JOURNAL OF
EMERGING INVESTIGATORS

Article

Evaluating the clinical applicability of neural networks for meningioma tumor segmentation on 3D MRI

Diya Sreedhar[1], Anahita Fathi Kazerooni[2]

[1] Troy High School, Fullerton, California

[2] Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

SUMMARY

Magnetic resonance imaging (MRI), in both 2D and 3D configurations, serves as the leading non-invasive and non-ionizing technique for the detection and clinical assessment of brain tumors. 3D multiparametric MRIs (mpMRIs) offer enhanced spatial and biological context over other MRIs, which can help clinicians plan tailored treatment and therapy. Presently, surgical intervention remains the primary treatment method for brain tumors, requiring accurate segmentation of tumor tissue on MRI scans. However, manual tumor segmentation and traditional machine learning techniques are labor intensive and subject to observer bias. Deep learning models can provide enhanced accuracy and precision in tumor segmentation due to their ability to automatically extract intricate features and patterns from MRI scans. In this study, MRI scans from 358 meningioma patients were used to train a series of deep learning models to segment the tumor subregions. We hypothesized that a deep learning framework designed for meningioma tumor segmentation and trained utilizing 3D mpMRIs would demonstrate superior segmentation accuracy compared to a framework trained on 2D MRIs or single-sequence 3D MRIs. We compared the model predictions with the ground truth labels using similarity metrics to assess accuracy and clinical applicability. The model trained on 3D mpMRI scans showed reliable performance for meningioma subregion segmentation with a Dice Similarity Coefficient score of 0.91 and a median Sensitivity of 91.38%. It outperformed the other models due to the additional spatial context provided by 3D mpMRIs, suggesting it could be readily used in clinical practice to help guide treatment strategy.

INTRODUCTION

Magnetic Resonance Images (MRIs) have seen increased adoption as the first-line imaging test of choice for diagnosis of brain tumors among radiologists (1). MRIs can be two-dimensional (2D), where each radiofrequency pulse excites a small slice of the brain, or three-dimensional (3D), where the pulse excites the entire imaging volume (2). Multiparametric MRIs (mpMRIs), which can be in 2D or 3D configurations, consist of different sequences that provide a complementary profile of tumor subregions (3). Typical tumor subregions include the enhancing tumor (ET), which indicates active tumor tissue associated with increased tumor activity, and edema (ED), which is the accumulation of fluid in the interstitial spaces of the brain tissue (4). It is important for physicians to differentiate ET tissue from healthy brain tissue and ED from the surrounding cerebrospinal fluid (CSF) to better plan surgical treatment (5). mpMRIs consist of four individual sequences: T1, T1 Contrast Enhanced (T1-CE), T2, and Fluid Attenuated Inversion Recovery (FLAIR) (6, 7). T1, the longitudinal proton relaxation time, also known as the spin-lattice sequence, and T1-CE, the gadolinium-enhanced T1-weighted sequence, show the relative enhancement of the tumor region compared to normal brain tissue, indicating the extent of ET (8, 9). T2, the transverse relaxation time, also known as the spin-spin sequence, and FLAIR show subregion abnormalities and differentiate the ED from the CSF (8, 9). Each of these individual sequences are referred to as single-sequence MRIs (8, 9).

Medical image segmentation is the process of identifying and demarcating anatomical regions of interest (ROI) on MRI scans so physicians can focus their diagnosis on these regions (10). The segmentation process also involves classifying the ROI into diagnostic classes that aid in disease prognosis and treatment planning (11). Tumor subregion segmentation on MRI scans is important for the assessment of treatment response and for the continued surveillance of tumor recurrence (12). However, manual segmentation of MRIs is time-consuming and subject to intra- and inter-observer variations and bias (12). Automated segmentation of tumor subregions can enable faster, more accurate diagnosis (12).

Meningioma is the most common primary central nervous system (CNS) tumor, accounting for 40% of all adult brain tumors (13). Meningiomas are often identified when the patient starts having symptoms, and imaging tests are recommended to assess the extent and grade of the tumor (13). Surgery is the most common treatment for meningiomas (13). Meningiomas are slow growing tumors that start in the meningeal tissues, the thin membranes that protect the brain and spinal cord (14). Although 80% of meningioma tumors are benign, they may cause significant secondary symptoms through local brain compression and cerebral edema (15). Meningioma occurs primarily in older patients and is observed to affect females more frequently than males, with an estimated 10-year overall survival rate between 57.1 and 77.7% (15). The 2016 World Health Organization report classifies meningioma tumors into three grades based on their histopathological features (16). Grade 1 tumors constitute 80% of all meningioma tumors and are characterized by the abnormal growth of meningothelial cells (17). Grade 2 tumors are atypical lesions and characterized by sheet-like growth, nuclei prominence and increased cellularity (17). Grade 3 tumors are anaplastic,

https://doi.org/10.59720/23-265

malignant lesions with a high rate of metastases, like high-grade sarcomas, carcinomas, or melanomas (17).

Deep learning for biomedical segmentation has seen increased use as a method to augment radiologists' expert segmentation and minimize observer bias (18, 19). While traditional machine learning techniques have been used for image segmentation, they still require manual feature extraction and engineering, negating the acceleration, bias minimization, and automation advantages (18, 19). Deep learning models, also known as neural network models, have emerged as a leading technique to overcome these challenges due to their dynamic feature selection and feature engineering capabilities (18, 19). However, medical image datasets vary significantly in their characteristics, such as cohort size, voxel intensity, class label, and image modalities, resulting in customized implementations of neural network pipelines that are not generalizable to tasks outside of the specific datasets on which they were trained (20). Model performance also varies unpredictably when custom models are used on unseen datasets. For example, AlBadawy, et al. used convolutional neural network (CNN) models to segment brain tumors of 44 glioblastoma patients from four different institutions (21). The authors observed that the prediction accuracy decreased significantly when the model was trained on data from a different institution compared to training with data from the same institution.

While there have been several studies on using deep learning to segment and diagnose malignant glial tumors such as glioblastoma, there are a limited number of studies on leveraging deep learning segmentation frameworks for meningioma tumors (22–26). Meningiomas are particularly suitable for automated segmentation using deep learning due to their well-defined boundaries (26). The clear demarcation between the tumor regions and surrounding brain tissue provides the optimal imaging characteristics to train deep learning segmentation models (26). As the most frequently occurring CNS tumor, its high prevalence places a significant diagnostic burden on physicians, which can be mitigated through automated segmentation (26). Meningiomas also display unique contrast enhancement patterns and signal intensity on T1, T2 and T1-CE MRI sequences, which can serve as training features for the deep learning models (26). Automated meningioma segmentation has previously been performed using the U-Net architecture and its variants using single-sequence MRIs, with DSC scores ranging from 0.78 to 0.85 (26).

We hypothesized that due to the additional spatial context and image intensity characteristics available in 3D mpMRIs, a meningioma-specific deep learning model trained using 3D mpMRIs would perform tumor segmentation more accurately than those trained on single-sequence 3D MRIs or 2D MRIs. This study used a self-adapting, deep learning framework called nnU-Net to train and test the comparative accuracy of a series of neural network segmentation models **(Figure 1)**. nnU-Net was chosen as the deep learning framework due to its ability to standardize and automate the neural network pipeline, including data preprocessing, dynamic adaptive model training, and inference (27). nnU-Net is a semantic segmentation framework that automatically adapts to the MRI dataset, configures a matching U-Net pipeline, extracts a data profile from the scans, and dynamically trains the neural network model by optimizing the hyperparameters (27).

Eight deep learning models, based on the input MRI sequences used for training, were trained using the nnU-Net framework, and model performance on an unseen test dataset was compared to the ground truth MRI scans segmented by expert radiologists. Statistical metrics, which included the DSC score, sensitivity, and Hausdorff-95 distance were utilized to measure model performance. The best performing neural network model was further evaluated for clinical applicability using volumetric analysis techniques.

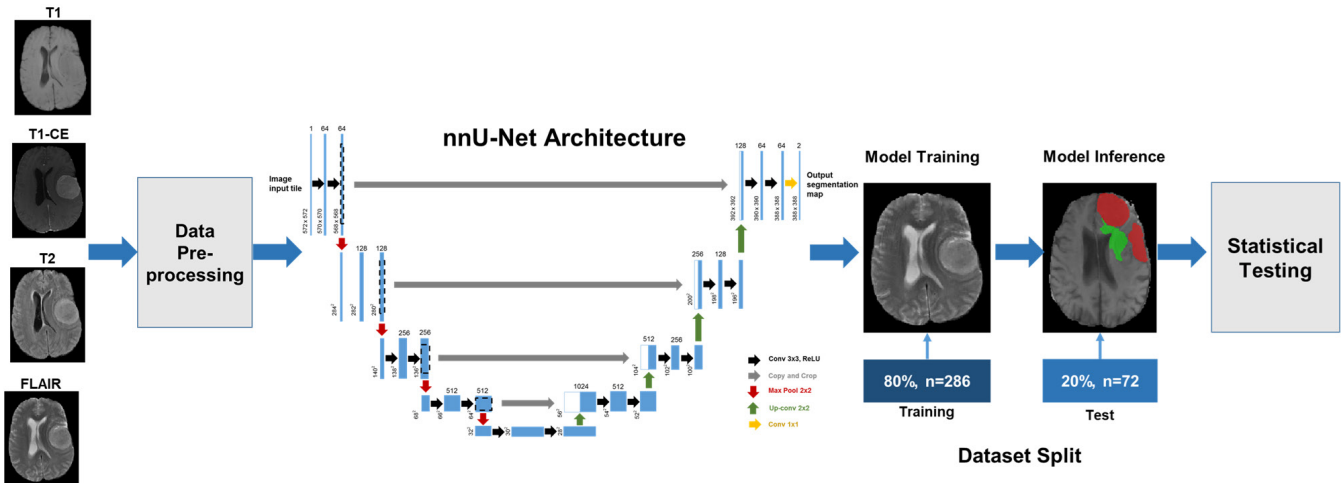

Figure 1: Deep neural network processing pipeline. The end-to-end pipeline shows the ingestion of MRI scans, data preprocessing, model training, model inference, and post-processing stages. MRI scans were ingested into the pipeline and normalized, then split into training (n=286) and test (n=72) cohorts. The training cohort was used to train the nnU-Net model, and the trained model was used to conduct automated segmentation on the unseen test cohort. The model performance was then evaluated using quantitative and volumetric similarity metrics. The nn-Unet architecture was adapted from the original U-Net paper by Ronneberger et al (40).

https://doi.org/10.59720/23-265

RESULTS

In this study, an automated deep learning pipeline was developed using nnU-Net to perform automated segmentation of meningioma tumor subregions. Eight neural network models were trained on 3D mpMRIs; 2D mpMRIs; individually on T1, T1-CE, T2, and FLAIR 3D single-sequence MRIs; and on T1/T1-CE and T2/FLAIR combined 3D MRI sequence pairs. Our dataset was primarily comprised of Grade 1 patients (85.5%), reflecting the naturally occurring frequency of Grade 1 tumors that make up 80% of all meningioma cases diagnosed. To compensate for this skew in the dataset towards Grade 1 tumors, and to account for outlier cases caused by the presence of image artifacts in certain images, the median values of the statistical metrics were considered in evaluating comparative model performance instead of the mean values (28). The models' performance in predicting the tumor segmentation were measured using the DSC score, Sensitivity and Hausdorff-95 distance. The DSC score, also known as the proportion of specific agreement, is a statistical validation metric that measures the spatial overlap accuracy between model predicted and ground truth segmentation maps. It ranges in value from 0 to 1, with 0 indicating no similarity between the two images and 1 indicating that the two images are identical (29). The DSC is set to 1 if a label is present in neither the prediction nor the ground truth images. The Sensitivity, or the true positive result, is the ratio of the number of true positive tumor voxels in the predicted image to the number of tumor voxels in the ground truth image (30, 31). The greater the Sensitivity, the more accurate the automated segmentation. The Hausdorff-95 distance evaluates the overlap of the predicted segmentation with the ground truth. The 95th percentile Hausdorff distance was chosen as it is not sensitive to outliers compared to the maximum Hausdorff distance measure (30, 31).

Comparative model performance

The model trained on 3D mpMRI sequences performed the best among all eight models in identifying and segmenting Whole Tumor (WT) tissue from healthy brain tissue, achieving a median DSC score of 0.91 (WT), a median Sensitivity of 91.38% (WT), and a median Hausdorff-95 distance of 3.00 (WT) (**Table 1**). This model also performed significantly better than the other models, except the model trained on the 2D mpMRI sequences in predicting the segmentation map for tumor subregions (Mann-Whitney U-test, $p<0.001$). For the enhancing tumor subregion, the model achieved a median DSC score of 0.91, a median Sensitivity of 92.69%, and a median Hausdorff-95 distance of 2.24 (**Table 1**). For the edema subregion, the model achieved a median DSC score of 0.85, a median Sensitivity of 89.95%, and a median Hausdorff-95 distance of 2.83 (Table 1). While the model trained on the 3D T1-CE sequence had a higher DSC score (0.92) than the model trained on all 3D mpMRI sequences (0.91), its sensitivity and Hausdorff-95 distance metrics were lower in comparison (**Table 1**). The model with the lowest predictive segmentation performance was the one trained only on the 3D T1 sequence, with a median DSC score of 0.55 (WT), a median sensitivity of 54.08% (WT), and a median Hausdorff-95 distance of 15.24 (WT) (**Table 1**). It also performed the lowest among all models in predicting the tumor subregion segmentation.

The best-performing model was further evaluated for clinical

MRI Sequence	Tumor Region	Median DSC Score	Median Sensitivity	Median Hausdorff-95 Distance
3D MpMRI Model trained on T1, T1-CE, T2 and FLAIR	Whole Tumor	0.91	91.38%	3.00
	Enhancing Tumor	0.91	92.69%	2.24
	Edema	0.85	89.95%	2.83
2d Model trained on T1, T1-CE, T2 and FLAIR	Whole Tumor	0.81	83.73%	6.00
	Enhancing Tumor	0.69	63.78%	8.31
	Edema	0.78	89.10%	7.48
3D MpMRI Model trained on T1-CE Only	Whole Tumor	-	-	-
	Enhancing Tumor	0.92	92.44%	2.45
	Edema	-	-	-
3D MpMRI Model trained on T1 and T1-CE Only	Whole Tumor	-	-	-
	Enhancing Tumor	0.92	91.49%	2.45
	Edema	-	-	-
3D MpMRI Model trained on T2 Only	Whole Tumor	0.74	71.44%	7.18
	Enhancing Tumor	-	-	-
	Edema	0.65	74.13%	8.61
3D MpMRI Model trained on FLAIR Only	Whole Tumor	0.82	82.96%	5.87
	Enhancing Tumor	-	-	-
	Edema	0.79	83.08%	6.08
3D MpMRI Model trained on T2 and FLAIR Only	Whole Tumor	0.81	83.73%	6.00
	Enhancing Tumor	-	-	-
	Edema	0.78	89.10%	7.48

Table 1: Performance metrics for the nnU-Net neural network models on the test dataset. Summary statistics of the models' prediction performance on the test dataset across the whole tumor (WT), enhancing tumor (ET) and edema (ED) subregions. Model inference was run on the test dataset and the summary statistics for the DSC score, Sensitivity, and the 95% Hausdorff distance were calculated. Since MRI sequences are specific to certain tumor regions, results were not computed for non-relevant regions when not all MRI sequences were utilized.

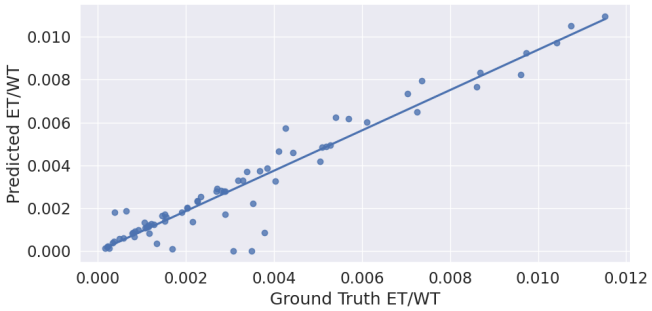

Figure 2: Correlation of VASARI volumetric features between the predicted segmentation and ground truth segmentation for the 3D mpMRI model on the enhancing tumor subregion using the test dataset. Correlation plot showing linear trend of predicted voxel volume to ground truth voxel volume for the enhancing tumor subregion (n=72). The ratio of the enhancing tumor volume to the whole tumor volume (ET/WT) was calculated for both the predicted segmentation (predicted ET/WT) and the ground truth segmentation (ground truth ET/WT) of each patient in the test dataset and plotted in a correlation plot. The Pearson's correlation coefficient was calculated for the predicted and ground truth segmentation volumes ($r2 = 0.92$, p-value = 0.000995).

https://doi.org/10.59720/23-265

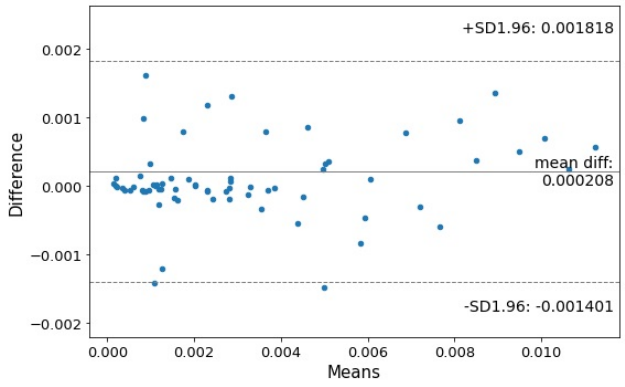

Figure 3: Mean difference between the ground truth and predicted segmentation voxel volumes for the 3D mpMRI model on the enhancing tumor subregion. Bland-Altman plot showing mean difference ± SD of voxel volumes for the enhancing tumor subregion (n=72). Voxel volumes were calculated for the predicted segmentation and ground truth segmentation of the enhancing tumor subregion for each patient in the test dataset and plotted.

applicability using Pearson's correlation and Bland-Altman plots between the predicted and ground truth segmentation. Accurate subregion segmentation is an important factor in the Response Assessment in Neuro-Oncology criteria for assessing treatment response (32). This study used the Visually Accessible Rembrandt Images (VASARI) features, a standardized method to evaluate tumor characteristics developed by multi-institutional neuroradiologists to assess the model's capabilities in segmenting ET and ED subregions (33). Per the VASARI MR Feature Guide V1.1, features f5 (proportion of tumor that is enhancing), and f14 (proportion of tumor that is edema), were calculated. To assess clinical applicability, the voxel volumes of the ET and ED subregions were computed for both the ground truth and the predicted segmentations drawn from patients in the test cohort and volumetric analysis was conducted.

Volumetric results for the enhancing tumor subregion of the 3D mpMRI model

A comparison between the voxel volumes from the predicted ET maps and ground truth ET maps indicated they were highly correlated, indicating strong spatial alignment between them ($r^{2=0.92}$, $p<0.001$) (**Figure 2**). A low mean difference of 0.000208 [+ 0.0018, - 0.0014] between the two maps further supports the accuracy of the model predicted segmentation for the ET subregion. (**Figure 3**).

Volumetric results for the edema subregion of the 3D mpMRI model

A correlation between the voxel volumes from the predicted ED maps and ground truth ED maps indicated they were highly correlated, indicating strong spatial alignment between them (r2=0.95, p<0.001) (**Figure 4**). A low mean difference 0.000291 [+ 0.0030, - 0.0025] between the two maps further supports the accuracy of the model predicted segmentation for the ED subregion. (**Figure 5**).

DISCUSSION

This study trained and evaluated a self-adapting, deep learning, biomedical image segmentation model for the automated segmentation of meningioma tumors. The prediction accuracy of the model trained on 3D mpMRI sequences, as measured by the median DSC scores was observed to be superior to those models trained on single sequence 3D MRIs, partially supporting our research hypothesis. The comparison in performance between the model trained on 3D mpMRI sequences and the model trained on 2D mpMRI sequences was not found to be statistically significant, indicating there is not enough statistical evidence to conclude that the performance of the model trained on 3D mpMRIs was better (p>0.001) (**Table 2**). To our knowledge, this is the first study of its kind to utilize all four standard MRI sequences and accurately segment the whole abnormal meningioma tumor and its component subregions, including the ET and ED subregion, which is critical in assessing tumor progression. The final model performed reliably in segmenting MRI scans and strongly agreed with manually segmented scans by expert clinicians, indicating that it is generalizable to new, unseen MRI scans.

The prediction performance of the model trained on 3D mpMRI sequences surpassed that of all the other models evaluated in this study as measured by the median DSC score across the whole tumor. Our initial hypothesis was that the 3D mpMRI sequences would provide additional spatial context to the deep learning model during training when compared to the 2D mpMRI sequences, potentially resulting in better segmentation accuracy. However, the model trained on 3D mpMRIs did not produce a statistically significant difference in performance compared to the model trained on 2D mpMRIs, which could have been due to more noise in the 3D mpMRIs (**Table 2**). This suggests that mpMRIs trained models outperform models trained on single sequences. An additional consideration is that models can be trained on 2D mpMRIs as they would be faster and less expensive, while producing an essentially similar tumor segmentation performance as models trained on 3D mpMRIs.

The model trained on only the 3D T1-CE sequence achieved the highest median DSC score in predicting the ET

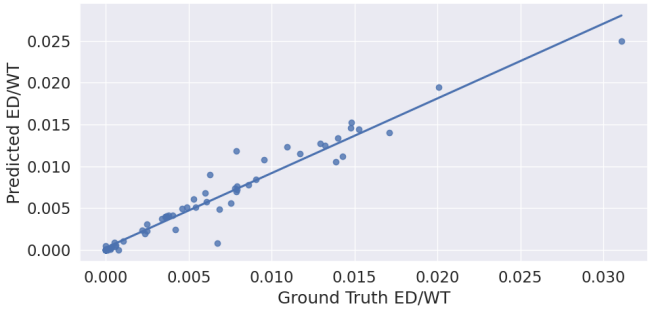

Figure 4: Correlation of VASARI volumetric features between the predicted segmentation and ground truth segmentation for the 3D mpMRI model on the edema subregion using the test dataset. Correlation plot showing linear trend of predicted voxel volume to ground truth voxel volume for the edema subregion (n=72). The ratio of the edema volume to the whole tumor volume (ED/WT) was calculated for both the predicted segmentation (predicted ED/WT) and the ground truth segmentation (ground truth ED/WT) of each patient in the test dataset and plotted in a correlation plot. The Pearson's correlation coefficient was calculated for the predicted and ground truth segmentation volumes (r2 = 0.95, p-value = 0.000469).

https://doi.org/10.59720/23-265

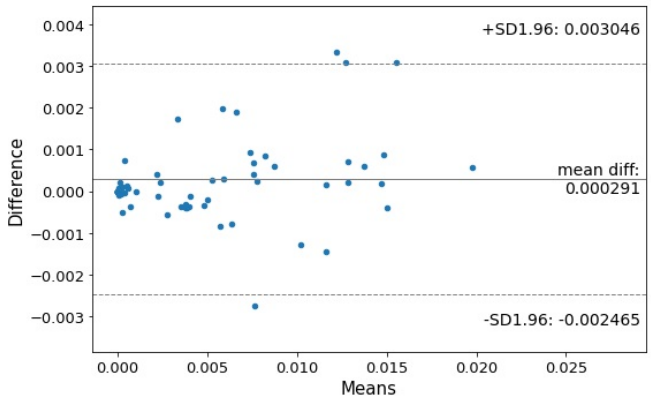

Figure 5: Mean difference between the ground truth and predicted segmentation voxel volumes for the 3D mpMRI model on the edema subregion. Bland-Altman plot showing mean difference ± SD of voxel volumes for the edema subregion (n=72). Voxel volumes were calculated for the predicted segmentation and ground truth segmentation of the edema subregion for each patient in the test dataset and plotted.

subregion. This is likely because the T1-CE sequence shows the relative enhancement of the tumor region in contrast with the surrounding normal brain tissue (34). Since the T1-CE sequence does not show the differentiation between the edema and the cerebrospinal fluid, the model was not evaluated on the edema subregion (34). Its sensitivity and Hausdorff-95 distance were lower than the model trained on all four 3D mpMRI sequences.

This supports our hypothesis that a deep learning-based semantic segmentation model that is trained on 3D mpMRI sequences would perform better than models trained on 3D single-sequence MRIs.

A high correlation coefficient (r2 = 0.92, p=0.000995) for the ET subregion confirms the spatial alignment between ground truth and predicted segmentation maps, indicating strong clinical applicability of the model's predictions. The agreement between the volumetric measures of ground truth and predicted segmentation in the test cohort was also confirmed by the low mean difference in the Bland-Altman plots for the ET (0.000208 [+ 0.0018, - 0.0014]).

Similarly, the high correlation coefficient for the ED subregion (r2 = 0.95, p=0.000469) supports the agreement between predicted and ground truth observations. The low mean difference of 0.000291 [+ 0.0030, - 0.0025] for the ED subregion further confirms this agreement between the volumetric measures of the predicted and ground truth observations, indicating clinical applicability of the model's predictions.

Taken together, the evaluation metrics affirm that the 3D mpMRI model is able to accurately segment meningioma tumors close to expert human segmentation, and that the model can be leveraged to segment MRI scans from large patient cohorts. The potential clinical applications of this model include its use in pre-surgical planning, post-operative tumor assessment, evaluation of subregion treatment response, and assessment of tumor recurrence.

Although the patient cohort of 358 patients used in this study was relatively large compared to other studies, the model could be improved by training it on a larger, more diverse dataset of meningioma patients from different demographic, geographic and clinical populations. The dataset was also heavily imbalanced, with Grade 1 meningioma patients making up 85.5% of the cohort. While most meningioma tumors are grade 1, acquiring a dataset with higher proportions of Grade 2 and Grade 3 tumors would also help improve model accuracy in tumor subregion segmentation. All patients in this study were from a single institution. Testing the model on external datasets from multiple institutions would also further validate the model's performance in a real-world clinical setting, enhancing its clinical applicability.

Automated segmentation of tumor subregions from MRI scans using deep learning models can enable faster, more accurate diagnosis of meningioma tumors and subregions (35). However, as surgery remains the primary method of treatment, inaccurate demarcation of tumor tissue could cause extended manual validation of model outputs by the clinician prior to treatment, loss of trust in model outputs, and potentially harmful patient outcomes (35). A generalizable, automated meningioma segmentation solution with clinical applicability as shown in this study can potentially be instrumental in reducing diagnostic burden, minimizing observer bias, and delivering tailored treatment to patients. The neural network pipeline developed in this study provides the ability to automatically segment MRI scans from large cohorts of meningioma patients, improving physician experience and resulting in favorable patient outcomes through focused treatment.

Future direction for this study could include the development of similar deep learning pipelines to segment multiple CNS tumors by training the model on mpMRI scans from different tumor types such as astrocytoma, oligodendroglioma and glioblastoma. The solution proposed in this study can be further improved by training the model on 3D mpMRIs from multiple institutions and populations, further expanding its reach and value to lower income and developing areas that currently may not have access to advanced diagnostic approaches for meningioma tumor assessment.

Model Pair	p-value of DSC Scores (α = 0.001)
3D mpMRI and 2D mpMRI	0.809
3D mpMRI and T1 single-sequence MRI	6.585e-17
3D mpMRI and T1-CE single-sequence MRI	2.491e-06
3D mpMRI and T2 single-sequence MRI	3.536e-11
3D mpMRI and FLAIR single-sequence MRI	2.255e-06
3D mpMRI and T1+T1-CE MRI	6.956e-05
3D mpMRI and T2+FLAIR MRI	3.146e-06

Table 2: p-values of predicted DSC scores between model pairs for patients in the test cohort. A Mann-Whitney U-test was conducted on DSC scores produced by the 3D mpMRI model and the DSC scores produced by each of the other models for the 72 patients in the test cohort. p-values indicate statistically significant difference for all model pairs except the 3D mpMRI and 2D mpMRI models.

https://doi.org/10.59720/23-265

MATERIALS AND METHODS

Data collection

The MRI scans were accessed through the University of Pennsylvania's meningioma tumors dataset after appropriate approvals were received by the University Ethics Committee. mpMRI scans of 358 de-identified patients with pathologically confirmed Grade 1 to Grade 3 meningioma tumors were collected. The MRI scans were processed and skull-stripped for automated segmentation using the BraTS preprocessing pipeline provided via the CaPTk software (36). Skull-stripping is an image preprocessing step that separates the brain from extra-cranial or non-brain tissues (37). The tumor regions and subregions were manually segmented and labeled by clinical experts to identify the ET and ED subregions to determine the ground truth for training and testing.

The dataset was then split into a training subset (80%, n=286) and a test subset (20%, n=72). A separate validation subset of data was not created as the nnU-Net framework utilizes a 5-fold cross-validation technique, which is a resampling procedure to estimate the predictive power of the model (38). Pre-processing of the ground truth segmentation files was performed to assign classification labels to tumor subregions (0=Clear, 1=Enhancing Tumor, 2=Edema) for use during model training.

Data preprocessing

MRI data was initially structured in a specific format and folder hierarchy as required by nnU-Net. nnU-Net preprocessing steps were performed to extract a data fingerprint, including image features such as image size, voxel spacing, and intensity information (38). Two U-Net configurations (2D and 3D MpMRI) and training plans for each of the two configurations were created.

Neural network model training and inference

Eight models were trained as part of this study, using different MRI sequences and configurations as training inputs: 3D mpMRI using T1, T1-CE, T2 and FLAIR sequences; 2D using T1, T1-CE, T2 and FLAIR sequences; 3D mpMRI using T1 sequence only, 3D mpMRI using T1-CE sequence only, 3D mpMRI using T2 sequence only, 3D mpMRI using FLAIR sequence only, 3D mpMRI using T1 and T1-CE sequences, and 3D mpMRI using T2 and FLAIR sequences. Model training hyperparameters such as batch size, patch size, learning rate, weight decay and the number of pooling operations per axis were dynamically sampled and set by nnU-Net based on the shape of the preprocessed training data and training loss trends (38). The sum of the cross-entropy loss and the DSC loss were used as the loss function (38). The Adam optimizer was used for stochastic gradient descent with an initial learning rate of 3 x 10-4 and weight decay of 3 x 10-5, which are standard hyperparameters within the nnU-Net framework (38). If the exponential moving average of the training loss did not improve within the last 30 epochs, the learning rate was decreased by a factor of 0.2 (38). The model was trained for five validation folds, with a period of 1000 epochs per validation fold, and the model checkpoint saved every 50 epochs (38). After each epoch, the training loss, validation loss, and an average global DSC score were charted and saved in an image file to assess model training progress (38). The model's softmax outputs were stored during the final validation to enable nnU-Net to identify the best performing ensemble of configurations (38). Inference, which is the process of using trained models to make predictions, was run on the unseen test data with all five available folds for all the eight models using the nnU-Net inference package. The resulting segmentation files were stored in an output folder defined as part of the pipeline configuration.

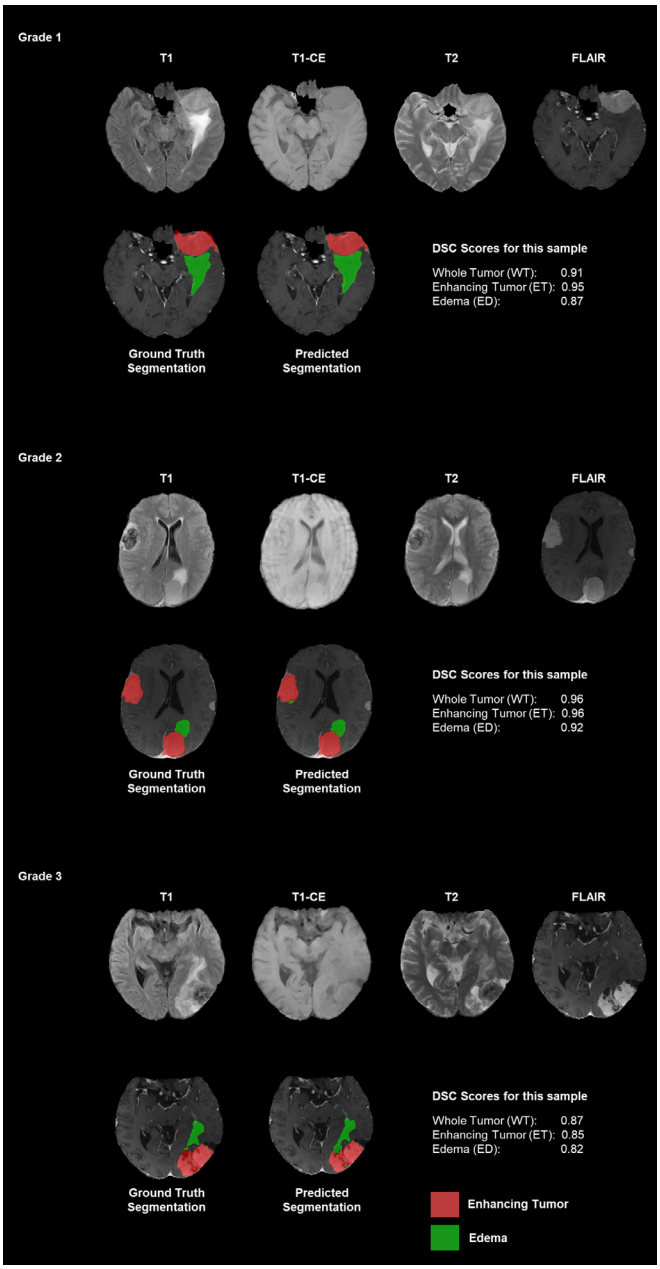

Figure 6: Sample predicted and ground truth segmentation of Grade 1, Grade 2 and Grade 3 meningiomas. Individual samples of patient MRI sequences for each meningioma tumor grade, along with the ground truth segmentation map and the predicted segmentation map with associated DSC scores for the whole tumor, enhancing tumor and edema subregions. Model inference was run on three patients' MRIs with Grade 1, Grade 2 and Grade 3 meningioma tumors, and the DSC scores were calculated on the predicted segmentation. The ground truth and predicted segmentation maps were visualized using ITKSnap software.

https://doi.org/10.59720/23-265

Neural network model performance metrics calculation

DSC score, Sensitivity, and 95% Hausdorff distance were calculated for the ET, ED and WT (defined as the union of all subregions) regions for all eight models using custom Python code that leveraged the *Medpy* package (**Figure 6**). A Kolmogorov-Smirnov (K-S) test was conducted for the DSC scores from each model to test for normality using the Python *scipy.stats.kstest* package. A Mann-Whitney U-test was then conducted to compare the DSC scores from the 3D mpMRI modelto the DSC scores of each of the other models using the Python *scipy.stats.mannwhitneyu* package. The non-parametric Mann-Whitney U-test was used to assess significance instead of a t-test since the Kolmogorov-Smirnov (K-S) test for normality indicated that the DSC scores produced by the models' predictions do not follow a normal distribution (39).

The Python package *nibabel* was used to compute volumetrics of the predicted and ground truth segmentation maps for the test cohort by reading each predicted and ground truth segmentation map and aggregating the number of voxels under each of the label classes corresponding to the tumor subregions ET and ED. The voxel ratios ET/WT and ED/WT were calculated for both the ground truth and predicted cases and plotted using the Python *pearsonr* package. The p-value was calculated using the Python *scipy.stats.mannwhitneyu package*. The Bland-Altman plot was generated using the Python *statsmodels.graphics.mean_diff_plot* package.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Omaditya Khanna for his review of the results and his valuable feedback.

Received: November 09, 2023
Accepted: January 04, 2024
Published: June 26, 2024

REFERENCES

1. Singh, Satya P., et al. "3D Deep Learning on Medical Images: A Review" Sensors, vol. 20, no. 18, Sept. 2020, p. 5097, https://doi.org/10.3390/s20185097.
2. Johnson, Glyn et al. "2D Multislice and 3D MRI Sequences Are Often Equally Sensitive" Magnetic Resonance in Medicine, vol. 41, no. 4, pp. 824–828, Apr. 1999, https://doi.org/10.1002/(sici)1522-2594(199904)41:4<824::aid-mrm23>3.0.co;2-1.
3. Guo, Peiying, et al. "Multiparametric Magnetic Resonance Imaging Information Fusion Using Graph Convolutional Network for Glioma Grading" Journal of Healthcare Engineering, vol. 2022, May 2022, https://doi.org/10.1155%2F2022%2F7315665.
4. Ellingson, Benjamin M., et al. "Modified Criteria for Radiographic Response Assessment in Glioblastoma Clinical Trials" Neurotherapeutics, vol. 14, no.2, pp. 307–320, Apr. 2017, https://doi.org/10.1007/s13311-016-0507-6.
5. Sun, Li, et al. "Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans With Deep Learning" Frontiers in Neuroscience, vol. 13, Aug. 2019, https://doi.org/10.3389/fnins.2019.00810.
6. Bakas, Spyridon, et al. "Overall survival prediction in glioblastoma patients using structural magnetic resonance imaging (MRI): advanced radiomic features may compensate for lack of advanced MRI modalities" Journal of Medical Imaging (Bellingham), vol. 7, no.3, May 2020, https://doi.org/10.1117/1.JMI.7.3.031505.
7. Gahrmann, Renske, et al. "Comparison of 2D (RANO) and volumetric methods for assessment of recurrent glioblastoma treated with bevacizumab—a report from the BELOB trial", Neuro-oncology, vol. 19, no.6, pp. 853–861, Jun. 2017, https://doi.org/10.1093/neuonc/now311.
8. Roozpeykar, Saeid, et al. "Contrast-enhanced weighted-T1 and FLAIR sequences in MRI of meningeal lesions" American Journal of Nuclear Medicine and Molecular Imaging, vol. 12, no.2, pp. 63–70, Apr. 2022, www.ajnmmi.us/ISSN:2160-8407/ajnmmi0141723.
9. Demirel, Hüseyin C. and John W. Davis. "Multiparametric magnetic resonance imaging: Overview of the technique, clinical applications in prostate biopsy and future directions" Turkish Journal of Urology, vol.44, no.2, pp. 93–102, Mar. 2018, https://doi.org/10.5152/tud.2018.56056.
10. Bhalodiya, Jayendra M., et al. "Magnetic resonance image-based brain tumour segmentation methods: A systematic review" Digital Health, vol. 8, pp. 1-19, Mar. 2022, https://doi.org/10.1177%2F20552076221074122.
11. Gryska, Emilia, et al. "Automatic brain lesion segmentation on standard magnetic resonance images: a scoping review" BMJ Open, vol. 11, no.1, Jan. 2021, https://doi.org/10.1136%2Fbmjopen-2020-042660.
12. Kazerooni, Anahita Fathi et al. "Characterization of 344 Active and Infiltrative Tumorous Subregions From Normal Tissue in Brain Gliomas Using Multiparametric MRI" Journal of Magnetic Resonance Imaging, vol. 48, no.4, pp. 938–950, Feb. 2018, https://doi.org/10.1002/jmri.25963.
13. Buerki, Robin A., et al. "An Overview of Meningiomas" Future Oncology, vol. 14, no. 21, Aug. 2018, https://doi.org/10.2217/fon-2018-0006.
14. Holleczek, Bernd, et al. "Incidence, mortality and outcome of meningiomas: A population based study from Germany" Cancer Epidemiology, vol. 62, Oct. 2019, https://doi.org/10.1016/j.canep.2019.07.001.
15. Ostrom, Quinn T., et al. "CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2009–2013" Neuro-Oncology,vol. 18, suppl. 5, pp. 1–75, Oct. 2016, https://doi.org/10.1093/neuonc/now207.
16. Louis, David N. et al. "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary" Acta Neuropathologica, vol. 131, no.6, pp. 803-820, Jun. 2016, https://doi.org/10.1007/s00401-016-1545-1.
17. Alruwaili AA, De Jesus O. "Meningioma" StatPearls [Internet], Treasure Island (FL): StatPearls Publishing, Aug. 2023, www.ncbi.nlm.nih.gov/books/NBK560538/, Accessed Dec. 20, 2023.
18. Tandel, Gopal S., et al. "Role of Ensemble Deep Learning for Brain Tumor Classification in Multiple Magnetic Resonance Imaging Sequence Data" Diagnostics (Basel), vol. 13, no.3, Feb. 2023, https://doi.org/10.3390/diagnostics13030481.
19. Nam, Seojin et al. "Understanding the Research Landscape of Deep Learning in Biomedical Science: Scientometric Analysis" Journal of Medical Internet

https://doi.org/10.59720/23-265

Research, Apr. 2022, https://doi.org/10.2196/28114.
20. Isensee, Fabian et al. "nnU-Net for Brain Tumor Segmentation", Nov. 2020, https://doi.org/10.48550/arXiv.2011.00848.
21. AlBadawy, Ehab A., et al. "Deep learning for segmentation of brain tumors: Impact of cross institutional training and testing" Medical Physics, vol. 45, no. 3, pp. 1150-1158, Mar. 2018, https://doi.org/10.1002/mp.12752.
22. Bouget, David et al. "Fast meningioma segmentation in T1-weighted magnetic resonance imaging volumes using a lightweight 3D deep learning architecture" Journal of Medical Imaging (Bellingham), Mar. 2021, https://doi.org/10.1117/1.JMI.8.2.024002.
23. Bouget, David et al. "Meningioma Segmentation in T1-Weighted MRI Leveraging Global Context and Attention Mechanisms" Frontiers in Radiology, vol. 1-2021, Sep. 2021, https://doi.org/10.3389/fradi.2021.711514.
24. Laukamp, Kai R., et al. "Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI" European Radiology, vol. 29, no.1, pp. 124–132, Jan. 2019, https://doi.org/10.1007/s00330-018-5595-8.
25. Boaro, Allesandro, et al. "Deep neural networks allow expert-level brain meningioma segmentation and present potential for improvement of clinical practice" Scientific Reports, vol. 12, Sep. 2022, https://doi.org/10.1038/s41598-022-19356-5.
26. Chen, Chaoyue, et al. "Automatic Meningioma Segmentation and Grading Prediction: A Hybrid Deep-Learning Method" Journal of Personalized Medicine, vol. 11, no. 8, pp. 786, Aug. 2021, https://doi.org/10.3390/jpm11080786.
27. Isensee, Fabian, et al. "nnU-Net for Brain Tumor Segmentation", Nov. 2020, https://doi.org/10.48550/arXiv.2011.00848.
28. Skarsø, Emma Riis, et al. "Development of a national deep learning-based auto segmentation model for the heart on clinical delineations from the DBCG RT nation cohort" Acta Ontologica, Sep. 2023, https://doi.org/10.1080/0284186X.2023.2252582.
29. Zou, Kelly H., et al. "Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index" Academy of Radiology, vol. 11, no.2, pp. 178-189, Feb. 2004, https://doi.org/10.1016/S1076-6332(03)00671-8.
30. Pei, Linmin, et al. "Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images" Scientific Reports, vol. 10, Nov. 2022, https://doi.org/10.1038/s41598-020-74419-9.
31. Livne, Michelle, et al. "A U-Net Deep Learning Framework for High Performance Vessel Segmentation in Patients With Cerebrovascular Disease" Frontiers in Neuroscience, vol. 13, no. 97, Feb. 2019, https://doi.org/10.3389/fnins.2019.00097.
32. "VASARI Research Project" The Cancer Imaging Archive. wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project. Accessed Aug. 1, 2023.
33. Gemini, Laura, et al. "Vasari Scoring System in Discerning between Different Degrees of Glioma and IDH Status Prediction: A Possible Machine Learning Application?" vol. 9, no. 4, pp. 75, Mar. 2023, https://doi.org/10.3390/jimaging9040075.
34. Vasantachart, April, et al. "Automatic differentiation of Grade I and II meningiomas on magnetic resonance image using an asymmetric convolutional neural network" Scientific Reports, vol. 12, Mar. 2022, https://doi.org/10.1038%2Fs41598-022-07859-0.
35. Akinyelu, Andronicus A., et al. "Brain Tumor Diagnosis Using Machine Learning, Convolutional Neural Networks, Capsule Neural Networks and Vision Transformers, Applied to MRI: A Survey" Journal of Imaging, vol. 8, no. 8, pp. 205, Aug. 2022, https://doi.org/10.3390%2Fjimaging8080205.
36. Pati, Sarthak et al. "The Cancer Imaging Phenomics Toolkit (CaPTk): Technical Overview" Brainlesion, vol. 11993, pp. 380-394, May 2020, https://doi.org/10.1007/978-3-030-46643-5_38.
37. Kalavathi, P. and Surya Prasath, "Methods on Skull Stripping of MRI Head Scan Images—a Review" Journal of Digital Imaging, vol. 29, no. 3, pp. 365-379, Jun. 2016, https://doi.org/10.1007/s10278-015-9847-8.
38. Isensee, Fabian et al. "nnU-Net for Brain Tumor Segmentation", Nov. 2020, https://doi.org/10.48550/arXiv.2011.00848.
39. Bridge, Patrick D. and Shomo Sawilowsky. "Increasing Physicians' Awareness of the Impact of Statistics on Research Outcomes", vol. 52, no. 3, pp. 229-235, Mar. 1999, https://doi.org/10.1016/S0895-4356(98)00168-1.
40. Ronneberger, Olaf, et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation", May 2015, https://doi.org/10.48550/arXiv.1505.04597.