Feasibility of nnUNet Model in MR Image Segmentation: A Case of Glioma Segmentation in Sub-Saharan Africa

Rancy Chepchirchir
Institute of Mathematical Sciences,
Strathmore University, Kenya.
kosgeyrancy2@gmail.com

Jill Sunday
Department of Medical Engineering,
Technical University of Mombasa,
Mombasa, Kenya
jillselesa35@gmail.com

Raymond Confidence,
Medical Artificial Intelligence Laboratory (MAI Lab),
Lagos, Nigeria
spark@mailab.io

Dong Zhang,
Medical Artificial Intelligence Laboratory (MAI Lab),
Lagos, Nigeria
spark@mailab.io

Udunna C Anazodo,
Medical Artificial Intelligence Laboratory (MAI Lab),
Lagos, Nigeria
udunna.anazodo@mcgill.ca

Kendi Muchungi
Brain & Mind Institute, The Aga Khan University, Nairobi, Kenya
kendi.muchungi@aku.edu

Yujing Zou *
Medical Physics Unit, McGill University, Montreal, Canada
yujing.zou@mail.mcgill.ca

Abstract

The study looks into Brain Tumor Segmentation with Optimized UNet as its baseline model. The study also employs both versions 1 and 2 of the nnUNet model for 2, 5, 10, 30, and 300 epochs across the BraTS 2021 dataset and the BraTS 2023 African dataset. The results have shown a slight reduction in performance on the African dataset though this has been improved by further fine-tuning of the model. In fact, the study achieved a 92.56% dice similarity score on the African dataset for 300 epoch size.

*Corresponding author: yujing.zou@mail.mcgill.ca

5th Deep Learning Indaba Conference (DLI 2023).

1 Introduction

Brain tumors present a substantial health challenge in Africa. Statistically, the Sub-Saharan Africa experiences a percentage of 0.14% in disability-adjusted life years (DALYs) and 0.17% of overall mortality rates.North Africa is more impacted as opposed to the sub Saharan with brain tumors reckoning 0.44% of DALYs and 0.62% of deaths [1]. 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 are among the highest in the world, with the sub saharan experiencing a rise of 25% [2]. General medical providers need to possess a fundamental comprehension of the diagnosis and treatment of brain tumors because they are frequently encountered. The most prevailing types of brain tumors are the intracranial metastases originating from cancers in other parts of the body, meningiomas, and gliomas, notably glioblastoma.Central nervous system metastases can develop at several locales throughout the neuroaxis and mandate intergrated multidisciplinary management involving neurosurgery, radiation oncology, and medical oncology.[3]

Precise segmentation of distinct sub-regions within gliomas such as peritumoral edema, necrotic core, enhancing, and non-enhancing tumor core, based on multimodal MRI scans, hold material clinical relevance for the diagnosis, prognosis, and treatment of brain tumors. However, the segmentation of these sub-regions presents a formidable challenge due to the heterogeneity of brain tumors[4] . The brain's heterogeneity makes procuring the complete tumor volume arduous [5]. Additionally, severe data imbalance exists not only between tumor and non-tumor tissues but also among different sub-regions within the tumor [6]. Imaging assumes a primary role in diagnosing and strategizing treatment for brain tumors. Accurate brain tumor segmentation is crucial, more so when subtle, irregular morphological changes within the tumor make assessment taxing through medical evaluation[7]. 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 [8] .This information guides surgical planning, enhances the fine-tumig of radiation therapy, and facilitates the evaluation of treatment response over time.

Brain MR images acquired in the sub Saharan Africa experience suboptimal image contrast and resolution [2] which may be due to use of low field MRI scanners [9]. Magnetic resonance images obtained using low magnetic field strength, typically below 1 Tesla (T), usually have impaired resolution along the slice direction and lower contrast. This is principally attributed to the relatively small signal-to-noise ratio (SNR) compared to MR images acquired at higher field strengths, typically 1.5T and 3T [10].

The manual segmentation of brain tumors from MRI images for cancer diagnosis is a challenging, laborious, and time-intensive process, prone to inter-observer variability [11] [12].This discrepancy can greatly impact the integrity and reproducibility of segmentation outcomes. A single brain MRI scan encompasses multiple slices capturing the 3D anatomical perspective. This adds complexity to the manual segmentation of brain tumors MR images, making it a daunting task [12]. Furthermore, the automated classification of brain tumors from MRI scans offers a non-invasive approach, eliminating the need for biopsies and enhancing the safety of the diagnosis process.[13].Within these images, the tumor structure is solely defined by variations in grayscale, and the resulting images obtained from different equipment and under different conditions may exhibit dissimilarities. Consequently, traditional image segmentation methods face challenges in effectively handling the segmentation of brain tumor images. [13].Due to the limited amount of brain tumor information present in traditional single-mode MRI images, segmenting these images alone proves challenging in meeting clinical requirements.[13]. Given the cumbersome and dragging nature of manual segmentation, there is a need for automatic segmentation methods to ensure precise and prompt treatment [12].This leads to improved clinical decision-making and patient care. Automated segmentation of brain tumors(BraTS), using MRI scans, plays a crucial role in the accurate diagnosis and effective treatment of brain tumors [14]. Deep learning models can computerize the segmentation process, eliminating use of manual segmentation. Accurate and robust segmentation of brain tumors is of utmost importance for diagnosis, treatment planning, and evaluating treatment outcomes. Many automatic brain tumor segmentation methods rely on manually crafted features. Likewise, traditional deep learning approaches, like convolutional neural networks, necessitate substantial amounts of annotated data for effective learning, posing challenges in obtaining such data within the medical field [11]. Other challenges are low-contrast imaging and annotation bias [15].

Differences in image quality and imaging protocol can arise due to gaps in machines or resources [16]. For instance, machine variability can lead to variations in image quality, resolution, and levels of noise [17]. Another research limitation pertains to resource constraints, which can hinder the acquisition of high-resolution images. Research gaps that impact image quality include imaging artifacts, the absence of standardized imaging protocols, inconsistencies in contrast agents, variations in technical expertise, and the use of outdated or subpar imaging equipment [18].

The use of lower quality MRI technology[19] in the SSA raises uncertainty about the feasibility of implementing ML methods for clinical purposes [9] [20]. Brain MRI scans typically obtained in SSA exhibit inadequate image contrast and resolution [10], necessitating advanced image pre-processing to enhance their resolution before employing ML techniques for tasks like tumor segmentation, classification, or outcome prediction [2]. The intent of our method is to solve the problems mentioned. We aim to further pre-process the image data in order to improve image quality and improve generalizability.

2 Methods

2.1 Implementation

Our solution, which is implemented with PyTorch, is an extension of NVIDIA's nn-UNet, which is publicly available on GitHub. 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.

2.2 Datasets

This dataset represents the largest publicly-available retrospective cohort of adult Africans with pre-operative glioma, covering both low-grade glioma (LGG) and high-grade glioma (GBM/HGG). It comprises routine multi-parametric MRI (mpMRI) scans acquired during standard clinical care, encompassing T1, post-contrast T1 (T1Gd), T2, and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) MRI. The data underwent standardized pre-processing using the BraTS workflow, involving conversion to the NIfTI format, co-registration to the SRI24 template, isotropic resolution resampling (1mm3), and skull-stripping. Researchers can access the pre-processing pipeline through CaPTk and FeTS. The ground truth labels were generated following the BraTS annotation protocol. Initial automated segmentations were manually refined by trained radiology experts and further reviewed by senior attending radiologists. This iterative process ensured the accuracy and reliability of tumor sub-region segmentations approved by experienced neuro-radiologists.

Figure 2.2, is a representation of a sample MRI image across the 4 modalities (Flair, T1, T1CE and T2-weighted respectively) a ground truth mask from the 2021 subset employed in the study.

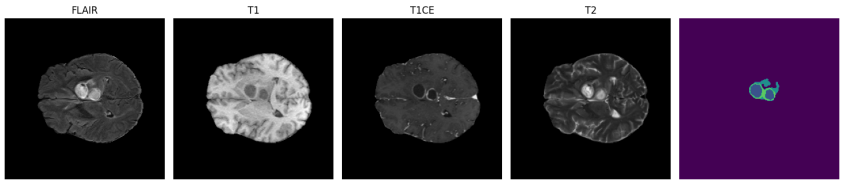

Figure 1: Ground Truth from 2021 subset

Figure 2.2 is a representation of a sample MRI image across the 4 modalities (Flair, T1, T1CE, and T2-weighted respectively) and a ground truth mask from the SSA subset employed in the study.

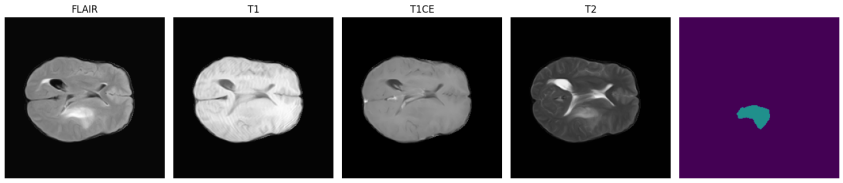

Figure 2: Ground Truth Mask from SSA subset

The variation in pixel values between the GLI and SSA datasets is further graphically demonstrated in figure 3

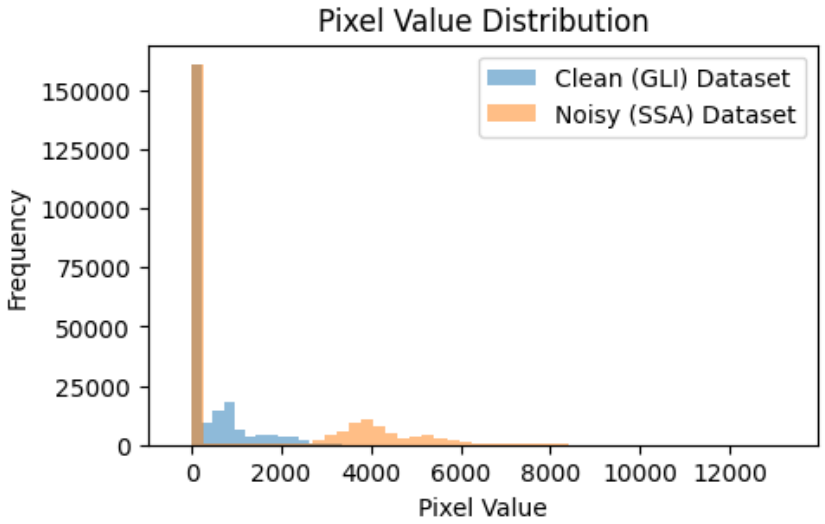

Figure 3: Variation in pixel values between GLI and SSA datasets

2.3 Data Preprocessing

The preprocessing step [21] employed by our group involves several key operations to prepare medical imaging data for subsequent analysis. The preprocessing pipeline is designed to handle various tasks and is flexible to accommodate different modalities, such as CT or MRI scans. The goal of the preprocessing step is to ensure data consistency, enhance image quality, and standardize the format for further processing.

The process begins by loading the dataset, which is organized according to a specified data path. Metadata associated with the dataset is extracted from a JSON file, providing crucial information about the modality, labels, and other relevant details. The modality information guides subsequent preprocessing steps, as certain operations differ depending on whether the modality is CT or MRI.

To improve the quality and consistency of the images, a crop foreground operation is applied. This step removes unnecessary background regions, focusing on the anatomical structures of interest. Additionally, intensity normalization is performed to ensure consistent intensity ranges across the images. For MRI scans, normalization is applied only to the non-zero regions, while for CT scans, additional intensity clipping is performed to limit values within the pre-defined minimum and maximum ranges.

Resampling is a critical step to address variations in voxel spacing among different scans. The target spacing, determined based on the dataset's median spacing, is used to resample the images and labels to a common voxel resolution. This step helps to ensure uniformity and facilitates comparison and analysis across different scans.

2.4 Baseline Model (Optimized U-Net)

The baseline for our model is inspired by the work of Futrega et al [21], the Optimized U-net model with deep supervision. Deep supervision is a method that improves gradient flow by calculating the loss function at various decoder levels. The Optimized U-Net model with deep supervision has an encoder depth of 7; the number of convolutional channels at each encoder level are: 64, 96, 128, 192, 256, 384, 512; has 3 output channels; and has 4 input channels with an additional one-hot-encoding channel for the foreground voxels to input data (5, 240, 240, 155) (C, H, W, D).

2.5 nnU-Net

nnU-Net is an image segmentation approach that independently conforms to a designated dataset [22]. The nnU-Net analyzes the given training dataset and autonomously sets up a U-Net- based segmentation pipeline that is similar to the data. Subsequently, nnU-Net generates multiple U-Net configurations for each dataset. This system basically simplifies the training of models. It is capable of processing both 2D and 3D images, regardless of the input modalities or channels. Even in situations with highly imbalanced classes, it can effectively comprehend voxel spacings and anisotropies with robustness. While the performance of segmentation remains the same, the version 2 of nnU-Net has some more features that have been added. These additional features make it easier to use as a development framework and also simplifies manually adjusting its configuration to new datasets. These new features include hierarchical labels, unified trainer class, cross-platform support specifically Cuda, mps (Apple M1/M2) and CPU support. Furthermore, nnU-Net version 2 offers expanded support for input and output data formats using ImageIO classes. To extend the I/O formats, you can implement new Adapters based on `BaseReaderWriter`. Moreover, it is faster to save cropped npz files. Upon unpacking, the preprocessed data and segmentation are stored in separate files. In version 1, the segmentation data was stored as int32, whereas in the current version, it is stored as int8. Consequently, the segmentation now occupies only one-fourth of the disk space per pixel, resulting in improved I/O throughput. In the current version, the framework offers built-in support for multi-GPU training using Distributed Data-Parallel (DDP). All functionalities of nnU-Net are now accessible through the API as well.Additionally, the dataset fingerprint is now generated and explicitly saved in a JSON file, just to name a few.

2.6 Training Schedule

Each experiment was trained for 2, 5, 10, 30, and 300 epochs using the Adam optimizer with varying learning rates. We also employed 5-fold cross-validation in model evaluation based on the mean dice score after each epoch. We then stored both the latest or final checkpoint and the best checkpoint model based on the Dice score on the validation dataset for use during inference.

2.7 Performance Evaluation

We used the dice similarity score to evaluate the accuracy of our models and also took into account also the model complexity in terms of both computational time and the size of memory storage required.

3 Results

3.1 Baseline Model (Optimized U-Net)

3.1.1 2021 Dataset

Table 3.1.1 represents the dice score, training, and validation loss from the 2021 dataset:

3.1.2 GLI + SSA dataset

Table 3.1.2represents the dice score, training, and validation loss from both the GLI and SSA 2023 dataset:

Epochs	Dice score	Train loss	Val loss
2	79.1	1.8259	0.209
5	84.87	1.1213	0.1513
10	86.33	0.8512	0.1367
30	89.69	0.5864	0.1031

Table 1: Results from training the Optimized U-Net on the 2021 dataset

Epochs	Dice score	Train loss	Val loss
2	77.07	1.8987	0.2293
5	84.59	1.0124	0.1541
10	86.46	0.7566	0.1354
30	88.16	0.5685	0.1184

Table 2: Results from training the Optimized U-Net on the GLI+SSA training dataset

From the above results, we have demonstrated a decrease in the model performance despite there being a considerably lower number of the Sub-Saharan African dataset (i.e. 60 cases) which accounts for only 4.6% of the total 1,311 training dataset.

Having gone through the SSA dataset case-by-case, we tried to also exclude the cases where the MR images were taken from a coronal plane section as opposed to the rest which were taken at axial plane sections. These four unique cases included; 00051, 00097, 00041, and 00068 cases as also visually represented in figure 4;

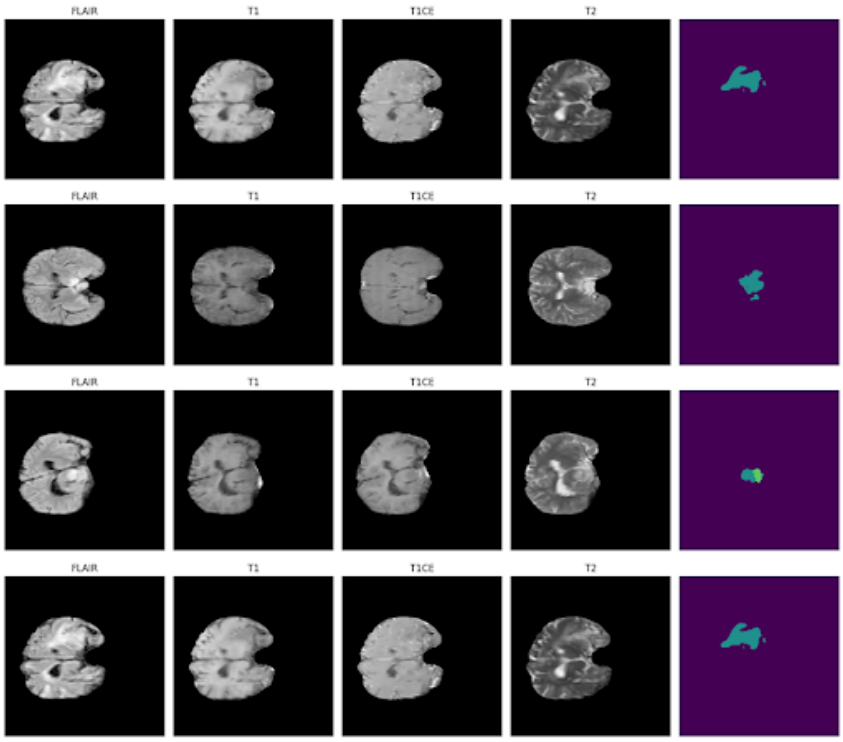

Figure 4: Excluded SSA Cases (from top to bottom) 00051, 00097, 00041, & 00084

Table 3.1.2 represents the results obtained after the exclusion of the above 4 cases on the SSA dataset. From the table, we see a slight increase in dice score for 30, 10, and 2 epochs from 88.16 to 88.58;

86.46 to 86.62; and 77.07 to 79.56 respectively. However, this was only an experiment and all cases were used in all the trainings done on this study.

Epochs	Dice score	Train loss	Val loss
2	79.56	1.581	0.2044
5	83.16	1.0121	0.1684
10	86.62	0.5574	0.1338
30	88.58	0.5539	0.1142

Table 3: Results from training the Optimized U-Net on the GLI+SSA dataset minus 4 cases

3.2 nnUNet version 2

We did a couple of ablation studies in order to select the most optimal neural network architecture for 0, 1, 2, 3, and 4 folds under the nnUNet V2 given below. For the 2021 dataset, the `splits_final.json` was employed with a split of 801 cases for training and the remaining 200 cases for validation. The split contains 5 splits with a desired fold for training as 4 for 3d full resolution configuration.

The graph shown below is a representation of the prediction mask overlayed on the four MR Images:

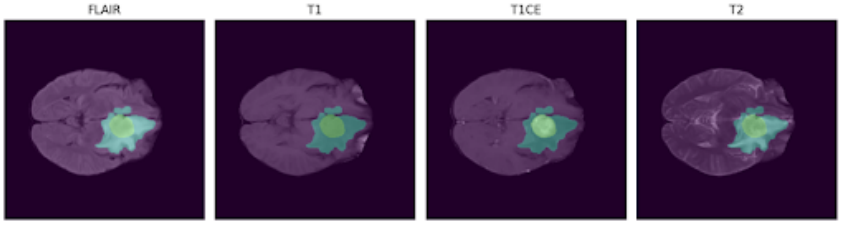

Figure 5: The Predicted mask under the 300 epoch model overlayed on the four MR Images

3.2.1 Configuration name: 3d full resolution

Here, we have employed a 3d full resolution with a batch_size of 2, a patch_size of [128, 128, 128], 32 UNet 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 the 2021 and the 2023 GLI and SSA dataset, as represented in table 3.2.1 and 3.2.1 respectively.

Epoch	Learning rate	Train_loss	Val_loss	Pseudo dice	Epoch time
2	0.01	0.00536	-0.6746	0.7833	209.28 s
5	0.00235	-0.7353	-0.7411	0.8146	191.05 s
10	0.00126	-0.7959	-0.7784	0.8305	194.02 s
30	0.00047	-0.8341	-0.8523	0.8873	190.9 s
300	6e-05	-0.8844	-0.8664	0.9471	201.65 s

Table 4: 3D full resolution on the GLI+SSA 2021 Dataset

From the table below; we can see that the 300 epoch model achieved the best dice score at 0.9256 followed by the 150 and 30 epoch models at 0.9198 and 0.8965 respectively.

The graphs shown in figure 6 are also a representation of the training progress for the best epoch for 3d configuration.

3.2.2 Configuration name: 2d full resolution

Here, we have employed a 2d full resolution with a batch_size of 105, a patch_size of [192, 160], 32 UNet 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]]

Epoch	Learning rate	Train_loss	Val_loss	Pseudo dice	Epoch time
2	0.01	0.00536	-0.6746	0.7833	209.28 s
5	0.00235	-0.7353	-0.7411	0.8146	191.05 s
10	0.00126	-0.7959	-0.7784	0.8305	194.02 s
30	0.00047	-0.8341	-0.8523	0.8965	190.9 s
150	0.00536	-0.8668	-0.8695	0.9198	495.48 s
300	6e-05	-0.8891	-0.8561	0.9256	196.42 s

Table 5: A representation of the average of class prediction dice scores per epoch number

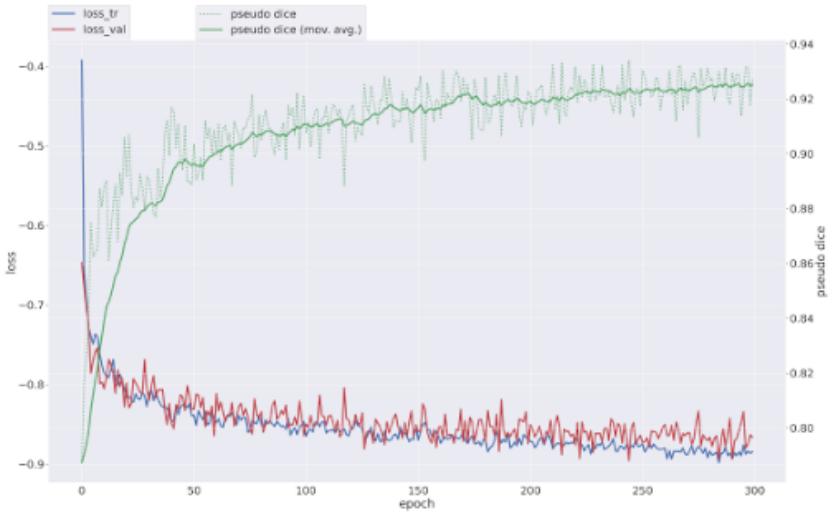

Figure 6: Graphical Representation of training progress for 300 epochs

on the GLI and SSA dataset. As seen in the table below, the 30-epoch model achieves the best dice score at 0.8965, followed by the 5-epoch model at 0.8629.

Epoch	Learning rate	Train_loss	Val_loss	Pseudo dice	Epoch time
2	0.00536	-0.7974	-0.8236	0.8432	337.17 s
5	0.00235	-0.8625	-0.8687	0.8629	165.65 s
10	0.00126	-0.7959	-0.7784	0.8305	195.28 s
30	0.00047	-0.8341	-0.8523	0.8965	190.9 s

Table 6: Comparison of performance with epoch size

These are the results for the African dataset (GLI + SSA). From the table below; we can observe that a higher epoch size, in this case 30, achieves the best performance across the 5-fold validation; followed by the 10 epoch size models; then the 5-epoch models, and lastly the 2-epoch size models.

4 Discussion and Future Work

From Tables 1, 2, and 3 above, we have seen that as the number of training epochs increases, the Dice score improves significantly, demonstrating that the model's segmentation performance gets better with more training. The reduction in both training and validation loss over epochs indicates that the model is effectively learning and generalizing well to unseen data. Moreover, training the model for approximately 300 epochs strikes a balance between achieving good segmentation performance and minimizing training time from Table 5. Fine-tuning the learning rate or exploring other optimization techniques could potentially further improve the model's performance.

Epoch	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
2	0.7658	0.7466	0.8175	0.7882	0.8432
5	0.8158	0.813	0.78	0.7916	0.7945
10	0.8823	0.8784	0.8832	0.8795	0.8778
30	0.895	0.8895	0.9006	0.8972	0.9244

Table 7: A comparison of performance across the folds

Additionally, it's essential to consider the trade-off between training time and performance when deciding the optimal number of epochs for model training. The training done with 2d configuration also tends to outperform that with 3d configuration for a smaller number of epochs (at least that was our case for up to 30 epochs from the tables above). A higher epoch size translates into better performance as the model trains more on the dataset and is able to increase its accuracy in the segmentation as it learns more patterns.

For future work, we can tackle this as a domain adaptation problem that can arise from any multi-center/institutional data. More specifically, the problem arises from the low-field MR producing lower quality images, we therefore considered the baseline nnUnet version 2 for both 2D and 3D methods but with the variation of self-supervised style transfer (this has been recently proposed, but have not been employed on the African dataset.

We could also employ the ensembling methods; style transfer methods, low to high-resolution quality transfer learning approach from model weights from Bouter et.al. Having validated the nnUnet v2 feasibility on the African dataset in this study, we will test our approach to a larger African dataset cohort (currently there are only 60 cases). Since we were able to achieve a 92.56% dice similarity score with 300 epochs, a higher epoch size could still be trained to check for further improvements in the score.

5 Conclusion

We have experimented with various U-Net variants including the base U-Net-V1, Optimized U-Net, and the nnUNet-V2. Based on our experiments, training a model with a higher epoch size, say 300, yields better results as opposed to a smaller epoch size though we still also have to take into consideration the model complexity to avoid overfitting.

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; [add all instructors of the hackathon & Yahoo Liu in alphabetical order, without their salutations]. The authors would also like to thank Talhah 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 which the authors immensely benefited from. 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).

References

[1] Setthasorn Zhi Yang Ooi, Rosaline de Koning, Abdullah Egiz, David Ulrich Dalle, Moussa Denou, Marvin Richie Dongmo Tsopmene, Mehdi Khan, Régis Takoukam, Jay Kotecha, Dawin Sichimba, et al. Management and outcomes of low-grade gliomas in africa: a scoping review. *Annals of Medicine and Surgery*, page 103246, 2022.

[2] Maruf Adewole, Jeffrey D Rudie, Anu Gbadamosi, Oluyemisi Toyobo, Confidence Raymond, Dong Zhang, Olubukola Omidiji, Rachel Akinola, Mohammad Abba Suwaid, Adaobi Emegoakor, 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] J Ricardo McFaline-Figueroa and Eudocia Q Lee. Brain tumors. *The American journal of medicine*, 131(8):874–882, 2018.

[4] Xue Feng, Nicholas J Tustison, Sohil H Patel, and Craig H Meyer. Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features. *Frontiers in computational neuroscience*, 14:25, 2020.

[5] Mukul Aggarwal, Amod Kumar Tiwari, M Partha Sarathi, and Anchit Bijalwan. An early detection and segmentation of brain tumor using deep neural network. *BMC Medical Informatics and Decision Making*, 23(1):1–12, 2023.

[6] Pei Wang and Albert CS Chung. Relax and focus on brain tumor segmentation. *Medical Image Analysis*, 75:102259, 2022.

[7] RB Dubey, Madasu Hanmandlu, and Shantaram Vasikarla. Evaluation of three methods for mri brain tumor segmentation. In *2011 eighth international conference on information technology: new generations*, pages 494–499. IEEE, 2011.

[8] Jin Liu, Min Li, Jianxin Wang, Fangxiang Wu, Tianming Liu, and Yi Pan. A survey of mri-based brain tumor segmentation methods. *Tsinghua science and technology*, 19(6):578–595, 2014.

[9] Udunna C Anazodo, Jinggang J Ng, Boaz Ehiogu, Johnes Obungoloch, Abiodun Fatade, Henk JMM Mutsaerts, Mario Forjaz Secca, Mamadou Diop, Abayomi Opadele, Daniel C Alexander, et al. A framework for advancing sustainable magnetic resonance imaging access in africa. *NMR in Biomedicine*, 36(3):e4846, 2023.

[10] Hongxiang Lin, Matteo Figini, Ryutaro Tanno, Stefano B Blumberg, Enrico Kaden, Godwin Ogbole, Biobele J Brown, Felice D'Arco, David W Carmichael, Ikeoluwa Lagunju, 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 2*, pages 58–70. Springer, 2019.

[11] Muhammad Imran Razzak, Muhammad Imran, and Guandong Xu. Efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks. *IEEE journal of biomedical and health informatics*, 23(5):1911–1919, 2018.

[12] Jiawei Sun, Wei Chen, Suting Peng, and Boqiang Liu. Drrnet: dense residual refine networks for automatic brain tumor segmentation. *Journal of medical systems*, 43:1–9, 2019.

[13] Erena Siyoum Biratu, Friedhelm Schwenker, Yehualashet Megersa Ayano, and Taye Girma Debelee. A survey of brain tumor segmentation and classification algorithms. *Journal of Imaging*, 7(9):179, 2021.

[14] Nima Hashemi, Saeed Masoudnia, Ashkan Nejad, and Mohammad-Reza Nazem-Zadeh. A memory-efficient deep framework for multi-modal mri-based brain tumor segmentation. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3749–3752. IEEE, 2022.

[15] Zhihua Liu, Lei Tong, Long Chen, Zheheng Jiang, Feixiang Zhou, Qianni Zhang, Xiangrong Zhang, Yaochu Jin, and Huiyu Zhou. Deep learning based brain tumor segmentation: a survey. *Complex & intelligent systems*, 9(1):1001–1026, 2023.

[16] Mariana Bento, Irene Fantini, Justin Park, Leticia Rittner, and Richard Frayne. Deep learning in large and multi-site structural brain mr imaging datasets. *Frontiers in Neuroinformatics*, 15:805669, 2022.

[17] Colin J Holmes, Rick Hoge, Louis Collins, Roger Woods, Arthur W Toga, and Alan C Evans. Enhancement of mr images using registration for signal averaging. *Journal of computer assisted tomography*, 22(2):324–333, 1998.

[18] Damon M Chandler. Seven challenges in image quality assessment: past, present, and future research. *International Scholarly Research Notices*, 2013, 2013.

[19] Naoto Hayashi, Yasushi Watanabe, Tomohiko Masumoto, Harushi Mori, Shigeki Aoki, Kuni Ohtomo, Osamu Okitsu, and Tetsuhiko Takahashi. Utilization of low-field mr scanners. *Magnetic resonance in medical sciences*, 3(1):27–38, 2004.

[20] Dong Zhang, Raymond Confidence, and Udunna Anazodo. 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*, pages 87–96. Springer, 2022.

[21] Michał Futrega, Alexandre Milesi, Michał Marcinkiewicz, and Pablo Ribalta. Optimized u-net for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*, pages 15–29. Springer, 2021.

[22] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.