Characterizing Domain Shift in Glioma Segmentation Between Global North and African MRI Datasets

Abraham Ajeolu Oluwasegun[1(✉)], Amy Cheruto[2], Mwansa Kawesha[3], Sammy Wanjohi Kiboi[4], Oyetayo Oyebisi[1], Dishan Otieno[5], Aondona Moses Iorumbur[6], Raymond Confidence[7,8], Udunna Anazodo[7,8], and Kerol Djoumessi[9(✉)]

[1] Pan African University Institute for Basic Sciences, Technology and Innovation, Kenya
ajeoluo@pau-au.africa,
[2] RFH Healthcare, Nairobi, Kenya
[3] Cancer Diseases Hospital, The University Teaching Hospitals, Zambia
[4] Kenyatta University Teaching, Referral and Research Hospital, Kenya
[5] ARTOG Center for Biomedical Engineering Research, Bern, Switzerland
[6] College of Medicine, Department of Medical Physics, University of Lagos, Nigeria
[7] Department of Biomedical Engineering, McGill University, Montreal, Canada
[8] Montreal Neurological Institute, McGill University, Montreal, Canada
[9] Hertie Institute for AI in Brain Health, University of Tübingen, Germany
kerol.djoumessi-donteu@uni-tuebingen.de

Abstract. Glioma is the most common primary brain tumour, yet patients in Sub-Saharan Africa face survival outcomes far worse than those in high-income settings, compounded by a severe shortage of neuro-oncology expertise and imaging infrastructure. While deep learning models trained on Global North datasets achieve state-of-the-art performance, their transferability to African clinical acquisitions remains unquantified. In this study, we benchmark three representative architectures—3D U-Net, DynUNet, and Swin UNETR—trained on BraTS 2021 and evaluated zero-shot on BraTS-Africa ($n = 146$). All three architectures achieve in-domain mean Dice scores of 81.2–83.1%, with performance drops (ΔDice) of 13.8–15.0 percentage point (pp) on BraTS-Africa (DynUNet zero-shot mean Dice: 69.0%, 95% CI: [67.8%, 70.2%]). Stratifying by pathology isolates disease shift from acquisition shift: glioma-only subjects ($n = 95$), achieve 69.9% mean Dice (95% CI: [68.6%, 71.3%]), reducing the domain gap to 10.3 pp, while non-glioma cases ($n = 51$) drop to 63.7% (95% CI: [61.5%, 65.9%]). We further introduce a convention-aware label-harmonization transform to resolve silent channel erasure errors. These results provide an unassisted lower bound for direct deployment, motivating lightweight domain adaptation strategies in resource-constrained settings.

Keywords: Glioma segmentation · Domain generalization · BraTS-Africa · 3D U-Net · Swin UNETR · nnU-Net · Resource-constrained settings

1 Introduction

Gliomas are the most prevalent and lethal primary brain tumours worldwide, with fewer than 20% of patients surviving beyond two years of diagnosis [29], and this burden is distributed highly inequitably: glioma mortality fell by up to 30% in the Global North between 1990–2016 while rising by an estimated 25% in Sub-Saharan Africa (SSA) over the same period [2]. This divergence is driven by health-system factors rather than tumour biology: Ghana and Nigeria have roughly one neurosurgeon per 2.4 million people [30], fewer than half of African countries operate functional radiotherapy services [19], histopathological confirmation is absent in up to 40% of reported Nigerian glioma cases [19], and MRI/neuropathology services are unevenly distributed across the continent [22,19], even as new SSA cancer cases are projected to rise 70% by 2030 [30].

Automated MRI segmentation is one of the few interventions that scales to this gap, with tumour-burden pipelines shown to match or exceed human readers in large multi-centre studies [2]. Progress here has been driven for over a decade by the Brain Tumor Segmentation (BraTS) challenge [20], which by 2021 comprised 2,040 studies drawn almost exclusively from North American and Western European institutions [1]. The applicability of BraTS-trained models to African populations was, until recently, simply untested [2,1] due to the lack of available African dataset. Futhermore, Brain MRI acquired in SSA differs systematically (1.5T scanners, heterogeneous protocols, more advanced disease at presentation [1,22]), creating a domain gap that can silently degrade models that appear state-of-the-art on Western benchmarks. To address the lack of african cohort, the BraTS-Africa dataset [1] was released. Including African data in training has already been shown to help [1], but the zero-shot domain gap of standard architectures has not been systematically characterised across multiple network families under a controlled protocol.

This paper presents a controlled, reproducible cross-domain benchmark establishing, for the first time under a multi-architecture protocol, the zero-shot performance drop when standard glioma segmentation models trained on Global-North data are applied to African clinical scans, and discusses the implications for designing lightweight, uncertainty-aware models suited to resource-constrained deployment. Our contributions are: (1) a controlled, reproducible cross-domain benchmark of three architecturally distinct networks, trained exclusively on BraTS 2021 and evaluated both in-domain and zero-shot on BraTS-Africa; (2) an explicit, convention-aware label-harmonisation layer unifying the two datasets' and (3) quantified evidence of a consistent 13.8–15.0 pp mean-Dice domain gap across all three architectures, with TC and ET most affected.

2 Previous Work

Deep learning for brain tumour segmentation. Brain tumour segmentation has progressed from handcrafted features to fully convolutional models with U-Net [26] and its 3D extension [7], produced a step change in performance [18], later extended with attention gates [23] and residual connections [12]. Self-configuring frameworks such as nnU-Net and transformer-based architectures

such as Swin UNETR [11]] have since pushed performance further, establishing the current state of the art on BraTS benchmarks. Across this work, a persistent weakness is generalisation: models strong on their training distribution are known to degrade under different scanners, protocols, or populations [18,4].

Domain shift, adaptation, generalization, and uncertainty. Domain shift is widely recognised as the principal barrier to clinical translation of medical imaging models [10,24]. Adaptation methods reduce source/target mismatch using target-domain data, via adversarial domain-invariant features or GAN-based translation [10], while federated learning avoids centralising sensitive multi-institutional data [6], relevant where SSA institutions are wary of data-sharing constraints. Domain *generalization* instead trains for unseen domains *without* target-domain data, via randomisation, augmentation, meta-learning, or domain-agnostic normalisation [24]; both require transparent acquisition reporting and external validation across distinct cohorts [24], the standard this benchmark meets by separating in-domain and zero-shot evaluation. Reliable RCS deployment further requires models to communicate *when* they are unreliable: aleatoric (data) and epistemic (model) uncertainty are both relevant to glioma segmentation [31], and voxel-wise uncertainty maps can flag low-confidence regions for review, though estimates require validation under controlled distribution shift before clinical trust is warranted [31,13].

The African data gap. Before BraTS-Africa [1], no glioma segmentation dataset was publicly available from an SSA population, so generalizability claims to African scans were untested. Volumetric tumour-burden estimates, informing radiation planning, surgical decisions, and response monitoring, are clinically important yet rarely a primary evaluation metric [25], motivating our use of Dice *and* boundary-sensitive HD95 (Sec. 3.5). The literature thus offers mature segmentation backbones, a developed domain-shift toolkit, and a case for uncertainty-aware evaluation, but no controlled benchmark has compared multiple standard 3D architectures, trained identically on Global-North data and evaluated zero-shot on BraTS-Africa under one protocol.

Table 1. Dataset summary. Only the enhancing-tumour (ET) integer label differs between cohorts.

	BraTS 2021 Task 1 [3]	**BraTS-Africa** [1]
Subjects	1,251	146 (95 glioma, 51 other CNS)
Acquisition sites	North America / Europe	Multi-site, Nigeria
Field strength	1.5T–3T	1.5T
Modalities	T1, T1ce, T2, FLAIR	T1, T1ce, T2, FLAIR
ET integer label	4	3

3 Methodology

3.1 Datasets, label harmonisation, and splitting

Experiments were conducted using the BraTS 2021 task 1 (mixed 1.5T/3T) and the BraTS-Africa dataset, (1.5T;table 1) summarises the two cohorts. Both pro-

vide four co-registered, skull-stripped MRI sequences (T1, T1ce, T2, FLAIR) and voxel-wise labels for necrotic core (NCR, 1), oedema (ED, 2), and GD-enhancing tumour (ET), but BraTS 2021 [3] encodes ET as label **4** while BraTS-Africa [1] uses label **3**. This Project trains exclusively on BraTS 2021, split 70/15/15 into train/validation/test. BraTS-Africa is reserved in full as an independent zero-shot evaluation cohort.

3.2 Label harmonisation

Both are mapped to the three standard, overlapping BraTS subregions, whole tumour (WT = NCR ∪ ED ∪ ET), tumour core (TC = NCR ∪ ET), and ET, via a custom MONAI `MapTransform` (`ConvertBraTSLabelsd`) that reads a per-subject `label_convention` field and applies the correct ET label accordingly, guarding against a documented failure mode of MONAI's built-in converter, which hard-codes the BraTS 2021 convention and silently zeroes the ET channel for unmodified BraTS-Africa subjects.

3.3 Preprocessing and augmentation

In the two datasets both datasets pass through an identical MONAI pipeline (Fig. 1): reorientation to RAS, resampling to isotropic $1\,\mathrm{mm}^3$ spacing, intensity clipping at [0.5, 99.5] pct, per-channel z-score normalisation ($I' = \frac{I-\mu}{\sigma}$), and foreground-aware crop/pad to 128^3, the standard nnU-Net default for BraTS [15]. Per-channel Z-score normalization standardizes contrast scales between the mixed 1.5T/3T training set and 1.5T test scans. Bias-field correction is omitted, already applied upstream by both providers. For training only, we apply foreground-biased random cropping ($p=0.80$ centred on tumour voxels, which occupy 1–4% of brain volume), random flips, mild rotation ($\pm 15°$, $p=0.20$), Gaussian noise, and random intensity scale/shift – the latter simulating scanner-to-scanner contrast variation relevant to cross-domain evaluation. Validation/test transforms omit all stochastic augmentation.

3.4 Network architectures

Three architectures, spanning the dominant convolutional and transformer-based design families in current BraTS literature, are compared under an identical protocol (Figs. 2–4). All three take the same 4×128^3 input and produce a 3×128^3 multi-label *sigmoid* output (one channel per subregion), rather than softmax, since WT/TC/ET are nested and not mutually exclusive.

3D U-Net [7] (Fig. 2) is a five-level residual encoder–decoder, channels $(32, 64, 128, 256, 512)$, stride-2 down-/up-sampling, two residual units [12] per level, PReLU, batch norm, dropout p=0.10.

DynUNet [15] (Fig. 3) follows the nnU-Net template: identical channel progression with residual blocks and *deep supervision*, via auxiliary 1×1×1 heads at the two decoder stages nearest the output (k=1, 2), supervised against downsampled ground truth and combined with the main output via weights w_k=2^{-k}.

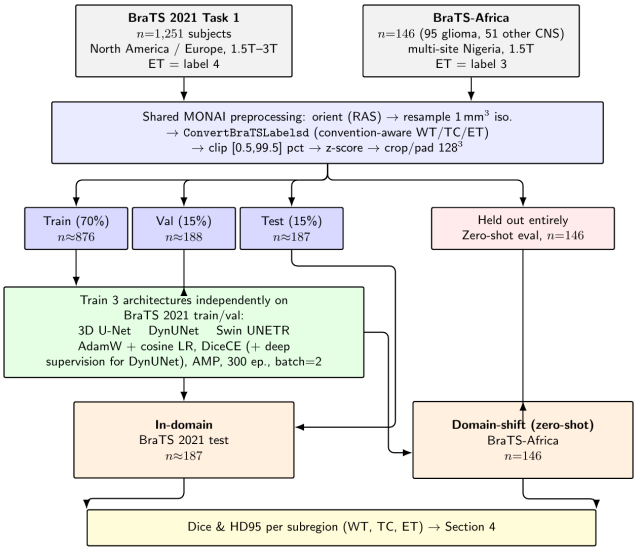

Fig. 1. End-to-end Phase 1 pipeline: identical, label-convention-aware preprocessing for both datasets; BraTS 2021 train/val/test (70/15/15); BraTS-Africa withheld in full as zero-shot domain-shift evaluation.

Swin UNETR [11] (Fig. 4) replaces the convolutional encoder with a hierarchical Swin Transformer: 2^3 patch embedding (48 ch.), four windowed/shifted-window self-attention stages (window 7^3) with channel-doubling patch merging, and a convolutional decoder with multi-stage skips. Gradient checkpointing keeps training within a single GPU's VRAM budget at batch size 2.

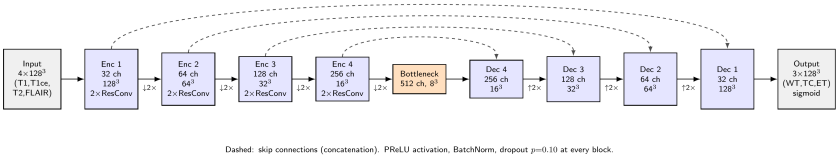

Fig. 2. 3D U-Net (MONAI `UNet`). Five-level residual encoder-decoder, channels $(32, 64, 128, 256, 512)$, stride-2 down-/up-sampling, concatenative skips; 3-channel sigmoid output.

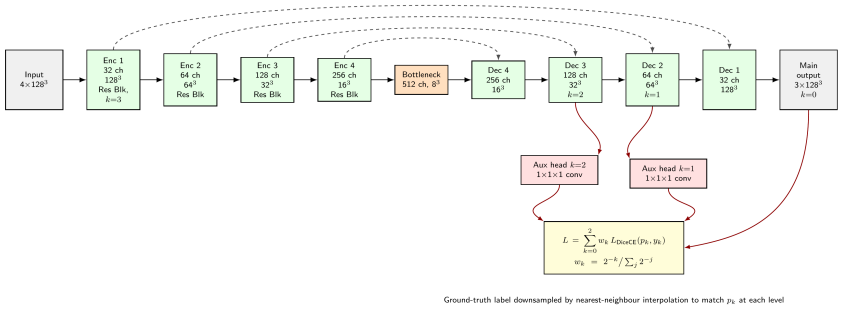

Fig. 3. DynUNet (MONAI `DynUNet`, nnU-Net-inspired). Deep supervision: auxiliary 1×1×1 heads at Dec2 (k=1, half-res) and Dec3 (k=2, quarter-res), combined with the main output via w_k=2^{-k} (normalised), following the nnU-Net schedule [15].

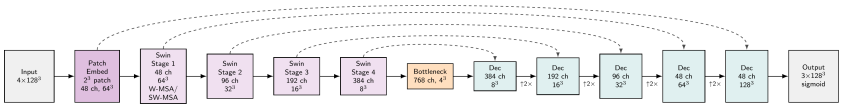

Fig. 4. Swin UNETR: hierarchical Swin Transformer encoder (4 shifted-window stages, patch 2^3, feature_size=48) with a convolutional decoder and multi-stage skips [11,28].

3.5 Loss function and hyperparameter selection

All architectures are trained with a compound Dice and cross-entropy loss function (DiceCE) [27], batch-wise (batch=True) for stability at small batch size, equally weighted ($\lambda_{\text{dice}} = \lambda_{\text{ce}} = 1$) following nnU-Net [15]; sigmoid is applied inside the loss, so all networks output raw logits. For DynUNet, DiceCE is wrapped in a deep-supervision module (Fig. 3) computing a 2^{-k}-weighted sum over the main and two auxiliary outputs. Optimisation uses AdamW [17] (lr $= 10^{-4}$, wd $= 10^{-5}$) with cosine-annealing decay [16] to 10^{-6} over 300 epochs, gradient clipping at 1.0, and AMP [21] throughout, batch size 2 for all three architectures. Validation runs every 10 epochs; the checkpoint with the highest mean validation Dice (WT/TC/ET) is retained, consistent with the BraTS primary ranking metric [4]. At evaluation, two complementary metrics are computed per subregion via MONAI's `DiceMetric` and `HausdorffDistanceMetric`: Dice similarity (DSC$= 2|P \cap G|/(|P| + |G|)$) [8] for volumetric overlap, and 95th-percentile Hausdorff distance (HD95) for boundary agreement, robust to isolated mis-predictions [14], the standard secondary and primary BraTS metrics [20,4]. Predictions are binarised at threshold 0.5; values are reported NaN-safe so subregions absent from a subject do not contaminate the aggregate.

3.6 Compute resources and inference protocol

This work was trained on **Narval**, a GPU cluster operated by the Digital Research Alliance of Canada (École de technologie supérieure, Montréal) [9]. Each

job requests a single NVIDIA A100-40GB GPU under PyTorch/MONAI [5]; because compute nodes have no internet access, the orchestration script applies two small version-specific patches at job start: a NumPy `ptp()` fix in `monai.data.utils` and a 32-bit `MAX_SEED` overflow fix in MONAI's transform-compose module. Each architecture trains independently (one A100, batch 2, 300 epochs), so the three runs can be distributed across separate single-GPU jobs rather than requiring multi-GPU parallelism.

Because 128^3 input size matches the training input, inference uses a single forward pass per preprocessed volume rather than sliding-window aggregation. The implications of this single-crop inference strategy are considered when interpreting the relatively lower WT Dice observed on the in-domain test set.

4 Results

Table 2. Performance comparison on in-domain BraTS 2021 ($n \approx 187$) and zero-shot BraTS-Africa ($n = 146$) with 95% CIs and pathology stratification.

Metric / Sub-Cohort	3D U-Net	DynUNet	Swin UNETR
BraTS 2021 (In-Domain Baseline, $n \approx 187$)			
Dice WT (%) ↑	66.8	67.1	67.3
Dice TC (%) ↑	90.5	90.3	90.5
Dice ET (%) ↑	87.4	88.2	87.9
HD95 WT / TC / ET (mm) ↓	8.6 / 4.9 / 4.4	8.8 / 5.4 / 4.3	8.3 / 6.0 / 4.5
BraTS-Africa Zero-Shot (Full Cohort, $n = 146$)			
Dice WT [95% CI] ↑	59.0 [56.8, 61.2]	60.6 [58.4, 62.8]	59.7 [57.5, 61.9]
Dice TC [95% CI] ↑	72.9 [70.9, 74.9]	73.9 [71.9, 75.9]	72.3 [70.2, 74.4]
Dice ET [95% CI] ↑	71.7 [69.4, 74.0]	69.0 [66.7, 71.3]	68.9 [66.5, 71.2]
HD95 WT / TC / ET (mm) ↓	17.9 / 18.2 / 10.6	15.4 / 14.2 / 9.4	16.8 / 15.9 / 12.3
BraTS-Africa Stratification (Mean Dice [95% CI])			
Glioma ($n = 95$)	68.1 [66.8, 69.5]	69.9 [68.6, 71.3]	68.8 [67.4, 70.2]
Non-Glioma ($n = 51$)	61.8 [59.6, 64.0]	63.7 [61.5, 65.9]	62.4 [60.2, 64.6]
ΔDice Full Cohort	−13.8	−14.1	−15.0
ΔDice Glioma Only	−10.1	−10.3	−11.4

The code[10] is implemented in PyTorch and MONAI [5]. All architectures perform comparably in-domain (mean Dice 81.6–81.9%). Lower in-domain WT Dice (66.8–67.3%) reflects single-pass 128^3 central cropping without sliding-window aggregation. On zero-shot BraTS-Africa ($n = 146$), mean Dice decreases by 13.8–15.0 pp across models, primarily driven by TC and ET degradation (> 15 pp drops).

Pathology stratification isolates disease shift: glioma-only cases ($n = 95$) yield higher zero-shot accuracy (DynUNet mean Dice 69.9%, 95% CI: [68.6%, 71.3%]), narrowing its domain gap from 14.1 pp to 10.3 pp. Non-glioma cases ($n = 51$) drop to 63.7% (95% CI: [61.5%, 65.9%]), confirming non-glioma pathology adds a ~ 6.2 pp penalty. DynUNet maintains the smallest overall gap and best HD95 (13.0 mm), while Swin UNETR shows the largest gap (15.0 pp); specific archi-

[10] Available at `https://github.com/abrahamsegun001/glioma-segmentation`

tectural drivers (deep supervision vs. lack of pre-training) remain hypotheses for future empirical validation.

5 Discussion

Observed patterns. The 13.8–15.0 pp mean-Dice drop across all three architectures confirms a substantial and consistent domain gap when transferring from BraTS 2021 to BraTS-Africa. The persistence of this degradation across architecturally distinct networks a convolutional baseline, a deep-supervised nnU-Net-style network, and a CNN-Transformer hybrid argues that the gap is distributional in nature rather than architecture-specific, and that architecture choice alone cannot close it without incorporating African training data or explicit adaptation. The HD95 results further underscore the clinical severity of this gap: boundary errors of 15–18 mm on BraTS-Africa vs. 4–9 mm in-domain represent clinically meaningful differences in tumour boundary delineation relevant to radiation field planning.

Label-convention pitfall. A notable engineering finding is that the BraTS-Africa ET label convention (integer 3) differs from BraTS 2021 (integer 4). An unmodified MONAI pipeline silently zeros the ET channel for African subjects, masking the true domain gap and reporting artificially low ET Dice. Our `ConvertBraTSLabelsd` transform guards against this and should be adopted by any group combining both datasets.

Limitations. BraTS-Africa's 146 subjects remain small relative to BraTS 2021; results are appropriate for characterising the zero-shot gap but not for high-confidence subgroup analysis. Histopathological confirmation is absent in a substantial fraction of SSA glioma cases [19], so ground truth reflects expert consensus throughout. Phase 1 deliberately omits African data from training, a methodological strength for *measuring* domain gap. So reported Dice/HD95 are a lower bound on what Phase 2 can achieve in future work.

6 Conclusion and Impact in Resource-Constrained Settings (RCS)

This benchmark is designed to directly inform deployment in the clinical settings it studies. The ΔDice values in Table 2 quantify, for the first time under a controlled, multi-architecture protocol, how much accuracy is lost when a model trained only on Global-North data is deployed in an SSA clinical context without adaptation. With fewer than half of African countries operating functional radiotherapy infrastructure and neurosurgical capacity on the order of one specialist per several million people [19,30], a tool that silently mis-segments tumour volume risks miscalibrating radiation-field planning or surgical referral [25], a patient-safety question, not only a benchmark statistic.

From benchmarking to deployable, sustainable AI. Phase 1's comparison is the empirical foundation for a Phase 2 model purpose-built for RCS deployment: depthwise-separable convolutions for CPU-based inference, attention gates for boundary delineation on lower-resolution scans, and Monte Carlo

Dropout for uncertainty maps flagging low-confidence predictions for review, benchmarked against the GPU baselines established here (Sec. 3.6).

Acknowledgements

The authors would like to thank the instructors of the Sprint AI Training for African Medical Imaging Knowledge Translation (SPARK) Academy 2026 summer school on deep learning in medical imaging for providing insightful background knowledge on brain tumours that informed the research presented here. The authors acknowledge the computational infrastructure support from the Digital Research Alliance of Canada (the Alliance) and the University of Washington Azure GenAI for Science Hub. Finally, we would like to thank the Lacuna Fund for Health and Equity (PI: Udunna Anazodo, 0508-S-001), the National Science and Engineering Research Council of Canada (NSERC) Discovery Launch Supplement (PI: Udunna Anazodo, DGECR-2022-00136), and the McGill University Doctoral Internship Program for making the SPARK Academy possible via research grant support.

References

1. Adewole, M., Rudie, J.D., Gbadamosi, A., Zhang, D., Raymond, C., Bakas, S., Anazodo, U.C., et al.: The BraTS-Africa dataset: Expanding the brain tumor segmentation data to capture african populations. Radiology: Artificial Intelligence **7**(4), e240528 (2025). https://doi.org/10.1148/ryai.240528
2. Adewole, M., Rudie, J.D., Gbdamosi, A., Toyobo, O., Raymond, C., Zhang, D., et al.: The brain tumor segmentation (BraTS) challenge 2023: Glioma segmentation in sub-Saharan Africa patient population (BraTS-Africa). arXiv preprint arXiv:2305.19369 (2023)
3. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Bakas, S., et al.: The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021)
4. Bakas, S., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:1811.02629 (2018)
5. Cardoso, M.J., et al.: MONAI: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
6. Chowdhury, A., Kassem, H., Padoy, N., Umeton, R., Karargyris, A.: A review of medical federated learning: Applications in oncology and cancer research. In: International MICCAI Brainlesion Workshop. pp. 3–24 (2021)
7. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), LNCS 9901. pp. 424–432 (2016). https://doi.org/10.1007/978-3-319-46723-8_49
8. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
9. Digital Research Alliance of Canada: Narval cluster technical documentation. https://docs.alliancecan.ca/wiki/Narval (2024)
10. Guan, H., Liu, M.: Domain adaptation for medical image analysis: A survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2022). https://doi.org/10.1109/TBME.2021.3117407
11. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: BrainLes Workshop, MICCAI, LNCS 12962. pp. 272–284 (2022). https://doi.org/10.1007/978-3-031-08999-2_22
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
13. Huang, L., Ruan, S., Xing, Y., Feng, M.: A review of uncertainty quantification in medical image analysis: Probabilistic and non-probabilistic methods. Medical Image Analysis **97**, 103223 (2024)

14. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the Hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence **15**(9), 850–863 (1993). https://doi.org/10.1109/34.232073
15. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z
16. Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. In: International Conference on Learning Representations (ICLR) (2017)
17. Loshchilov, I., Hutter, F.: Decoupled weight decay regularisation. In: International Conference on Learning Representations (ICLR) (2019)
18. Magadza, T., Viriri, S.: Deep learning for brain tumor segmentation: A survey of state-of-the-art. Journal of Imaging **7**(2), 19 (2021). https://doi.org/10.3390/jimaging7020019
19. Mensah, G., Lehleng, A., Belo, M., Komla, A., Adama, E.g., Ayelola, B.: Neuro oncology in Sub-Saharan Africa from 1990 to 2025: Current status, diagnostic gaps, and strategic directions for care and research. International Journal of Neurology Sciences **8**, 05–08 (2026). https://doi.org/10.33545/26646161.2026.v8.i1a.53
20. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694
21. Micikevicius, P., et al.: Mixed precision training. In: International Conference on Learning Representations (ICLR) (2018)
22. Ogbole, G.I., Adeyomoye, A.O., Badu-Peprah, A., Mensah, Y., Nzeh, D.A.: Survey of magnetic resonance imaging availability in West Africa. Pan African Medical Journal **30**, 240 (2018). https://doi.org/10.11604/pamj.2018.30.240.14000
23. Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning where to look for the pancreas. In: Medical Imaging with Deep Learning (MIDL) (2018)
24. Rafi, T.H., Mahjabin, R., Ghosh, E., Ko, Y.W., Lee, J.G.: Domain generalization for semantic segmentation: A survey. Artificial Intelligence Review **57**(9), 1 (2024)
25. Ralph-Okhiria, O., Alonge, I.: Leveraging artificial intelligence to strengthen surgical systems in Sub-Saharan Africa. Academia Medicine **2**(2) (2025). https://doi.org/10.20935/AcadMed7735
26. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), LNCS 9351. pp. 234–241 (2015)
27. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA), MICCAI Workshop, LNCS 10553. pp. 240–248 (2017). https://doi.org/10.1007/978-3-319-67558-9_28
28. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of Swin transformers for 3D medical image analysis. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20730–20740 (2022). https://doi.org/10.1109/CVPR52688.2022.02007
29. Weller, M., Wen, P.Y., Chang, S.M., et al.: Glioma. Nature Reviews Disease Primers **10**, 33 (2024). https://doi.org/10.1038/s41572-024-00516-y
30. Yevudza, W.E., Buckman, V., Darko, K., Banson, M., Totimeh, T.: Neuro-oncology access in Sub-Saharan Africa: A literature review of challenges and opportunities. Neuro-Oncology Advances **6**(1), vdae057 (2024). https://doi.org/10.1093/noajnl/vdae057
31. Zou, K., Chen, Z., Yuan, X., Shen, X., Wang, M., Fu, H.: A review of uncertainty estimation and its application in medical imaging. Meta-Radiology **1**(1), 100003 (2023)