Generative Style Transfer for MR Image Segmentation: A Case of Glioma Segmentation in Sub-Saharan Africa

Rancy Chepchirchir[1,2], Jill Sunday[3], Raymond Confidence[4,5], Dong Zhang[6], Talha Chaudhry[7], Udunna C. Anazodo[4,5,8,9], Kendi Muchungi[10], and Yujing Zou[11(✉)]

[1] Institute of Mathematical Science, Strathmore University, Nairobi, Kenya
[2] Faculty of Science and Engineering, University of Hull, Hull, UK
r.chepchirchir-2023@hull.ac.uk
[3] Department of Medical Engineering, Technical University of Mombasa, Mombasa, Kenya
jillselesa35@gmail.com
[4] Medical Artificial Intelligence Lab, Lagos, Nigeria
raymondconfidence@gmail.com
[5] Lawson Health Research Institute, London, ON, Canada
[6] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
donzhang@ece.ubc.ca
[7] University of Nairobi, Nairobi, Kenya
talhahchaudhry99@gmail.com
[8] Department of Clinical and Radiation Oncology, University of Cape Town, Cape Town, South Africa
udunna.anazodo@mcgill.ca
[9] Montreal Neurological Institute, McGill University, Montreal, Canada
[10] Brain Mind Institute, The Aga Khan University, Nairobi, Kenya
kendi.muchungi@aku.edu
[11] Medical Physics Unit, McGill University, Montreal, Canada
yujing.zou@mail.mcgill.ca

Abstract. In Sub-Saharan Africa (SSA), the utilization of lower-quality Magnetic Resonance Imaging (MRI) technology raises questions about the applicability of machine learning (ML) methods for clinical tasks. This study aims to provide a robust deep learning-based brain tumor segmentation (BraTS) method tailored for the SSA population using a threefold approach. Firstly, the impact of domain shift from the SSA training data on model efficacy was examined, revealing no significant effect.Secondly, a comparative analysis of 3D and 2D full-resolution models using the nnU-Net framework indicates similar performance of both the models trained for 300 epochs achieving a five-fold cross-validation score of 0.93. Lastly, addressing the performance gap observed in SSA validation as opposed to the relatively larger BraTS glioma (GLI) validation set, two strategies are proposed: fine-tuning SSA cases using the GLI + SSA best-pretrained 2D fullres model at 300 epochs, and introducing a novel neural style transfer-based data augmentation technique for the SSA cases. This investigation underscores the potential of enhancing brain tumor prediction within SSA's unique healthcare landscape.

U. Anazodo et al. (Eds.): MImA 2024/EMERGE 2024, CCIS 2240, pp. 110–123, 2025.
https://doi.org/10.1007/978-3-031-79103-1_12

Keywords: Brain Tumor Segmentation · Neural style transfer · nnU-Net

1 Introduction

Brain tumors present a substantial health challenge in Africa. The efforts of research on brain tumors have barely made any positive change in survival rate in low-and middle-income countries (LMICs). The rate of mortalities from glioma is among the highest in the world, with the Sub-Saharan Africa experiencing a rise of 25% [1].

Accurate segmentation of distinct sub-regions within gliomas such as peritu- moral edema, necrotic core, enhancing, and non-enhancing tumor core, based on multimodal MRI scans, hold clinical relevance for the diagnosis, prognosis, and treatment of brain tumors. Accurately delineating the regions of interest within a tumor provides essential insight about its size, location, and shape, enabling the determination of the extent of tumor involvement [2].

However, the segmentation of these sub-regions presents a formidable challenge due to the heterogeneity of brain tumors [3] and in resource-constrained settings such as the SSA, due to the propensity for suboptimal image contrast and resolution [1] from lower quality MRI scanners and lack of availability of advanced imaging techniques [4]. These challenges raise uncertainty about the feasibility of implementing ML methods for clinical purposes [4, 5]. Furthermore, suboptimal image contrast and resolution [6], necessitates advanced image pre-processing to enhance their resolution before employing ML techniques for tasks like tumor segmentation, classification, or outcome prediction [1].

In response to these challenges, this study aims to develop a generalizable deep learning-based brain tumor segmentation method. Our approach addresses the challenge of lower quality MRI scans in the SSA by developing an effective model that handles data variability. Through advanced preprocessing and optimized segmentation, we aim to create a robust model that improves diagnostic outcomes in under-resourced settings.

1.1 Related Work

Neural style transfer (NST) has been employed as a data augmentation technique in various medical domains, including COVID-19 diagnosis classification [7] and 3D cardiovascular MR image segmentation [8]. However, this study marks the first application of NST to brain tumor segmentation within a Sub-Saharan Africa context. Moreover, Tomar et.al. [9] have explored a self-supervised style transfer technique as data augmentation to improve brain tumor segmentation performances. This comprehensive methodology was, however, more computationally expensive than the NST approach. Bouter et al. [10] demonstrated the feasibility of artificially creating super-resolution MR images from low-resolution counterparts, indicating that such an approach could be leveraged for our 2021 BraTS dataset. Lastly, the work conducted by Sendra et al. [11] addressed similar domain-shift challenges using comparable approaches. Their findings suggested that employing transfer learning for domain adaptation could integrate modest-sized African samples into extensive databases of developed nations. Notably, both studies highlighted the imbalance between African and high-resource country cases, with

Sendra et al. studying 25 patients from five African centers, while our study examined 60 African cases.

2 Methods

Implementation
Our solution, which is implemented with PyTorch, is an extension of NVIDIA's nnUNet [12], which is publicly available on GitHub[1]. The baseline model training and inference were done with mixed precision to minimize costs (i.e. time and memory). The experiments were run on Tesla T4 Turing GPUs and NVIDIA's V100 system on Compute Canada cluster. We then stored both the latest and best checkpoint models based on the Dice score on the validation dataset for use during inference.

Datasets
This work was conducted as part of the BraTS 2023 Challenge, where the datasets preselected by the organizers were used. The datasets were multi-center MRI scans of 1251 adult glioma (GLI) cases from the 2021 Continuous Evaluation sub-challenge [13] and 60 adult glioma cases acquired in SSA (SSA) from the BraTS-Africa sub challenge [1] - the largest publicly available African adult glioma MRI data. Thus, a total of 1311 MRI scans of adults with pre-operative glioma including both low-grade glioma (LGG) and high-grade glioma (GBM/HGG) were used to train and validate the proposed model. The MRI scans comprised of routine T1-weighted (T1), post-contrast T1-weighted (T1ce), T2-weighted (T2) and T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) images acquired as part of standard of care [1, 13]. Each case also contained pre-labelled brain tumor sub-region masks, namely, necrotic tumor core (NCR), enhancing tumor (ET), and peritumoral edematous tissue (ED).

The datasets were split into GLI (1251 cases) and GLI + SSA (1311) and used separately to train the model, while a five-fold cross-validation approach was used to evaluate the performance of the model. Four cases from the SSA dataset were excluded as outliers based on image quality inconsistencies. To investigate the impact of the outliers in real-world applications, the GLI + SSA data were also trained excluding the four outlier cases (GLI + SSA2; 1307 cases).

Data Preprocessing
The data were preprocessed following Futrega et.al Optimized U-Net approach [14] and involved foreground cropping operation, intensity normalization, and resampling. These preprocessing steps were taken to establish data coherence among the multi-center data, enhance image quality, and standardize the format for subsequent processing steps. Commencing with loading the dataset, organized in alignment with a designated data path, we extracted crucial metadata from a JSON file. Enhancing image quality and uniformity was realized through the application of a crop foreground operation. This step effectively removed extraneous background regions. Additionally, the process encompassed intensity normalization. Notably, for MRI scans, normalization was confined to non-zero regions. Addressing variations in voxel spacing among diverse scans, resampling played

[1] https://github.com/MIC-DKFZ/nnUNet.

a pivotal role. The aforementioned measures are crucial in reducing variability in input data and improving reliability of the model's predictions [14].

2.1 Baseline Model (Optimized U-Net)

The baseline for our model was inspired by the work of Futrega et al. [14], the Optimized U-Net model with deep supervision to improve the gradient flow by calculating the loss function at various decoder levels. Here, each experiment was trained for 2, 5, 10, and 30 epochs using the Adam optimizer with varying learning rates for the three experiments: 1) GLI, 2) GLI + SSA, and 3) GLI + SSA2.

2.2 nnU-Net Model

nnU-Net, an image segmentation method introduced in [12], adapts to specific datasets by autonomously configuring a U-Net-based segmentation pipeline. It simplifies model training by creating multiple U-Net configurations for different datasets, effectively handling diverse input modalities and class imbalances. Notably, nnU-Net version 2 brings enhancements, offering a user-friendly development framework.

Two- and Three-Dimensional (2D & 3D) Configuration
We have employed a 3D full resolution with a batch size of 2, a patch size of [128, 128, 128], 32 U-Net base features, a per-stage encoder and decoder of [2, 2, 2, 2, 2, 2], kernel sizes of [[1, 1, 1], [2, 2, 2], [2, 2, 2], [2, 2, 2], [2, 2, 2], [2, 2, 2]], and convolution kernel sizes of [[3, 3, 3], [3, 3, 3], [3, 3, 3], [3, 3, 3], [3, 3, 3], [3, 3, 3]] on both the GLI dataset and the GLI+SSA dataset. We also employed a 2D full resolution with a batch size of 105, a patch size of [192, 160], 32 U-Net base features, a per-stage encoder and decoder of [2, 2, 2, 2, 2, 2], kernel sizes of [[1, 1], [2, 2], [2,2], [2, 2], [2, 2], [2, 2]], and convolution kernel sizes of [[3, 3], [3, 3], [3, 3], [3, 3], [3, 3], [3, 3]] on the GLI and GLI+SSA dataset. Each experiment was trained for 2, 5, 10, 30, and 300 epochs using the Adam optimizer with varying learning rates

2.3 Proposed Methods: Neural Style Transfer Augmentation and 2D Full-Res Model Finetuning

In the context of resource-constrained settings where limited training data continues to pose a challenge, we employed neural style transfer (NST), first proposed by Gatys et al. [15], as a data augmentation technique to enhance the effectiveness of our model training process. Neural style transfer is a technique rooted in deep learning that enables the separation and combination of content and style aspects within images. Specifically, for our application, we leveraged neural style transfer to enhance the SSA MR images. This technique involved using an SSA MR image as the content (or source) image and performing a random pairing with a GLI MR image as the style (or target) image. The neural style transfer process entailed adapting the stylistic features of the GLI image onto the SSA image, thereby creating new, augmented training samples for SSA cases. Using the Keras functional API, the intermediate layers of a pretrained VGG19 image classification network [16] were used as a feature extractor to obtain the content and style representations of an image. The neural style transfer algorithm computes a

loss by evaluating the discrepancies between the stylistic and content features of the generated image and target images. The style loss measures stylistic differences using Gram matrices of feature maps, while the content loss quantifies content dissimilarity through feature map comparisons.

Overall Workflow. The proposed methodology is illustrated below in Fig. 1. At the time of submission for this short paper, we hereby present results for the nnU-Net best 2D and 3D models, as well as the best pre-trained 2D[2] fullres nnU- Net model trained from GLI + SSA training data fine-tuned on the original and NST-augmented SSA training data. The 5-fold cross validation model evaluation was based on mean Dice Similarity Coefficient (DSC) after each epoch. A Paired Samples t-test was conducted to compare model performances between datasets and as a function of training time/epochs (significance at $p < 0.05$).

3 Results

3.1 Baseline Model (Optimized U-Net)

To establish a robust evaluation, we conducted experiments using the Optimized U-Net [14] on both the GLI and GLI + SSA datasets as part of our foundational benchmarking process. This comparison aimed to elucidate how distributional shifts within training data can influence cross-validated model performance. Table 1 presents outcomes from both datasets. It includes their Dice Similarity Coefficient (DSC), training, and validation

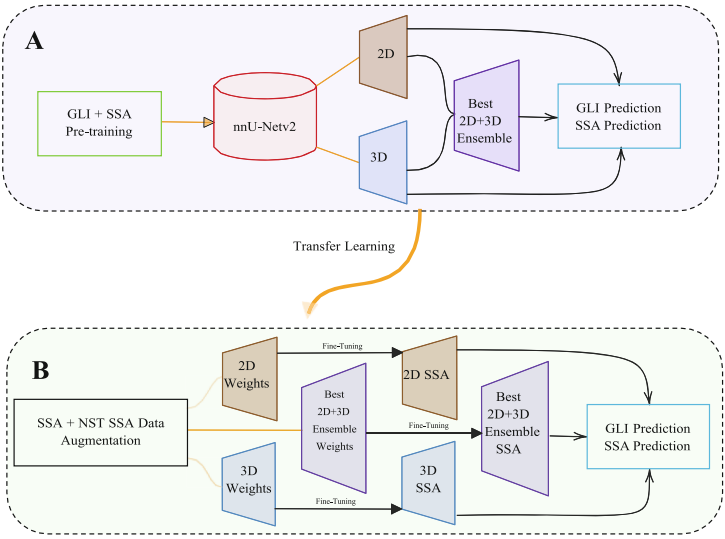

Fig. 1. Fine tuning small-scale SSA data with neural style transfer data augmentation techniques (B) with pretrained weights from large-scale GLI data via transfer learning (A).

[2] https://github.com/CAMERA-MRI/SPARK2023/tree/main/SPARK_BTS_KIFARU.

losses for the GLI and GLI + SSA datasets, with or without four outlier cases (00051, 00097, 00041, and 00084), as shown in Fig. 1 in Appendix. We halted training at 30 epochs for comparative purposes with an experiment using the complete dataset for the same epoch count. From the aforementioned results, a decrease in model performance was observed from a Dice score of 0.89 for the GLI trained dataset to 0.88 for the GLI + SSA dataset at 30 epochs (Table 1). There was no statistically significant difference between the Dice scores of the GLI and GLI + SSA models ($p > 0.05$) or between models when outliers were excluded **($p > 0.05$)** (Table 1). However, this exclusion was specific to this experiment, aimed at probing the domain shift issue in multi-institutional data, particularly in the Sub-Saharan Africa context.

3.2 nnU-Net Model (Version 2)

We compared the 3D fullres versus 2D fullres configurations of nnU-Net trained on the GLI + SSA dataset. The best model was considered to have obtained the best 5-fold cross-validation averaged Dice Similarity Coefficient.

Configuration Name: 3D Full Resolution. Here, we employed the 3D configuration models for the GLI and GLI + SSA datasets, as represented in Table 2. It is evident that the 300 epoch model achieved the best pseudo Dice score at 0.95 and 0.93 for the GLI and GLI + SSA datasets respectively; followed by the 30 epoch model at 0.90 for both the datasets. The paired t-test comparisons between models, showed no statistically significant difference ($p > 0.05$) between the pseudo Dice scores (Table 2).

Configuration Name: 2D Full Resolution. Table 3 reveals that the 2D full-resolution nnU-Net model, trained for 300 epochs, achieved five-fold cross- validation scores of 0.91 and 0.93 for the GLI and GLI + SSA training datasets respectively. Despite the pseudo Dice score of the GLI + SSA dataset being slightly higher than that of the GLI

Table 1. Comparison of the GLI, GLI + SSA, and GLI + SSA2(Excl. 4 SSA outliers) Datasets for the Optimized U-Net.

Datas	Results			Paired t-test
	Dice Similarity Coefficient	Train Loss	Val Loss	
GLI	0.79	1.83	0.21	GLI vs. GLI+SSA: **t-stat=1.82, p=0.17, df=3**
	0.84	1.12	0.15	
	0.86	0.85	0.14	
	0.89	0.59	0.10	
GLI+SSA	0.77	1.90	0.23	GLI+SSA vs. GLI+SSA2: **t-stat=-0.51, p=0.65, df=3**
	0.84	1.01	0.15	
	0.86	0.76	0.14	
	0.88	0.57	0.12	
GLI+SSA2	0.79	1.58	0.20	
	0.83	1.01	0.17	
	0.86	0.56	0.14	
	0.88	0.55	0.11	

Val loss = validation loss

model, the paired t-test revealed that the difference between the pseudo Dice scores is not statistically significant ($p > 0.05$) (Table 3). Also, at 300 epochs of training, for the GLI + SSA dataset, the five-fold cross-validation performance of 2D fullres nnU-Net is equal to that of 3D (0.93). However, the GLI + SSA dataset has significantly longer training times than that of the GLI dataset, with the 3D fullres nnU-Net training faster. The leision-wise Dice Similarity Coefficients for each fold of the 2D fullres nnU-Net model are presented in Table 4 of the Appendix.

Table 2. 3D fullres nnU-Net configuration: A representation of the average of Pseudo Dice Score per epoch number for the GLI and the GLI + SSA datasets, with Paired t-test results included.

Dataset	Epoch	Learning Rate	Train Loss	Val Loss	Pseudo Dice	Epoch Time (s)	Paired t-test
GLI	2	0.00536	-0.67	-0.71	0.78	169.28	GLI vs GLI+SSA Pseudo Dice:
	5	0.00235	-0.86	-0.87	0.86	165.65	
	10	0.00126	-0.80	-0.78	0.83	195.28	
	30	0.00047	-0.83	-0.85	**0.90**	190.9	
	300	6e-05	-0.88	-0.87	**0.95**	201.65	
GLI+SSA	2	0.00536	-0.42	-0.64	0.78	449.16	**t-stat: 1.12**
	5	0.00235	-0.72	-0.72	0.79	467.69	
	10	0.00126	-0.78	-0.79	0.84	427.12	**p-value: 0.33**
	30	0.00047	-0.84	-0.85	**0.90**	427.3	**df:4**
	300	6e-05	-0.89	-0.86	**0.93**	496.42	

Val loss = validation loss

Table 3. 2D fullres nnU-Net: A representation of the average of prediction Pseudo Dice Scores per epoch number for the GLI and the GLI + SSA datasets, with t-test results.

Dataset	Epoch	Learning Rate	Train Loss	Val Loss	Pseudo Dice	Time (s)	Paired t-test
GLI	2	0.00536	-0.79	-0.82	0.85	193.98	GLI vs GLI+SSA Pseudo Dice:
	5	0.00235	-0.86	-0.87	0.85	190.19	
	10	0.00126	-0.89	-0.89	0.88	192.5	
	30	0.00047	-0.83	-0.83	0.89	294.5	
	300	6e-05	-0.89	-0.90	**0.91**	330.0	
GLI+SSA	2	0.00536	-0.80	-0.82	0.84	337.17	**t-stat: -1.43**
	5	0.00235	-0.86	-0.87	0.86	488.99	
	10	0.00126	-0.89	-0.88	0.88	300.19	**p-value: 0.22**
	30	0.00047	-0.92	-0.91	0.89	230.32	**df: 4**
	300	6e-05	-0.91	-0.90	**0.93**	501.19	

Val loss = validation loss

3.3 Fine-Tuning and Neural Style Transfer on SSA Training Data as a Data Augmentation Technique Together Improves SSA Validation Results

Despite the 2D fullres nnU-Net trained using GLI + SSA data displaying good performance in five-fold cross-validation and generating satisfactory predictions for previously unseen GLI validation data as illustrated in Fig. 3 in the Appendix, its performance on SSA validation data was weaker. It particularly performed badly on the SSA MR images that appeared to be incomplete. For example in the left image of Fig. 4 in the Appendix, an empty mask was predicted. The neural style transfer technique used as a data augmentation method helped solve this problem. Furthermore, the best 2D fullres nnU-Net trained model at 300 epochs was used as a pretrained model to fine-tune the original 60 cases of SSA training data as well as its stylized augmented SSA data. This directly resulted in an improvement in prediction on the same unseen SSA validation example shown in the right image of Fig. 4 of the Appendix whereas originally an empty mask was predicted before the NST data augmentation and fine-tuning step.

4 Discussion

In summary, at the time of this short paper submission, we have demonstrated the following results. ***Firstly***, we highlighted the similarity in the performance of the model with or without the SSA datasets. The results of the Paired Samples t-test between datasets revealed that there is no significant statistical difference between the performance of the three models as shown in Table 1. The results showed that the model performs equally on both the GLI and GLI + SSA datasets, demonstrating the its generalizability. ***Secondly***, we conducted a comparative analysis between the performance of 3D and 2D full-resolution models, utilizing the latest iteration (version 2) of nnU-Net [12]. Our investigation revealed that both the 2D and 3D full-res nnU-Net models trained for 300 epochs yielded an average pseudo Dice score of 0.93 for the GLI + SSA training data via a 5-fold cross-validation strategy. ***Thirdly***, our validation process on the provided GLI and SSA cases - without ground truth annotations - revealed a significant performance discrepancy between the GLI and SSA validation sets. This can be visually inspected by the relatively good prediction for the GLI validation data shown in Fig. 3 as compared to Fig. 4A in the Appendix. The fusion of the neural style transfer data augmentation with subsequent fine-tuning targeted specifically at SSA cases have demonstrated significant improvements in results for the SSA validation set (Fig. 4B).

An important limitation of our study was the scarcity of African datasets. Future work should extend this approach to a larger African dataset to enhance its applicability. Thus, as proposed in the overall methodology workflow in Fig. 1, in the near future, we will: ensemble the best 2D fullres and 3D full res nnU-Net model trained from the combined GLI and SSA training data, before repeating the fine-tuning experiment with the neural style transfer data augmented SSA plus the original SSA training data. Moreover, we will capitalize on the availability of higher-quality GLI data that could result in more neural style transfer random pairing with the limited SSA training data. We will then find an optimal number of data augmentation pairs while examining model performances. Further post- processing methods will also be investigated.

5 Conclusion

In this investigation, we have established the viability of enhancing brain tumor prediction within the limited-resource context of Sub-Saharan Africa (SSA). By utilizing a pretrained and high-performing 2D fullres nnU-Net model, we achieved refinement through fine-tuning using SSA training data augmented via neural style transfer. This methodology underscores the potential for notable performance improvements within SSA's unique healthcare setting.

Acknowledgments. The authors would like to thank the following instructors of the Sprint AI Training for African Medical Imaging Knowledge Translation (SPARK) Academy 2023 summer school on deep learning in medical imaging for providing insightful background knowledge on brain tumors that informed the research presented here; Craig Jones, Eranga Ukwatta, Esin Ozturk-Isik, Evan Calabrese, Jeff Rudie,

Konstantinos Gousias, MacLean Nasrallah, Malhar Patel, Mueez Waqar, Nicole Levy, Peizhi Yan, Piotr Pater, Ujjwal Baid, and Yihao Liu. The authors would also like to thank Talha Chaudhry for his clinical input with regard to the Sub- saharan Africa dataset and Linshan Liu for administrative assistance in supporting the SPARK Academy training and capacity-building activities. The authors acknowledge the computational infrastructure support from the Digital Research Alliance of Canada (The Alliance) and knowledge translation support from the McGill University Doctoral Internship program through student exchange program for the SPARK Academy. The authors are grateful to McMedHacks for providing foundational information on python programming for medical image analysis as part of the 2023 SPARK Academy program. This research was funded by the Lacuna Fund for Health and Equity (PI: Udunna Anazodo, grant number 0508-S-001) and National Science and Engineering Research Council of Canada (NSERC) Discovery Launch Supplement (PI: Udunna Anazodo, grant number DGECR-2022-00136).

Appendix

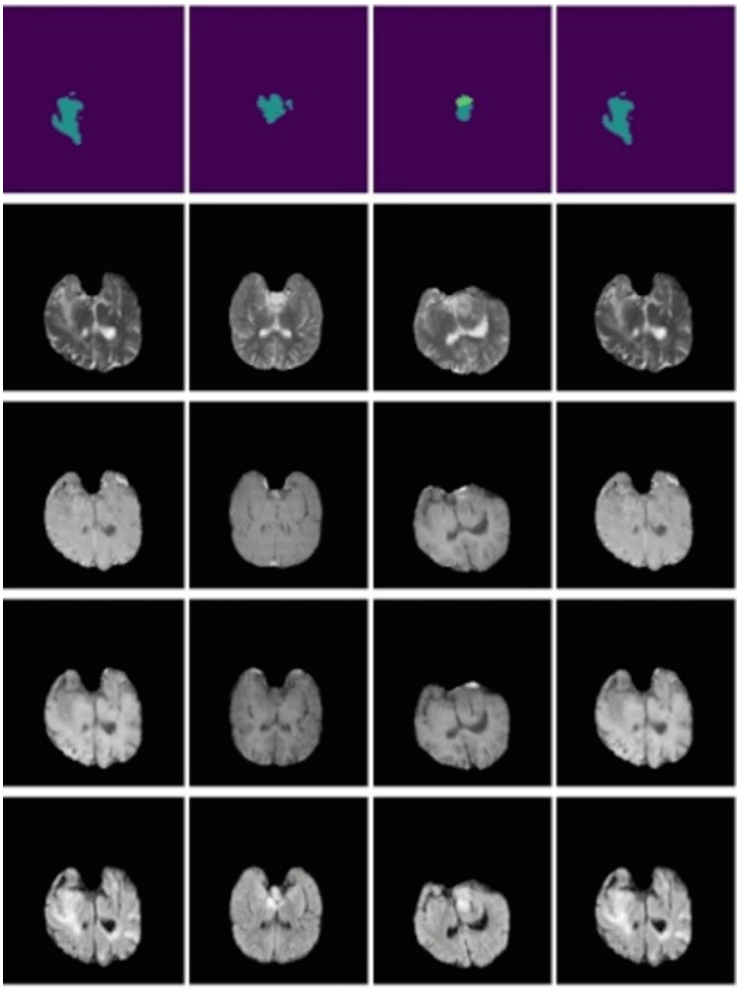

Fig. 2. Excluded SSA training Cases A) 00051, B) 00097, C) 00041, & D) 00084 only for the optimized UNet baseline experiment [Predicted masks (top row), T2 (second row), T1ce (third row),T1 (fourth row), and T2-FLAIR (bottom row)]. This was not employed for the rest of the experiments. The predicted masks are shown with color coding as follows: background: purple, necrotic tumor core (NCR): blue, enhancing tumor (ET): yellow, peritumoral edematous tissue (ED): turquoise. (color figure online)

Table 4. Lesion-wise dice score from five-fold cross-validation for the best 2D fullrest nnU-Net trained mdoel.

Model	Dice Score			Epochs
	Dice_ET	Dice_TC	Dice_WT	
Fold 0	0.8689	0.8205	0.8082	2
	0.9131	0.8534	0.8304	5
	0.8963	0.8231	0.8415	10
	0.9388	0.9031	0.8991	30
	0.9471	**0.9179**	**0.9179**	300
Fold 1	0.8745	0.7932	0.776	2
	0.9234	0.866	0.8509	5
	0.9327	0.8839	0.8922	10
	0.9294	0.8949	0.8958	30
	0.9488	**0.9051**	**0.8890**	300
Fold 2	0.8805	0.8024	0.8212	2
	0.9178	0.8568	0.8352	5
	0.9388	0.9062	0.8977	10
	0.9469	0.9145	0.902	30
	0.9369	**0.8955**	**0.8989**	300
Fold 3	0.8197	0.7734	0.7608	2
	0.9129	0.837	0.8247	5
	0.9303	0.8709	0.8414	10
	0.9395	0.9096	0.9009	30
	0.9441	**0.9101**	**0.9011**	300
Fold 4	0.8706	0.8028	0.7922	2
	0.8914	0.8247	0.8004	5
	0.9248	0.8637	0.8592	10
	0.9294	0.8949	0.8958	30
	0.9459	**0.9222**	**0.9211**	300

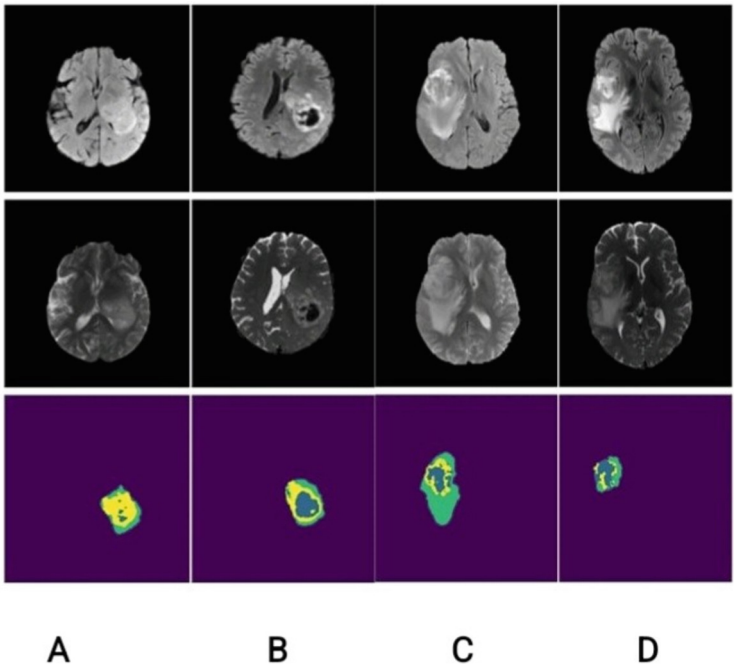

Fig. 3. Predicted Masks for validation data BraTS-GLI-00001-000 (A), BraTS-GLI- 00001-001 (B), BraTS-GLI-00013-000 (C), and BraTS-GLI-00013-001 (D), [T1 (top row), T2 (bottom row)] cases using the well-performing best 2D fullres nnUNet model without fine-tuning. The predicted masks (bottom row) are shown with color coding as follows: background: purple, necrotic tumor core (NCR): blue, enhancing tumor (ET): yellow, peritumoral edematous tissue (ED): turquoise. (color figure online)

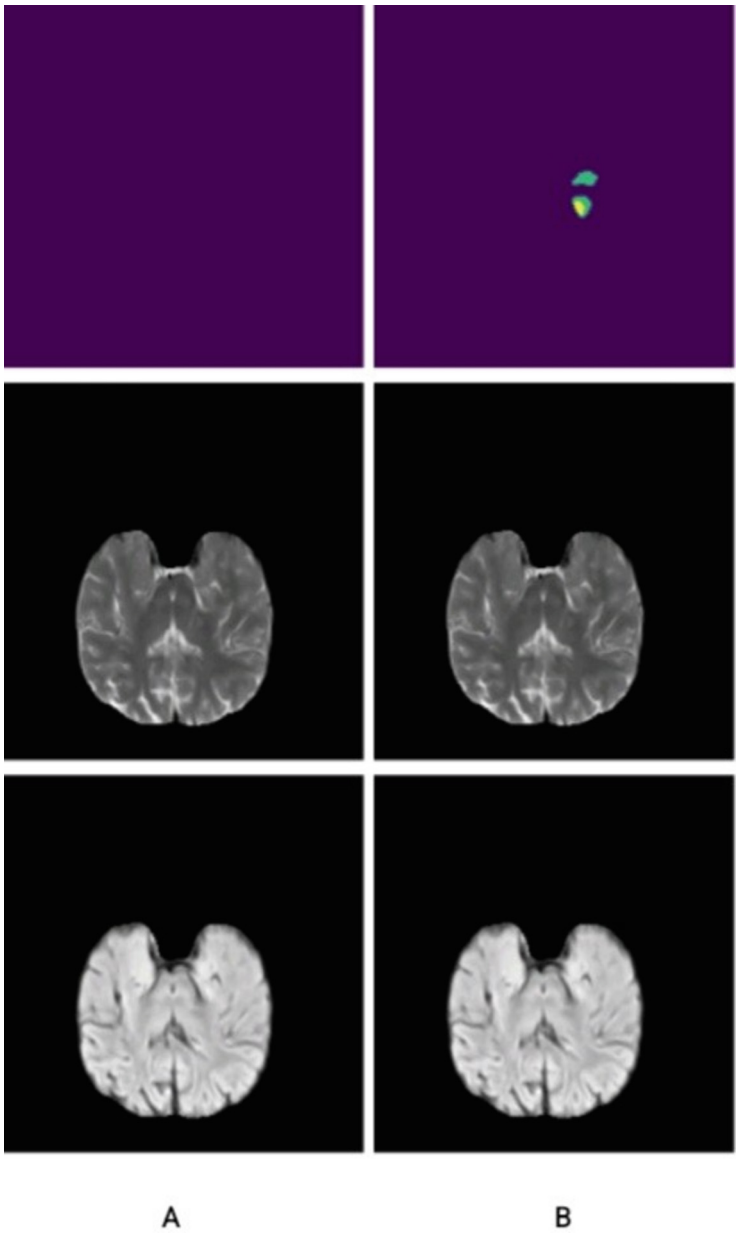

Fig. 4. Tumor segmentation improvement for SSA validation case BraTS-SSA-00192- 000 before (A) and after (B) neural style transfer data augmentation. [T1 (bottom row), T2(middle row)]. This was also after fine-tuning on SSA training data only from the best GLI pretrained model at the best 2D fullres nn-Unet. The color coding is the following: background: purple, necrotic tumor core (NCR): blue, enhancing tumor (ET): yellow, peritumoral edematous tissue (ED): turquoise. (color figure online)

References

1. Adewole, M., 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)
2. Liu, J., Li, M., Wang, J., Fangxiang, W., Liu, T., Pan, Y.: A survey of mri-based brain tumor segmentation methods. Tsinghua Sci. Technol. **19**(6), 578–595 (2014)
3. Feng, X., Tustison, N.J., Patel, S.H., Meyer, C.H.: Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features. Front. Comput. Neurosci. **14**, 25 (2020)
4. Anazodo, U.C., et al.: A framework for advancing sustainable magnetic resonance imaging access in Africa. NMR Biomed. **36**(3), e4846 (2023)
5. Zhang, D., Confidence, R., Anazodo, U.: Stroke lesion segmentation from low-quality and few-shot mris via similarity-weighted self-ensembling framework. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 87–96. Springer, Cham (2022)
6. Lin, H., et al.: Deep learning for low-field to high- field MR: image quality transfer with probabilistic decimation simulator. In: Machine Learning for Medical Image Reconstruction: Second International Workshop, MLMIR 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings, vol. 2, pp. 58–70. Springer, Cham (2019)
7. Hernandez, A.G., et al.: Improving image quality in low-field MRI with deep learning. In: International Conference on Information Photonics (2021)
8. Ma, C., Ji, Z., Gao, M.: Neural style transfer improves 3d cardiovascular MR image segmentation on inconsistent data (2019). arXiv:1909.09716 arXiv:1909.09716
9. Tomar, D., Bozorgtabar, B., Lortkipanidze, M., Vray, G., Rad, M.S., Thiran, J.P.:Self- supervised generative style transfer for one-shot medical image seg- mentation. In: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1737–1747. IEEE. Waikoloa (2022)
10. de Leeuw, M.L., den Bouter, G., Ippolito, T.P.A.O., Remis, R.F., van Gijzen, M.B., Webb, A.G.: Deep learning-based single image super-resolution for low-field mr brain images. Sci. Rep. **12**(1), 6362 (2022)
11. Sendra-Balcells, C., et al.: Generalisability of fetal ultrasound deep learning models to low-resource imaging settings in five African countries. Sci. Rep. **13**(1), 2728 (2023)
12. 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. Nat. Methods, 18(2), 203–211 (2021)
13. BraTS Challenge Organizers. Brats continuous evaluation sub-challenge 2021 (2021). https://www.synapse.org/Synapse:syn27046444/files/. Accessed 19 Aug 2024
14. Futrega, M., Milesi, A., Marcinkiewicz, M., Ribalta, P.: Optimized u-net for brain tumor segmentation. In: International MICCAI Brainlesion Workshop, pp. 15–29. Springer (2021)
15. Gatys, L.A., Ecker, A.S., Bethge, M.: A Neural Algo- rithm of Artistic Style (2015). arXiv:1508.06576
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (ICLR) (2015)