Hezekiah Branch
CS 135: Machine Learning
Professor Marty Allen
Tufts University

Computational Alchemy:
Outperforming Gold Standard Tumor Segmentation in GBM via Deep Learning

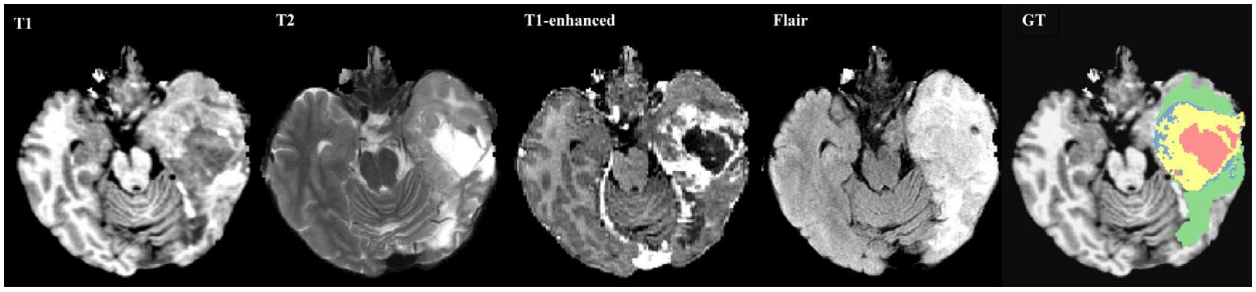

Figure 1: MRI modalities (n=4) used as input for various CNNs with ground truth labels on last image

In this paper, we will discuss the application of machine learning methods, particularly deep convolutional neural networks (dCNNs), to the problem of tumor segmentation in patients with glioblastoma multiforme (GBM). GBM is a grade IV glioma with largely unknown causes that is typically explored in radiogenomics, a field that bridges imaging and the science of genes (Tian et al., 2021). Glioma is the most common primary brain tumor that results from mutated glial cells in the spinal cord and brain (Ranjbarzadeh, 2021). The grade IV designation categorizes GBM as a member of the most destructive class of glioma. At the time of this paper, GBM is considered to be the most aggressive form of brain cancer with minimal life expectancy after prognosis. Currently, there is no cure or effective drug therapy for GBM given the difficulty of predicting the loci of tumorigenesis (i.e. tumor growth) on canonical signalling pathways associated with GBM (Holtzapple et al., 2021). However, tumor

segmentation is believed to be a possible route of predicting the growth of malignancy associated with GBM. Tumor segmentation deals with extracting the spatial location of a tumor from an inputted brain image, usually through multi-modal magnetic resonance imaging (MRI). It is also considered to be one of the most challenging tasks in the field of medical image analysis (Zhao et al., 2019). MRI is the standard data collection method for this task, often outperforming CT scans given the ability to detect blood movement without extraneous artifacts (Ranjbarzadeh et al., 2021). Despite the recent deep learning revolution in the world of neuroimaging, the “gold standard” of tumor segmentation has historically been manual segmentation where medical experts build handcrafted annotations of brain regions where tumor growth is observed. However, these manual annotations are highly prone to interobserver and intraobserver variability, leading to dramatic results in the individualized treatment of patients, not to mention an incredibly inefficient analysis workflow (Botvinik-Nezer et al., 2020). Given this high amount of variability in the approach of tumor segmentation, the problem is categorized into three classes: (1) “gold standard” manual segmentation, (2) semi-automatic segmentation, and (3) fully-automatic segmentation. While the discussion in this paper narrows the focus of segmentation to the neuropathology of GBM, this three-way categorization is often applied to any segmentation challenge for a given disease. Current applications of deep learning to GBM brain imaging data focus primarily on the third category of tumor segmentation, motivated by the goal of creating robust survival prediction pipelines that can lead to the first cure for GBM. The most widely accepted benchmark for the task of automatic tumor segmentation is the BRATS dataset which has released an annual challenge for automatic segmentation since 2012 (Menze et al.,

2014). The most widely accepted metric for evaluating segmentation performance is the Dice Score which is a statistical validation metric that quantifies the similarity of a segmentation prediction and its ground truth annotation. Secondary metrics include cross-entropy, weighted cross-entropy, Dice loss, and focal loss (Zhou et al., 2019). In the next section, we discuss the role of preprocessing brain imaging data on creating robust deep learning pipelines for overall survival (OS) analysis.

One of the most important findings in the application of deep learning methods to GBM brain imaging data is the critical role of detailed preprocessing techniques for preoperative data. A major contribution to the world of cancer research was the Woletz et al. (2019) contribution of the optimized preprocessing pipeline used for 569 datasets. This pipeline experimented with varying combinations of brain image preprocessing techniques including: (1) bias-field correction, (2) polynomial detrending (first and second orders), (3) white matter (WM) signal regression, (4) cerebrospinal fluid (CSF) signal regression, (5) realignment parameter (RP) regression. This team also provided a novel metric for assessing preprocessing quality for resting-state functional magnetic resonance imaging (rs-fMRI), a recently popular imaging modality that “offers the possibility to assess brain function independent of explicit tasks and individual performance.” The work by Lindquist et al. (2019) of how modular preprocessing reintroduces artifacts downstream of rs-fMRI is incredibly timely for the increase of preprocessing publications in the last decade, allowing researchers to dedicate more time to building robust preprocessing pipelines. Işın and Şah (2016), in more detail, discuss the variations of brain images between patients including 3D features before processing that include tumor shape, size, location, boundary in relation to healthy

tissue, and the need to not only segment the surface but also sub-regions of the brain. These complexities often surpass the capabilities of more edge-based, linear systems that are unable to capture the dimensions of such data. Remembering that preoperative imaging sets are the norm for this task, the argument of focusing on tumor tissue samples to avoid such complexities ignores the human challenges of acquiring such samples.

Presently, tumor tissue samples require intense radiation therapy and surgical treatment that can cause severe damage in the body. As a result, most deep learning approaches focus on preoperative data given the life span for a patient with GBM is around 14 months (Liu et al., 2021). The most common modalities of MRI that appear in literature on deep learning for preoperative tumor segmentation are T1, T1c, T2, and FLAIR (Brown et al., 2014). T1 and T2 definitions originate from tissue relaxation timings where T1 timing monitors longitudinal relaxation time and T2 monitors transverse relaxation time. Both modalities converge on the rate of relaxing tissue to equilibrium after activation. FLAIR is then an additional modality useful for analyzing brain images when the effect of fluid movement is noticeable. As mentioned by Brown et al. (2014), The T1 and T2 MRI modalities are highly practical given that there are easily differentiable differences in the way that cerebrospinal fluid (CSF) appears in their image domains, typically dark for T1-weighted images and bright for T-2 images. Isensee et al. (2019) have presented arguably the most noteworthy contribution to automatic preprocessing with the HD-BET algorithm, relying on a U-Net variant (Ronneberger et al., 2015) to process damaged brains. Previous to this contribution, brain extraction algorithms were only optimized to handle healthy brains and struggled

to successfully extract features on more realistic validation sets. Additionally, U-Net was accepted as the most popular architecture for deep learning in the medical imaging community, bolstering the place of CNNs for brain tumor segmentation with limited annotations. These findings indicate that, similar to machine learning problems outside of cancer research, preprocessing of training data in cancer research is critically important for classification of GBM and other forms of cancer. Next we discuss the application of deep convolutional neural networks (dCNNs) to image segmentation, first focusing on dCNN as an end-to-end predictor and later, as a member of an ensemble of network-based predictors.

On the topic of deep learning for predicting overall survival (OS) of GBM patients, both preprocessing and automatic segmentation is often carried out with dCNNs for medical analysis. As previously discussed, brain tumor segmentation is not only important for analyzing neural differences between multiple samples but also for expediting timely diagnosis for future treatment. Transfer learning is also not off the table in such approaches. A notable achievement from Lao et al. (2017) extracts deep features using transfer learning to predict OS via genomic signatures. The team uses a LASSO model as input into a Cox regression model and combines clinical risk factors with genomic signatures for feature engineering. The pre-trained model used for deep feature extraction is a dCNN with five convolution layers and three fully-connected layers. The hyper-parameters of the dCNN (CNN_S) were weight decay 5×10^{-4}, momentum 0.9, and an initial learning rate of 10^{-2}. Necrosis, tumor core, and tumor subregions were used as input for a total of 98304 (4 x 3 x 2 x 4096) features were extracted for each patient using forward propagation. The team evaluated prognostic

performance using concordance index (C-index) which generalizes area under the receiver operating characteristic curve (AUROC). Top features were then selected after thresholding the C-index, followed by removing any feature with the lowest-valued correlation from pairs of highly correlated features. The value of this paper is the advantage of constructing deep genomic signatures and using these signatures as signals for biomarkers that can lead to robust OS prediction models, relying only on preoperative data. With a slightly different objective in mind, Fu et al. (2021) offers an automatic deep learning-based workflow for predicting OS, prioritizing feasibility with a 3-dimensional CNN. Similarly, Bae et al. (2020) indicates the power of using deep neural networks on handcrafted genomic data to distinguish GBM tumors from solitary brain metastases, outperforming traditional machine learning models such as NB naïve Bayes, RF random forest, Ada adaptive boosting, L-SVM linear support vector machine, R-SVM radial basis function support vector machine, and LDA linear discriminant analysis. A slightly different approach to this problem is explored by Havaei et al. (2017) which proposes a number of CNN architectures for brain tumor segmentation before detailing a novel *Two-Path* dCNN architecture that captures local and global contexts of brain images. The two-phase training procedure that the team includes also manages imbalanced datasets well and at a much higher order than competing state-of-the-art models. The team makes reference to the Menze et al. (2014) distribution of papers that quantifies the number of publications focused on automatic brain tumor segmentation over the past few decades. The *Two-Path d*CNN recognizes the lack of resolution in the third dimension for BRATS images and handles preprocessing using a sequence of 2D axial slices, associating each pixel with a different MRI modality (i.e. T1, T2, T1C, and

FLAIR). Training for an imbalance of healthy and unhealthy voxels, followed by regularization and dropout allow for the state-of-the-art results that the team describes before mentioning the usefulness of cascading neural networks. In the next section, we discuss the use of cascading and fusion neural network models that combine CNNs.

Given the complexity of brain imaging data, particularly the images found in the benchmark BRATS dataset, multiple researchers have found ensembles of deep networks to be useful for brain tumor segmentation. Fusing CNNs as seen in the Hao et al. (2021) paper achieved an overall accuracy of 96.34% for GBM tumor identification, relying on higher-quality training data such as hyperspectral imaging for intraoperative diagnosis. Calabrese et al. (2020) mentions the usefulness of cascading instances of 2-dimensional dCNN for automatic 3D segmentation of three key components of GBM that are seen on MRI: enhancing, non-enhancing tumor, and surrounding tumor related edema. These results are generated on preoperative MRI data, allowing for the prediction of genetic alterations without the need for invasive data collection methods. Further, Zhao et al. (2019) describes, in addition to a variety of deep learning architectures, the ways in which later fusions can learn the complex relationship between different modalities. The contributions to the topic of multi-modal fashions and fusion strategies for deep learning, particularly for brain tumor segmentation, are still fairly recent. Several of the papers with the highest reported metrics were published this year and still require additional review for practicality. However, the explosion of interest in this topic provides support toward the idea of this new “alchemy” of brain tumor segmentation, redefining our understanding of medical image analysis. In conclusion, we have discussed the many facets that compose the difficult challenge of brain tumor

segmentation including (1) the critical role of preprocessing, (2) the effectiveness of deep learning for robust overall survival (OS) prediction with deep convolutional neural networks (dCNNs), and (3) the ongoing search for ensemble strategies that fuse deep models for improved performance. And while these methods are not without limitations, they may transform the way in which we target complex diseases and more intentionally therapize treatment for individuals with limited time for a cure.

References

Bae, S., An, C., Ahn, S. S., Kim, H., Han, K., Kim, S. W., ... & Lee, S. K. (2020). Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: model development and validation. *Scientific reports*, *10*(1), 1-10.

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., ... & Rieck, J. R. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84-88.

Brown, R. W., Cheng, Y. C. N., Haacke, E. M., Thompson, M. R., & Venkatesan, R. (2014). *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons.

Calabrese, E., Villanueva-Meyer, J. E., & Cha, S. (2020). A fully automated artificial intelligence method for non-invasive, imaging-based identification of genetic alterations in glioblastomas. Scientific reports, 10(1), 1-11.

Fordham, A. J., Hacherl, C. C., Patel, N., Jones, K., Myers, B., Abraham, M., & Gendreau, J. (2021). Differentiating Glioblastomas from Solitary Brain Metastases: An Update on the Current Literature of Advanced Imaging Modalities. *Cancers*, *13*(12), 2960.

Fu, J., Singhrao, K., Zhong, X., Gao, Y., Qi, S. X., Yang, Y., ... & Lewis, J. H. (2021). An Automatic Deep Learning–Based Workflow for Glioblastoma Survival Prediction Using Preoperative Multimodal MR Images: A Feasibility Study. Advances in radiation oncology, 6(5), 100746.

Hao, Q., Pei, Y., Zhou, R., Sun, B., Sun, J., Li, S., & Kang, X. (2021). Fusing Multiple Deep Models for In Vivo Human Brain Hyperspectral Image Classification to Identify Glioblastoma Tumor. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-14.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., ... & Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical image analysis*, *35*, 18-31.

Holtzapple, E., Cochran, B., & Miskov-Zivanov, N. (2021). Automated verification, assembly, and extension of GBM stem cell network model with knowledge from literature and data. *bioRxiv*.

Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., ... & Kickingereder, P. (2019). Automated brain extraction of multisequence MRI using artificial neural networks. *Human brain mapping*, *40*(17), 4952-4964.

Işın, A., Direkoğlu, C., & Şah, M. (2016). Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Computer Science*, *102*, 317-324.

Lao, J., Chen, Y., Li, Z. C., Li, Q., Zhang, J., Liu, J., & Zhai, G. (2017). A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. Scientific reports, 7(1), 1-8.

Lindquist, M. A., Geuter, S., Wager, T. D., & Caffo, B. S. (2019). Modular preprocessing pipelines can reintroduce artifacts into fMRI data. *Human brain mapping*, *40*(8), 2358-2376.

Liu, Z., Jiang, Z., Meng, L., Yang, J., Liu, Y., Zhang, Y., ... & Zhou, R. (2021). Handcrafted and Deep Learning-Based Radiomic Models Can Distinguish GBM from Brain Metastasis. *Journal of Oncology*, *2021*.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... & Van Leemput, K. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, *34*(10), 1993-2024.

Mohan, G., & Subashini, M. M. (2019). Medical imaging with intelligent systems: a review. *Deep learning and parallel computing environment for bioengineering systems*, 53-73.

Park, J. E., Ham, S., Kim, H. S., Park, S. Y., Yun, J., Lee, H., ... & Kim, N. (2021). Diffusion and perfusion MRI radiomics obtained from deep learning segmentation provides reproducible and comparable diagnostic model to human in post-treatment glioblastoma. *European Radiology*, *31*(5), 3127-3137.

Ranjbarzadeh, R., Kasgari, A. B., Ghoushchi, S. J., Anari, S., Naseri, M., & Bendechache, M. (2021). Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. *Scientific Reports*, *11*(1), 1-17.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Tian, J., Dong, D., Liu, Z., & Wei, J. (2021). *Radiomics and Its Clinical Application: Artificial Intelligence and Medical Big Data*. Academic Press.

Wan, Y., Rahmat, R., & Price, S. J. (2020). Deep learning for glioblastoma segmentation using preoperative magnetic resonance imaging identifies volumetric features associated with survival. *Acta neurochirurgica*, *162*(12), 3067-3080.

Woletz, M., Hoffmann, A., Tik, M., Sladky, R., Lanzenberger, R., Robinson, S., & Windischberger, C. (2019). Beware detrending: Optimal preprocessing pipeline for low‐frequency fluctuation analysis. *Human brain mapping*, *40*(5), 1571-1582.

Zhou, T., Ruan, S., & Canu, S. (2019). A review: Deep learning for medical image segmentation using multi-modality fusion. Array, 3, 100004.