Title

Predicting Newly Diagnosed Glioma Pathology with MRI and Deep Learning

Author

Singh, Paramjot

Predicting Newly Diagnosed Glioma Pathology with MRI and Deep Learning

by
Paramjot Singh

THESIS
Submitted in partial satisfaction of the requirements for degree of
MASTER OF SCIENCE

in

Biomedical Imaging

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Janine Lupo
F1E4E52A4E3D4D8...

Janine Lupo
Chair

DocuSigned by:
Yang Yang
[illegible]471...

Yang Yang

DocuSigned by:
Yan Li
[illegible]4F9...

Yan Li

DocuSigned by:
DEA76591766D41E...

Dr. Javier E. Villanueva-Meyer□

Committee Members

Dedication and Acknowledgement

I wish to dedicate this thesis to my family, who have been my support throughout my degree and thesis research.

I want to thank Jacob Ellison, whose support throughout this thesis was invaluable. I also wish to thank Oluwaseun Adegbite for her help in the thesis writing process. Finally, I want to thank Professor Janine Lupo and the rest of the Lupo Lab for their support.

Predicting Newly Diagnosed Glioma Pathology with MRI and Deep Learning

Paramjot Singh

Abstract

Current methods of glioma pathology assessment using tumor score metric rely on the extraction of a biopsy sample for evaluation by a pathologist. This method is limited by the fact that tumor score can vary within a glioma and that it only gives information regarding glioma pathology at one time point. An approach in which allows for the assessment of glioma pathology at various timepoints and in the entire brain is thus desirable.

We explored such a method of glioma pathology prediction with machine learning, using both traditional and deep learning approaches. Using a dataset of patient information (MRI images and corresponding tumor scores, we performed several experiments with traditional machine learning models to explore the potential benefits of a deep learning based approach. We then developed, trained, and tuned a deep learning model that predicted tumor score from MRI data, and experimented with various forms of transfer learning to evaluate the impact of loading weights from different autoencoders.

We determined the results of our traditional machine learning experiments showed a potential for a deep learning model's ability to predict tumor score from MRI data. When evaluating our deep learning model, we found that domain shift played a significant role in affecting our results in terms of testing accuracy, and we explored several methods to alleviate this issue. That said, our deep learning approach did outperform our traditional machine learning models, indicating the effectiveness of this approach.

Table of Contents

List of Figures

List of Tables

Introduction

Gliomas are a form of cancer in the central nervous system which emerge due to the abnormal growth and division of glial cells (1). It is estimated that 6 out of every 100,000 Americans develop gliomas every year (1). Life expectancy for glioma patients vary based on patient characteristics and the nature of the tumor, such as patient age, glioma location, molecular nature, and tumor aggressiveness (4). In the case of glioblastomas, the most severe type of glioma, patient life expectancy has a mean of 14 months (4). Current methods of glioma diagnosis rely on surgical biopsies to extract tissue with which pathologists determine the presence of glioma. The rapid progression and low life expectancy of this disease emphasizes the importance of rapid access to diagnostic information at all stages of the treatment process, as well as for an understanding of the nature of gliomas and how a glioma's composition should inform patient treatment. As a part of this information is the category which the glioma belongs to, in terms of severity.

Traditionally, the WHO has classified gliomas based on the cellular origin of the tumor (2). This led to the creation of the categories of astrocytomas (of which glioblastomas are the most severe), oligodendrogliomas, and mixed gliomas (1). Alongside this anatomical classification exists the WHO grading system for gliomas, in which tumor samples are evaluated and categories based on histological features such as "increased cellular density, nuclear atypias, mitosis, vascular proliferation and necrosis" (2). In the 2021 WHO criteria for classifying gliomas, molecular markers are what determine glioma grade (10). Rather than relying on histological analysis, such novel glioma classification methods focus on traits such as mutations of the IDH1/IDH2 gene, ATRX gene mutation, and telomerase reverse transcriptase levels (5). IDH1/IDH2 is a gene which plays a role in cell development, and mutations in it can lead to tumorigenesis in glial cells, as well as in cases of leukemia (8). Likewise, telomerase reverse transcriptase is associated with the development of tumors in cells, as it lengthens the telomeres at the end of chromosomes and allows for abnormal patterns of cell division (9). The old 2007 WHO glioma grading system ranged from 1 to 4, with 1 corresponding to the least aggressive forms of glioma and 4 corresponding to the most aggressive and the overall grade for a patient determined by the worst grade obtained by several tissue samples and prone to sampling errors due to the heterogeneity of these lesions (6). The revised 2021 WHO criteria of gliomas of the IDH-mutant type are considered lower grade, while gliomas with the IDH-wildtype type are of considered glioblastoma, the most malignant subtype (10). This system of grading and classification is the current standard in terms of glioma pathology. Thus, the molecular and genetic aspects of a glioma directly impact the nature of the growth and disease progression. A molecular classification system provides a way to classify gliomas which could have direct implications for disease therapy, as the molecular nature of a glioma can help determine eligibility for new targeted therapies to improve patient outcomes (5).

An additional method alongside the WHO grading system for quantifying tumor pathology is the assignment of tumor scores, which evaluate tumor samples on a histological basis. Tumor scores are assigned by a neuropathologist upon review of H&E-stained sections and range from 0 to 3 (3). A score of 0 corresponds to a sample in which no cells contain tumor, while a 1 corresponded to a sample containing an “infiltrating tumor margin” (3). A tumor score of 1 corresponds <10% tumor cells in a tissue sample. A tumor score of 2 is given to a biopsy sample containing between 10% and 75% cells of a “infiltrating cellular tumor”, and a tumor score of 3 corresponds to more than 75% of the cells in a sample being indicative of tumor, with no non-tumor neuropil (3). This system of scoring tumors is used at UCSF to evaluate glioma aggressiveness from pathology, and thus during this project, tumor score was used to classify patient biopsy samples on a voxel or region based level rather than the WHO grading system which is lesion based.

Machine learning techniques and their applications to the analysis of gliomas is an area of active research. There has historically been significant focus and progress on the problem of segmenting glioma lesions on MRI images of the brain, but recent efforts have focused more on the prediction of tumor characteristics such as molecular composition (6). This work has relied on a variety of machine learning algorithms and techniques, including deep learning. These developments form a major subfield of the emerging field of radiomics, which seeks to integrate machine learning into the medical image processing pipeline to help extract more information from medical images, such as MRI images. While current methods of glioma classification rely on neuropathologist analysis, the classification of gliomas through deep learning methods applied to MRI images is an area of active research which could help speed up the classification of patient samples. Most development in this field has attempted to create models which attempt to classify tumors into the categories laid out by the older WHO grading system (6), or into a binary high- and low-grade labeling system which group together several WHO glioma grades. Applying deep learning on small patches of images surrounding the coordinates from which tissue has been sampled has yet to be explored in classifying glioma based on pathology. Thus, there is an opportunity for novel research into the application of deep learning to solve this problem.

That said, there remains a great deal of progress in applying deep learning technologies such as convolutional neural networks have been applied to predicting other characteristics of gliomas and their outcomes, with one example of such applications being the segmentation of regions of interest within MRI images of gliomas (6). However, as noted by the literature, “no consensus has been reached regarding the optimal ML algorithm for image-based glioma classification” (6). Therefore, the creation of a deep learning algorithm which classifies gliomas based on tumor score could help serve as a valuable tool in glioma classification. It would complement similar work being done to classify gliomas based on tumor grade while using a metric that is spatially varying throughout the tumor to assess lesion heterogeneity and spatial extent.

One case of deep learning being applied to generate spatial maps of tumor pathology is at the time of suspected recurrence, where a deep learning model designed to categorize glioma lesions on MRI images post-treatment into the categories of being either recurrent tumor or treatment effect (7). This model formed the basis for our exploration of deep learning applications for glioma pathology analysis. We borrowed from this model in terms of architecture and overall boosting strategies, which were all modified and tuned heavily during the project in order to yield better performance when applied to our problem case. The dataset used by Ellison et al. was also derived from UCSF patients, and while the project’s prediction aims differed from our own, a similar method of data acquisition was used for both projects.

The motivation for this project stems from the potential impact that a method to predict the pathology of glioma would have in terms of informing patient response to treatment. Gliomas can be highly heterogenous, and as a result delineating the exact boundaries of tumor regions over time can be difficult with current-day approaches. A non-invasive imaging method like MRI which could be used on patients between surgical treatments to collect images, which could then be fed into a deep learning model to predict glioma pathology at that point in time could help surgeons and oncologists plan out further surgery and potentially more accurately define response to cancer treatment. In the ideal scenario, this model would generate spatial maps, which would yield probabilities of cells being cancerous throughout the entire brain.

Methods

Data Acquisition

The MRI data used in this project was all acquired on 3T GE Healthcare scanners. Patients who were scanned as part of this study had newly diagnosed gliomas before receiving any surgical treatment. MRI data was collected in the following sequences: T1-weighted pre- and post-contrast, T2-weighted FLAIR, DSC perfusion, and DWI. Fractional anisotropy and apparent diffusion coefficient maps were generated from the DWI sequence data, while peak height and percent recovery (of the delta R2* curve) maps were generated from the DSC perfusion data. In addition, tissue samples of the tumor region were obtained during surgery immediately before tumor resection and were assigned a tumor score by a neuropathologist. This led to an initial dataset of 1009 samples from 396 patients. This dataset was then filtered to ensure that all members of the dataset contained image data for all six needed MRI sequences, as well as tumor score information. Samples with an unassigned tumor score or an indeterminable tumor score were excluded from the dataset. Our final dataset consisted of 206 samples (see **Table 2.1**), with an additional 144 samples being of acceptable quality but lacking a pathologist assigned tumor score. Of the 206 samples used during the project, 10.19% has a tumor score of 0, 15.05% had a tumor score of 1, 27.67% had a tumor score of 2, and 47.09% had a tumor score of 3. Additional information regarding the nature of our dataset is shown in **Table 2.2**, revealing that of the 97 patients, 82 had gliomas which were categorized as glioblastomas, 10 had astrocytomas, and 5 had oligodendrogliomas. While we did not take this information into account in the course of our work, it may have had an impact on the performance of the deep learning model in terms of testing set accuracy.

Table 2.1. Dataset Breakdown by Tumor Score. The breakdown of the final dataset by tumor score.

Tumor Score	# of Samples	Percentage
0	21	10.19%
1	31	15.05%
2	57	27.67%
3	97	47.09%

Table 2.2. Dataset Breakdown by Tumor Type. Breakdown of the final dataset by glioma type.

Glioma Type	Count
Glioblastoma	82
Astrocytoma	10
Oligodendroglioma	5

We also performed minor preprocessing on the labels for the dataset, which originally corresponded to the tumor score. Due to the imbalanced nature of the dataset, tumor scores of 2 and 3 were a much larger share of the overall dataset than 0s and 1s. To alleviate this, a binary labeling method was employed, where tumor scores of 0s and 1s were considered as one group (with label 1) and tumor scores of 2 and 3 were labelled with 0. (This convention followed the convention used in the recurrent glioma project, which facilitated an easy transfer of weights in later transfer learning).

Preprocessing

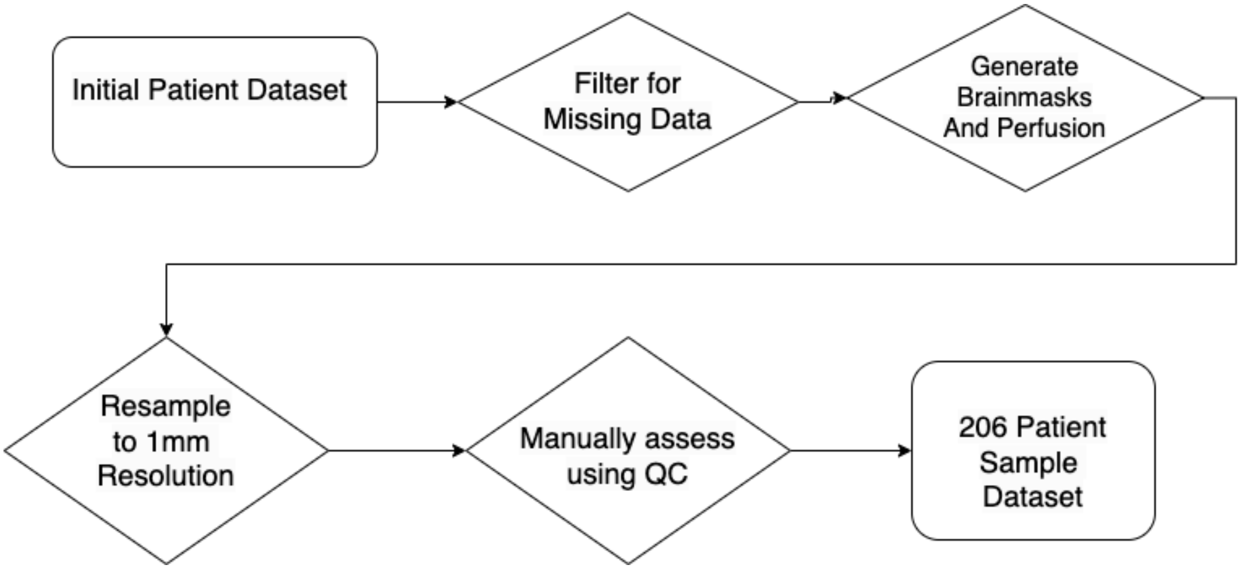
Initial Patient Dataset
Filter for Missing Data
Generate Brainmasks And Perfusion
Resample to 1mm Resolution
Manually assess using QC
206 Patient Sample Dataset

Figure 2.1. MRI Image Preprocessing Pipeline. The preprocessing pipeline developed and used to filter, align and resample our dataset.

We began our work by preprocessing the MRI data as shown in **Figure 2.1**. After filtering the dataset to ensure all needed images were present for the samples, we performed perfusion processing to generate maps of the percent recovery and peak height. All MRI images were first converted from dicom images to INT2 or BYT volume files, with corresponding IDF header files. Using Python scripts, all images were then aligned with the T1 image, ensuring that all MRI data was in the same space and that cubes of MRI data from different MRI images corresponded to the same location in the brain. After alignment, these images were resampled to a 1 mm cubic resolution and standardized to the mode of the histogram of normal-appearing brain tissue, before being saved as NIFTI files, a process which was done using Python and shell scripts. Python scripts were then used to both generate masks to represent brain regions across all images, as well as to generate cubes masks. Finally, a Python script was run to extract the 10x10x10 mm cube of MRI image data around the coordinates from which the tissue was sampled during surgery. These cubes were saved in the NIFTI format. While the tissue sample itself was much smaller than 10 mm, a 10 mm cube ensured that the tissue sample location would be captured within the cube regardless of minor shifts in MRI imaging.

We then shifted to manual inspection of all images, regions of interest, and brainmasks, using a Python-based QC tool developed previously in our lab. We manually checked for alignment and image coverage of ROIs as shown by the tool. Each sample was scored from 0 to 7 on a scale which categorized the quality of sample MRI images, with samples scored as 0 and 1 being of usable quality for machine learning. Higher scores indicated misalignment of one or more images, or missing images, and so were excluded from our final dataset.

Traditional Machine Learning

Once the preprocessing was complete, we moved to the problem of predicting tumor score with the usage of traditional machine learning approaches. The motivation for this stage of the project was to see if the outcomes of experiments with traditional machine learning indicated that a deep-learning based approach had potential to result in improved accuracy.

The first phase of our experiment was to preprocess the 10x10x10 pixel cubes of MRI data in three different ways as shown in **Figure 2.2**, to evaluate how different methods of treating the data would affect the accuracy of the model predictions. We were interested in seeing how maintaining the cube as an atomic piece of information affected performance when compared against methods which considered only subsets of the cube data. The first method of cube preprocessing was flattening, in which the 10x10x10 cube was flattened into a one-dimensional, 1,000 element array. This maintained the cube as a contiguous block of data, but it destroyed any patterns which were encoded in the spatial relationships between pixels. The second form of data preprocessing was to apply principal component analysis onto the cube, which would return the ten elements within the cube which contributed most to the overall variance of the sample. The final form of cube preprocessing was a pixel-by-pixel weak labeling approach. In this approach, each pixel in the 10x10x10 pixel cube was treated as its own sample, with a corresponding label and 6 channels of image information. With this approach, we had 1,000 times more samples than before, but all information encoded in the cube was lost. This approach is called “weak-labeled” because not all pixels in the original cube contributed equally to the overall cube label, so the cube label, when applied to each pixel individually may not have been accurate for some pixels.

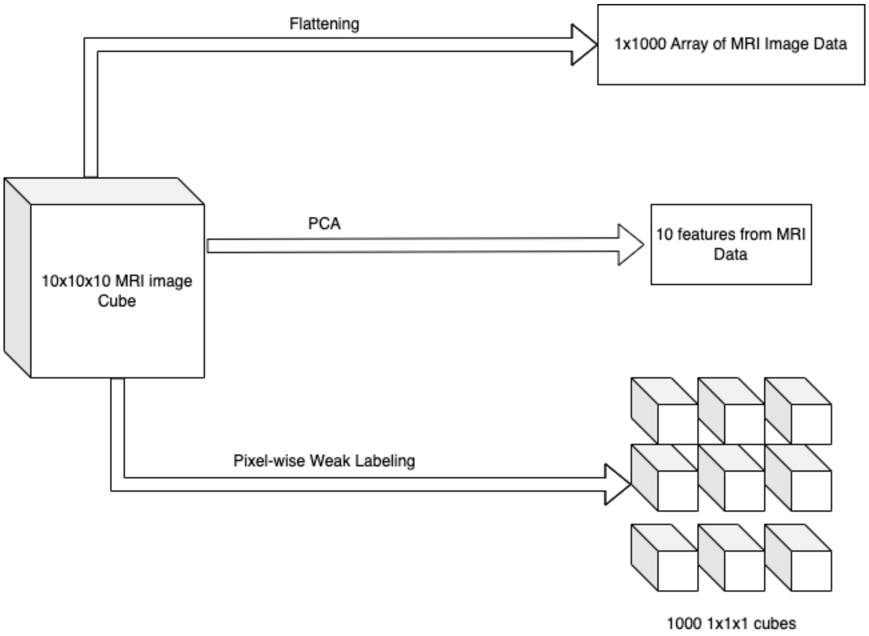

Figure 2.2. MRI Image Cube Preprocessing Approaches. The preprocessing pipeline developed and used to filter, align and resample our dataset.

Oversampling was used to deal with the unbalanced nature of the dataset using sklearn’s Synthetic Minority Oversampling Technique (SMOTE) algorithm to synthetize new samples which would match the data samples within the “low-score” label of 1. The parameters of the algorithm were specified so that it would use the K-Means algorithm with a neighborhood of 3 to generate new synthetic datapoints.

The traditional machine learning models used in this project were all implemented with the use of sklearn packages on Jupyter Notebook. A total of four different classification models were evaluated: logistic regression, AdaBoost, support vector classifier (SVC), and random forest classification. Each model was trained using 4-fold cross-validation with an overall 70/10/20 split between training, validation, and testing. All models were tuned, with hyperparameters modified to the values which experimentation had shown were the most effective in terms of in increasing validation accuracy. For example, the logistic regressor was set to classify with a one-over-rest multiclass paradigm and balanced class weight. The C regularization parameter for the SVC was set to 0.001 and the model was trained with a linear kernel. The random forest classifier was extensively tuned, with a max tree depth of 50, 100 estimators, a minimum sample split of 5, a maximum of 5 leaf nodes, and balanced class weight.

Deep Learning

The deep learning model used in this project was implemented in TensorFlow, and allowed the architecture shown in **Figure 2.3**. The architecture of the model was designed after the model by Ellison et al at UCSF, which classified regions of interest into the categories of recurrent glioma and treatment effect (7). The first portion of the model was an autoencoder, which was trained with three filtering layers followed by three encoding layers and MSE loss. The weights trained in the autoencoder were then used as a starting point to finetune the classifier model. The classifier consisted of dense layers with dropout and L1/L2 regularization after flattening the output of the encoding blocks from the autoencoder. The convolutional weights are trainable during fine tuning.

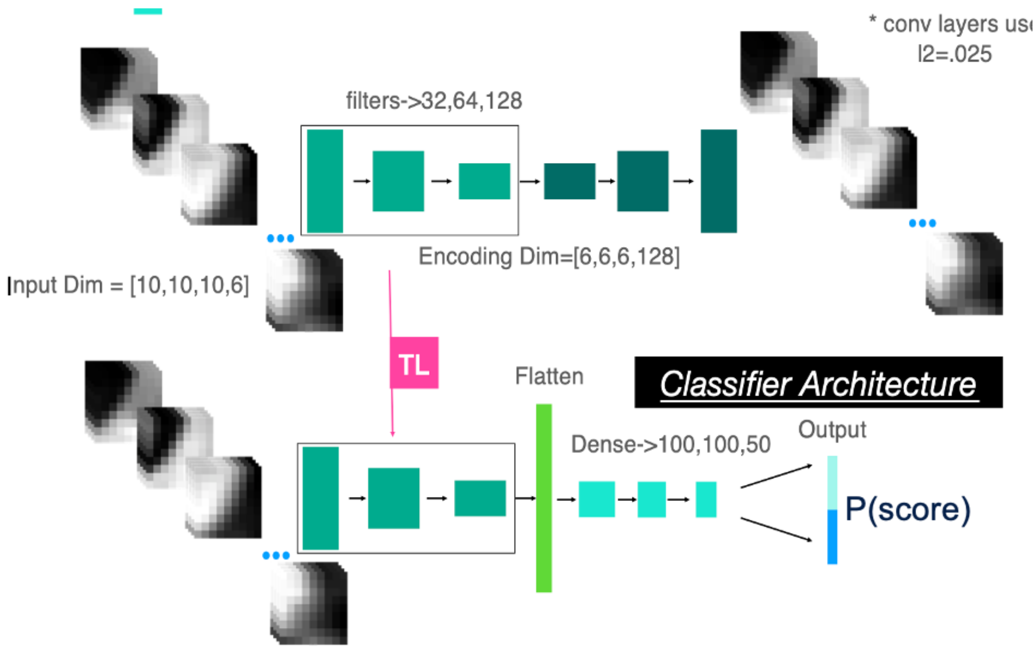

Figure 2.3. Deep Learning Classifier Architecture. The overall classifier architecture used in the deep learning portion of our project. Note the transfer learning shown from an autoencoder above into our classifier model.

The deep learning model was heavily tuned to ensure maximum accuracy. The L2 regularization for the autoencoder was set to 0.025, and the autoencoder was trained both with recurrent glioma data and our own dataset to see if there was any value in training on a larger (albeit qualitatively different) dataset. The classification model itself was trained with a learning rate of .000001 and a decay rate of 0.1. The classification model was also regularized, with an L2 term of 0.6 and a L1 term of 0.06.

Like the traditional models, the deep learning model was trained using 4-fold crossfold validation. A 70/10/20 group stratified split was used for training, validation, and testing. During each epoch, the epoch validation AUC was checked against the maximum encountered validation AUC. If greater, the model would save the epoch's weights for testing. The model was training for 500 epochs and was evaluated for saving at the end of the epoch.

Results

Traditional Machine Learning

The results of the traditional machine learning approach are listed in **Table 3.1.** Results are recorded as the mean of the validation AUCs reported for each of the four folds of the cross-fold validation setup. The random forest classifier with a flattened cube preprocessing approach performed best in terms of validation AUC, with a mean validation AUC of 0.6926. The random forest classifier maintained this lead in performance over the other models across all three methods of cube preprocessing. Apart from PCA cube preprocessing, AdaBoost performed the worst out of all classifiers for all preprocessing methods.

Table 3.1. Mean Validation AUCs for Traditional Approach. Mean Validation AUCs for traditional machine learning models, across the three forms of cube preprocessing.

Model	Flattened Mean Val. AUC	PCA Mean Val. AUC	Pixel-Based Mean Val. AUC
Random Forest	0.6926	0.5979	0.5593
SVC	0.6169	0.5798	0.484
Logistic Regression	0.5752	0.4473	0.508
AdaBoost	0.4978	0.501	0.228

Deep Learning

The results for the deep learning approach are broken down in **Table 3.2**, where the maximum validation AUC for each fold of cross-validation is shown for both cases of transfer learning: one with an autoencoder trained on our dataset, the other with weights derived from an autoencoder trained on recurrent glioma data which was part of the work done by Ellison et al. on the classification of recurrent gliomas versus treatment effect. The mean validation AUC for this setup is 0.4853, markedly lower than the mean validation AUCs of the two model setups in which the autoencoder was trained on the newly diagnosed glioma data.

Table 3.2. Mean Validation AUCs for Deep-Learning Approach. Validation AUCs across the four folds of our cross-validation setup for deep learning, as well as the mean AUC across all folds. The top row corresponds to the model which loaded weights from the recurrent glioma autoencoder, while the second row loaded weights from the autoencoder trained on our newly diagnosed glioma dataset. The third row records validation AUC results for an untuned model with TL trained on our dataset, and the final row records the test AUC results for a tuned model with TL trained on our dataset.

	Fold 1 AUC	Fold 2 AUC	Fold 3 AUC	Fold 4 AUC	Mean AUC
Val. AUCs with TL From Recurrent Glioma Dataset	0.5923	0.3788	0.4981	0.4712	0.4853
Untuned Val. AUC with TL from Newly Diagnosed Glioma Dataset	0.589	0.573	0.739	0.773	0.6678
Val. AUCs with TL from Newly Diagnosed Glioma Dataset	0.7038	0.5455	0.7713	0.7500	0.6926

The mean validation AUC across the four folds for the DL model trained with our autoencoder was 0.6926. If the outlier of the second fold is removed from consideration, the mean validation AUC rises to 0.7417. The mean AUCs for all three other setups are much lower than this value, ranging from 0.485 for the model with TL weights trained on the recurrent glioma dataset to 0.6678 for the untuned model with TL trained on our newly diagnosed glioma dataset.

We also present the validation ROC curves for the three models, with their corresponding ensemble AUC values, as shown in **Figure 3**. The ensemble AUC value for the tuned model with weights loaded from the recurrent glioma autoencoder is 0.43. The untuned model with weights loaded from the newly diagnosed glioma autoencoder's ensemble validation AUC value is 0.62, while the tuned model's is 0.64. Finally, the mean test AUC score for the tuned model with weights loaded from the newly diagnosed autoencoder was 0.4545, which was also much lower than the validation AUC for the same model.

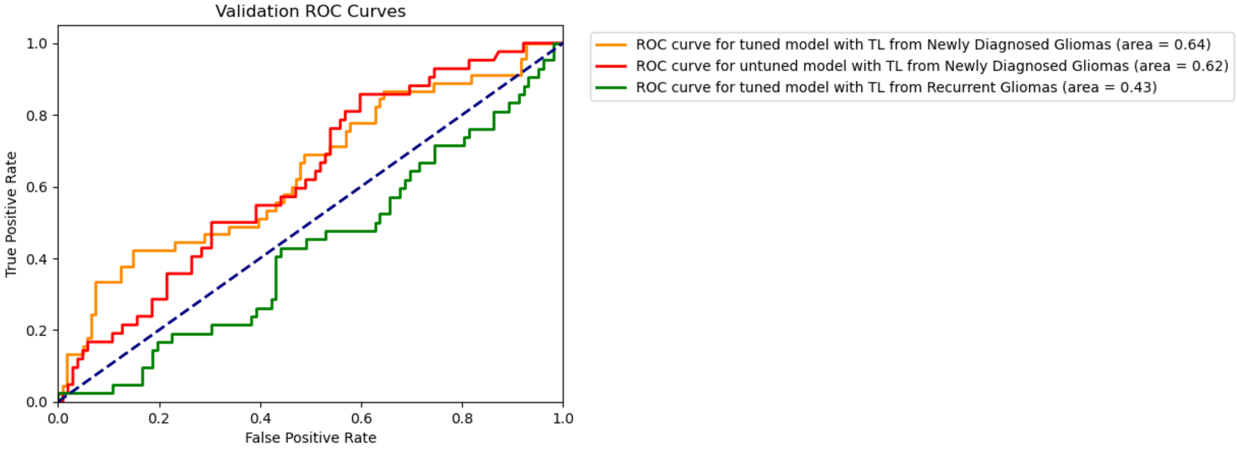

Figure 3.1. Validation ROC Curves. The validation ROC curves for the three deep learning models trained during this project.

Discussion

The results from our research into the traditional machine learning approach demonstrated the potential for a deep-learning based approach. In particular, the fact that the highest performance was seen when using the flattened cube preprocessing approach told us that there was predictive value in maintaining the contents of the cube of MRI data as a contiguous block of information. This aspect of the data is lost in PCA, which selects only the ten features which contribute most to the sample's variance, and in pixel-based weak labeling, in which the cube is broken into its pixels. A deep-learning based approach would maintain the contiguous nature of the cube data, but it would also maintain the three-dimensional nature of the cube itself, which we were unable to do with a traditional machine learning approach. Thus, any pattern encoded in the spatial relationship between pixels would be maintained with deep learning. We expected this to result in an additional boost in performance of the deep learning-based model, and the results showed that after excluding outliers, such a boost occurred.

Moreover, the fact that the best classifier in terms of performance was the random forest classifier gave additional confirmation that a deep learning approach would have merit. This is because the random forest classifier had more parameters than the other types of models and so was more tunable, allowing it to fit more to a more complex dataset. A deep learning model like ours has far more parameters (in our case, the model had a total of 2,525,783 trainable parameters) and so can be expected to perform better than traditional machine learning approaches.

Table 3.2 reveals that the transfer learning performed with the two autoencoders had a marked impact on the performance of the model. Our experiment showed that transfer learning with a dataset with the confounds of various treatment resulted in extremely poor performance. Indeed, the impact of the TL outweighed the impact of tuning the model, as shown in the relative similarity of the mean validation AUCs for the tuned and untuned models which used TL from the newly diagnosed glioma dataset. Reasons for the importance of the TL dataset choice likely include clinical differences in the effects of treatment on MRI characteristic of tumor, which was reflected in the cubes of MRI data we used for training. When we instead trained the autoencoder on the newly diagnosed dataset, the weights loaded in from transfer learning corresponded to images of gliomas that were of the same clinical stage as ours. It is likely that if another dataset of newly diagnosed gliomas was used to train the autoencoder, we would see the same sort of improved performance as we saw when we trained the autoencoder on our dataset. If a larger dataset was used we might see some additional improved performance due to an improved autoencoder yielding better weights.

The results from our deep learning model trained on the newly diagnosed glioma autoencoder reveal that there was a boost in performance when compared to the traditional models, but it also highlight the impact of outliers on our results. In particular, fold 2 of our four-fold cross-validation scheme performed much worse in terms of validation AUC than the other folds, with a gap of 0.15 when compared to the next lowest fold validation AUC. In contrast, the other three validation AUCs are all within 0.05 of one another. This gap was consistent across various tuning attempts and re-splitting of the training and testing sets. Therefore, we concluded the outlier effect is likely a consequence of the small size of the dataset. With a larger dataset the outlier effect observed in fold 2 may disappear, leading to a higher overall validation accuracy and an improvement in terms of mean AUC. The results as shown by the ROC curves confirms the same trend we observed in the mean AUCs across the four-fold cross-validation for the three deep learning model setups. The ensemble AUC score for the model with weights loaded from the autoencoder trained on the recurrent glioma data lagged in performance compared to the other two models, which matched what we observed with the mean validation AUCs. The small gap between the untuned and tuned models might be indicative of suboptimal tuning, and with another set of parameters the tuned model's performance in terms of validation AUC might increase.

That said, the wide gap between the validation and testing AUCs indicated that our model experienced significant domain shift. We attempted to resolve this by first modifying our regularization terms to prevent overfitting, and later by repeatedly splitting the dataset into test and train datasets in the hopes of ensuring our test and train sets would be representative of one another. Despite these changes, the gap in performance indicates that domain shift remained. One possible source of the domain shift could be the glioma type. We did not control or group by glioma type when dividing the dataset into test and train datasets, and if aspects of the histology of the tumor was not captured on MRI we would have missed those differences. This could have resulted in a lower testing accuracy.

Conclusion

The results of our two approaches to the problem of predicting tumor score from MRI data reveal the potential of applying deep learning for this application once more data is included. Problems remain in terms of managing domain shift to ensure that the high performance observed in the validation phase of the model is replicated in the test set, which will likely be solved once more pathology from the rest of the dataset is evaluated the main bottleneck. Nonetheless, the deep learning results suggest that separate models need to be trained for newly-diagnosed and post-treatment data to account for the effects of treatment affecting imaging metrics.

References

1. Mesfin FB, Al-Dhahir MA. Gliomas. [Updated 2023 May 20]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK441874/
2. Figarella-Branger D, Colin C, Coulibaly B, Quilichini B, Maues De Paula A, Fernandez C, Bouvier C. Classification histologique et moléculaire des gliomes [Histological and molecular classification of gliomas]. Rev Neurol (Paris). 2008 Jun-Jul;164(6-7):505-15. French. doi: 10.1016/j.neurol.2008.03.011. Epub 2008 Jun 10. PMID: 18565348.
3. Ramon F. Barajas, Jr and others, Regional variation in histopathologic features of tumor specimens from treatment-naive glioblastoma correlates with anatomic and physiologic MR Imaging, Neuro-Oncology, Volume 14, Issue 7, July 2012, Pages 942–954, https://doi.org/10.1093/neuonc/nos128
4. Mohammed S, Dinesan M, Ajayakumar T. Survival and quality of life analysis in glioblastoma multiforme with adjuvant chemoradiotherapy: a retrospective study. Rep Pract Oncol Radiother. 2022 Dec 29;27(6):1026-1036. doi: 10.5603/RPOR.a2022.0113. PMID: 36632307; PMCID: PMC9826661.
5. Kenta Masui, Paul S. Mischel, Guido Reifenberger. Chapter 6 - Molecular classification of gliomas. Editor(s): Mitchel S. Berger, Michael Weller. Handbook of Clinical Neurology, Elsevier, Volume 134, 2016, Pages 97-120. https://doi.org/10.1016/B978-0-12-802997-8.00006-2.
6. Marquet G, Dameron O, Saikali S, Mosser J, Burgun A. Grading glioma tumors using OWL-DL and NCI Thesaurus. AMIA Annu Symp Proc. 2007 Oct 11;2007:508-12. PMID: 18693888; PMCID: PMC2655830.

7. Ellison, Jacob. Tissue-level probabilistic mapping of treatment-induced effects in recurrent glioblastoma. ISMRM. 2022.

8. Yang H, Ye D, Guan KL, Xiong Y. IDH1 and IDH2 mutations in tumorigenesis: mechanistic insights and clinical perspectives. Clin Cancer Res. 2012 Oct 15;18(20):5562-71. doi: 10.1158/1078-0432.CCR-12-1773. PMID: 23071358; PMCID: PMC3897211.

9. Dratwa Marta, Wysoczańska Barbara, Łacina Piotr, Kubik Tomasz, Bogunia-Kubik Katarzyna. TERT—Regulation and Roles in Cancer Formation. Frontiers in Immunology. 2020. https://www.frontiersin.org/articles/10.3389/fimmu.2020.589929

10. Huang, L.E. Impact of *CDKN2A/B* Homozygous Deletion on the Prognosis and Biology of IDH-Mutant Glioma. *Biomedicines* 2022,*10*,246. https://doi.org/ 10.3390/biomedicines10020246

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Paramjot Singh

4F8CE72B997E4B1...

Author Signature

8/27/2023

Date