Pathologically-Validated Tumor Prediction Maps in MRI

Alex Barrington
Marquette University

Pathologically-Validated Tumor Prediction Maps in MRI

By

Alex Barrington, B.S.

A Thesis Submitted to the Faculty of the Graduate School, Marquette University,

in Partial Fulfillment of the Requirements for the Degree of Master of Science

Milwaukee, Wisconsin

December 2019

ABSTRACT

Pathologically-Validated Tumor Prediction Maps in MRI

Alex Barrington

Marquette University, 2019

Glioblastoma (GBM) is an aggressive cancer with an average 5-year survival rate of about 5%. Following treatment with surgery, radiation, and chemotherapy, diagnosing tumor recurrence requires serial magnetic resonance imaging (MRI) scans. Infiltrative tumor cells beyond gadolinium enhancement on T1-weighted MRI are difficult to detect. This study therefore aims to improve tumor detection beyond traditional tumor margins. To accomplish this, a neural network model was trained to classify tissue samples as ‘tumor’ or ‘not tumor’. This model was then used to classify thousands of tiles from histology samples acquired at autopsy with known MRI locations on the patient’s final clinical MRI scan. This combined radiological-pathological (rad-path) dataset was then treated as a ground truth to train a second model for predicting tumor presence from MRI alone. Predictive maps were created for seven patients left out of the training steps, and tissue samples were tested to determine the model’s accuracy. The final model produced a receiver operator characteristic (ROC) area under the curve (AUC) of 0.70. This study demonstrates a new method for detecting infiltrative tumor beyond conventional radiologist defined margins based on neural networks applied to rad-path datasets in glioblastoma.

ACKNOWLEDGEMENTS

I would like to thank my friends and family, all those who assisted and mentored me during my time at Marquette, my professors, committee, and director. I would like to thank Dr. Elizabeth Cochran for taking the time to go through thousands of tiles. I would like to thank Dr. John Bukowy for his assistance setting up the networks. I would like to thank Allison Lowman for all the incredible work she does in the LaViolette Lab. I would like to thank Sam Bobholz for his assistance in dataset creation and Sean McGarry for allowing me to bounce ideas off them regularly. I would like to thank the Medical College of Wisconsin and all of the resources that were provided to me through them and the Research Computing Center. And I would like to thank Dr. Peter LaViolette for all his help and allowing me to do the work I have done in his lab.

TABLE OF CONTENTS

I.INTRODUCTION

According to the American Brain Tumor Association (ABTA), almost 80,000 people will be diagnosed with a brain tumor this year, and 16,000 people will die from a brain tumor this year (Brain tumor FAQs | ABTA.). Glioblastoma (GBM), a World Health Organization (WHO) grade IV astrocytoma, is an aggressive glial tumor with a poor 5-year survival rate of 5% (Brain tumor – statistics, 2018; Tamimi & Juweid, 2017). These tumors are diagnosed with a combination of imaging techniques and tissue samples acquired during surgery or biopsy (Brain tumor – diagnosis, 2018). Research on the exact initial cause of GBMs is still limited, however, these tumors can develop from lower-grade gliomas, such as grade II or grade III astrocytomas (Development of glioblastoma multiforme.; Forst, Nahed, Loeffler, & Batchelor, 2014).

Glioblastoma Multiforme

GBMs are classified as one of three genetic subtypes, based variations of isocitrate dehydrogenase (IDH): IDH-wildtype, IDH-mutant, and not otherwise specified (NOS). Additionally, IDH-wildtype has three subtypes: giant cell glioblastoma, gliosarcoma, and, most recently added, epithelioid glioblastoma. IDH-wildtype accounts for around 90% of GBMs diagnosed (Louis et al., 2016). These are generally de novo cases, while the IDH-mutant varieties commonly develop from lower grade gliomas (LGGs) (Ohgaki & Kleihues, 2013). These two varieties also show trends in the ages when they develop. IDH-wildtype is generally diagnosed in patients with a median age around 62, while the mutant

variety has a median age of 44 (Louis et al., 2016). Additional differences between the two include median survival, where IDH-wildtype is about 15 months and IDH-mutant is 31 months, the location of the tumor, with the IDH-wildtype developing most anywhere in the cerebrum and the IDH-mutant developing in the frontal lobe, and apparent necrosis, which is widespread in IDH-wildtype and seen much less in IDH-mutant (Ellingson, B. M. et al., 2013; Louis et al., 2016).

Invasive GBM tumor cells spread in a diffuse manner lacking a defined tumor boundary beyond the core (The life of a brain tumor: How does glioblastoma grow? – penn medicine.), often following white matter tracts (Mickevicius et al., 2015). Abnormal tissue outside of contrast enhancing margins can be seen as hyperintense regions on a T2-weighted fluid attenuated inversion recovery (FLAIR) MRI scan. These hyperintense regions are thought to be a combination of non-contrast-enhancing tumor (nCET), and vasogenic edema associated with an inflammatory response. Another treatment related imaging artifact is known as pseudo-progression, which occurs during the first 12 weeks after the standard of care treatment administration of radiation and temozolomide, and manifests as contrast enhancement not associated with viable tumor (Melguizo-Gavilanes, Bruner, Guha-Thakurta, Hess, & Puduvalli, 2015; Villanueva-Meyer, Mabray, & Cha, 2017). Although there have been studies showing the efficacy of multi-parametric MRI techniques, they generally use biopsy samples, which is not necessarily representative of the entire underlying tissue (Kazerooni et al., 2018; Kimura & da Cruz, L. Celso Hygino, 2016; Yang, X. et al., 2019). Additionally, areas adjacent to the bright area on an

MRI may contain viable tumor or pseudoresponse (Nguyen et al., 2016; Villanueva-Meyer et al., 2017). Differentiating these processes from tumor growth, response to treatment, and treatment effects is difficult and puts increasing pressure on optimizing surgical resection to be correct and sufficient (Li, Suki, Hess, & Sawaya, 2016; Villanueva-Meyer et al., 2017). Therefore, an understanding of how multimodal imaging signatures relate to the underlying tissue pathology is critical to better treatment direction.

Research has shown that poorer prognosis, IDH-mutant status, and decreased survival are all correlated with the presence of nCETs (Cohen-Gadol, DiLuna, Bannykh, Piepmeier, & Spencer, 2004; Jain et al., 2014; Kotrotsou et al., 2018; Lasocki, Gaillard, Tacey, Drummond, & Stuckey, 2016; Lasocki, Gaillard, Tacey, Drummond, & Stuckey, 2018; Li et al., 2016). These nCETs are regions of tumor that do not show up on T1-weighted with gadolinium contrast (T1C) scans but can usually be seen in FLAIR images. These have a less distinct boundary, similar to what is seen in the histology of the tumor. Because of the importance of the nCETs in prognosis, imaging, and treatment, capturing this information is critical to accurate surgical resection and estimating the spread of the tumor (Li et al., 2016). Therefore, an understanding of how multimodal imaging signatures relate to the underlying tissue pathology is critical to better treatment direction.

The Mayo Clinic states that "a biopsy is the only way to definitively diagnose a brain tumor and give a prognosis to guide treatment decisions" (Glioma | diagnosis and treatment, 2019). However, when a tumor cannot be

biopsied, a diagnosis is made on the imaging alone, which could miss key areas such as nCETs (How we diagnose brain tumors.). Despite the pathological aspect of this tumor being key to an exact diagnosis, many tumor prediction and segmentation algorithms that exist only use MRI to predict tumor (University of Pennsylvania, 2018).

Machine Learning

The 2018 MICCAI Multimodal Brain Tumor Segmentation Challenge used board-certified neuroradiologist-annotated MRI scans to produce a dataset of T1-weighted (T1), T1C, T2-weighted (T2), and fluid-attenuated inversion recovery (FLAIR) images with edema, tumor core, necrosis, and whole tumor annotations (University of Pennsylvania, 2018). While this type of challenge is good in practice to assist in annotation and speed up a radiologist's heavy workload, the underlying pathology may not be adequately identified. The 2014 MICCAI Brain Tumor Digital Pathology Challenge proposed two challenges, one for classifying between LGG and GBM on whole slide images (WSI), and the other for segmenting necrotic regions (MICCAI 2014 grand challenges, 2016). Neither challenge involved segmenting tumor from no tumor in brain tissue, so none of the resulting models from that challenge will segment brain tumor without additional training. In the paper by Xu, AlexNet transfer learning was proved to be useful with histological images, particularly with brain tumor pathology (Xu et al., 2017). Many other examples of AlexNet's performance has been demonstrated in open datasets and clinical datasets (Lu, S., Lu, & Zhang, 2019; Yang, Y. et al., 2018). These examples of successful transfer learning using

AlexNet as well as MATLAB's ease of use to implement a pre-trained AlexNet model were the main reason that AlexNet was used as the architecture for this histological prediction model. However, while Xu demonstrated good results with LGG vs GBM classification, necrosis segmentation, and colon cancer segmentation, tumor segmentation in brain tissue was not demonstrated due to the limitation of 2014 MICCAI dataset (MICCAI 2014 grand challenges, 2016; Xu et al., 2017). Because of this, there is no specific model to be able to compare our results to, rather just a speculative comparison between existing AlexNet transfer learning models.

Machine learning is widely used in both cancer research and the clinical setting for its ability to quickly and reliably identify tumor masses as well as potential outcomes for a variety of cancer types (Bakas et al., 2018; Kourou, Exarchos, Exarchos, Karamouzis, & Fotiadis, 2015; Lu, C. et al., 2018; Manogaran et al., 2018; Rathore et al., 2019; University of Pennsylvania, 2018; Varuna Shree & Kumar, 2018; Zijlstra, Novitskaya, Vizio, Reis- Sobreiro, & Freeman, 2019). There are a seemingly myriad number of articles describing machine learning development, uses, and applications, but most do not use a histological bases for their studies. For example, Kumar uses radiological scans only and uses a texture analysis to segment tumor (Vijay Kumar & GV Raju, 2010). Dong also uses radiological scans, borrowing the 2015 BRATS dataset (Dong, Yang, Liu, Mo, & Guo, 2017). However, as stated previously, samples of tissue are the only way to exactly diagnose a brain tumor (Brain tumor - diagnosis, 2018; Glioma | diagnosis and treatment, 2019). In order to have more

accurate models being trained on MRI data, the underlying histology ought to be included. There is therefore a gap in the validity of the underlying truth used to train current tumor prediction models. This underlying histology can be registered and brought into MRI-space to align with the MRI scans using previously established methods (LaViolette et al., 2014; McGarry et al., 2016). Once a histological dataset is created, improved MRI predictions based on underlying histology can occur, which is the goal of this project.

This project creates such a dataset and then makes use of a machine learning classification scheme where a small dataset is used to create a much larger dataset, which is then used as ground truth for a final classification. Described in the methods, this project takes a set of 1402 tiles and labels and produces a model. Then that model is used to create ground truth for a new model to be trained using MRI. While there are many existing articles using weakly supervised learning (Bukowy et al., 2019; Ge, Yang, & Yu, 2018; Mlynarski, Delingette, Criminisi, & Ayache, 2018) similar to the methods used in this project, few demonstrate the ability to use a small initial dataset to create ground truth to classify a much larger dataset.

Scope

The long-term goal of this project is to generate models trained with datasets from patient imaging and aligned histology to better predict the location of tumor invasion beyond conventionally defined margins. We expect that these models will then provide oncologists with more accurate tumor detection for targeting with additional radiation, or for a better targeted biopsy procedure.

II.METHODS

This study is comprised of two distinct experiments. Experiment 1 deals solely with histology from autopsy samples, testing the hypothesis that machine learning models trained with pathologist annotated samples are able to differentiate regions of viable tumor. The second experiment expands these models onto a large dataset of histology aligned to MRI scans, to test the hypothesis that a second algorithm trained with the combined rad-path dataset is able to predict tumor presence with MRI alone.

Study Population

This IRB approved study, (MCW-PRO17446) included 23 total subjects (16 male and 7 female). Experiment 1 used the tissue samples and histology from 15 patients, while Experiment 2 used the imaging from 16 patients. Nine of the patients overlapped between experiments (Table 1 and 2). Patients in this study ranged from 41 to 88 years old (median=62yo). Of the 23 patients, 21 had a final diagnosis of GBM. The other two non-GBM patients (included in Experiment 2 only to demonstrate model generalizability) had a final diagnosis of a WHO grade 3 oligodendroglioma and WHO grade 3 astrocytoma.

Table 1. Per Subject Breakdown of Histological Dataset for Experiment 1.

Subject	Images	Test Set	Tumor	No Tumor
1	39		19	20
2	19		1	18
3	105		91	14
4	433		294	139
5	66		32	34
6	26		16	10
7	79		73	6
8	27		10	17
9	18		3	15
10	24		12	12
11	66		24	42
12	39		32	7
13	87	X	57	30
14	328	X	221	107
15	45	X	19	26

Table 2. Per Subject Breakdown of the MRI Dataset

Subject	Images	Test Set	Tumor	No Tumor
2*	1369	X	0	1369
3*	173	X	173	0
5*	406	X	201	205
6*	578		364	214
11*	575	X	3	572
12*	1486		853	633
14*	1257		777	480
15*	3753		2716	1037
16	954	X	326	628
17	156	X	0	156
18	4144		782	3362
19	1969		432	1537
20	9842	X	9741	101
21	6429		2960	3469
22	4571		1033	3538
23	16405		4295	12110

* Used in experiment 1.

Histological Processing (Experiment 1 & 2)

Patients who have a GBM diagnosis are first consented to undergo donation of tissue to the Brain Bank in the LaViolette Lab at the Medical College of Wisconsin. The following process is the same as (Nguyen et al., 2016). At autopsy, the brains were removed and placed in a 3D printed brain cage during formalin fixation to maintain structural integrity. The brain tissue fixes in formalin solution for about two weeks. Once fixed, the brain is then sliced in the same orientation as the axial MRI using a custom 3D printed slicing jig. Tissue samples are then taken from those slices. These samples are put onto slides and stained using hematoxylin and eosin (H&E) to highlight cell nuclei blue/purple and cytoplasm and extracellular matrix pink. Once these slides are created, they are then digitized using a slide scanner at 40X magnification. This image is then down-sampled to 10X magnification. The resulting RGB image is tiled into 2000x2000 pixel images. This process of using smaller, tiled images is similar to a process Barker uses in the digital processing of their 2016 paper (Barker, Hoogi, Depeursinge, & Rubin, 2016).

These images produced a dataset that needed to be organized before proceeding. Nuclei, red blood cells, cytoplasm, and whitespace were extracted from these images. To do this, original images were scaled to range from 0 to 1 instead of 0 to 255. The image was then transformed into hue, saturation, and intensity (HSV) space. Any pixels with saturation less than 0.1 and intensity greater than 0.7 were considered whitespace. Any pixels with saturation greater than 0.6 and hue greater than .85 or less than 0.05 was considered to be red

blood cells. The reason a pixel that was either greater than 0.85 or less than 0.05 was used is due to how hue is represented in HSV space. The values from 0 to 1 represent red to yellow to blue to red. Red occupies the regions where hue is less than .85 and greater than 0.05, so these values were used. Then the Euclidean distant to red (HSV values of [1 1 1]) from the HSV image was found. This resulting image was used in part to extract the nuclei. An image of the Euclidean distance from RGB pink (RGB values of [1 0 1]). The equation to extract nuclei took the extracted pink and subtracted the extracted red. This was then multiplied by the inverse of the original scaled RGB image. The red blood cells were masked out of this image. This image was then converted to grayscale to create an image that highlighted only nuclei. The remaining extracellular matrix was extracted as any pixel less than 40% of the maximum value possible in the pink image. The thresholds for creating these images were set so that there is minimal to no overlap between what is highlighted in each of the images. All four of these images, whitespace, red blood cells, nuclei, and extracellular matrix, were resized to 1% of the original size, which essentially created density maps of each of these images. This was done in preparation for histological registration to MRI scans.

A selection of 1402 tiles that were free of artifacts were then chosen as the dataset. This is significantly smaller than the total number of tiles in the dataset, which ranges in the millions. Using this small set, a weak classification scheme was applied to classify the rest of the dataset.

Dataset Curation and Histology Annotation (Experiment 1)

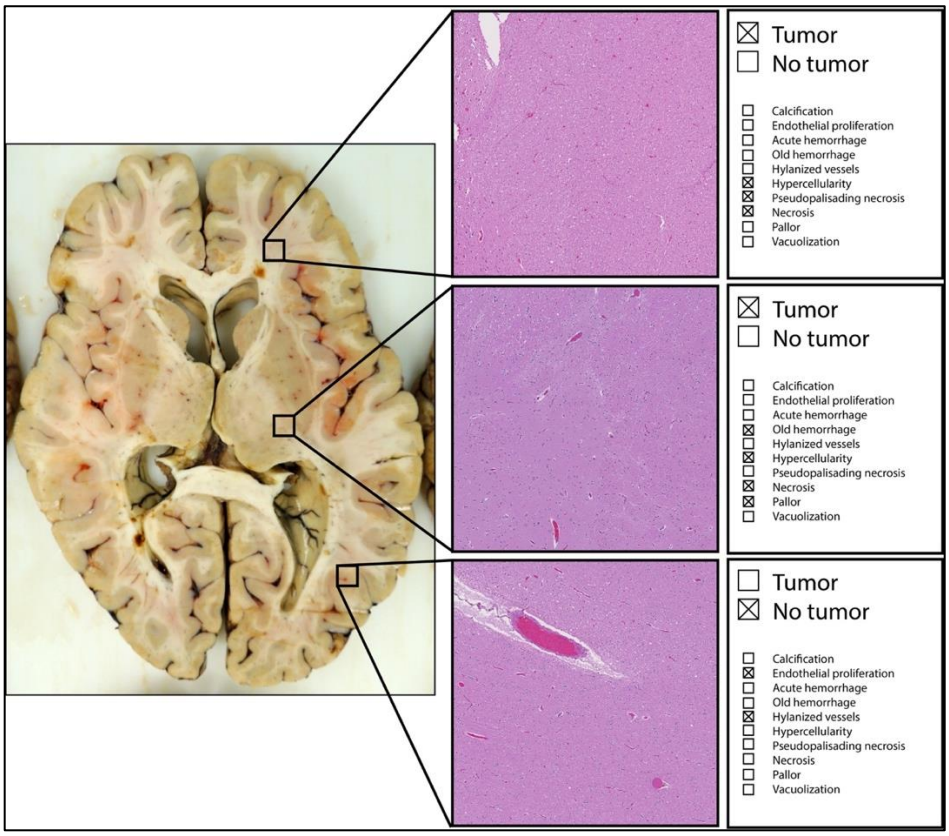

Figure 1. Example of a slice of brain tissue and its underlying tiles from the digitized histology. On the far right is an example of how a pathologist might annotate each tile.

A MATLAB graphical user interface (GUI) was created in order for a pathologist to classify the selected tiles. This GUI had multiple checkboxes with descriptive pathological features, such as pseudopalisading necrosis, hypercellularity, and calcification, as well as whether or not the tile contained tumor. A board-certified pathologist classified each of the 1402 tiles as containing tumor or not containing any tumor (EJC). The pathological features were not used for this study but will most likely be used in future research.

Figure 2. Screenshot of the GUI created for pathologist annotations.

After the 1402 images were classified by the pathologist, three patients were randomly selected to be left out of the training dataset. 163 images without tumor and 298 images with tumor were left out to be tested on from these three left-out patients. One subject had a high number of images (294 total images from one subject) which caused the number of images in the tumor-containing training dataset to be much higher than the no-tumor-containing set. After class balancing the training dataset, 334 no tumor class images and 334 tumor class images were then used to train a pretrained AlexNet convolutional neural network (CNN) to identify if an image contained tumor (Krizhevsky, Sutskever, & Hinton, 2012). This CNN was pretrained using the ImageNet dataset (ImageNet, 2016), which is a set of more than 14 million images that are categorized in 27

high-level natural image categories and have up to several thousand subcategories. The architecture of this network is shown in Figure 3. 25% of the remaining images were used for validation. Data augmentation for this dataset included rotations of 90, 180, and 270 degrees, as well as flipping vertically and horizontally.

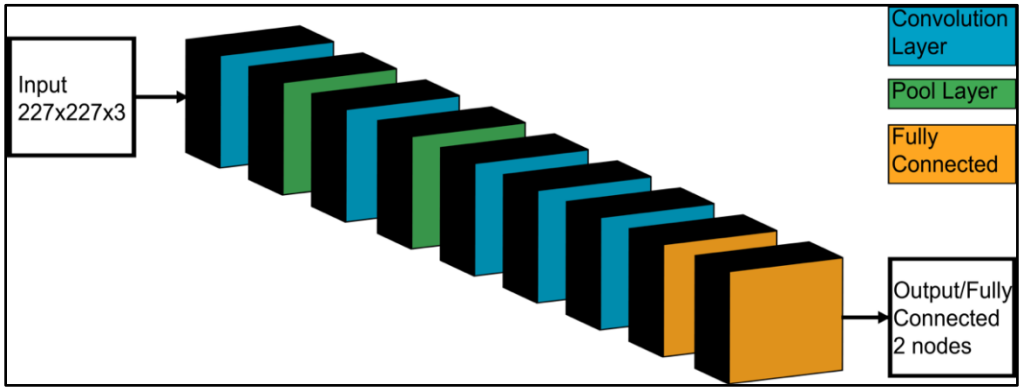

Figure 3. AlexNet architecture.

Table 3. Breakdown of Histological Dataset.

	Train	**Test**	**Total Images**	**Used in Training**
Tumor	607	298	905	334
No Tumor	334	163	497	334
Total Images	941	461	1402	668

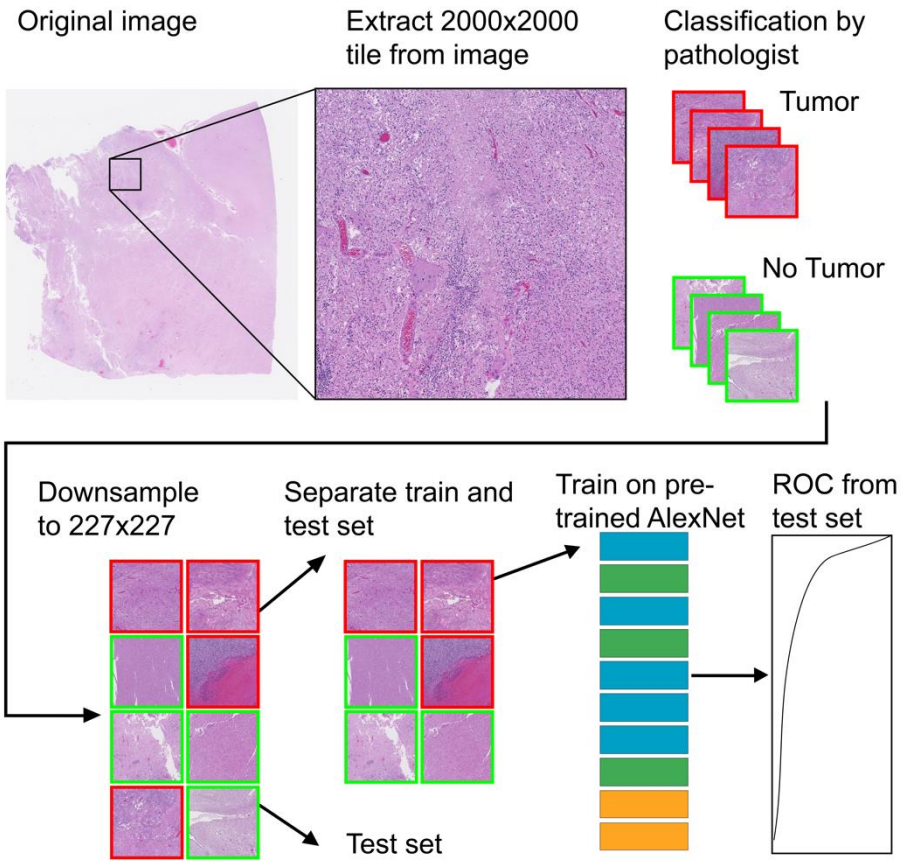

Figure 4. Design diagram for experiment 1 and the training and testing of the histological tumor classifier.

Training Histological Model

The final CNN output layer gave a percent confidence ranging from 0 to 1 with 0 indicating no tumor presence and 1 indicating high confidence in tumor presence in the input tile. The input, as constrained by the pretrained AlexNet network, was a 227x227 RGB image. Since the images were 2000x2000, they were down-sampled to the AlexNet input size. The network, which was trained across 4 K80 GPUs using MATLAB 2018a, trained with a learning rate of .001 over the course of 500 epochs. The training completed in about 6 hours. After training finished, it was applied across all digitized slide samples. The input was

a 2000x2000 tile that was extracted from the image and then down-sampled. Then a stride of 200 pixels was used to classify areas across the image, which effectively created a predicted heatmap that was .5% of the original 10X image. This is shown in Figure 5.

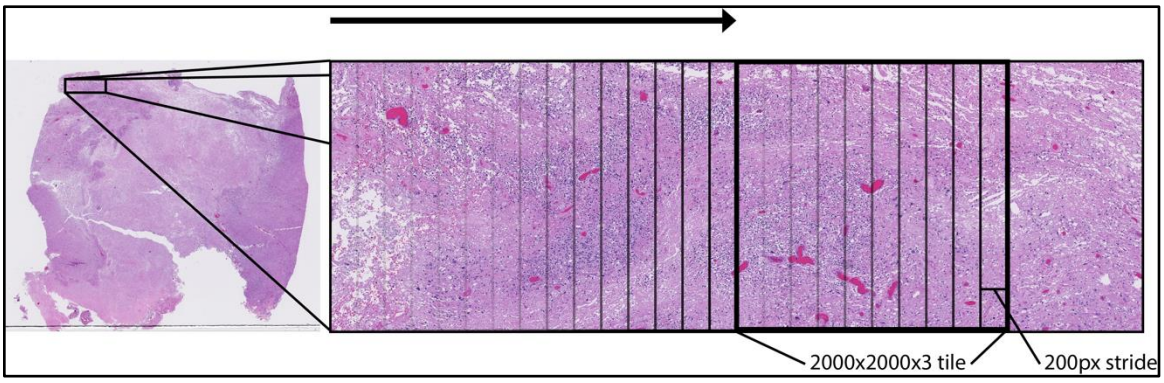

Figure 5. Example of how the prediction map is created. Each black bounded box is predicted, then the next image 200 pixels over is predicted and so on.

In order to have a test-set on an entire slide, a board-certified pathologist drew ROIs of tumor on slides. Binary masks of tumor were created as ground truth. These masks were binarily dilated with a square structure element with a width of 1000 pixels. This was done because if the images fed into the network may contain tumor in half of the image, it will classify the image as tumor, even if the underlying pixel in the ground truth is not in the ROI drawn. In this way, a dilated image is most representative of what the model will predict.

MRI Processing (Experiment 2)

MRIs were acquired using a GE MR system or Siemens system. The MRI closest to a patient's date of death was used for analysis. These varied in magnetic strength due to the scans being on different machines. 9 of the scans

were at 3T, 5 of these scans were at 1.5T, and 2 were on a Siemens system at 1.5T; the two sessions from a Siemens system were not used during training. T1, T1C, FLAIR, and diffusion weighted images (DWI) (diffusion weights or b-values: b = 0 and b = 1000 s/mm^2) were all acquired during these scans. The time between imaging and death ranged from 5 days to 184 days (median=42 days).

The scans from these sessions required preprocessing. First, all scans were converted from DICOM images to NIfTI files. Apparent diffusion coefficient (ADC) maps were then created from the DWI images. This is done by taking the logarithm of the DWI image with b-value > 0 divided by the DWI image with b-value = 0. Then dividing this by 1 over the b-value > 0, shown in Equation 1.

$$Equation\ 1.\ \ ADC = \log \frac{DWI_{1000}}{DWI_0}$$

All images were then registered to each subject's FLAIR images, which were 512x512x22 volumes of .43x.43x5mm^3. Brain masks were extracted from the registered images. To extract the brain masks, ROBEX, a robust skull stripping algorithm for T1-weighted images, was used (Iglesias, Liu, Thompson, & Tu, 2011). This was chosen over the conventional brain extraction tool (BET) or skull stripping because of its better performance, specifically on populations with abnormal anatomy such as a tumor or enlarged ventricles (Smith, 2002). After a brain mask was created, the standard deviation within the brain mask was calculated to normalize the dataset. This was done in the same process as the 2018 Ellingson paper (Ellingson, Benjamin M. et al., 2018). After normalization, T1-subtraction maps were created, again in the same process as the 2018

Ellingson paper (Ellingson et al., 2018). Used in this project were the T1, T1 plus contrast (T1C), FLAIR, and ADC scans.

The images created from the histology classifier were then co-registered to the processed MRI images using custom MATLAB software. These steps are the same processing steps as in the 2014 LaViolette paper (LaViolette et al., 2014). Because the tissue was cut in approximately the same plane as the MRI scans, these can be compared and co-registered through this process. This process warps the tissue into MRI-space and aligns it with the last scan of the patient's brain. Regions of interest (ROIs) are drawn on the tissue and matched MRI so that artifacts can be avoided. This produces a set of ROIs of both the tissue and MRI and their coordinates. These were aggregated into 10x10 voxel ROIs in MRI-space.

Training MRI Model (Experiment 2)

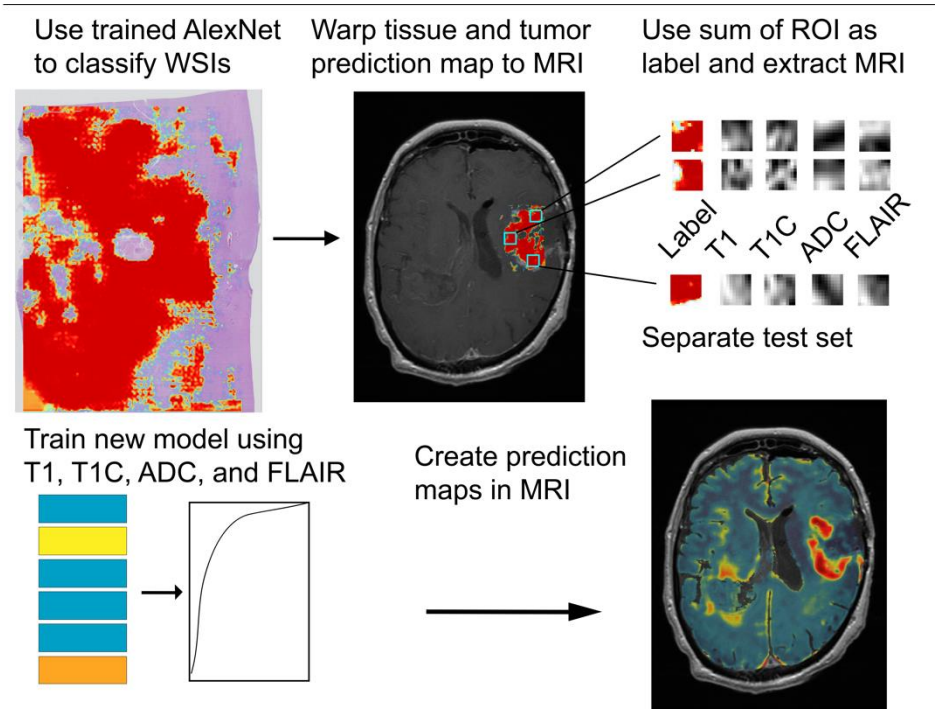

Figure 6. Illustration of the experimental design for Experiment 2. The model trained in Experiment 1 is first applied to whole slide images. The prediction maps are then brought into MRI space for comparison to the co-registered MRI patches. These patches and the tumor prediction values were then used to train a second classifier which was then mapped into MRI space.

Using the tumor/no tumor classification, generated from the histology model predicting iteratively across the slides, as ground truth, an additional CNN was trained to find tumor in MRI. The sum of all histological tumor predictions in an ROI was used as the label. Because the size of the tissue ROI was 10x10 pixels, the underlying value for the tumor prediction was a 100-value max, the sum of those pixels. Tumor was considered to be anything that was above a sum of 10 in total. 10x10 voxel ROIs of T1, T1C, FLAIR, and ADC were input to the network. The network architecture is shown in Figure 8. This network was trained on 9 different subjects and tested on 7 additional subjects for a total of 16

subjects in the dataset. A total number of 54067 ROIs were given from this dataset, with 32473 ROIs used for training and 21594 ROIs used for testing. 7 subjects were removed from the training set for testing as well as 20% of the training dataset.

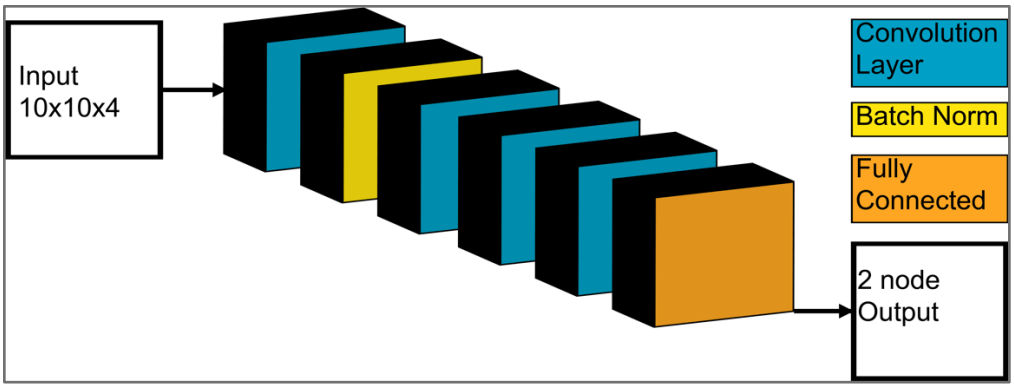

Figure 7. MRI model architecture.

Table 4. Breakdown of the MRI Dataset

	Train	**Test**	**Total**
Subjects	9	7	16
Tumor ROIs (sum of pixels > 10)	11369	13287	24656
No Tumor ROIs (sum of pixels < 10)	21104	8307	29411
Total ROIs	32473	21594	54067

The training used a learning rate of 0.01 and trained over 100 epochs. The trained network was then applied voxel-wise with T1, T1C, FLAIR, and ADC as the inputs to produce the final output, which is the same size as the input. This effectively gives a heatmap of active tumor regions with a score of 0 to 1, 0 being confidently not tumor and 1 being confidently tumor.

III.RESULTS

Histology

The histology tumor classifier ran for 500 epochs and ended with a final training accuracy of 89.6% and a loss of .201. Within the test-set, the area under the curve (AUC) for the receiver operating characteristic (ROC) curve was .745, shown in Figure 88, and an accuracy of 70.1%. The test-set contained only naïve subject images. Additionally, entire slides were tested from the same naïve test-set of subjects. Using pixel-wise comparison of annotations done by the pathologist and the tumor predicted in the WSI, three slides were analyzed. Only pixels within the tissue were used for comparison. 25749 pixels were extracted from the images. This resulted in an overall accuracy of 90.15% and an AUC of .92, shown in Figure 9.

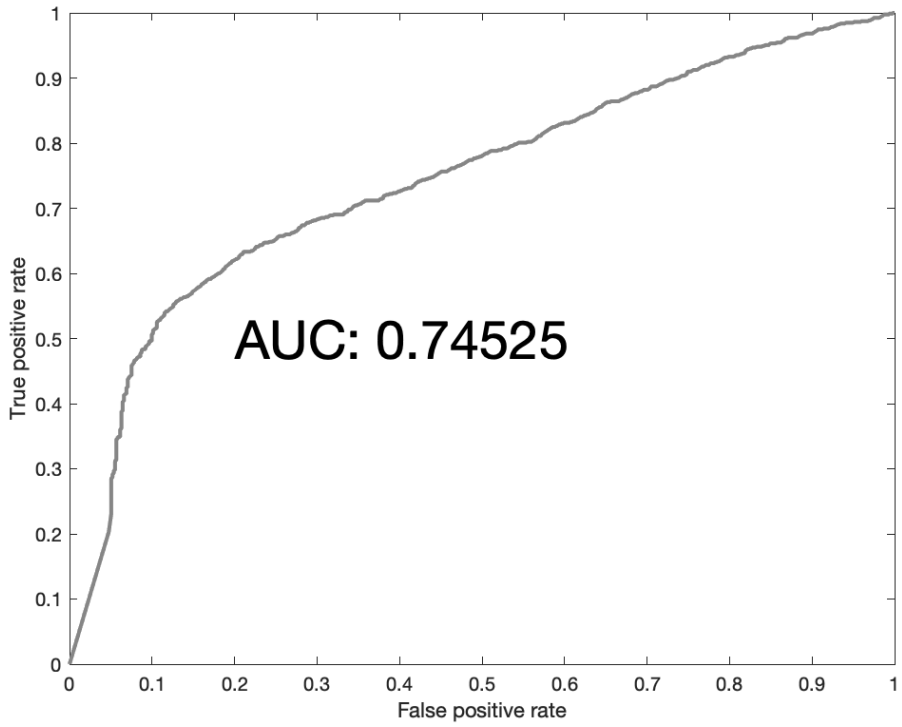

Figure 8. ROC for the test set of the histological classifier. An AUC of .745 was achieved.

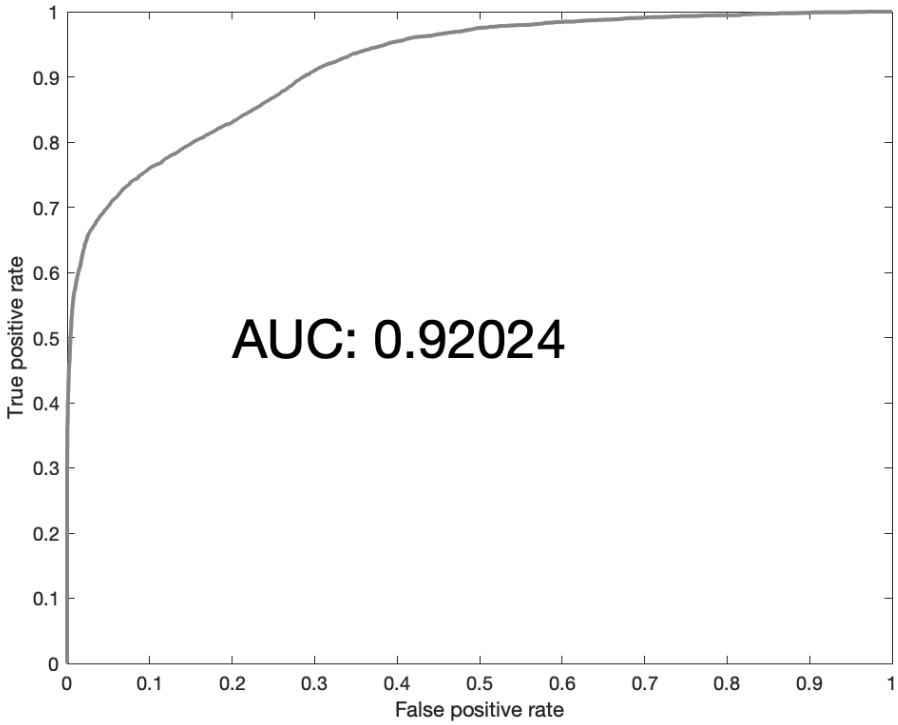

Figure 9. ROC for the histological classifier within the pixelwise comparison of multiple test-set images that a pathologist had annotated. Here, AUC of .920 was achieved.

Figure 10 shows an example of the input, output, and both images combined. This is an example of a novel slide from a patient that was within the training dataset. This is a good example of the classifier avoiding regions of necrosis, seen in the middle of the tissue, but highlighting areas of active tumor. The classifier also did not consider areas of artifact as tumor. However, it did wrongly classify some regions of an artifact in the image. The artifact, at the bottom of the image, is the edge of the coverslip. This type of image artifact was not something presented to the classifier during training because all artifacts were discarded.

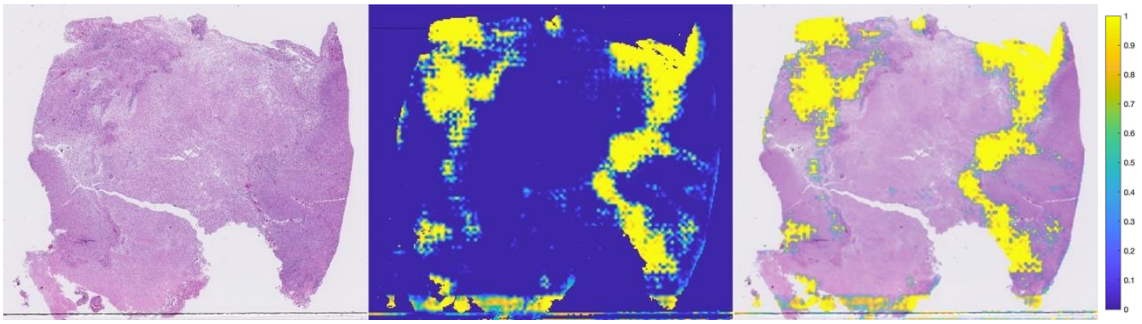

Figure 10. Digitized tissue sample (left), histological tumor prediction heatmap (middle), and the heatmap overlaid on the tissue (right). This is an example of a slide that was processed using the histological classifier.

Two examples of the test-set are shown below in Figure 11 and Figure 12. The first example shows how well the model did on a naïve subject. The second example shows a naïve subject where the model performed poorly.

Shown in Figure 11, the area of tumor, annotated by the pathologist, seem to be picked up well by the classifier. While the entire right piece of tissue was tumor, the classification on the left piece of tissue followed the same pattern as the annotation by the pathologist. Figure 12 shows some areas where the prediction follows the annotations, but many areas where it missed.

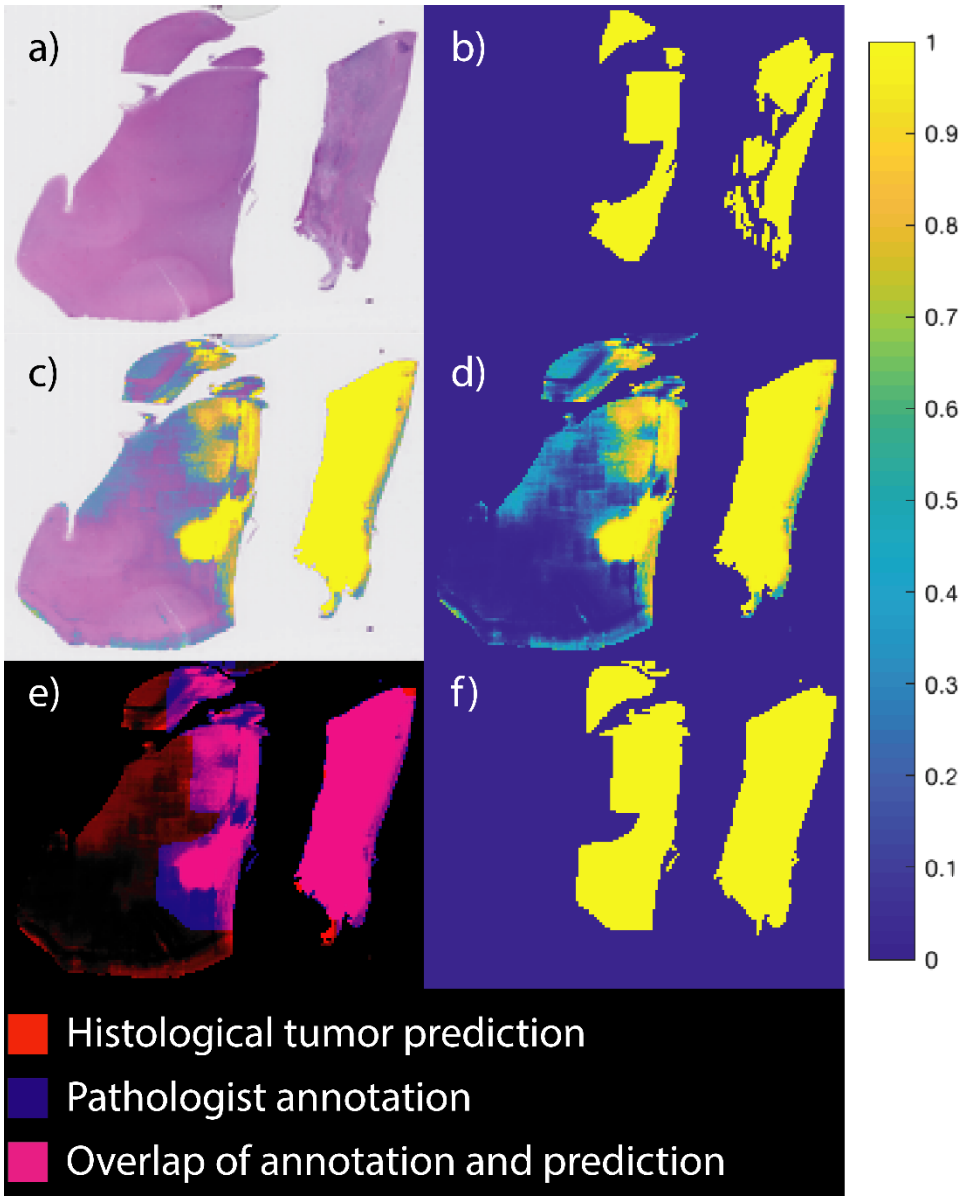

Figure 11. Original tissue a) from a test subject shown with pathologist's annotation of areas of tumor b). The prediction overlaid on the tissue c) and the prediction itself d) are on a scale of dark blue to yellow: dark blue meaning high confidence that the area contains no tumor and yellow meaning high confidence that tumor is in this area. Shown in e), both the annotation by the pathologist (blue) and the prediction (red) are combined to show how well the two images align. The image in f) is the pathologist's annotations binarily dilated by a square of 1000x1000pixels. This image is the same image in blue seen in e).

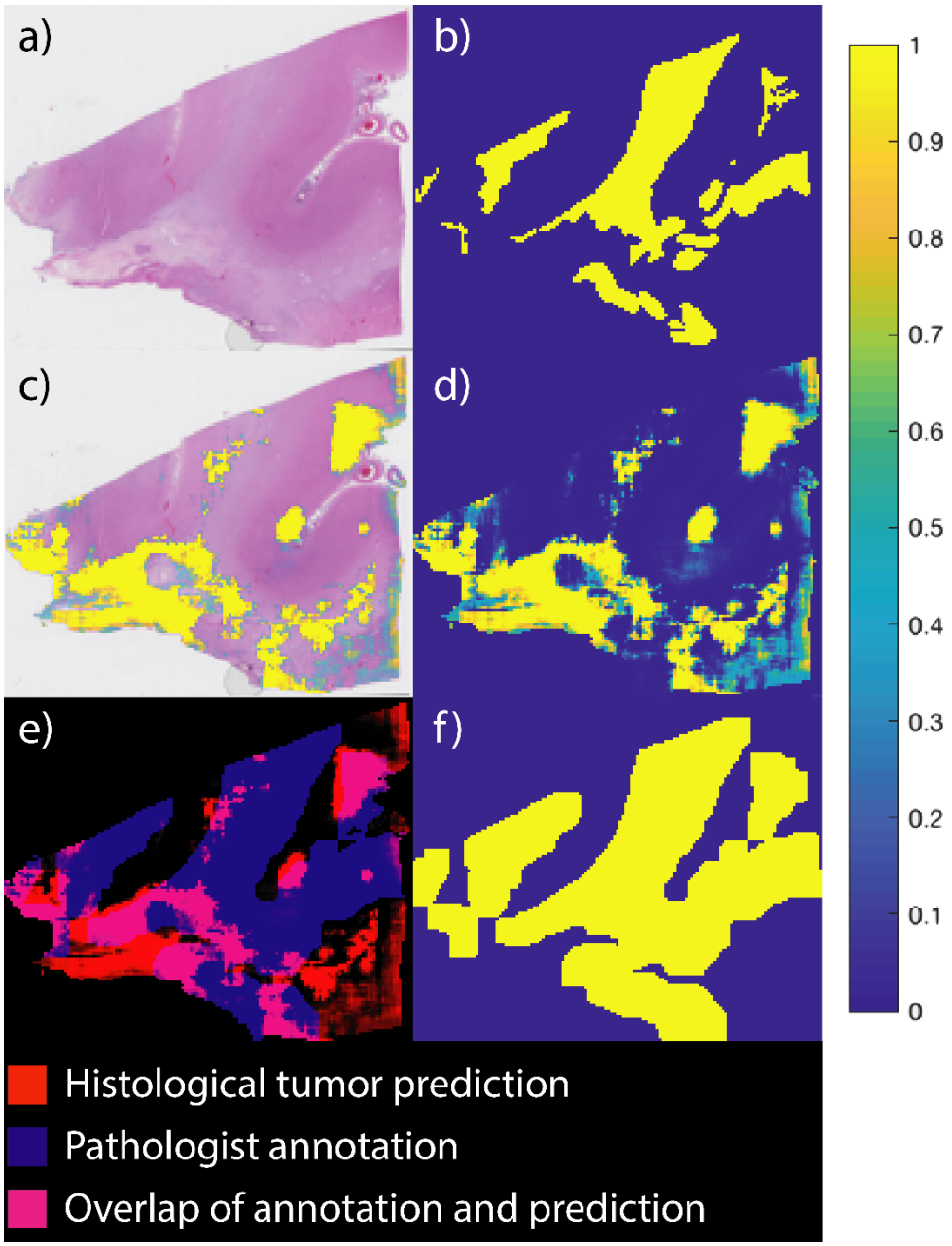

Figure 12. Another example of tissue from a test subject, displayed in the same fashion as Figure 11. This was an example where the model failed in multiple areas. Figure 13 goes into detail about where the model failed in this example.

Figure 13 goes into detail where the model failed for the example tissue in Figure 12. There are areas of necrosis where the model tends to be about 50% accurate. Additionally, while the model was correct in this image around areas of

sulci and dura, this is usually variable upon qualitative inspection. These areas are usually masked out if possible. Seen in Figure 13 b), an area of tumor was also misclassified, which the model predicted as mostly not containing tumor. The model also completely avoided the bubble artifact on the lower middle of the image. Because the model avoided this area, it did miss areas of tumor in this region.

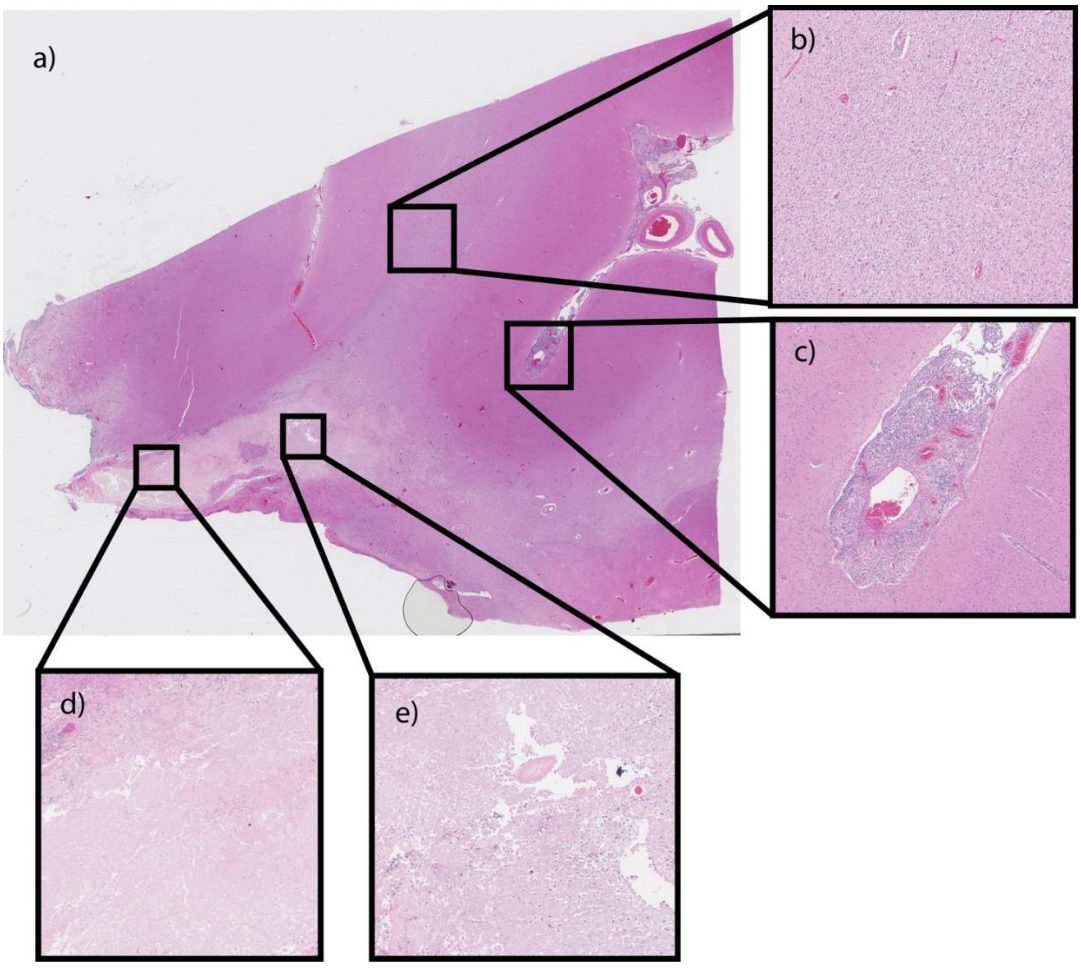

Figure 13. Examples from the overall tissue a) are representative of where the model tends to fail. The area of b) was an area that the model missed an area of tumor. In c), the model was mostly correct, though areas in the bottom right of this image and below this image were tumor and the model did not predict tumor. Both d) and e) are areas of necrosis that the model predicted contained tumor. The model incorrectly predicted tumor in d) where there was none and correctly in e).

The part of the MRI training set produced by this image would not include the artifact region because of how the ROIs are selected during tissue registration to MRI, therefore, this misclassification should not persist in the MRI training. The ROI selection process is a manual process by which the researcher

marks boxes that are clear of artifacts and parts of tissue that are not useful or not relevant (e.g. an area of dura).

MRI

The MRI classifier was not quite as accurate as the histological classifier. The training accuracy converged to 83.1% over 100 epochs with a loss of .430. Checking against a test-set, the AUC was nearly .70, shown in

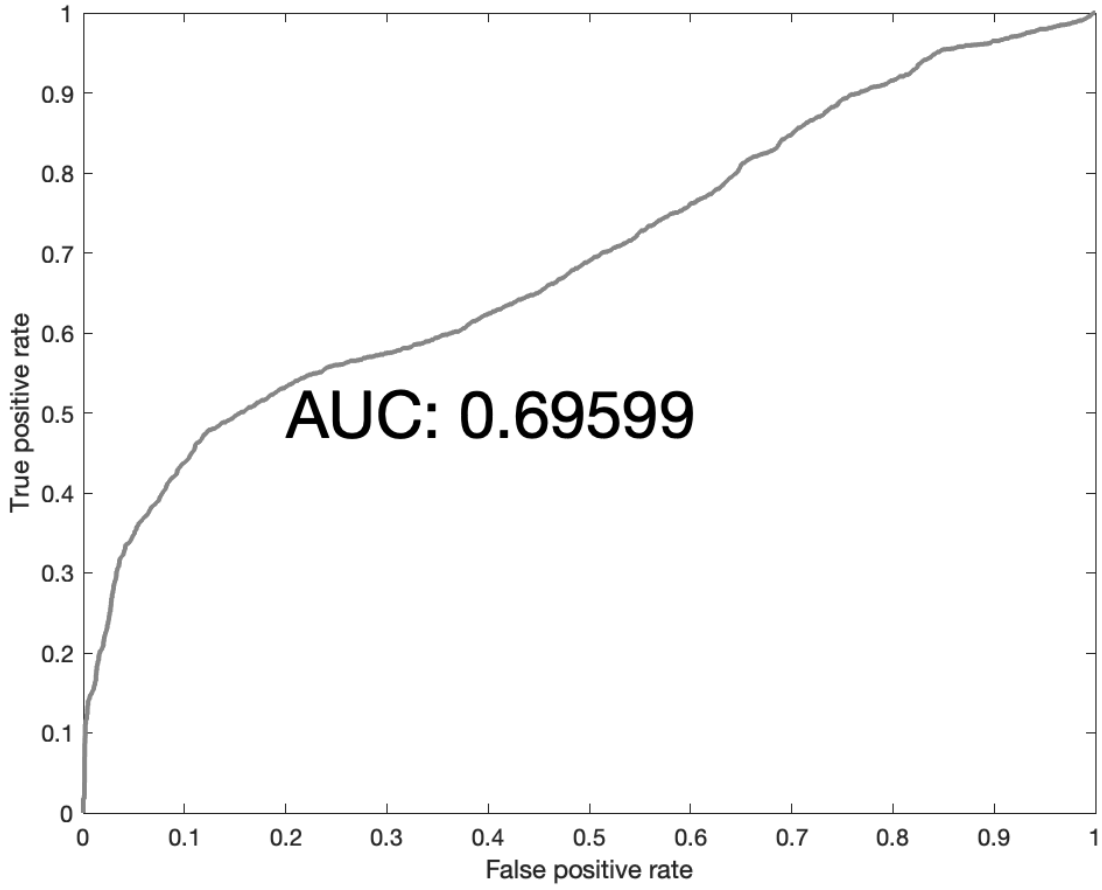

Figure 14, and an accuracy of 73.3%.

Table 5. Breakdown of Training Information on the Histological and MRI Models

	Epochs	Final training accuracy (%)	Final loss	Test accuracy (%)	Test AUC
Histology	500	89.6	.201	70.1	.745
MRI	100	83.1	.430	73.3	.696

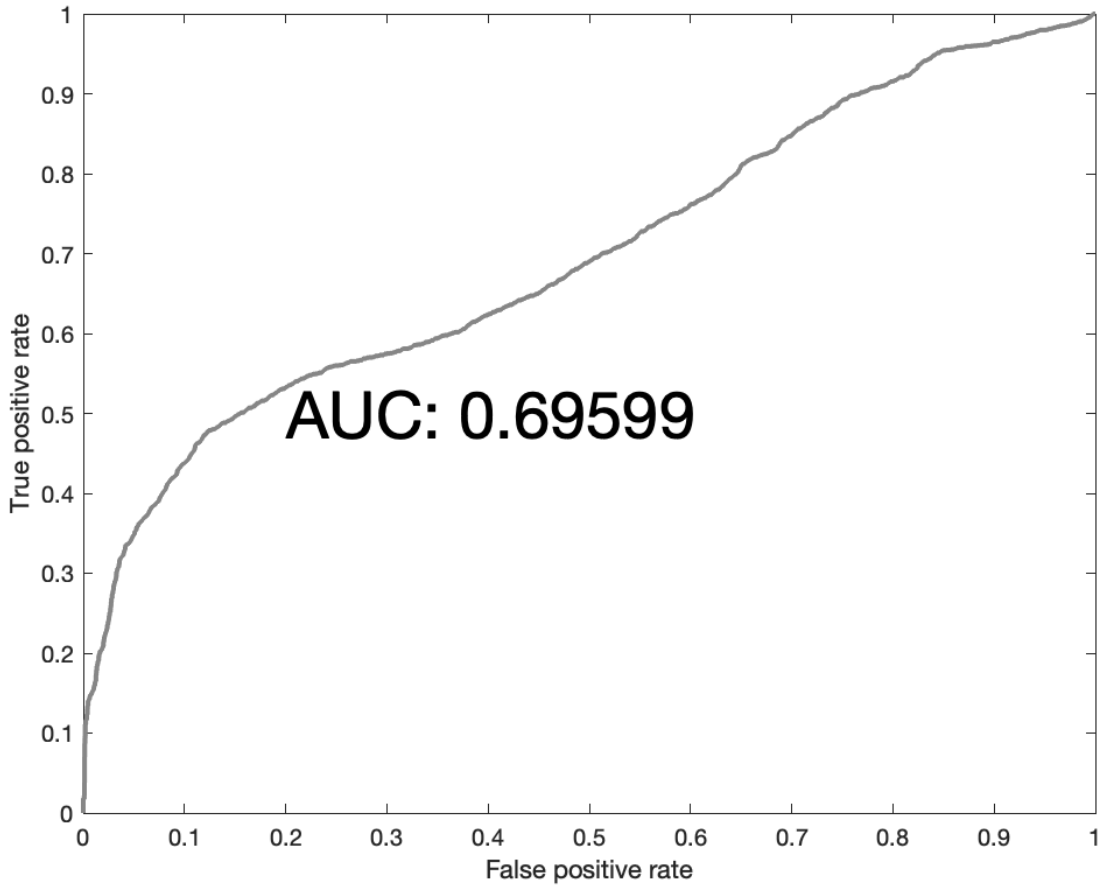

Figure 14. ROC for the MRI classifier. AUC of .696 was achieved.

This ROC curve shows subtle predictive utility for the MRI classifier within the test-set. However, as with the histological classifier, when applying the classifier to an entire image, it is clear that the test-set ROC is not entirely representative.

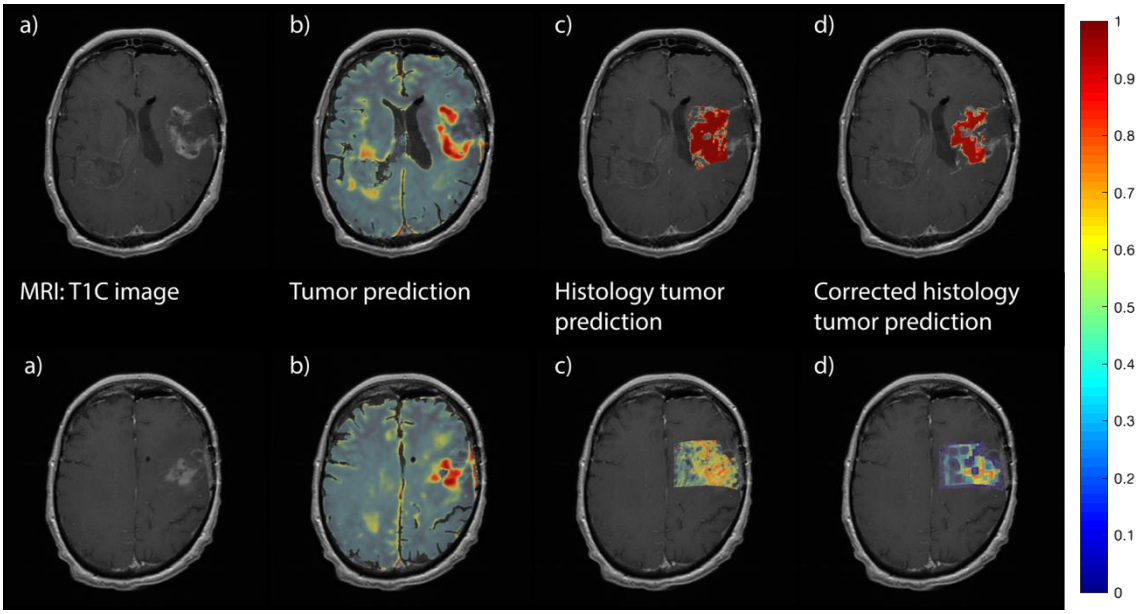

Figure 15 shows two examples of the final output. This is an example of tissue that has a clearly defined tumor, highlighted in the T1C scan, but also identified by the MRI model.

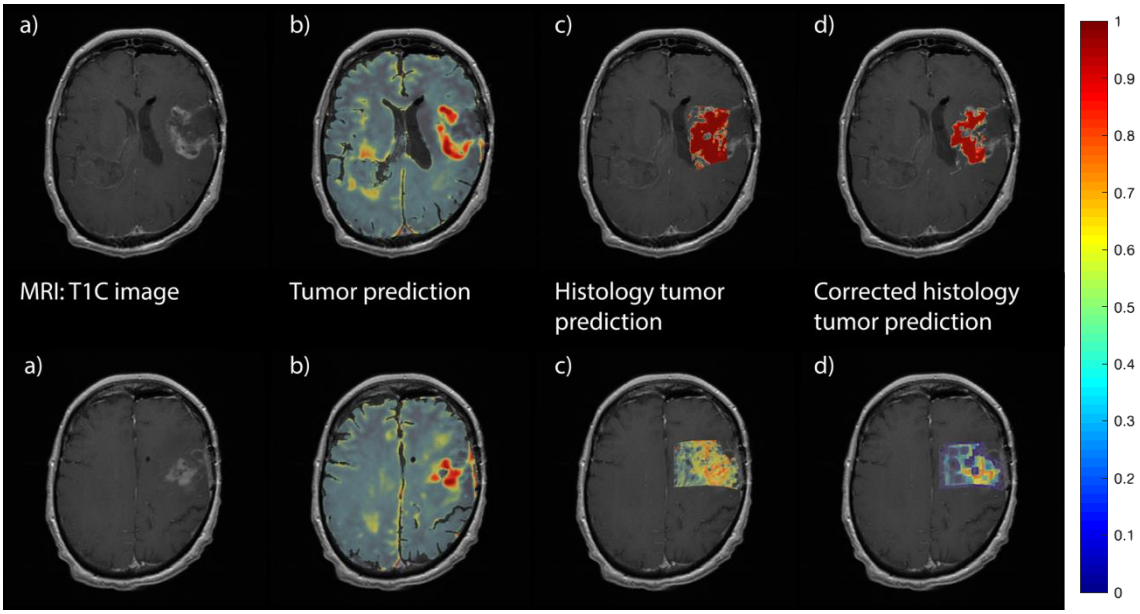

Figure 15. Example of two different MRI slices from a subject within the test-set a). MRI tumor prediction heatmap b) shows areas of red that are confidently tumor and areas of dark blue are confidently not tumor. This is the same for the images on the right, which are the underlying histological tumor prediction c) and the histological tumor prediction binarily eroded by a square of 10x10pixels d).

Similar to the analysis of the histology classifier, there was a morphological operation performed before comparing the images. In

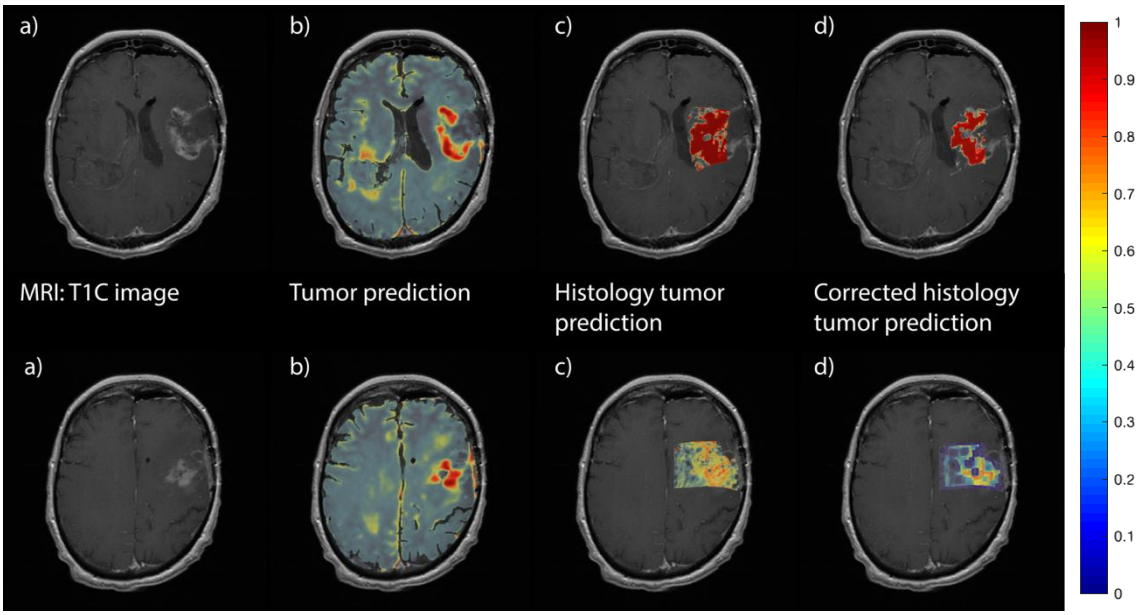

Figure 15 d), the ground truth of the histological prediction was grayscale-eroded by a 10x10 square. Since the histological prediction is a much larger region that what is actually tumor, this was done to show the accuracy of the MRI predictor. And while the images shown in

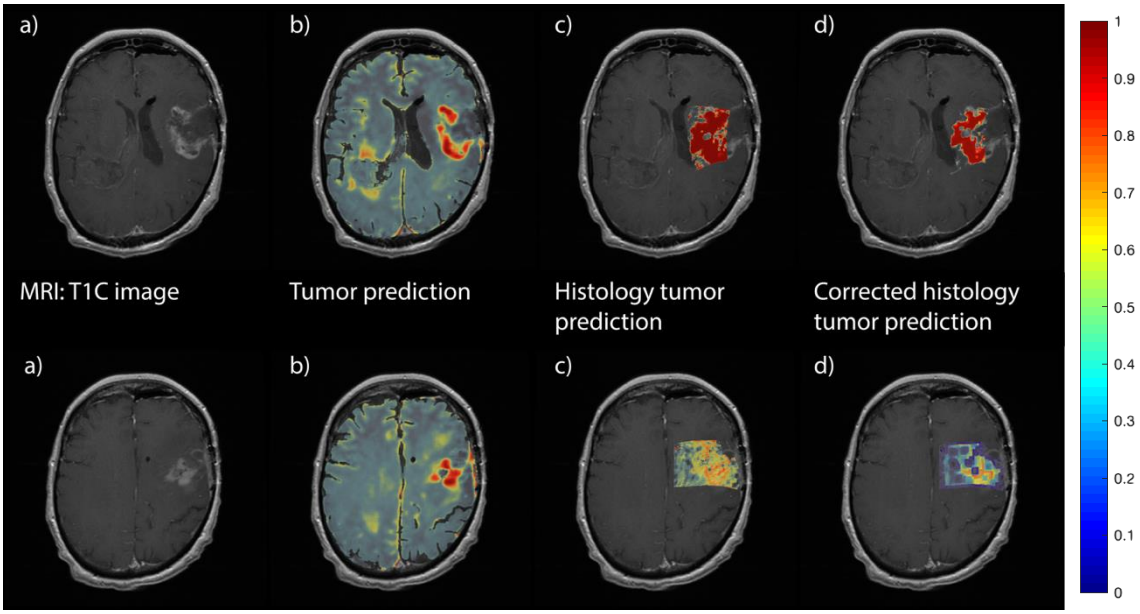

Figure 15 c) of the original histological tumor prediction are fairly widespread, eroding them shows their similarity to the tumor prediction in MRI.

A representative slice from each of the test patients, along with their conventional imaging is shown in Figure 16. Figure 17 shows a validation sample taken from a test patient indicating where tumor was found compared to the predictive map. The annotation by the pathologist on the tissue sample is also shown. This is an example where tumor was found outside of conventionally defined radiologic boundaries, with no contrast enhancement and very little FLAIR hyperintensity, having areas of tumor confirmed by the pathologist and by both models.

.

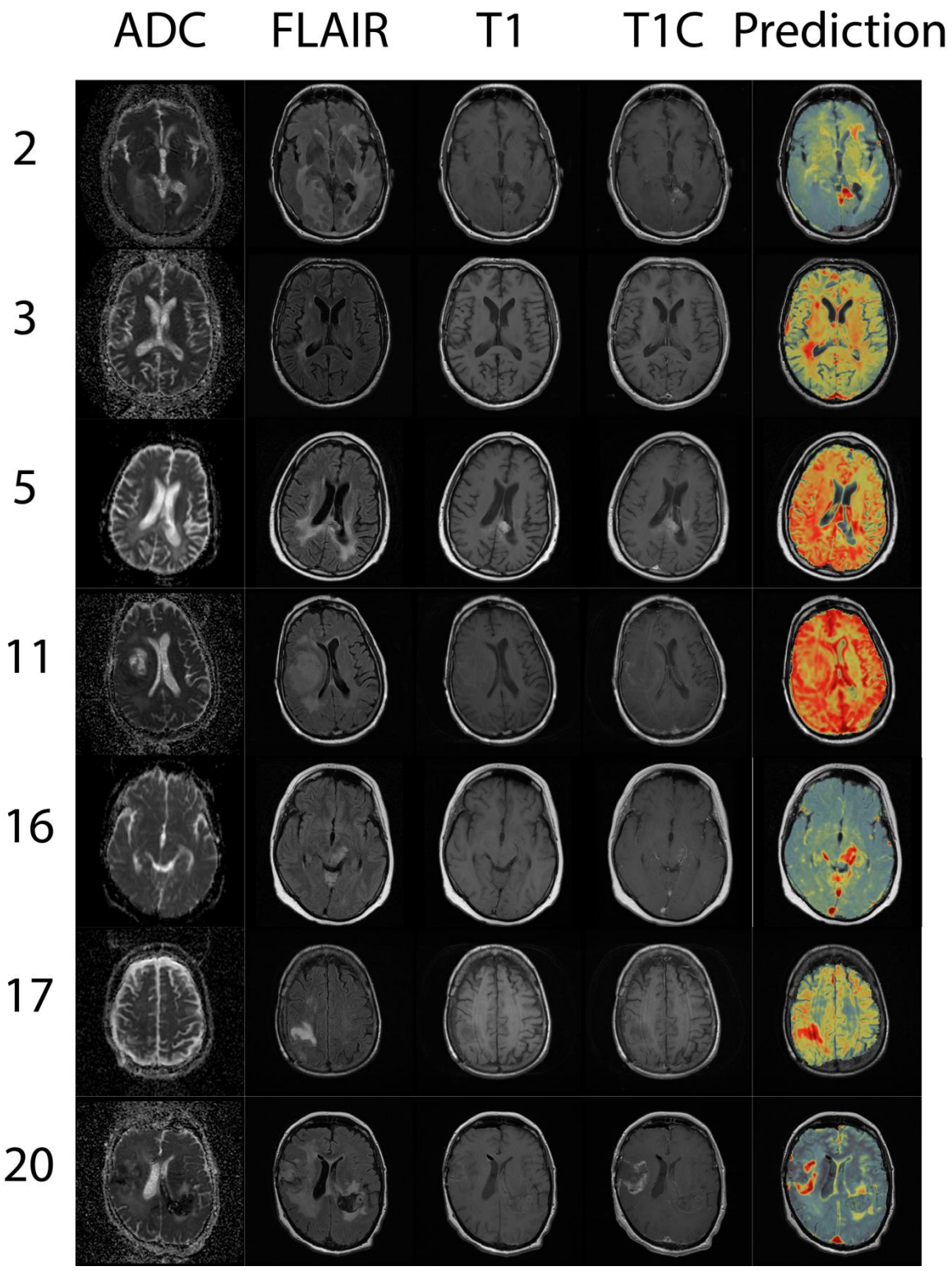

Figure 16. A representative MRI slice from each of the test set subjects from the MRI portion of this study.

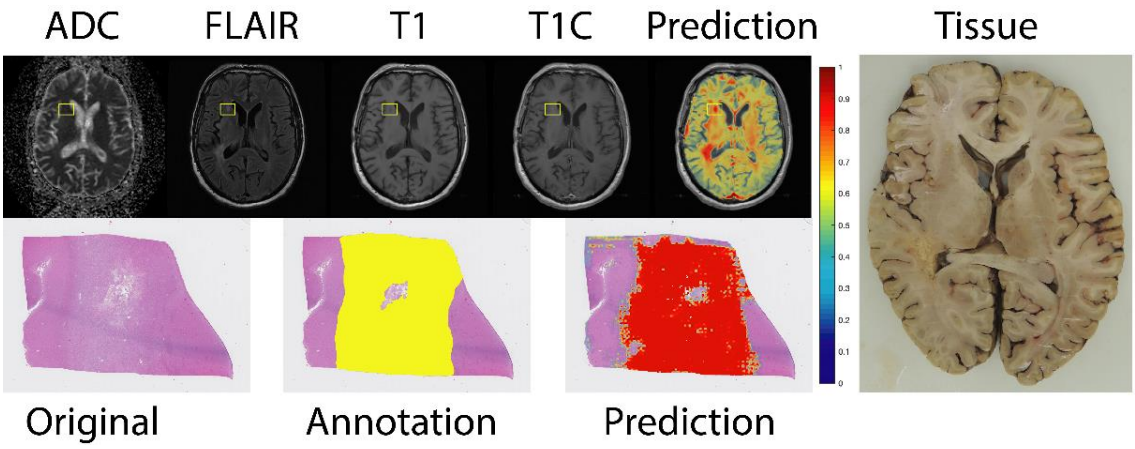

Figure 17. A representative sample of tissue, histology, and MRI (tissue outlined in a yellow box). This sample shows non-enhancing tumor that was predicted in both the histological and MRI models and was annotated by our pathologist.

IV. DISCUSSION

This study found that deep learning models were able to identify areas of tumor in both histological samples and MRI images. The analysis of the whole slide images showed how well the classifier was able to predict various areas of tumor with an accuracy of 90.15%, which was well above the accuracy of 70.1% on the test set images alone. The MRI classifier had poor test set results, with an accuracy of 73.3%, but well-represented the underlying tissue upon qualitative inspection.

The histological model performed very well on areas of hypercellularity and infiltrative tumor. Though the model had good accuracy when it comes to the whole slide images, there were a number of specific areas where the classifier tended to fail. The histological model tended to avoid areas of gray matter, despite some areas of tumor. This is consistent with the fact that GBMs are generally found in areas of white matter. Because of this, the training set did not contain many areas of gray matter infiltration. While areas of white matter were usually classified accurately, the model did not generalize well enough to identify tumor appearing in gray matter within the test-set. The model also classified areas of tissue artifacts (tears, blurs, etc.) as not tumor, despite some areas of tumor being present. This is difficult to avoid without giving the model training images with artifacts present. However, this runs the risk of the model identifying all areas containing artifacts as containing tumor. One last area where the model occasionally failed was areas that contained both edema and high vascularity,

classified as not tumor, which in some cases was wrong. These areas are particularly important to be able to be correctly identified, particularly because they have imaging biomarkers that can be shown in MRI (Batchelor et al., 2013; Lin, 2013; McGahan et al., 2017). These cases can be seen in Figure 13.

While a segmentation model may have been able to show better accuracy on a test set, a segmentation model is not a practical or effective solution to this problem. A pathologist would have to create annotations on hundreds or thousands of images. This is much more difficult than simply saying an image does or does not contain tumor. Additionally, a pathologist will be hesitant to create a definitive tumor boundary on a small image. GBMs tend to have diffuse boundaries that are more like a gradient than a boundary. Because of this, a dataset produced from many small segmented tiles will be much more inaccurate than the dataset produced here from many categorized tiles. A segmentation model could be produced from the outputs of this model in the future, but currently, this method of creating maps of active tumor in histology using a well-categorized dataset ought to be used as the new standard of GBM tumor prediction.

The MRI model achieved a test set AUC of .70. While this is not exceptional, this should not be discounted given the context and the limitations of the noisy and clinical dataset. Additionally, the tumor prediction maps generated by this model creates images that accurately represent the histology. This is much more representative of the model's performance than just a test set accuracy. This low AUC leaves room for improvement while establishing

methods to be used in the future, especially considering this AUC may not be representative of the model's performance. However, generalizability is a concern. MRIs put into this network ought to be processed in the same fashion and have similar ranges of values.

The final AUC of the MRI model could be influenced and hindered by the accuracy of the original histological classifier. For example, looking at only the data that the histological model classified incorrectly, if the MRI classifier was equally wrong and right for those examples where the histology classifier was wrong, at best, the AUC may actually be much higher. However, if the MRI model is classifying the incorrectly classified portions in the same incorrect fashion (tumor where the original ground truth says there is none and vice versa), then the .70 AUC may be artificially inflated and may actually be much lower.

In comparison to existing models, some studies demonstrating transfer learning methods using AlexNet achieve AUCs around .90. However, in many cases, these are difficult to compare to this project. Yang's 2018 paper yields an AUC of .966 for their test set using AlexNet (Yang et al., 2018). While this shows AlexNet's efficacy, their dataset only differentiated LGG from HGG tumor. Chang's 2018 paper uses AlexNet to differentiate tumor, necrosis, and transition to necrosis (Chang, Han, Zhong, Snijders, & Mao, 2018). These images are visually much different from each other, so their accuracy of their models being above 90% is expected. Another paper shows that AlexNet can achieve an AUC of .94 in discriminating lung disease using a patch-based input (Hoo-Chang et al., 2016). Many of these papers are also based on the MICCAI dataset, which

as stated previously, does not have GBM histology tumor segmentation nor an MRI dataset verified beyond radiology.

Unconventional Methodology

There were a few instances throughout this project were unconventional methods were used. A morphological operation of binarily dilating the pathologist's annotations was done in order to account for the fact that an image given to the network may contain tumor on the edge of the image. This is because the center pixel of that image might in fact be distant from the tumor boundary itself. A dilation of 1000x1000 pixels was done to approximate what the tumor would capture if the tumor were in about half of the image. This made the pathologist's annotations more representative of what the classifier can capture since it is only predicting one pixel at a time. A morphological erosion of the predicted histology in MRI was also performed to qualitatively compare the MRI prediction maps. A 10x10 grayscale erosion helped account for the same issue stated previously in the histology model.

Future Work

Future work will include refining both models to increase accuracy. This will be accomplished by increasing the size of the training dataset given to the models as well as increasing the resolution of the data itself. For histology, a resolution of 10X gives a good sense of the general texture of the tissue and hypercellularity of an area, but a resolution of 40X may be able to show morphological information of nuclei that a model could use to better predict tumor. Recent publications have made use of a multi-pathway model for better

classification (Havaei et al., 2017; Pereira, Pinto, Alves, & Silva, 2016). This could be employed such that multiple magnifications are represented while highlighting the center of the image. This would ideally eliminate the need for any post-processing morphological operations. If a magnification greater than 10X is included, this may also alleviate some of the issues with the necrosis and edema misclassification. Having a narrower window of classification may allow for better detection of areas of tumor.

The threshold for classification in MRI could also be adjusted. A mean value greater than 10 is considered to contain tumor, but there are regions where there is plenty of classified tumor in histology where the MRI may not predict tumor. During a post hoc analysis of this threshold, it was found that using 10 as a threshold for training and 50 for a threshold for testing dramatically improved the test set AUC. This could mean that while the model does not perform exceedingly well on the test set, it is good at finding areas of high tumor concentration. However, this results in poor sensitivity as evidenced by the test set AUC.

Multi-class, rather than binary classification, could be employed to avoid the issues described in Figure 13. This could be done in a hierarchical way. Each area would be first classified as having artifact or not, and then any areas without artifacts would be further classified based on tissue type, and then finally classified based on whether it contains tumor. The final classification would be a combination of the previous classifications, e.g. white matter containing tumor free from artifact, gray matter containing no tumor free from artifact, or artifact.

The dilation and erosion portion of this analysis would also be revisited. The MRI AUC could be significantly improved by training on an eroded MRI dataset. Since this is more accurate to what the underlying tumor is, the signature of tumor within MRI may be able to be found more easily by the model. Since the post-hoc analysis of eroding the predicted images aligns better with the predicted tumor in MRI, it is hypothesized that training on these eroded images may yield better results for the MRI model.

In terms of real-world applications, if the MRI classifier would have had better generalizable results, this model could be potentially used as a radiological plug-in at a workstation. This would highlight areas of suspected tumor for the radiologist to then confirm or reject. This could be used in combination with existing radiologically based models and would allow a radiologist more confidence in the underlying pathology.

Limitations

While the accuracy of this model and ROC curve shown in the results do look promising, as with any study, there are limitations. First, histology slides contain gigabytes of data that were distilled down to a set of 227x227 RGB tiles. This is representative of only about 1% of the total data within the slide. While some of this data may get filtered out anyway, radiomics and pathomics research has shown that much of a dataset's texture can be used as a good predictor of labels and outcomes (Barajas et al., 2012; Ellingson, Benjamin M., 2015; Kickingereder et al., 2012; Kniep et al., 2018; Lu et al., 2018; Rathore et al., 2018; Tixier, Um, Young, & Veeraraghavan, 2019; Zacharaki et al., 2009; Zhou

et al., 2018). Ideally, in the future, more of this information will be preserved and used, but current limitations on processing speed and power hinder development using such large images in a dataset.

Additionally, the 10x10 ROIs have the opposite problem of the previous limitation. There is not much data that is fed into the network. Ideally, a better method would be to use a 10x10 or larger input to a pretrained network. Since there are not 10x10 pretrained networks that exist, a transfer learning methodology could not be used directly on this dataset. The only way this could have been possible would have been to upscale the input data to a size that something like AlexNet could handle, which would be 227x227 (Krizhevsky, Sutskever, & Hinton, 2012). This would have created a great deal of artificial data that would have been given to the network. An attempt was made to locally train a MNIST dataset CNN model, which would then be used as a transfer learning model on the MRI dataset. The idea was that since the images are 28x28 in the MNIST dataset, there would have been minimal upscaling and therefore minimal artificial data being input to the network. This showed poorer results within the test set than the final model, most likely due to the upscaling that was performed, so this method was discarded. Additionally, the fact that this was trained locally and an established MNIST classifier was not used could have affected the performance of the model. Another problem with using existing models like AlexNet was that they can only accept 3 channels as input (Krizhevsky et al., 2012), but additional MRI volumes could have been added. During tests with these existing models, the T1 scan was left out because of its generally low

contrast. Other volumes could include T1 subtraction maps, additional T2 scans, or diffusion imaging. This would provide more information to the network, which would be particularly helpful with such a small initial ROI size.

An additional limitation regarding image artifacts should also be noted. In Figure 10, at the bottom of the image, the edge of the coverslip can not only be seen but overlaps with the tissue. Because of this overlap, this artifact cannot be easily removed from processing. While the artifacts in the training images were removed from the dataset, artifacts that persist in the overall dataset have not yet been removed. In this case, the model will understandably perform unpredictably and will have to be masked out. This can be difficult in some cases where, for example, cover slip edges go across tissue or when marker is drawn on the slide itself. While the tissue will ideally be processed without artifacts, this is not a practical assumption.

Once this image is co-registered in MRI-space, ROIs are placed to avoid areas such as this. Ideally, no images with significant artifacts persisted in the MRI dataset when that model trained. However, nothing is perfect, and the fact that it could not be removed in this histological processing should be noted. Artifacts within the MRI dataset also were not removed prior to training. These consisted of some minor ringing artifacts, but this nonetheless could be affecting the MRI model.

The histological model also did not add rotational or translational image augmentation to the dataset, which could have potentially increased the AUC of the model. This was done to prevent adding information to an image given to the

model. This type of image augmentation adds dummy values in for areas created when rotating or translating. Having a set value being added to areas of the image is not only adding artificial data into the training set, it could skew the model depending on what value is chosen to fill these gaps. These dummy values could potentially skew the training, so those two augmentation methods were not used.

Pitfalls and Challenges

Originally, the histology tumor classifier was trained on naïve networks with the input size changed to be 2000x2000x3. This included simple convolutional networks, a variation of InceptionV4, and a variation of AlexNet. Despite multiple attempts at training networks in this fashion, they did not produce an AUC as good as the pretrained AlexNet. They generally ranged from .6 to .7 for an AUC. In addition to the subpar AUC, upon testing across entire images, the resulting tumor prediction maps were much more varied and distributed than the pretrained AlexNet model, showing a lack of generalizability.

This issue persisted with the MRI network as well, but to a greater degree. The naïve networks used to train on the MRI data tended to overclassify to one output or one value (e.g. either one category, “tumor”, or one specific prediction value, .6). This issue was greater in the MRI model due to the limited amount of data going into the network, particularly because the original MRI dataset used 5x5 voxel ROIs. In all, there were only 100 data points per sample to be classified (5x5 images from the four scans). This was upped to 400 data points per example by using 10x10 voxel ROIs. The issue with using naïve networks in

MRI was overcome by training many different models until one converged properly.

The histology dataset also changed from the original training set to the final training set. At first, subjects were dispersed between both the training and test set, only leaving out a percentage of the images and not images from specific subjects. In order to accurately test generalizability of the model, images from some subjects were left out of the training set. The images in the training set were then class balanced and the remainder of the images were used for validation during training.

At first, a simple classification tree model was attempted with the MRI dataset, but while this model achieved a high test-set AUC, it did not generalize well to MRI slices. This was done either pixel-wise or with a vectorized version of the 10x10 images. The predictions only tended to highlight edges of the brain and areas of blood flow. Bagged trees and random forest models were also tested with similar results. In general, the images produced from these basic machine learning techniques were highly varied and did not generalize or highlight areas of tumor, even if they showed good performance on the test-set.

Local development was also a challenge with the histological network. The size of the images and the number of images needed to train the network were too large for a single computer to handle. This yielded a necessity for the use of the MCW Research Computing Center (RCC). The RCC was able to handle training the AlexNet model and the hundreds of images associated with it in a matter of hours instead of days. This allowed for many iterations of the model to

be trained, rather than training one or a few models and hoping that one would be sufficient. Due to the randomness of shuffling the training dataset, initializing gradient changes, and changing the hyperparameters of the model, training many models will generally yield one with much better results than training a few. Roughly 100 different models were trained on the RCC before one was chosen that performed at this level. This simply could not have been achieved using local development only.

This was not an issue with the MRI model. Local development was used for this training because the tens thousands of images used were only 10x10 and the architectures for the models were not very deep. While many different architectures were tested, only a few performed well. Once an architecture was selected, this was again trained in many different iterations, producing about 50 different models. The best model was selected based on training and testing accuracy. None of the models achieved higher than 83% accuracy during training. This may have been due to the quality of the training dataset. Upon inspection of the last MRIs of subjects used in the training set, some contained ringing artifacts that may be preventing the model from reaching a higher training accuracy.

Summary

Test set accuracy of each model was 70.1% and 73.3%, respectively. However, comparing the histological model to a pathologist's annotations showed a significant improvement in accuracy to 90.15%. Likewise, qualitative

comparisons of the MRI predictions showed that the prediction maps characterize the histological prediction of the same tissue.

V. CONCLUSION

Test set accuracy of the histological and MRI model was 70.1% and 73.3%, respectively. However, comparing the histological model to a pathologist's annotations showed a significant improvement in accuracy to 90.15%. Likewise, qualitative comparisons of the MRI predictions showed that the prediction maps characterize the histological prediction of the same tissue.

This project found that it is possible to predict tumor presence in MRI using histological images from pathologically validated tissue. It also demonstated the efficacy of using a small, well-annotated dataset to classify a much larger dataset, as well as a novel approach to quickly creating deep annotations in histology. Using transfer learning, a pretrained AlexNet was able to classify regions of tumor in H&E images across several subjects with good accuracy. Additionally, despite the model not performing exceedingly well in the test-set alone, looking at the images in their entirety, the annotations from the pathologist and predictions align well with high accuracy.

This project accomplished the goal of showing that a histologically based and pathologically verified MRI tumor predictor can be used to accurately identify tumor. A weakly-classified dataset was used to generate this final model. The efficacy of this model with an AUC of .75 leaves both room for improvement and an impression of confidence in the methods presented. The limitations of this study allow for potential future solutions to reduce the errors in the final model and improve the overall accuracy.

The results of both classifiers tell an important story. While they have their limitations in their differing methods of data input, it is still noteworthy that the MRI classifier did not generalize as well as the histological classifier. This reiterates the point made at multiple institutions that a biopsy is the best way to precisely diagnose a brain tumor. Because of that, it is imperative that pathologically-validated datasets are used to train tumor predictions models.

REFERENCES

Aldape, K., Zadeh, G., Mansouri, S., Reifenberger, G., & von Deimling, A. (2015). Glioblastoma: Pathology, molecular mechanisms and markers. *Acta Neuropathologica, 129*(6), 829-848. doi:10.1007/s00401-015-1432-1

Arthur, F., & Hossein, K. R. (2019). Deep learning in medical image analysis: A third eye for doctors. *Journal of Stomatology, Oral and Maxillofacial Surgery,* doi:S2468-7855(19)30158-2

Bakas, S., Reyes, M., Battistella, E., Chandra, S., Estienne, T., Fidon, L., . . . Menze, B. (2018). *Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge*. Retrieved from https://hal.inria.fr/hal-01959023

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., . . . al., e. (2018). *Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge* Retrieved from http://arxiv.org/abs/1811.02629

Barajas, J., Ramon F, Phillips, J. J., Parvataneni, R., Molinaro, A., Essock-Burns, E., Bourne, G., . . . Nelson, S. J. (2012). Regional variation in histopathologic features of tumor specimens from treatment-naive glioblastoma correlates with anatomic and physiologic MR imaging. *Neuro-Oncology, 14*(7), 942-954. doi:10.1093/neuonc/nos128

Barker, J., Hoogi, A., Depeursinge, A., & Rubin, D. L. (2016). Automated classification of brain tumor type in whole-slide digital pathology images

using local representative tiles. *Medical Image Analysis, 30*, 60-71. doi:10.1016/j.media.2015.12.002

Batchelor, T. T., Gerstner, E. R., Emblem, K. E., Duda, D. G., Kalpathy-Cramer, J., Snuderl, M., . . . Jain, R. K. (2013). Improved tumor oxygenation and survival in glioblastoma patients who show increased blood perfusion after cediranib and chemoradiation. *Proceedings of the National Academy of Sciences, 110*(47), 19059-19064.

Brain tumor - diagnosis. (2018). Retrieved from https://www.cancer.net/cancer-types/brain-tumor/diagnosis

Brain tumor - grades and prognostic factors. (2018). Retrieved from https://www.cancer.net/cancer-types/brain-tumor/grades-and-prognostic-factors

Brain tumor - statistics. (2018). Retrieved from https://www.cancer.net/cancer-types/brain-tumor/statistics

Brain tumor FAQs | ABTA. Retrieved from https://www.abta.org/about-brain-tumors/brain-tumor-education/

Bukowy, J. D., Foss, H., McGarry, S. D., Lowman, A., Hurrell, S., Ickzkowski, K. A., . . . LaViolette, P. (2019). Accurate segmentation of prostate cancer histomorphometric features using a weakly supervised convolutional neural network. *The FASEB Journal, 33*(1), lb12. doi:10.1096/fasebj.2019.33.1_supplement.lb12

Byrne, J. V. (2008). Neuroradiology. *Medicine, 36*(10), 545-555. doi:10.1016/j.mpmed.2008.07.013

Chang, H., Han, J., Zhong, C., Snijders, A. M., & Mao, J. (2018). Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(5), 1182-1194. doi:10.1109/TPAMI.2017.2656884

Coroller, T. P., Bi, W. L., Huynh, E., Abedalthagafi, M., Aizer, A. A., Greenwald, N. F., . . . Aerts, Hugo J W L. (2017). Radiographic prediction of meningioma grade by semantic and radiomic features. *PloS One, 12*(11), e0187908. doi:10.1371/journal.pone.0187908

Deriu, J., Lucchi, A., Luca, V. D., Severyn, A., Müller, S., Cieliebak, M., . . . Jaggi, M. (2017). *Leveraging large amounts of weakly supervised data for multi-language sentiment classification* Retrieved from http://arxiv.org/abs/1703.02504

Diagnosis. Retrieved from https://www.abta.org/about-brain-tumors/brain-tumor-diagnosis/

Dong, H., Yang, G., Liu, F., Mo, Y., & Guo, Y. (2017). Automatic brain tumor detection and segmentation using U-net based fully convolutional networks. Paper presented at the *Medical Image Understanding and Analysis,* 506-517.

Ellingson, B. M., Lai, A., Harris, R. J., Selfridge, J. M., Yong, W. H., Das, K., . . . Cloughesy, T. F. (2013). Probabilistic radiographic atlas of glioblastoma phenotypes. *American Journal of Neuroradiology, 34*(3), 533. doi:10.3174/ajnr.A3253

Ellingson, B. M. (2015). Radiogenomics and imaging phenotypes in glioblastoma: Novel observations and correlation with molecular characteristics. *Current Neurology and Neuroscience Reports, 15*(1), 506. doi:10.1007/s11910-014-0506-0

Ellingson, B. M., Aftab, D. T., Schwab, G. M., Hessel, C., Harris, R. J., Woodworth, D. C., . . . Cloughesy, T. F. (2018). Volumetric response quantified using T1 subtraction predicts long-term survival benefit from cabozantinib monotherapy in recurrent glioblastoma. *Neuro-Oncology, 20*(10), 1411-1418. doi:10.1093/neuonc/noy054

Fathi Kazerooni, A., Nabil, M., Zeinali Zadeh, M., Firouznia, K., Azmoudeh-Ardalan, F., Frangi, A. F., . . . Saligheh Rad, H. (2018). Characterization of active and infiltrative tumorous subregions from normal tissue in brain gliomas using multiparametric MRI. *Journal of Magnetic Resonance Imaging, 48*(4), 938-950. doi:10.1002/jmri.25963

Forst, D. A., Nahed, B. V., Loeffler, J. S., & Batchelor, T. T. (2014). Low-Grade gliomas. *The Oncologist, 19*(4), 403-413. doi:10.1634/theoncologist.2013-0345

Fyllingen, E. H., Stensjøen, A. L., Berntsen, E. M., Solheim, O., & Reinertsen, I. (2016). Glioblastoma segmentation: Comparison of three different software packages. *PLoS ONE, 11*(10) doi:10.1371/journal.pone.0164891

Ge, W., Yang, S., & Yu, Y. (2018). Multi-evidence filtering and fusion for multi-label classification, object detection and semantic segmentation based on weakly supervised learning.

Glioma | diagnosis and treatment. (2019). Retrieved from https://www.mayoclinic.org/diseases-conditions/glioma/diagnosis-treatment/drc-20350255

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. doi:10.1016/j.media.2016.05.004

Hebli, A., & Gupta, S. (December 2017). Brain tumor prediction and classification using support vector machine. Paper presented at the *2017 International Conference on Advances in Computing, Communication and Control (ICAC3),* 1-6. doi:10.1109/ICAC3.2017.8318767

Hoo-Chang, S., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., . . . Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging, 35*(5), 1285-1298. doi:10.1109/TMI.2016.2528162

Horská, A., & Barker, P. B. (2010). Imaging of brain tumors: MR spectroscopy and metabolic imaging. *Neuroimaging Clinics, 20*(3), 293-310. doi:10.1016/j.nic.2010.04.003

How we diagnose brain tumors. Retrieved from https://www.dana-farber.org/brain-tumors/diagnosis/

Hutterer, M., Hattingen, E., Palm, C., Proescholdt, M. A., & Hau, P. (2015). Current standards and new concepts in MRI and PET response assessment

of antiangiogenic therapies in high-grade glioma patients. *Neuro-Oncology, 17*(6), 784-800. doi:10.1093/neuonc/nou322

Iglesias, J. E., Liu, C., Thompson, P. M., & Tu, Z. (2011). Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Transactions on Medical Imaging, 30*(9), 1617-1634. doi:10.1109/TMI.2011.2138152

Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., & Maier-Hein, K. (2018). Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge. Paper presented at the *Brainlesion: Glioma, Multiple Sclerosis,* 287-297.

Islam, A., Reza, S. M. S., & Iftekharuddin, K. M. (2013). Multifractal texture estimation for detection and segmentation of brain tumors. *IEEE Transactions on Biomedical Engineering, 60*(11), 3204-3215. doi:10.1109/TBME.2013.2271383

Jang, B., Jeon, S. H., Kim, I. H., & Kim, I. A. (2018). Prediction of pseudoprogression versus progression using machine learning algorithm in glioblastoma. *Scientific Reports, 8*(1), 12516-9. doi:10.1038/s41598-018-31007-2

Kamboj, A., Rani, R., & Chaudhary, J. (December 2018). Deep leaming approaches for brain tumor segmentation: A review. Paper presented at the 599-603. doi:10.1109/ICSCCC.2018.8703202

Kickingereder, P., Bonekamp, D., Nowosielski, M., Kratz, A., Sill, M., Burth, S., . . . Capper, D. (2012). Radiogenomics of glioblastoma: Machine learning–

based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology, 281*(3), 907-918. doi:10.1148/radiol.2016161382

Kickingereder, P., Bonekamp, D., Nowosielski, M., Kratz, A., Sill, M., Burth, S., . . . Capper, D. (2016). Radiogenomics of glioblastoma: Machine learning–based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology, 281*(3), 907-918. doi:10.1148/radiol.2016161382

Kimura, M., & da Cruz, L. Celso Hygino. (2016). Multiparametric MR imaging in the assessment of brain tumors. *Magnetic Resonance Imaging Clinics of North America, 24*(1), 87-122. doi:10.1016/j.mric.2015.09.001

Kniep, H. C., Madesta, F., Schneider, T., Hanning, U., Schönfeld, M. H., Schön, G., . . . Gellissen, S. (2018). Radiomics of brain MRI: Utility in prediction of metastatic tumor type. *Radiology, 290*(2), 479-487. doi:10.1148/radiol.2018180946

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal, 13*(C), 8-17. doi:10.1016/j.csbj.2014.11.005

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Paper presented at the *Advances in Neural Information Processing Systems,* 1097-1105.

LaViolette, P. S., Mickevicius, N. J., Cochran, E. J., Rand, S. D., Connelly, J., Bovi, J. A., . . . Schmainda, K. M. (2014). Precise ex vivo histological validation of heightened cellularity and diffusion-restricted necrosis in regions of dark apparent diffusion coefficient in 7 cases of high-grade glioma. *Neuro-Oncology, 16*(12), 1599-1606. doi:10.1093/neuonc/nou142

The life of a brain tumor: How does glioblastoma grow? – penn medicine. Retrieved from https://www.pennmedicine.org/updates/blogs/neuroscience-blog/2018/november/the-life-of-a-brain-tumor-how-does-glioblastoma-grow

Lin, Z. (2013). Glioma-related edema: New insight into molecular mechanisms and their clinical implications. *Chin J Cancer, 32*(1), 49-52. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3845586/

Liu, T. T., Achrol, A., Rubin, D., & Chang, S. (2017). Nimg-27. glioblastoma tumor segmentation using deep convolutional neural networks. *Neuro-Oncology, 19*(Suppl 6), vi147. doi:10.1093/neuonc/nox168.602

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., . . . Ellison, D. W. (2016). The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathologica, 131*(6), 803-820. doi:10.1007/s00401-016-1545-1

Lu, C., Hsu, F., Hsieh, K. L., Kao, Y. J., Cheng, S., Hsu, J. B., . . . Chen, C. (2018). Machine Learning–Based radiomics for molecular subtyping of gliomas. *Clinical Cancer Research, 24*(18), 4429-4436.

Lu, S., Lu, Z., & Zhang, Y. (2019). Pathological brain detection based on AlexNet and transfer learning. *Journal of Computational Science, 30*, 41-47. doi:10.1016/j.jocs.2018.11.008

Manogaran, G., Vijayakumar, V., Varatharajan, R., Malarvizhi Kumar, P., Sundarasekar, R., & Hsu, C. (2018). Machine learning based big data processing framework for cancer diagnosis using hidden markov model and GM clustering. *Wireless Personal Communications, 102*(3), 2099-2116. doi:10.1007/s11277-017-5044-z

McGahan, B., Neilsen, B., Kelly, D., McComb, R., Kazmi, S., White, M., . . . Aizenberg, M. (2017). Assessment of vascularity in glioblastoma and its implications on patient outcomes. *Journal of Neuro-Oncology, 132*(1), 35-44. doi:10.1007/s11060-016-2350-3

McGarry, S. D., Hurrell, S. L., Kaczmarowski, A. L., Cochran, E. J., Connelly, J., Rand, S. D., . . . LaViolette, P. S. (2016). Magnetic resonance imaging-based radiomic profiles predict patient prognosis in newly diagnosed glioblastoma before therapy. *Tomography (Ann Arbor, Mich.), 2*(3), 223-228. doi:10.18383/j.tom.2016.00250

Melguizo-Gavilanes, I., Bruner, J., Guha-Thakurta, N., Hess, K., & Puduvalli, V. (2015). Characterization of pseudoprogression in patients with glioblastoma: Is histology the gold standard? *Journal of Neuro-Oncology, 123*(1), 141-150. doi:10.1007/s11060-015-1774-5

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., . . . Van Leemput, K. (2015). The multimodal brain tumor image segmentation

benchmark (BRATS). *IEEE Transactions on Medical Imaging, 34*(10), 1993-2024. doi:10.1109/TMI.2014.2377694

MICCAI 2014 grand challenges. (2016). Retrieved from https://wiki.cancerimagingarchive.net/display/Public/MICCAI+2014+Grand+Challenges

Mickevicius, N. J., Carle, A. B., Bluemel, T., Santarriaga, S., Schloemer, F., Shumate, D., . . . LaViolette, P. S. (2015). Location of brain tumor intersecting white matter tracts predicts patient prognosis. *Journal of Neuro-Oncology, 125*(2), 393-400. doi:10.1007/s11060-015-1928-5

Mlynarski, P., Delingette, H., Criminisi, A., & Ayache, N. (2018). *Deep learning with mixed supervision for brain tumor segmentation* Retrieved from http://arxiv.org/abs/1812.04571

Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D. A., Barnholtz-Sloan, J. S., Velázquez Vega, J. E., . . . Cooper, L. A. D. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences, 115*(13), E297-E2979.

Moffat, B. A., Chenevert, T. L., Lawrence, T. S., Meyer, C. R., Johnson, T. D., Dong, Q., . . . Ross, B. D. (2005). Functional diffusion map: A noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Proceedings of the National Academy of Sciences, 102*(15), 5524-5529.

Nguyen, H. S., Milbach, N., Hurrell, S. L., Cochran, E., Connelly, J., Bovi, J. A., . . . LaViolette, P. S. (2016). Progressing bevacizumab-induced diffusion restriction is associated with coagulative necrosis surrounded by viable

tumor and decreased overall survival in patients with recurrent glioblastoma. *American Journal of Neuroradiology, 37*(12), 2201. doi:10.3174/ajnr.A4898

Ohgaki, H., & Kleihues, P. (2013). The definition of primary and secondary glioblastoma. *Clinical Cancer Research, 19*(4), 764-772.

Padma, M. V., Said, S., Jacobs, M., Hwang, D. R., Dunigan, K., Satter, M., . . . Mantil, J. C. (2003). Prediction of pathology and survival by FDG PET in gliomas. *Journal of Neuro-Oncology, 64*(3), 227-237. doi:1025665820001

Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging, 35*(5), 1240-1251. doi:10.1109/TMI.2016.2538465

Phillips, W. E., Velthuizen, R. P., Phuphanich, S., Hall, L. O., Clarke, L. P., & Silbiger, M. L. (1995). Application of fuzzy c-means segmentation technique for tissue differentiation in MR images of a hemorrhagic glioblastoma multiforme. *Magnetic Resonance Imaging, 13*(2), 277-290. doi:10.1016/0730-725X(94)00093-I

Porz, N., Bauer, S., Pica, A., Schucht, P., Beck, J., Verma, R. K., . . . Wiest, R. (2014). Multi-modal glioblastoma segmentation: Man versus machine. *Plos One, 9*(5), e96873. doi:10.1371/journal.pone.0096873

Rathore, S., Akbari, H., Rozycki, M., Abdullah, K. G., Nasrallah, M. P., Binder, Z. A., . . . Davatzikos, C. (2018). Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Scientific Reports, 8*(1), 5087-12. doi:10.1038/s41598-018-22739-2

Rathore, S., Bakas, S., Akbari, H., Nasrallah, M. P., Bagley, S., & Davatzikos, C. (2019). Abstract 1392: Machine learning radiomic biomarkers non-invasively assess genetic characteristics of glioma patients. *Cancer Research, 79*(13 Supplement), 1392.

Reyes, A. M. de los, Buemi, M. E., Alemán, M. N., & Suárez, C. (November 2018). Development of a graphic interface for the three-dimensional semiautomatic glioblastoma segmentation based on magnetic resonance images. Paper presented at the 1-6. doi:10.1109/CACIDI.2018.8584357

Roetzer, T., Leskovar, K., Peter, N., Furtner, J., Muck, M., Augustin, M., . . . Woehrer, A. (2019). Evaluating cellularity and structural connectivity on whole brain slides using a custom-made digital pathology pipeline. *Journal of Neuroscience Methods, 311*, 215-221. doi:10.1016/j.jneumeth.2018.10.029

S Tandel, G., Biswas, M., G Kakde, O., Tiwari, A., S Suri, H., Turk, M., . . . Suri, J. S. (2019). A review on a deep learning perspective in brain cancer classification. *Cancers, 11*(1), 111. doi:10.3390/cancers11010111

Sanghani, P., Ang, B. T., King, N. K. K., & Ren, H. (2018). Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning. *Surgical Oncology, 27*(4), 709-714. doi:10.1016/j.suronc.2018.09.002

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17*(3), 143-155. doi:10.1002/hbm.10062

Song, B., Wang, H., Chen, Y., Liu, W., Wei, R., Dai, Z., . . . Wang, L. (2018). Medicine. *Medicine, 97*(26) Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6039645/

Spille, D. C., Sporns, P. B., Heß, K., Stummer, W., & Brokinkel, B. (2019). Prediction of high-grade histology and recurrence in meningiomas using routine preoperative magnetic resonance imaging: A systematic review. *World Neurosurgery, 128*, 174-181. doi:10.1016/j.wneu.2019.05.017

Stadlbauer, A., Moser, E., Gruber, S., Buslei, R., Nimsky, C., Fahlbusch, R., & Ganslandt, O. (2004). Improved delineation of brain tumors: An automated method for segmentation based on pathologic changes of 1H-MRSI metabolites in gliomas. *Neuroimage, 23*(2), 454-461. doi:10.1016/j.neuroimage.2004.06.022

Szopa, W., Burley, T. A., Kramer-Marek, G., & Kaspera, W. (2017). Diagnostic and therapeutic biomarkers in glioblastoma: Current status and future perspectives. *BioMed Research International, 2017* doi:10.1155/2017/8013575

Tamimi, A. F., & Juweid, M. (2017). Epidemiology and outcome of glioblastoma. In S. De Vleeschouwer (Ed.), *Glioblastoma* (). Brisbane (AU): Codon Publications. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK470003/

Tixier, F., Um, H., Young, R. J., & Veeraraghavan, H. (2019). Reliability of tumor segmentation in glioblastoma: Impact on the robustness of MRI-radiomic features. *Medical Physics,* doi:10.1002/mp.13624

Toyonaga, T., Hirata, K., Yamaguchi, S., Hatanaka, K. C., Yuzawa, S., Manabe, O., . . . Tamaki, N. (2016). 18F-fluoromisonidazole positron emission tomography can predict pathological necrosis of brain tumors. *European Journal of Nuclear Medicine and Molecular Imaging, 43*(8), 1469-1476. doi:10.1007/s00259-016-3320-x

University of Pennsylvania. (2018). Multimodal brain tumor segmentation challenge 2018 | data. Retrieved from https://www.med.upenn.edu/sbia/brats2018/data.html

Varuna Shree, N., & Kumar, T. (2018). Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network. *Brain Informatics, 5*(1), 23-30. doi:10.1007/s40708-017-0075-5

Vergun, S., Suhonen, J. I., Nair, V. A., Kuo, J. S., Baskaya, M. K., Garcia-Ramos, C., . . . Prabhakaran, V. (2018). Predicting primary outcomes of brain tumor patients with advanced neuroimaging MRI measures. *Interdisciplinary Neurosurgery, 13*, 109-118. doi:10.1016/j.inat.2018.04.013

Vijay Kumar, G., & GV Raju, D. (2010). *Biological early brain cancer detection using artificial neural network*

Villanueva-Meyer, J. E., Mabray, M. C., & Cha, S. (2017). Current clinical brain tumor imaging. *Neurosurgery, 81*(3), 397-415. doi:10.1093/neuros/nyx103

Visser, M., Müller, D. M. J., van Duijn, R. J. M., Smits, M., Verburg, N., Hendriks, E. J., . . . de Munck, J. C. (2019). Inter-rater agreement in glioma segmentations on longitudinal MRI. *NeuroImage. Clinical, 22*, 101727. doi:10.1016/j.nicl.2019.101727

Wang, S., Lu, S., Dong, Z., Yang, J., Yang, M., & Zhang, Y. (2016). *Dual-tree complex wavelet transform and twin support vector machine for pathological brain detection* doi:10.3390/app6060169

Xu, Y., Jia, Z., Wang, L., Ai, Y., Zhang, F., Lai, M., & Chang, E. I. (2017). Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinformatics, 18*(1), 281-17. doi:10.1186/s12859-017-1685-x

Yang, X., Zhu, J., Dai, Y., Tian, Z., Yang, G., Shi, H., . . . Tao, X. (2019). Multi-parametric effect in predicting tumor histological grade by using susceptibility weighted magnetic resonance imaging in tongue squamous cell carcinoma. *BMC Medical Imaging, 19*(1), 24. doi:10.1186/s12880-019-0322-8

Yang, Y., Yan, L., Zhang, X., Han, Y., Nan, H., Hu, Y., . . . Wang, W. (2018). Glioma grading on conventional MR images: A deep learning study with transfer learning. *Frontiers in Neuroscience, 12* doi:10.3389/fnins.2018.00804

Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R., & Davatzikos, C. (2009). Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magnetic Resonance in Medicine, 62*(6), 1609-1618. doi:10.1002/mrm.22147

Zhe Zhao, Guan Yang, Yusong Lin, Haibo Pang, & Meiyun Wang. (2018). Automated glioma detection and segmentation using graphical models. *PLoS One, 13*(8), e0200745. doi:10.1371/journal.pone.0200745

Zhou, M., Scott, J., Chaudhury, B., Hall, L., Goldgof, D., Yeom, K. W., . . . Gatenby, R. (2018). Radiomics in brain tumor: Image assessment, quantitative feature descriptors and machine-learning approaches. *AJNR. American Journal of Neuroradiology, 39*(2), 208-216. doi:10.3174/ajnr.A5391

Zijlstra, A., Novitskaya, T., Vizio, D. D., Reis- Sobreiro, M., & Freeman, M. (2019). Abstract 4448: Machine-learning assisted histopathology (HistoMAP) links nuclear membrane instability to disease progression in prostate cancer. *Cancer Research, 79*(13 Supplement), 4448.