Automated Segmentation for the Brain Tumor Segmentation (BraTS) Metastases 2025 Challenge Using Multi-Architectural Deep Learning

Wes Krikorian[1], Fabian Umeh[2], Nikolay Yordanov[3], Crystal Chukwurah[4], Nazanin Maleki[5], Ananya Purwar[6], Mariam Aboian[5]

[1]Horace Mann School, Bronx, NY, USA, [2]Teesside University, Middlesbrough, UK, [3]Medical University of Sofia, Sofia, Bulgaria, [4]Yale University, New Haven, CT, USA, [5]University of Pennsylvania, Philadelphia, PA, USA, [6]Harvey Mudd College, Claremont, CA, USA

Abstract. Brain metastases are the most frequently diagnosed brain tumors and are associated with significant clinical and economic burdens. Accurate segmentation of metastases is essential for minimally invasive treatment but remains a time-consuming and labor-intensive process. This work presents a machine learning-based segmentation pipeline to compete in the Brain Tumor Segmentation (BraTS) Metastases 2025 Challenge, using the MONAI Auto3DSeg framework with three neural network architectures: SegResNet, DiNTS, and Swin-UNETR. The model was trained on 1,296 cases, each including four sequences (T1, T1-post contrast, T2, and FLAIR), with an NVIDIA ADA 6000 GPU. Training was done with two-fold cross-validation on 100 epochs with early stopping after 5 epochs. Once training was completed, the models each made predictions on 179 unlabeled cases, which were then post-processed by removing noise, enforcing a label hierarchy, and filling in holes. Then, the final predictions were submitted to the BraTS Synapse page for model evaluation. The final predictions achieved Dice Similarity Coefficients (DSC) of 0.87 for the resection cavity, 0.69 for the tumor core, 0.68 for the whole tumor, and 0.66 for the enhancing tumor. These results highlight the combination of multiple architectures and a robust post-processing pipeline to improve segmentation accuracy. Future work will focus on using more cross-validation folds, optimizing model architectures through neural architecture search, and expanding training data to improve the detection of smaller tumor regions.

Keywords: Brain Metastases, Residual Neural Network, Transformer, Neural Architecture Search

1. Introduction

Brain metastases are the most commonly diagnosed brain neoplastic lesions and occur as a complication in patients with primary solid tumors in the lung, breast, colon, etc [5,9]. In some studies, the percentage of patients with primary cancer who develop brain metastatic disease amounts to 30% with correspondingly low overall survival rates [3].

Accurate segmentation of brain metastases is crucial for diagnosis, treatment planning, and monitoring. However, this requires extensive manual effort to ensure the precision required for therapies, such as stereotactic radiosurgery, underscoring the need for a reliable automated solution. This scenario presents a significant opportunity for AI innovation.

The brain metastases segmentation model presented in this scientific work was created using MonAI's Auto3dSeg open-source project, specifically, the SegResNet, Swin-UNETR, and DiNTS architectures [4].

2. Methods

The BraTS-METs 2025 challenge includes two unique datasets: 1296 labeled cases and a 176-unlabeled dataset. The larger, labeled dataset was used for training, while the smaller, unlabeled dataset was used for validation.

The dataset includes pre-contrast T1-weighted (T1W), post-contrast T1-weighted (T1C), T2-weighted (T2W), and T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) sequences. Each scan is annotated for four key tumor subregions: the tumor core (TC), the whole tumor (WT), the enhancing tumor (ET), and the resection cavity (RC). The dataset used to train the

model represents a large-scale, multi-institutional collection of brain MRI scans for metastasis segmentation. Ultimately, the trained model generates the predictions that would be subsequently uploaded to the BraTS Synapse page for evaluation.

2.2 Annotation Protocol

Annotations were generated using a multi-step pipeline to address the size of the data and ensure clinical accuracy. An automated metastasis segmentation tool first produced initial labels, which were then manually refined by trained medical students. These refined labels were subsequently reviewed and verified by neuroradiologists. This approach ensures high-quality ground truth to train accurate models.

2.3 Architecture

To enhance the robustness of final predictions and leverage the complementary strengths of different architectures, we trained models under three distinct frameworks: SegResNet, base-level DiNTS, and Swin-UNETR. This multi-architecture approach allowed certain models to compensate for the weaknesses of others, improving overall segmentation performance. SegResNet, base DiNTS, and Swin-UNETR represent three neural network architectures commonly used in medical image segmentation, each suited to different use cases. SegResNet, a convolutional residual network, is known for its efficiency and strong performance on standard segmentation tasks, making it ideal for quick experimentation and deployment in resource-limited environments. The base DiNTS model serves as the starting point for neural architecture search (NAS), offering a generic UNet-style structure that can be optimized to fit the characteristics of a specific dataset better. While its baseline performance is modest, it lays the

groundwork for more powerful custom models. Swin-UNETR, a transformer-based architecture, excels in capturing long-range spatial dependencies and is particularly well-suited for complex, multimodal imaging tasks such as brain tumor segmentation. Although computationally intensive, it often delivers state-of-the-art results when sufficient data and GPU resources are available. Together, these models cover a wide spectrum of segmentation needs, from baseline efficiency to maximum accuracy.

2.4 Pre-processing

All cases in the BraTS-METs 2025 dataset underwent pre-processing by the BraTS organizers. The pipeline started with a format conversion where the original Digital Imaging and Communication in Medicine (DICOM) files were converted into the Neuroimaging Informatics Technology Initiative (NIfTI) format. During this step, all private patient information was removed. Next, spatial registration was conducted: all sequences were aligned so the tumors would be in the same spot for each sequence. Then the images were resampled to an isotropic resolution of 1 mm^3. Lastly, a final anonymization was done by using a deep learning model to skull-strip the image for further patient privacy.

Then, when training the model, MONAI's Autorunner pre-processed the data so it would be compatible with the SegResNet architecture. It changed the input dimensions from 3D to 5D; a channel dimension of four was added (one for each MRI modality), and a batch dimension of one was added. The Autorunner's pre-processing pipeline also resampled voxel spacing and voxel intensity. Lastly, it cropped the image to the annotated portions to remove unnecessary background.

The final step of pre-processing involved assembling all the data into a metadata.json file, allowing the model to easily access the dataset. This metadata file included three keys: training, validation (which was unused), and testing. All 1,296 labeled cases were assigned to the training key, which was then split into five folds. The 176 unlabeled cases were placed under the testing key.

2.5 Training

Model training was performed using an NVIDIA Ada 6000 GPU, chosen for its high memory capacity suitable for large 3D medical imaging workloads. Although the dataset was split into five folds for cross-validation, only models for folds 0 and 1 were trained due to time and computational resource constraints. Each model was trained for up to 100 epochs, with early stopping applied after five consecutive epochs without improvement in validation performance to prevent overfitting and reduce training time. A patch size of 160×160×160 and a batch size of 1 were used to fit the models within GPU memory limits. Additionally, a low cache rate was configured in the Auto3DSeg training pipeline to manage system memory during data loading. Lastly, the learning rate was set to 1e-4 with a warmup cosine scheduler to avoid large early updates, which could potentially harm the model. This setup balanced performance, resource usage, and time constraints while still allowing for meaningful model evaluation.

2.6 Testing

Building on the multi-architecture training approach, the post-processing pipeline further refines segmentation predictions to enhance accuracy and clinical relevance. First, ensembling combines outputs from the SegResNet, base DiNTS, and Swin-UNETR models to leverage their

complementary strengths, producing more robust and stable segmentations. Next, hierarchy enforcement ensures that the segmented regions respect anatomical relationships, such as containment or exclusion rules, which is critical in complex structures like brain tumors. Small, isolated clusters (five voxels or fewer), often representing noise or false positives, are then removed to clean the segmentation maps. Finally, any holes or gaps within segmented regions are filled to generate smooth, continuous structures. This carefully designed post-processing pipeline transforms the combined raw model outputs into precise and reliable segmentations.

2.7 Evaluation Metrics

For the validation and testing phases of the BraTS challenge, the model would be evaluated by the Dice Similarity Coefficient (DSC) and Normalized Surface Distance (NSD) on the tumor core, whole tumor, enhancing tumor, and the resection cavity.

3. Results

During model training, the SegResNet model performed a validation epoch every three training epochs (the Swin-UNETR and DiNTS models didn't perform any validation due to time constraints). The SegResNet model peaked at an average DSC of 0.64 across all classes. The fully ensembled model demonstrated improvement in performance: from the validation phase, the enhancing tumor had a DSC of 0.66 and an NSD of 0.47; the resection cavity had a DSC and NSD of 0.87; the tumor core had a DSC of 0.69 and an NSD of 0.47; the whole tumor had a DSC of 0.68 and an NSD of 0.35.

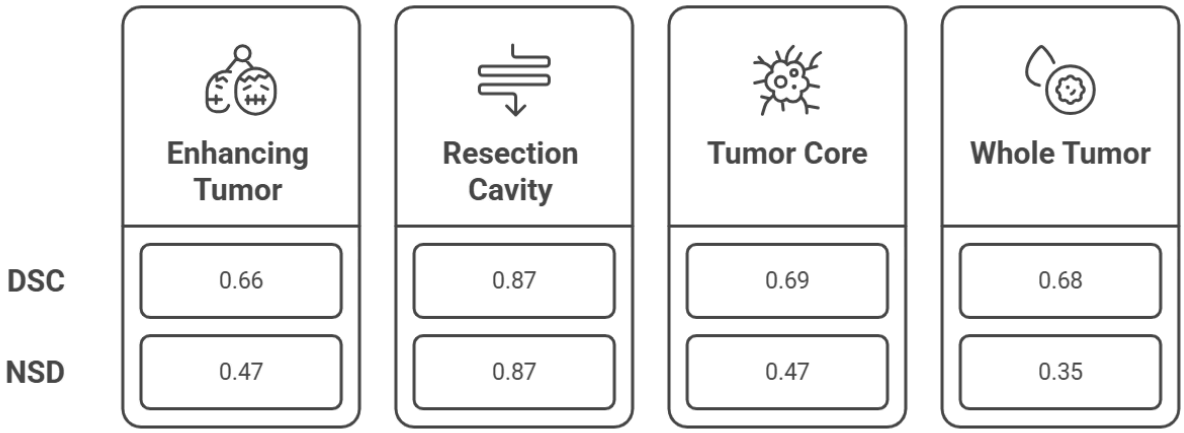

Figure 3.1 shows the dice similarity coefficient and normalized surface distance for each of the classes.

4. Discussion

Despite evolving treatment, brain metastatic disease is associated with substantial morbidity and mortality [6,9]. The Graded Prognostic Assessment (GPA) varies depending on the primary cancer site and the selected treatment approach and strategy [11]. In a recent Norwegian study of 912 patients with first-time brain metastasis, the mean overall survival for the entire cohort was 5.9 months, with 50% of the patients succumbing to their disease within 6 months after the diagnosis had been established [12]. In addition to the unfavorable prognosis, brain metastases incur a significant economic burden. A recent nationwide study conducted in the US found that patients with lung cancer with central nervous system (CNS) spread incurred healthcare costs that were approximately $13,000 higher than those with lung cancer but no brain metastases [10].

According to the Congress of Neurological Surgeons guidelines on brain metastases treatment, surgery plus whole-brain radiotherapy (WBRT) is the recommended initial approach

for patients with first-time diagnosed brain metastatic disease [8]. However, WBRT is associated with significant side effects such as, but not limited to, fatigue, nausea, anorexia, cognitive dysfunction, and hearing loss [1]. In contrast to WBRT, stereotactic radiosurgery (SRS) has a less negative effect on neurocognition and quality of life [6]. In addition, Level 3 evidence exists that stereotactic radiosurgery alone can be considered as an alternative treatment to surgery plus WBRT, with the caveat that the patient has oligometastatic disease, i.e., 1 to 4 lesions [2,8].

For the radiation delivered by SRS to be highly focused, precise contouring of the lesions is performed by radiation oncologists using a radiotherapy planning software. This approach is tedious and time-consuming [13]. Deep Learning (DL) models for brain metastases segmentation, such as the MetNet developed by Zhou et al. [13], have been proposed as a tool to mitigate the challenges with manual lesion delineation. However, to our knowledge, there is no widely distributed tool that is routinely used in clinical practice to autosegment and quantify the volume of intracranial lesions.

The ensemble model achieved the highest segmentation accuracy on the resection cavity (DSC of 0.87), while performance on the enhancing tumor was more modest (DSC of 0.66). These results highlight the effectiveness of both the post-processing pipeline and the ensembling strategy in improving segmentation quality. The strong performance on the resection cavity is likely due to its well-defined and consistent appearance across patients. In contrast, the lower Dice score for the enhancing tumor may stem from its smaller size, variability in appearance, and lower prevalence within the dataset.

In a clinical setting, this model has the potential to significantly reduce the time burden associated with manual segmentation. Rather than delineating lesions from scratch, radiologists

could use the model's outputs as a baseline and make minor adjustments as needed, streamlining the workflow and increasing efficiency.

A major limitation of this study was time. Due to computational and scheduling constraints, only two out of the five cross-validation folds were trained. Additionally, the DiNTS model was used in its unoptimized, base form, without applying neural architecture search (NAS) to tailor it to the dataset. Model parameters were also adjusted for faster training rather than optimal performance. For example, training was capped at 100 epochs with an aggressive early stopping criterion, which may have limited the model's learning capacity.

Future work will address these limitations by completing training on all five folds, conducting a full architecture search for DiNTS, and tuning hyperparameters for performance rather than speed. Moreover, we plan on expanding the training dataset with more cases containing smaller, less common labels, such as enhancing tumors, which may improve sensitivity and segmentation accuracy for underrepresented regions.

References

1. Aizer, A. A., Lamba, N., Ahluwalia, M. S., Aldape, K., Boire, A., Brastianos, P. K., Brown, P. D., Camidge, D. R., Chiang, V. L., Davies, M. A., Hu, L. S., Huang, R. Y., Kaufmann, T., Kumthekar, P., Lam, K., Lee, E. Q., Lin, N. U., Mehta, M., Parsons, M., … Wen, P. Y. (2022). Brain metastases: A Society for Neuro-Oncology (SNO) consensus review on current management and future directions. *Neuro-Oncology*, *24*(10), 1613–1646. https://doi.org/10.1093/neuonc/noac118
2. Brown, P. D., Ahluwalia, M. S., Khan, O. H., Asher, A. L., Wefel, J. S., & Gondi, V. (2018). Whole-Brain Radiotherapy for Brain Metastases: Evolution or Revolution? *Journal of Clinical Oncology*, *36*(5), 483–491. https://doi.org/10.1200/JCO.2017.75.9589
3. Brown, P. D., Gondi, V., Pugh, S., Tome, W. A., Wefel, J. S., Armstrong, T. S., Bovi, J. A., Robinson, C., Konski, A., Khuntia, D., Grosshans, D., Benzinger, T. L. S., Bruner, D., Gilbert, M. R., Roberge, D., Kundapur, V., Devisetty, K., Shah, S., Usuki, K., … for NRG Oncology. (2020). Hippocampal Avoidance During Whole-Brain Radiotherapy Plus Memantine for Patients With Brain Metastases: Phase III Trial NRG Oncology CC001. *Journal of Clinical Oncology*, *38*(10), 1019–1029. https://doi.org/10.1200/JCO.19.02767
4. Consortium, T. M. (2020). *Project MONAI* (Version 0.4.0) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.4323059
5. Lamba, N., Wen, P. Y., & Aizer, A. A. (2021). Epidemiology of brain metastases and leptomeningeal disease. *Neuro-Oncology*, *23*(9), 1447–1456. https://doi.org/10.1093/neuonc/noab101
6. Mansouri, A., Ozair, A., Bhanja, D., Wilding, H., Mashiach, E., Haque, W., Mikolajewicz, N., De Macedo Filho, L., Mahase, S. S., Machtay, M., Metellus, P., Dhermain, F., Sheehan, J., Kondziolka, D., Lunsford, L. D., Niranjan, A., Minniti, G., Li, J., Kalkanis, S. N., … Ahluwalia, M. S. (2025). Stereotactic radiosurgery for patients with brain metastases: Current principles, expanding indications and opportunities for multidisciplinary care. *Nature Reviews Clinical Oncology*, *22*(5), 327–347. https://doi.org/10.1038/s41571-025-01013-1
7. Myronenko, A., Yang, D., He, Y., & Xu, D. (2023). *Automated 3D Segmentation of Kidneys and Tumors in MICCAI KiTS 2023 Challenge* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2310.04110
8. Nahed, B. V., Alvarez-Breckenridge, C., Brastianos, P. K., Shih, H., Sloan, A., Ammirati, M., Kuo, J. S., Ryken, T. C., Kalkanis, S. N., & Olson, J. J. (2019). Congress of Neurological Surgeons Systematic Review and Evidence-Based Guidelines on the Role of Surgery in the Management of Adults With Metastatic Brain Tumors. *Neurosurgery*, *84*(3), E152–E155. https://doi.org/10.1093/neuros/nyy542
9. Ostrom, Q. T., Wright, C. H., & Barnholtz-Sloan, J. S. (2018). Brain metastases: Epidemiology. In *Handbook of Clinical Neurology* (Vol. 149, pp. 27–42). Elsevier. https://doi.org/10.1016/B978-0-12-811161-1.00002-5
10. Shim, Y.-B., Byun, J.-Y., Lee, J.-Y., Lee, E.-K., & Park, M.-H. (2022). Economic burden of brain metastases in non-small cell lung cancer patients in South Korea: A retrospective cohort study using nationwide claims data. *PLOS ONE*, *17*(9), e0274876. https://doi.org/10.1371/journal.pone.0274876

11. Sperduto, P. W., Mesko, S., Li, J., Cagney, D., Aizer, A., Lin, N. U., Nesbit, E., Kruser, T. J., Chan, J., Braunstein, S., Lee, J., Kirkpatrick, J. P., Breen, W., Brown, P. D., Shi, D., Shih, H. A., Soliman, H., Sahgal, A., Shanley, R., … Mehta, M. P. (2020). Survival in Patients With Brain Metastases: Summary Report on the Updated Diagnosis-Specific Graded Prognostic Assessment and Definition of the Eligibility Quotient. *Journal of Clinical Oncology*, *38*(32), 3773–3784. https://doi.org/10.1200/JCO.20.01255
12. Yri, O. E., Astrup, G. L., Karlsson, A. T., Van Helvoirt, R., Hjermstad, M. J., Husby, K. M., Loge, J. H., Lund, J.-Å., Lundeby, T., Paulsen, Ø., Skovlund, E., Taran, M.-I., Winther, R. R., Aass, N., & Kaasa, S. (2025). Survival and quality of life after first-time diagnosis of brain metastases: A multicenter, prospective, observational study. *The Lancet Regional Health - Europe*, *49*, 101181. https://doi.org/10.1016/j.lanepe.2024.101181
13. Zhou, Z., Sanders, J. W., Johnson, J. M., Gule-Monroe, M., Chen, M., Briere, T. M., Wang, Y., Son, J. B., Pagel, M. D., Ma, J., & Li, J. (2020). MetNet: Computer-aided segmentation of brain metastases in post-contrast T1-weighted magnetic resonance imaging. *Radiotherapy and Oncology*, *153*, 189–196. https://doi.org/10.1016/j.radonc.2020.09.016