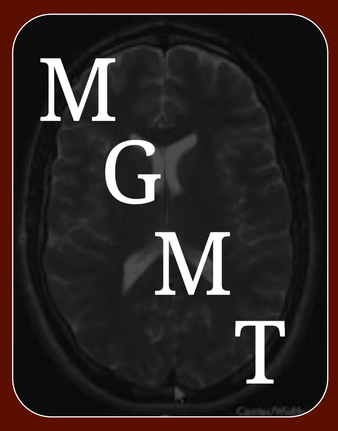
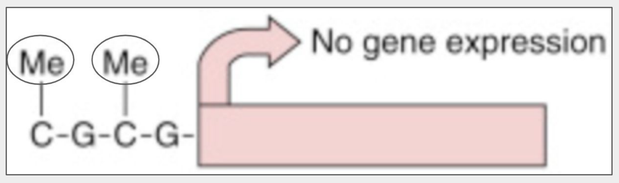
Creation of a Novel Machine Learning Model to Predict MGMT Promoter Methylation Status Using Multimodal MRI Images



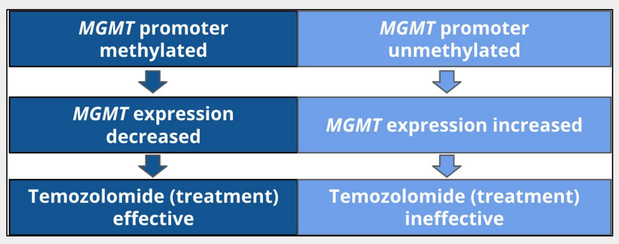
Project Overview

- While treating Glioblastoma multiforme (GBM), a rapidly-progressive brain cancer with an average survival of <1 year, doctors traditionally conduct biopsies of the patient's tumor to assess its microenvironment. One genetic feature is the methylation status of the O6-methylguanine-DNA--methyltransferase (MGMT) gene, which produces the MGMT.
- **Silencing of the MGMT gene decreases levels of MGMT protein and increases patient responsivity to therapies such as Temozolomide (TMZ).** However, conducting biopsy extractions can be life-threatening to patients already undergoing intense treatments.
- **In this project, an end-to-end ML pipeline was developed to noninvasively analyze GBM MRI scans** (with variations in patient orientation, 3D slice count, thickness and spacing), **and predict the MGMT methylation status of the cancer** presented, with an accuracy of ~70%.
- In the pre-processing stage, an affine transformation was applied to bring the images to the patient coordinate system. Black backgrounds were removed, and 32x32x32 3D-voxels were generated. Texture features such as T1c minimum and NGTDM values were extracted, and positional encoding of the voxels was applied to increase accuracy.
- A 3D-Unet model was used for tumor segmentation and a 3D Dense-Net was used for classification. The Dice Coefficient Loss function was used to account for the imbalance in the 3D voxels with tumors and healthy brain tissue.

Figure 1: Effects of methylation on treatment receptivity



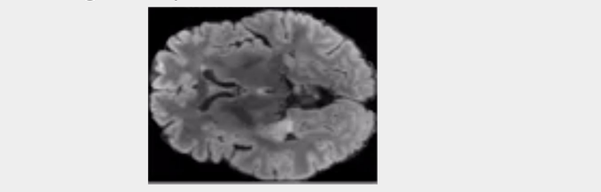
Source: https://https://www.sciencedirect.com/topics/engineering/methylation



Current Studies & Value Add

- Prior authors who only focus on classification of the MGMT promoter methylation status report lower accuracies than those incorporating segmentation. Additionally, fewer authors have attempted solutions in the 3D space.

Fig 2: Han *et* al.'s devised framework for classification (only) of the MGMT promoter methylation status.[6]



Source:https://psb.stanford.edu/psb-online/proceedings/psb18/han.pdf

- Segmenting the tumorous regions before passing the image to the DenseNet narrows the region the model needs to analyze. My use of a U-Net stacked with a Dense-Net incorporating positional encoding and texture features of T1c Minimum and Neighborhood Grey Tone Difference Matrix values, using a patch based approach, has not been reported.

Dataset

The 3D input files were retrieved from the Brain Tumor Segmentation Challenge dataset, provided by the Radiological Society of Neuroradiologists, which contains 1000+ high-resolution glioblastoma MRI scans. The dataset contained 4 modalities for each patient, (FLAIR, T1w, T1w-contrast-enhanced (T1wpCE), and T2w). The scans were annotated by board certified radiologists, and regions presenting necrosis, Gadolinium enhanced tumors, and edema were outlined. Additionally, each patient ID was associated with their promoter methylation status, indicated by a 1 or 0.

Fig 3: Sample MRI GBM scan from dataset

Source: https://www.kaggle.com/c/rsna-miccai-brain-tumor-radiogenomic-classification

Computational Challenges

Runtime limitations

- Users are given a 1 GB memory limit when running programs on Kaggle's GPU; Kaggle also restricts users to 37 hours of GPU usage a week
- To keep the model's memory footprint as low as possible, all values were converted to single precision in order to ensure it would run to completion

Input File Limitations

- When running the initial input MRIs (max size 512x512x512) through the model, it exceeded CPU limits on Kaggle's notebook.
- To account for this, the initial images were divided into 32x32x32 size voxels, to feed into the model for training
- Positional Encoding was used to associate patches with the same patient's scant, and indicate where the patch was located on the MRI scan
- As all the MRIs scans were taken from different MRI machines and in different coordinate systems, an affine transformation was applied.
- I plan to utilize Interpolation of the initial MRI scans to bring all dimensions to 512x512x512

Fig 4: Table of variations in scanner settings

Image	Slice Thickness	Spacing	Image Position	Image Orientation	Slice Location	Rows	Columns	Pixel Spacing
train/00203/T2w/Image-63.dcm	3.0'	1.0'	[-95.31, -162.98, 108.93]	[0.99, -0.002, 0, 0.002, 0.99 0]	108.93475341797'	256	208	[0.9375, 0.9375]
train/00203/T2w/Image-64.dcm	3.0'	1.0'	[-95.31, -162.98, 111.93]	[0.99, -0.002, 0, 0.002, 0.99 0]	111.93475341797'	256	208	[0.9375, 0.9375]
train/00009/T2w/Image-1.dcm	1.4'	1.0'	[102.887, -161.852, 157.644]	[-0, 1, 0, -0, -0, -1]	-102.8865051'	512	512	[0.5, 0.5]
train/00009/T2w/Image-2.dcm	1.4'	1.0'	[102.187, -161.852, 157.644]	[-0, 1, 0, -0, -0, -1]	-102.1865082'	512	512	[0.5, 0.5]
train/00009/T2w/Image-3.dcm	1.4'	1.0'	[101.487, -161.852, 157.644]	[-0, 1, 0, -0, -0, -1]	-101.4865036'	512	512	[0.5, 0.5]
train/00009/T2w/Image-4.dcm	1.4'	1.0'	[100.786, -161.852, 157.644]	[-0, 1, 0, -0, -0, -1]	-100.786499'	512	512	[0.5, 0.5]

Fig 5: Patch Creation Flowchart

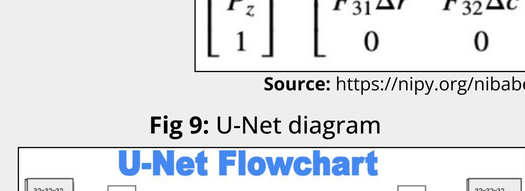
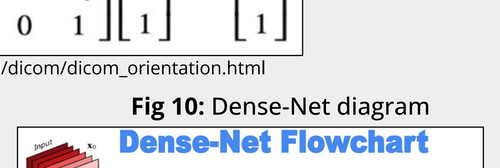


Fig 6: Banerjee *et* al's creation of patches for 2D MRIs to account for memory limitations[4]



Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6993215/

Framework

Fig 7: Devised end-to-end pipeline for tumor segmentation and status classification.

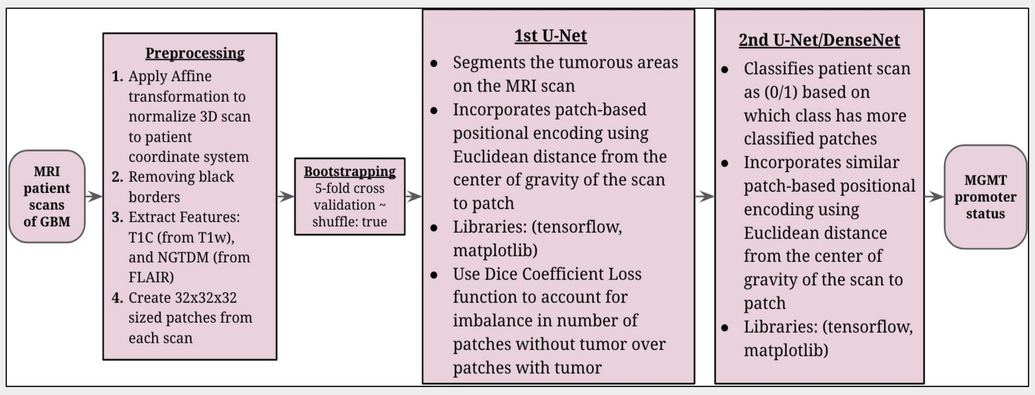
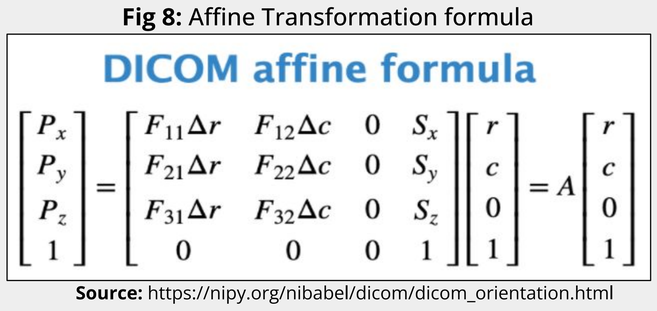


Fig 8: Affine Transformation formula



Source: https://nipy.org/nibabel/dicom/dicom_orientation.html

Fig 9: U-Net diagram

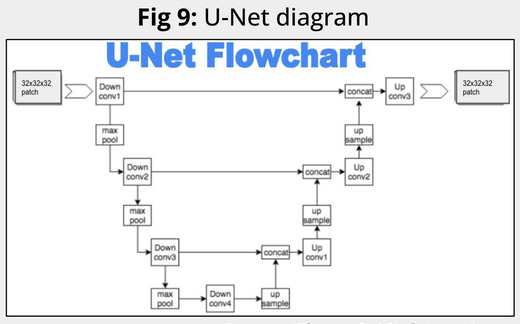
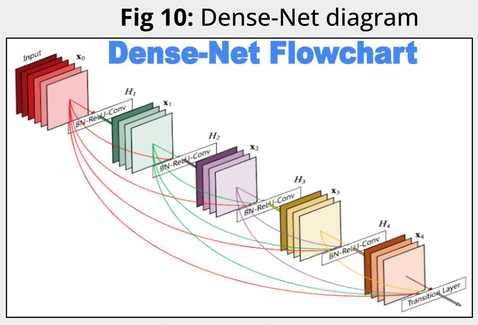


Fig 10: Dense-Net diagram

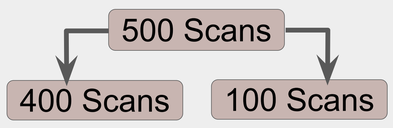


Source (Figure 9-10): https://www.coursera.org/specializations/ai-for-medicine

Cross Validation

- Five-fold cross validation with 100 epochs was used for training/testing the model
- The shuffle parameter was set to true in order to avoid overfitting of the algorithm to the dataset
- Dataset (500 images) was split 80/20 ~ 400 scans for training the model, and 100 scans for validation

Fig 11: Dataset Split

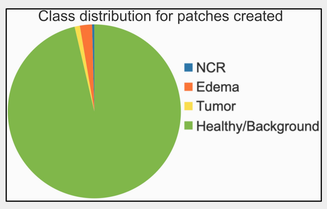


Revisions

Fig 12: Features Incorporated in each Iteration

Version	Features Computed and Incorporated into Model
1	Pixel values
2	Pixel values + T1c min
3	Pixel values + NGTDM coarseness
4	Pixel values + T1c min + NGTDM coarseness
5	Pixel values + T1c min + NGTDM coarseness + Positional encoding
6	Pixel values + T1c min + NGTDM coarseness + Positional encoding + Weighted Cross Categorical Entropy
7	Pixel values + T1c min + NGTDM coarseness + Positional encoding + Dice Coefficient

Fig 13: Dataset distribution



*** All images, diagrams, and charts created by Ananya Anand unless otherwise cited**

Results

- **U-Net:** 0.86 accuracy (F1 Score: 0.89, Precision: 0.91, Recall: 0.88)
- **Dense-Net:** 0.71 classification accuracy ~ 100 epochs ~ (F1 Score: 0.72, Precision: 0.70, Recall: 0.75)
- Runs under 1GB of memory

Fig 14: Completed Segmentation Labelled Results from the U-Net: (Original vs Segmented)

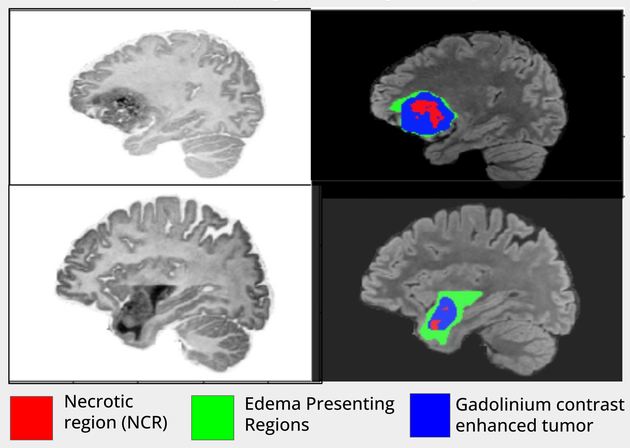


Fig 15: Resulting accuracies from each iteration

Features Computed and Incorporated into Model	Accuracy
Pixel values	.45
Pixel values + T1c min	.55
Pixel values + NGTDM coarseness	.53
Pixel values + T1c min + NGTDM coarseness	.57
Pixel values + T1c min + NGTDM coarseness + Positional encoding	.62
Pixel values + T1c min + NGTDM coarseness + Positional encoding + Weighted Cross Categorical Entropy Loss Function	.68
Pixel values + T1c min + NGTDM coarseness + Positional encoding + Dice Coefficient Loss Function	.71

Fig 16: Confusion matrix of results from model ~ 100 scans

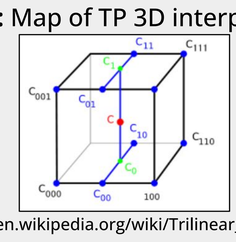
TN: 34	**FP:** 16
FN: 13	**TP:** 37

True Positive (TP)
True Negative (TN)
False Positive (FP)
False Negative (FN)

Future Work

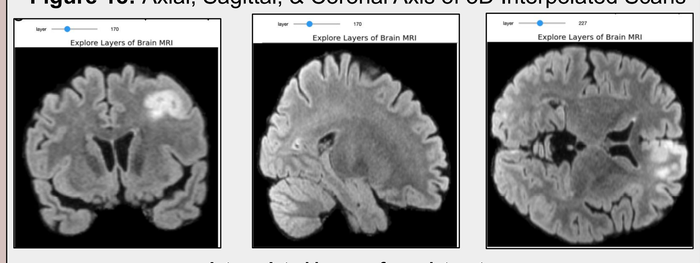
- I plan to continue exploring the effects of interpolating the MRI input scans to generate additional accurate data points to train my model. Currently, the MRI scans in the dataset have been taken with various scanner settings, with variations in the number of slices between images. I believe that interpolation can be used to bring all the MRI scans to the same dimensions of 512x512x512. Two methods of performing interpolation are the Nearest Neighbor (NN) method and Trilinear Interpolation (TP).
- Additionally, I plan on partnering with local hospitals to test out my model, and conduct a collaborative study

Fig 17: Map of TP 3D interpolation



https://en.wikipedia.org/wiki/Trilinear_interpolation

Figure 18: Axial, Sagittal, & Coronal Axis of 3D Interpolated Scans



Interpolated images from dataset:
https://www.kaggle.com/c/rsna-miccai-brain-tumor-radiogenomic-classification

References

1. Chen, X., Zeng, M., Tong, Y., Zhang, T., Fu, Y., Li, H., Zhang, Z., Cheng, Z., Xu, X., Yang, R., Liu, Z., Wei, X., & Jiang, X. (2020, September 23). *Automatic prediction of MGMT status in glioblastoma via Deep Learning-based Mr Image Analysis*. BioMed Research International. Retrieved January 13, 2022, from https://www.hindawi.com/journals/bmri/2020/9258649/
2. Yogananda, C. G. B., Shah, B. R., Nalawade, S. S., Murugesan, G. K., Yu, F. F., Pinho, M. C., Wagner, B. C., Mickey, B., Patel, T. R., Fei, B., Madhuranthakam, A. J., & Maldjian, J. A. (2021, May 1). *MRI-based deep-learning method for determining glioma MGMT promoter methylation status*. American Journal of Neuroradiology. Retrieved January 13, 2022, from http://www.ajnr.org/content/42/5/845
3. Saatci, O., Kaymak, A., & Ersan, P. G. (2020, May 15). *Targeting Lysyl oxidase (LOX) overcomes chemotherapy resistance in triple negative breast cancer*. Nature News. Retrieved January 15, 2022, from https://www.nature.com/articles/s41467-020-16199-4
4. Banerjee, S., & Mitra, S. (2020). Novel Volumetric Sub-region Segmentation in Brain Tumors. *Frontiers in computational neuroscience*, *14*, 3. https://doi.org/10.3389/fncom.2020.00003
5. Calabrese, E., Villanueva-Meyer, J. E., & Cha, S. (2020, July 16). *A fully automated artificial intelligence method for non-invasive, imaging-based identification of genetic alterations in glioblastomas*. Nature News. Retrieved January 13, 2022, from https://www.nature.com/articles/s41598-020-68857-8
6. Kaggle GPU: https://www.kaggle.com/code/dansbecker/running-kaggle-kernels-with-a-gpu/notebook
7. Mehta, V. (2021, January 20). *Glia-deep: Glioblastoma image analysis using deep learning convolutional neural networks to accurately classify gene methylation and predict drug effectiveness*. Cancer Biology and Therapeutic Oncology. Retrieved January 13, 2022, from https://www.imedpub.com/abstract/gliadeep-glioblastoma-image-analysis-using-deep-learning-convolutional-neural-networks-to-accurately-classify-gene-methylation-and-predict-drug-effectiveness-37929.html
8. Han, L., & Kamdar, M. R. (2018). MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, *23*, 331–342.