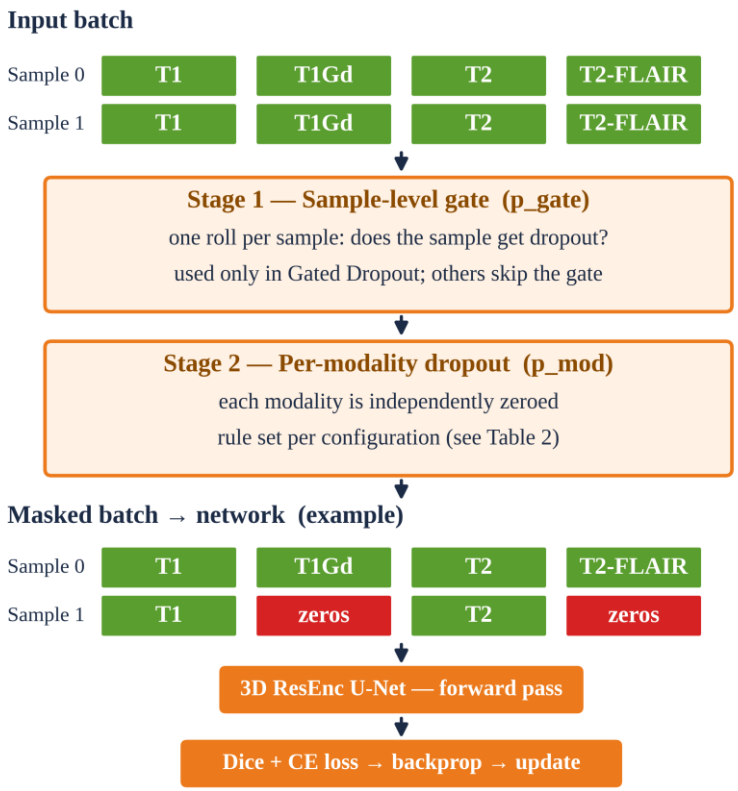

Fig. 1. Illustration of modality dropout applied at the network input stage.

3.6 Phase 2: Minimum Modality Training

A dedicated reduced-modality model was trained based on the best-performing in phase 1 models with the least number of modalities as input. Three models were compared: the baseline model from phase 1, the best generalising model from phase 1, and the dedicated reduced-modality model created in phase 2.

Table 2. Model Configuration

Model	Gate Dropout	Modality Dropout
Uniform Dropout	Always On	Equal Probability Dropout
Aggressive Dropout	Always On	1 modality always dropped
Always Dropout (p=0.25)	Always On	$p_{mod} = 0.25$
Always Dropout (p=0.5)	Always On	$p_{mod} = 0.50$
Gate dropout	$p_{gate} = 0.60$	$p_{mod} = 0.25$

4 Results

4.1 Phase 1: Probabilistic Modality Dropout Training

Under full four-modality input – experiment 1 (Figure 2), all six models performed comparably, with mean Dice scores of 0.89–0.91 across WT, TC and ET (Figure 2a). The full-modality baseline reached Dice 0.89 and HD95 12.74 mm. Boundary accuracy separated the models more clearly (Figure 2b): Always Dropout (p_{mod} = 0.25) gave the lowest mean HD95 (7.03 mm, SD 12.52 mm), versus 9.75–10.89 mm for the more aggressive dropout variants. Gate Dropout matched the highest mean Dice (0.91), statistically indistinguishable from the baseline (Wilcoxon signed-rank, p = 0.18). Its mean HD95 (9.81 mm, SD 19.64 mm) was significantly lower than the baseline's (p = 0.0007, r = 0.65) and statistically comparable to Always Dropout (p_{mod} = 0.25; 7.03 mm; p = 0.22).

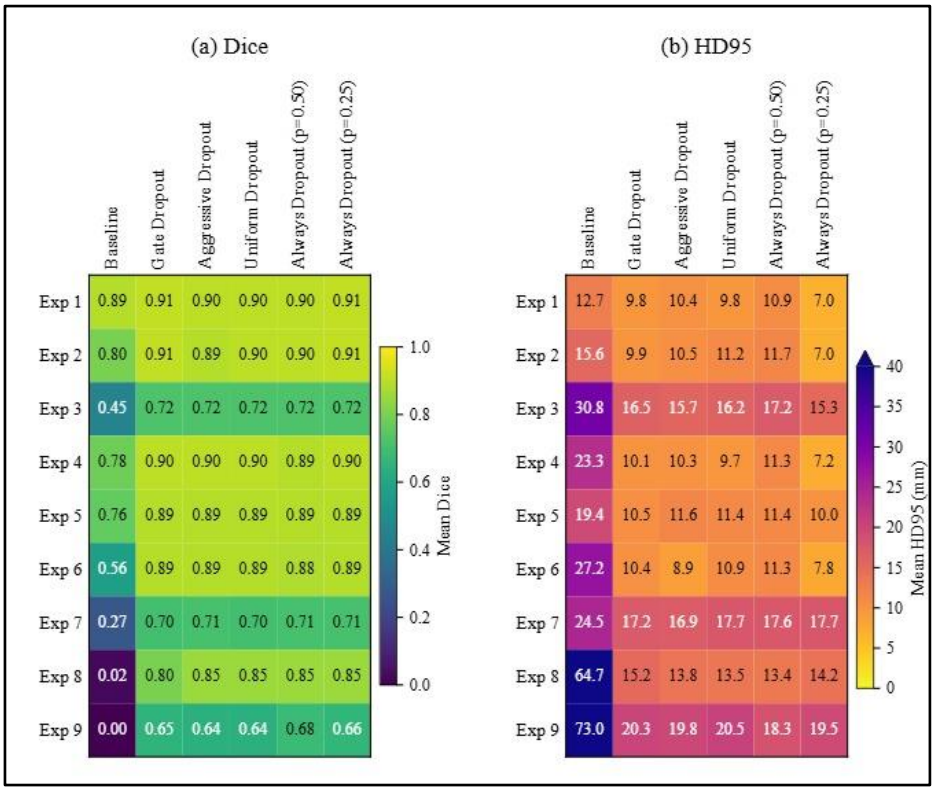

Figure 2. Mean Dice (a) and HD95 (b) per modality experiment (rows; see Table 1) for all six models. Brighter = better in both panels. Exp = experiment.

Across all nine inference combinations (Table 1), the baseline degraded as expected under constrained input, while every dropout configuration held a mean Dice of 0.82–0.83. Always Dropout (p_{mod} = 0.25) was strongest overall, with the highest mean Dice

(0.83), the lowest mean HD95 (11.68 mm) and the smallest boundary-error variance (SD 18.38 mm).

4.2 Phase 2: Minimum Modality Training

Under complete four-sequence input, the three models performed comparably, with mean Dice scores of 0.895 (Baseline), 0.895 (Minimum-Modality) and 0.908 (Always Dropout) (Figure 3). When input was restricted to T1Gd + T2-FLAIR, the baseline collapsed to mean Dice of 0.560 and HD95 of 27.24 mm, whereas the Minimum-Modality and Always Dropout models held mean Dice 0.901 and 0.894 at HD95 = 7.70 and 7.75 mm, respectively.

Restricting the Always Dropout model to T1Gd + T2-FLAIR cost only ~0.02 mean Dice (0.91→0.89) and ~0.8 mm HD95 (7.0→7.8 mm) relative to full input, within inter-rater annotation variability and unlikely to be clinically meaningful. Dropout did not significantly affect Dice versus baseline (all $p > 0.05$), but Always Dropout ($p_{mod} = 0.25$) and Gate Dropout significantly reduced HD95 ($p = 0.002$ and 0.004; rank-biserial −0.70 and −0.65, Holm-corrected). Full and reduced-modality predictions are compared against ground truth in Figure 4.

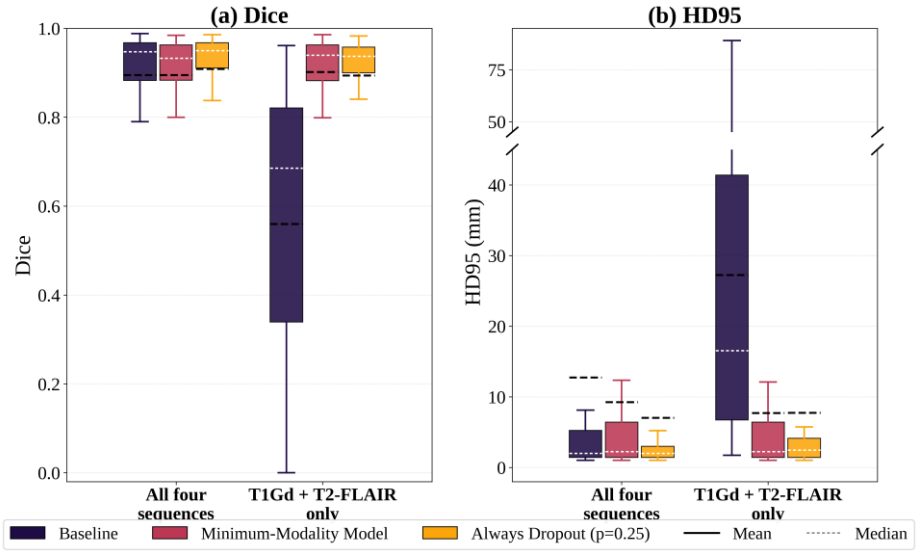

Figure 3. Phase 2 Dice (a) and HD95 (b) for the three models with all four sequences versus T1Gd + T2-FLAIR only, pooled over WT/TC/ET. Mean = solid line (annotated), median = dashed line. The Baseline collapses under the reduced protocol while the Minimum-Modality and Always Dropout models do not.

5 Discussion

These results indicate that moderate, consistent per-modality dropout improves generalisation even when complete input is available: Always Dropout (p_{mod} = 0.25) reduced boundary error by roughly 45% relative to the full-modality baseline without compromising Dice, while Gate Dropout's higher HD95 at equal Dice points to less precise spatial delineation despite comparable volumetric overlap. The same configuration proved most robust under the modality scarcity typical of the target clinical setting, motivating its adoption as the p = 0.25 strategy carried into Phase 2. This contrast was sharper still in the reduced-modality setting. Conventionally trained models stayed tied to the modality combination seen during training, whereas probabilistic modality dropout yielded a single generalist model with robust cross-sequence representations. That model reached specialist-level performance on the reduced T1Gd + T2-FLAIR protocol while remaining resilient across modality combinations. Together, the two phases show that p_{mod} = 0.25 dropout delivers a model both competitive under ideal acquisition and reliable under the minimum viable protocol for automated segmentation. These findings are bounded by the single-institution scale of the BraTS-Africa cohort and the absence of external multi-site validation, so the minimum viable protocol for automated segmentation`s generalisation to other scanners and populations remains to be confirmed.

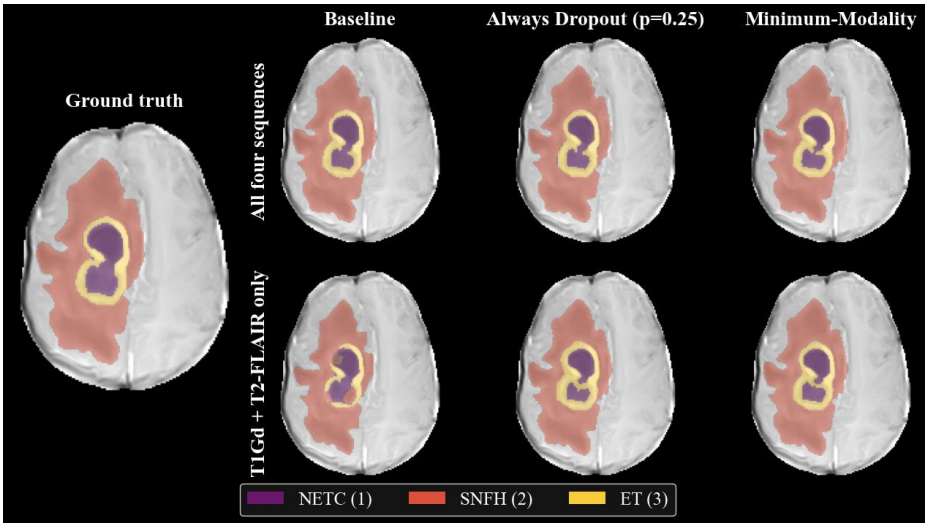

Figure 4. Qualitative comparison of brain tumour segmentation outputs on a representative BraTS validation case. Comparisons were made in reference to the ground truth annotations by experts based on access and review of the four sequences.

Limitations and future work. Although this work presents a minimum viable protocol for automated segmentation, this cannot empirically be extrapolated to other clinical processes like diagnosis or treatment planning, and the other modalities might also

be required for those processes. One of the clinical problems in low-resource settings remains the availability and the cost of contrast agents. Although this work explores three experimental cases where the T1Gd modality is dropped, the studied technique shows poor performance when T1Gd is absent. Future work could focus on evaluating the feasibility of synthesizing a contrast-enhanced image from a native T1 image to account for this gap in research. Finally, this work explores only validation fold-based verification. Independent validation on independent test data should be performed as part of future research before clinical implementation.

6 Impact in resource-constrained settings (RCS)

In many LMICs, comprehensive mpMRI is limited by cost, scanner availability, acquisition time, and inconsistent workflows. Our results suggest that a minimal T1Gd + T2-FLAIR protocol may be sufficient for reliable automated glioma segmentation in these settings, reducing scan time, protocol variability, and the rate of non-diagnostic or incomplete studies, ultimately improving access. Because robustness here comes from the training strategy rather than any architectural change, the approach adds no inference-time cost and can be deployed on the same hardware as a conventional model. A single dropout-trained model also removes the need to acquire, store, and maintain separate sequence-specific models, and degrades gracefully when a sequence is missing, corrupted, or unusable rather than failing outright. Together, these properties lower the barrier to clinically useful glioma automatic segmentation where full mpMRI is impractical or not clinically needed.

Acknowledgments. This work was part of the Sprint AI Training for African Medical Imaging Knowledge Translation (SPARK) Academy 2026 summer school on deep learning in medical imaging. The authors would like to thank the instructors of the summer for providing insightful background knowledge on brain tumours that informed the research presented here, most notably: Maruf Adewole, Mohannad Barakat, Craig Jones, Noha Magdy, Amal Saleh, Aondona Iorumbur, Bernes Atabonfack, Confidence Raymond, Craig Jones, Ilerioluwakiiye Abolade, Jeremiah Fadugba, Juampablo Heras Rivera, Kenneth Aguh, Lukman E. Ismaila, Maruf Adewole, Mohtady Barakat, Nourou Dine Bankole, Saikat Roy, Toufiq Musah, Willem Boonzaier.. The authors acknowledge the funding support provided to SPARK through the Lacuna Fund for Health and Equity (PI: Udunna Anazodo), the RSNA R & E Foundation (PI: Farouk Dako), the University of Washington Population Health Initiative Tier2 Grant (PI: Mehmet Kurt), McGill University Healthy Brain and Healthy Lives (HBHL, Anazodo), and the National Science and Engineering Research Council of Canada (NSERC) Discovery Launch Supplement (PI:UdunnaAnazodo, DGECR-2022-00136).

Artificial Intelligence Declaration. Artificial intelligence tools were used for language editing, proofreading and graphical editing of the manuscript. All study methodology, experiments, data analysis, interpretation, and scientific conclusions were performed by the authors.

7 References

1. Patel, A.P., Fisher, J.L., Nichols, E., Abd-Allah, F., Abdela, J., Abdelalim, A., Abraha, H.N., Agius, D., Alahdab, F., Alam, T.: Global, regional, and national burden of brain and other CNS cancer, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 18, 376–393 (2019).
2. Anazodo, U.C., Ng, J.J., Ehiogu, B., Obungoloch, J., Fatade, A., Mutsaerts, H., Secca, M.F., Diop, M., Opadele, A., Alexander, D.C., Dada, M.O., Ogbole, G., Nunes, R., Figueiredo, P., Figini, M., Aribisala, B., Awojoyogbe, B.O., Aduluwa, H., Sprenger, C., Wagner, R., Olakunle, A., Romeo, D., Sun, Y., Fezeu, F., Orunmuyi, A.T., Geethanath, S., Gulani, V., Nganga, E.C., Adeleke, S., Ntobeuko, N., Minja, F.J., Webb, A.G., Asllani, I., Dako, F.: A framework for advancing sustainable magnetic resonance imaging access in Africa. NMR Biomed. 36, e4846 (2023).
3. Upadhyay, N., Waldman, A.: Conventional MRI evaluation of gliomas. Br. J. Radiol. 84, S107–S111 (2011).
4. Adewole, M., Rudie, J.D., Gbadamosi, A., Zhang, D., Raymond, C., Toyobo, O., Omidiji, O., Akinola, R., Suwaid, M.A., Daji, F., Emegoakor, A., Aguh, K., Ojo, N., Kalaiwo, C., Babatunde, G., Ogunleye, A., Gbadamosi, Y., Iorpagher, K., Onuwaje, M., Betiku, B., Saluja, R., LaBella, D., Calabrese, E., Baid, U., Bakas, S., Menze, B., Fatade, A., Dako, F., Anazodo, U.C., Verdier, M.C. De, Raymond, H., Velichko, Y., Aboian, M., Moawad, A., Farahani, K., Albrecht, J., Chung, V., Wiestler, B., Linguraru, M.G., Kazerooni, A.F., Conte, G.M., Li, H., Iglesias, J.E., Anwar, S.M., Kofler, F., Aristizabal, A., Kassem, H., Karargyris, A.: Brain tumor segmentation in Sub-Saharan Africa patient population: The BraTS-Africa challenge. Neurooncol. Adv. 8, (2026). https://doi.org/10.1093/noajnl/vdag082.
5. Gervais, E.-M.: A review of MRI studies in Africa with special focus on quantitative MRI: Historical development, current status and the role of medical physicists. Physica Medica.
6. Rasool, N., Bhat, J.I.: A critical review on segmentation of glioma brain tumor and prediction of overall survival. Archives of Computational Methods in Engineering. 32, 1525–1569 (2025).
7. Azad, R., Dehghanmanshadi, M., Khosravi, N., Cohen-Adad, J., Merhof, D.: Addressing missing modality challenges in MRI images: A comprehensive review. Comput. Vis. Media (Beijing). 11, 241–268 (2025).
8. van Kempen, E.J., Post, M., Mannil, M., Witkam, R.L., Ter Laan, M., Patel, A., Meijer, F.J., Henssen, D.: Performance of machine learning algorithms for glioma segmentation of brain MRI: a systematic literature review and meta-analysis. Eur. Radiol. 31, 9638–9653 (2021).
9. Neverova, N., Wolf, C., Taylor, G., Nebout, F.: Moddrop: adaptive multi-modal gesture recognition. IEEE Trans. Pattern Anal. Mach. Intell. 38, 1692–1706 (2015).
10. Liu, H., Wei, D., Lu, D., Sun, J., Wang, L., Zheng, Y.: M3AE: multimodal representation learning for brain tumor segmentation with missing modalities. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 1657–1665 (2023).
11. Liang, G., Zhou, Q., Wang, Z., Chen, J., Gu, L., Yao, C., Wu, S., Huang, B., Chen, K.: Semantic-guided masked mutual learning for multi-modal brain tumor segmentation

with arbitrary missing modalities. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 5137–5145 (2025).

12. Adewole, M., Rudie, J.D., Gbadamosi, A., Zhang, D., Raymond, C., Ajigbotoshso, J., Toyobo, O., Aguh, K., Omidiji, O., Akinola, R., Suwaid, M.A., Emegoakor, A., Ojo, N., Kalaiwo, C., Babatunde, G., Ogunleye, A., Gbadamosi, Y., Iorpagher, K., Onuwaje, M., Betiku, B., Cakmak, J., Menze, B., Baid, U., Bakas, S., Dako, F., Fatade, A., Anazodo, U.C.: The BraTS-Africa Dataset: Expanding the Brain Tumor Segmentation Data to Capture African Populations. Radiol. Artif. Intell. 7, (2025). https://doi.org/10.1148/ryai.240528.
13. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods. 18, 203–211 (2021).