Improving diagnosis and management of patients with glioma using artificial intelligence and multi-parametric MRI

by

Nate Tran

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
AND
UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

DocuSigned by:
Janine Lupo
F1F4F52A4F3D4D8...

Janine Lupo

Chair

DocuSigned by:
Chunlei Liu
...

Chunlei Liu

DocuSigned by:
Annette Molinaro
1B4CF5317...

Annette Molinaro

Committee Members

Acknowledgements

I would like to thank first and foremost, Dr. Janine Lupo, for her mentorship in the last 3.5 years. I joined Janine's lab in 2019, already starting the 4th year of my Ph.D. but unable to continue with my previous P.I. due to a very unexpected relocation. It was also a particularly hard time for all of us due to the start of the Covid-19 pandemic. Janine is most definitely one of the most gentle and patient professors I've met. Her knowledge and passion in the field of radiology and brain tumors are infectious, and I've had so much fun working with and learning from her.

I would like to thank and give a special shoutout to everyone in the Lupo Lab (as well as everyone in the Brain group). None of this work would happen without their helps. In particular, I want to thank Dr. Tracy Luks for all your help with data annotation and analysis. Angela Jakary, you are truly a rockstar for all the lessons when I first joined the lab as well as all the instruction in the Wiki page, as well as all the alignment stuff. Jacob Ellison, you are seriously one of the most hard-working people I've known, and you've inspired me in so many ways; it's been such a wonderful experience working and learning with you. Oluwaseun Adegbite, Bo Liu, Jenny Lee, Dr. Melanie Morrison, Devika Nair, Marisa Lafontaine, Paul Rowley, James Golden, thank you so much for all your help in my projects. I want to thank Dr. Yan Li for all the help with the metabolic data and processing, Dr. Hui Lin for help with the radiation therapy project (among other things), Dr. Steve Braunstein, Dr. Annette Molinaro, Dr. Joanna Phillips for all the insights; Dr. Chunlei Liu and Dr. Annette Molinaro for agreeing to be a member of my dissertation committee.

I also want to give a shoutout to everyone in the UC Berkeley-UCSF Bioengineering program, especially Victoria Starrett and everyone in my 2016 cohort (who have all decided to graduate way before me). Thank you, Dr. Duan Xu, for being my graduate advisor. I want to also thank Dr. Xiao Hu, my previous P.I., as well as Dr. Michelle Pelter in the Department of

Nursing, and everyone else in the Hu Lab (especially Dr. Tânia Pereira, thank you for all the sleepless nights bouncing off deep learning ideas with me).

And last but not least, I want to thank my partner, Francis DiDonato, for being so supportive in the last 3 years. Your encouragement and support have allowed me to finish this dissertation. To my friend Wally Bee, thank you for proofreading my paper. Finally, a quick shout-out also to my friends in the Oakland Gay Men's Chorus, as well as my backpacking friends, who allow me to stay sane toward the end of my Ph.D.

Nate Tran

December 2023

Abstract

Improving diagnosis and management of patients with glioma using artificial intelligence and multi-parametric MRI

Nate Tran

Gliomas are highly infiltrative, heterogenous brain tumors with poorly defined margin, and varying overall survival based on molecular subtype and grade. Despite recent developments in new diagnostic and treatment tools for gliomas, progression free survival and overall survival has only improved marginally for patients with glioma. Furthermore, treatment of glioma tends to be "one size fits all", which can lead to either undertreating or overtreating the subclinical disease. Thus, the management of gliomas needs to be more patient-specific and more flexible over the course of the disease if the goal is to maximize both the longevity and quality of life of these patients.

Recent advancement in MRI and radiation therapy research has opened the door for many opportunities to answer these questions. While the use of MRI in the clinic has been mostly limited to anatomical imaging, other MRI modalities have been gaining a lot of traction and have been proven to be able to provide clinical information not available in anatomical MRI. However, incorporating multimodal MRI in glioma management is a difficult task, because more advanced MRI acquisitions are not consistently acquired across institutions, and effectively understanding the consequences of changes observed on multimodal MRI over time is difficult even for trained radiologist.

Artificial Intelligence has shown promise in making predictions from multi-parametric images, as multiple inputs can be given at the same time, and all processing and prediction tasks can be pre-trained and automatic applied. In this dissertation, we attempted to use multimodal

MRI and artificial intelligence to improve both the diagnosis and treatment planning for patients with glioma. First, we developed an 1D deep learning based model that can help predict tumor histopathology noninvasively using the full spectrum of ^{1}H MR Spectroscopic Imaging data. Then, we developed 3D segmentation-based deep learning model using multi-parametric MRI to redefine the clinical target volume for radiotherapy treatment planning and found that our model performed better than current practice, both in terms of better detecting subclinical disease and future progression, as well as sparing normal brain tissue. Finally, we highlighted the efficacy of using multi-parametric MRI in predicting a patient's progression free and overall survival and improved the model performance by applying different types of masks.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

Glioma is the most prevalent type of primary brain tumor that is thought to originate mostly from glial cells and accounts for 80% of all malignant primary brain tumors. Glioblastoma, the most common type of glioma, is also the deadliest kind, with the 5-year overall survival rate of only 0.05-4.7%[1]. While relatively rare, glioblastoma is often regarded as one of the most dangerous type of tumors due to its highly infiltrative and heterogenous nature, usually with poorly defined margins. Despite recent developments in new diagnostic and treatment capabilities for patient with glioma, progression free survival (PFS) and overall survival (OS) has only improved marginally for these patients. Many of these approaches tend to be very invasive, potentially worsening the patients' quality of life. Furthermore, non-invasive imaging signatures of which patients would do best on a certain therapy has potentially limited the effectiveness of phase III clinical trials. In the case of newly diagnosed glioblastoma, since Stupp et al. [2] in 2005 showed a prolonged survival benefit with the addition of temozolomide (TMZ) in addition to surgical resection and radiotherapy, the protocol for standard of care (SOC) treatment has remained relatively unchanged, with the only current option being the addition of an anti-tumor wearable device after radiation therapy (RT) known as tumor treating fields (TTF) [3]. Current research in glioma have shown that the management of gliomas needs to be more targeted spatially over the course of the disease if the goal is to maximize both the longevity and quality of life of these patients. What are some ways that we can achieve this?

Recent advancement in MRI and RT research has opened the door for many opportunities to answer these questions. While the use of MRI in the clinic has been mostly limited to anatomical imaging, other MRI modalities have been gaining a lot of traction and have been proven to be able to provide relevant clinical information not available in anatomical MRI. For example, diffusion-weighted MRI can identify subclinical tumor invasion not seen in anatomical imaging, which causes an increase in cellularity or edema and decrease in directionality along white matter

tracts. Metabolite levels estimated using the Proton Magnetic Resonance Spectroscopy (1H-MRS) and the derived Choline-to-NAA index (CNI) can help measuring the underlying cellular metabolism associated with tumor aggressiveness[4], [5], hypoxia [6], as well as tumor growth and progression[7]–[9]. However, incorporating multimodal MRI in glioma management has proven to be a difficult task, because MRI acquisition is not consistent across institutions, and visual interpretation of changes on these images in patients' multimodal MRI is difficult even for trained radiologist.

Artificial Intelligence (AI) has recently allowed these difficult tasks to be possible, as multiple inputs can be given at the same time, and all processing and prediction tasks can be pre-trained and automatic. The use of machine learning and deep learning in biomedical imaging research continues to be one of the most exciting developments in the last decade. In the field of MRI, AI has allowed tasks such as image contrast reconstruction, patient outcome prediction, lesion segmentation, etc., to be performed with high speed, accuracy, and precision. In this dissertation, we attempted to use multimodal MRI and AI to improve both the diagnosis and treatment planning for patients newly-diagnosed with glioma. This work included several background sections, followed by original research that dives deep into different stages of glioma management.

Chapter 2 provides detailed information on glioma, its prognosis, as well as the current practice in diagnosing and treating patients with glioma. In this chapter, we also highlighted areas that can be improved regarding glioma management and the significance of our work.

Chapter 3 provides background information on the physics of MRI, the generation of standard clinical anatomical MR Imaging contrasts, as well as introduced various advanced MRI techniques used in our research, including diffusion-weighted MRI and 1H-MRS. We emphasized the importance of incorporating these advanced MRI techniques in clinic for improving diagnosis and treatment planning of patients with glioma.

Chapter 4 introduces the fundamentals of AI, particularly machine learning and deep learning techniques, and how to best apply AI in tasks that involve brain tumor imaging. We also discussed several methods to improve model performance, as well as best practice for model evaluation.

Chapter 5 discusses “Machine learning for predicting voxel-wise histopathology of tumor cells in newly diagnosed glioblastoma patients using Proton Magnetic Resonance Spectroscopy”, where we developed a deep learning model that utilizes the wealth of information contained in the entire spectrum to predict voxel-wise histopathology of tumor cells, including tumor cellularity, mitotic activity (Ki-67), and a new composite tumor aggressiveness index (CTAI; defined as the sum of normalized cellularity and Ki-67) using tissue samples with spatially mapped coordinates on 3D 1H-MRSI.

Chapter 6 discusses “Defining radiation target volumes for glioblastoma (GBM) and predicting tumor recurrence with machine learning using pre-radiotherapy anatomical, diffusion & metabolic MRI”, where we used multiparametric MRI acquired within a week of beginning RT with machine learning to predict regions of subsequent tumor progression in order to ultimately guide precision-based RT planning, and then compare the resulting predicted map to the standard of care RT clinical target volume consisting of a 2cm uniform expansion of the T2-hyperintensity lesion volume.

Chapter 7 discusses " Early prediction of PFS and OS of patients with glioblastoma using machine learning and multi-parametric MRI”, where we developed a machine learning model that incorporates multi-parametric metabolic and physiologic MRI parameters from before and/or mid-therapy to predict OS and PFS in patients with GBM treated with upfront radiation, anti-angiogenic-, and cytotoxic-chemotherapy.

Chapter 8 provides a summary of the significant findings in this dissertation, limitations in our methods, as well as future directions of the presented work.

Chapter 2. Glioma

2.1. Overview of glioma

Brain tumors are usually classified as either primary brain tumors, which originate within the brain, or secondary brain tumor, which originate elsewhere in the body and migrate to the brain. Each year, there are about 90,000 new cases of primary brain tumor being diagnosed in the United States, 29% of which are malignant [10].

Of all primary malignant brain tumor types, adult-type diffuse gliomas are by far the most prevalent, accounting for about 80-85% [11]. Gliomas are defined as brain tumors that originated from the neuroglial cells or stem cells, and are known to be very heterogenous, highly infiltrative, and often with poorly defined margins [1]. It is very important to correctly diagnose and grade brain tumors using both the histological and genomic archetype of the tumor, as that will allow doctors to accurately determine the patients' prognosis, as well as determining the appropriate course of treatments.

Tumor grades are usually determined using histopathological assessment of tumor cells, or the appearance of the tumor cells under the microscope. Since 2016, thanks to the advances in DNA sequencing, more emphasis has been given to the genetic and molecular makeup of the tumor cells, and molecular markers such as IDH, 1p19q, ATRX, and TERT have been used to determine glioma subtype [12]. That said, histopathological assessment of tumor cells are still an important part of brain tumor diagnosis and management, especially in low-resource settings.

In the newest update to the 'World Health Organization (WHO) Classification of Tumors of the Central Nervous System', adult-type diffuse gliomas were further classified into three types: 1) astrocytoma, IDH-mutant (WHO CNS Grade 2-4); 2) oligodendroglioma, IDH-mutant, 1p/19q-codeleted (WHO CNS Grade 2-3); and 3) glioblastoma, IDH-wildtype [12]. Unsurprisingly, median overall survival (OS) of patients with lower grade gliomas (~10 years) is much longer than patients with glioblastoma (less than 2 years) [2], [3]. This poor outcome is partially due to difficulties in

defining and treating the full extent of these tumors. In the next few sections, I will discuss the challenges in diagnosis and management of gliomas in greater details, and approaches to improve patient outcome.

2.2. Challenges in glioma diagnosis

2.2.1. Glioma diagnosis

Patients newly-diagnosed with brain tumors usually presented clinical symptoms such as headache, nausea, vomiting, fatigue, etc., with a prolonged onset due to an increase in intracranial pressure [11], [13], [14]. Neurological symptoms can also arise, such as speech difficulties, hearing loss, seizure, etc [15]. Once a brain tumor is suspected, radiological evaluation is performed. Magnetic resonance imaging (MRI) of the brain is usually the primary imaging modality, and the standard protocol includes at least 3D T1-weighted IR-SPGR imaging pre- and post- the injection of a gadolinium-based contrast agent (T1-weighted pre-GD and T1-weighted post-GD), and T2-weighted axial fluid-attenuated inversion recovery (T2-weighted FLAIR) [16]. These sequences allowed radiologist to identify and evaluate the structure and architecture of tumor tissues, non-enhancing tumors, and edema [17]. Other advances MRI techniques, such as T2*-weighted susceptibility-weighted imaging (SWI), diffusion-weighted imaging (DWI), perfusion-weighted imaging (PWI), and magnetic resonance spectroscopy (MRS) can also be performed to aid in diagnosis. These imaging modalities will be discussed in greater details in Chapter 3.

Brain tumor diagnosis is determined from molecular and histopathology assessment of tissue samples taken during surgical resection of the tumor or a biopsy if resection is not possible due the location of the tumor in the brain. During this process, several small samples of tumor tissue are generally obtained at the time of surgery using a stereotactic needle. Advanced tissue navigation systems, such as the BrainLab (Vector Vision) neuro-navigational system, has since

allowed tissue samples to be taken based on imaging characteristics and record the coordinates of the exact location on MR images during surgery [18]. The tumor tissue collected is then observed by pathologists under a microscope to provide a full histopathology profile of the tumor needed for diagnosis. Not only do the histopathology features allow pathologists to determine the tumor grades, but they also provide information regarding tumor's behavior, such as its aggressiveness, proliferation, necrosis. One limitation of using relying on tumor's histopathology features for tumor diagnosis is that it can be subject to 'inter-rater reliability', as well as insufficient tissue sampling, or tissue sampled not representing the full range of tumor characteristics. Because of this, the new WHO criteria for tumor grading/classification have shifted to rely more on molecular/genetic features, such as IDH, 1p19q, ATRX, and TERT [12], [19]. This new classification has shown to reflect prognosis and OS of patients more accurately. The full flow-chart for diffuse glioma classification based on molecular/genetic features is presented in Figure 2.1. Based on this classification, adult-type diffuse gliomas were classified into three types: 1) astrocytomas, IDH-mutant (WHO CNS Grade 2-4); 2) oligodendrogliomas, IDH-mutant, 1p/19q-codeleted (WHO CNS Grade 2-3); and 3) glioblastoma, IDH-wildtype.

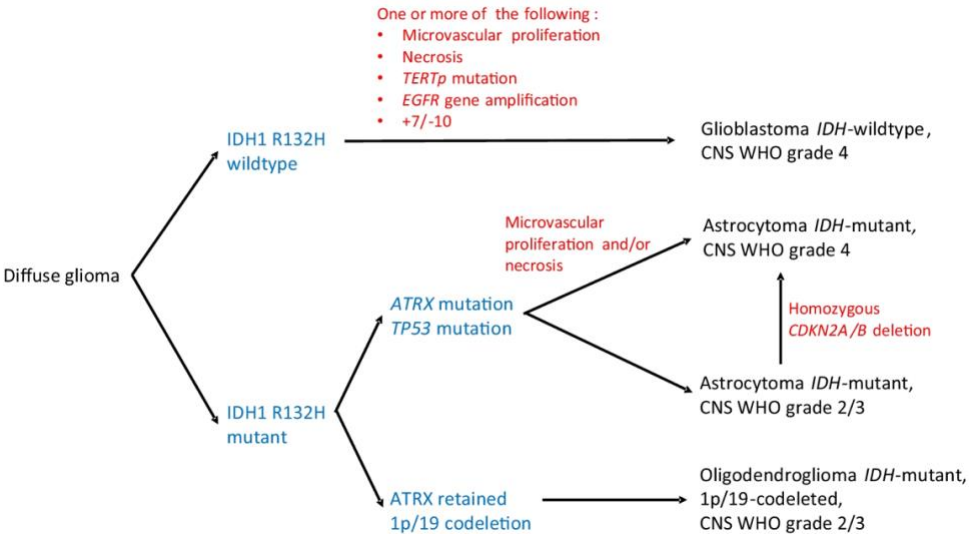

Figure 2.1. Classification of diffuse glioma based on molecular/ genetic features
All diffuse glioma are classified by IDH mutation status and 1p/19q codeletion/ ATRX mutation status respectively.

Besides MRI and histopathology/molecular assessment of tumor tissue, glioma diagnosis can also include (but not limited to) the following: CT scan, positron emission tomography (PET) scan, lumbar puncture (to assess cerebrospinal fluid - CSF), myelogram (if tumor spreads to CSF), eyes and neurological assessment [20].

2.2.2. Glioma histopathology assessment and challenges

Despite the shift in WHO's criteria for classifying glioma types to rely more on genetic markers, histopathology assessment remains the gold standard method for tumor diagnosis, as it can also still be performed in a low-resource setting. In addition, histopathology assessment of tumor cells, including microvascular hyperplasia, cellular proliferation, nuclear atypia, architectural disruption, and necrosis, etc., can reveal the heterogeneity of lesion within the same patient, allowing physician to identify tumor margin, personalize treatment, and target the most aggressive tumor tissue if the sampling strategy allows.

Table 2.1 shows a list of histopathology measures and their corresponding tumor biology [21]. Lower-grade tumors usually are associated with low cell density of larger, well-differentiated, less-proliferative cells, while higher grader tumors have high cellularity, uncontrollable cell proliferation, more necrosis, and more vascular hyperplasia. Among these features, two of the most important tumor biology are the tumor proliferation, and total cell density [22]. Total cell density is calculated by counting the total number of positive hematoxylin-stained nuclei per field (or per mm^2). Tumor proliferation is calculated by counting the tumor cells positive for KI-67 stain over total number of tumor cells [21]. This works because the KI-67 antigen is usually expressed in all phases of cell cycle (S, G1, G2 and M phases – phases with active cell proliferation) except G0 phase (resting phase) [23]. Both cell density and tumor proliferation are strongly correlated with tumor progression, OS, glioma type, and tumor aggressiveness [24]–[28]. Figure 2.2 shows an example of how mean KI-67 increases with histologic grade tissue samples.

Table 2.1. Histopathology measurement and corresponding tumor biology
Example of histopathology measurement and corresponding tumor biology [21]

Histopathology measures	Tumor Biology
KI-67	Tumor proliferation
Cell density, H&E	Cellularity
Necrosis, H&E	Necrosis
CA-9	Hypoxia
SMI-31	Architectural disruption
Factor VIII	Microvascular hyperplasia
GFAF, H&E	Gliosis formulation

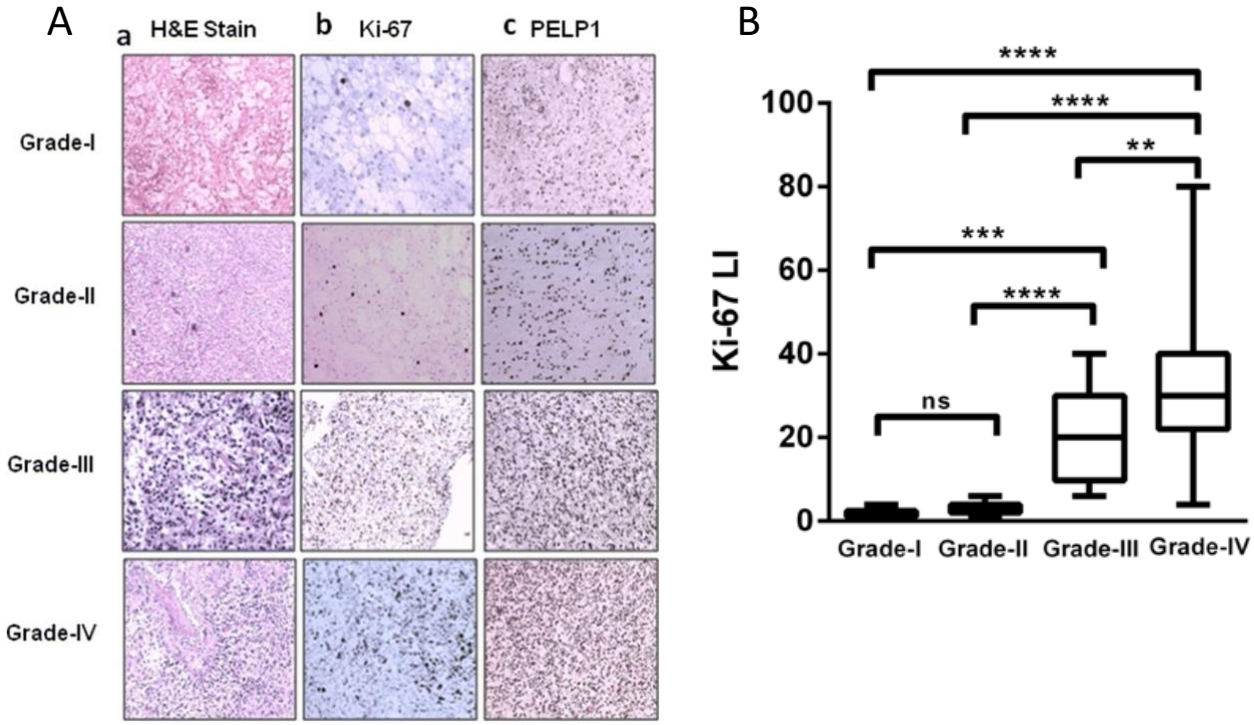

Figure 2.2. Histopathology can explains tumor grades
A. Representative slides of different grades of astrocytomas (a) H&E staining (b) Ki-67, and (c) PELP1 expression immunostained by IHC technique; B. KI-67 is significantly higher for higher grade astrocytomas (grade III, IV) compared to lower grade (grade I, II)
Figure adapted from Padmavathy, 2020 [25]

The previous section has shown how histopathology measures can depict the heterogeneous and infiltrative nature of glioma, especially glioblastoma. However, there are some

challenges that remain for tumor-wide assessment and characterization. First, it is hard to capture the tumors heterogeneity with biopsied tissue samples, because only a few tissue samples can be collected during the surgery, not fully representing the whole lesion [29], [30]. Furthermore, due to the biological heterogeneity of the tumor tissues, it is also challenging to identify the most malignant area of the lesion to obtain tissue samples from to make an accurate diagnosis. This gives rise to the need of a more non-invasive and systematic way to map tumor histopathology measure in real time. Chapter 5 will show how the advancement of multimodality MRI and artificial intelligence could be used to help solve this ongoing problem.

2.3. Prognosis and treatment for newly-diagnosed glioma patients and challenges

2.3.1. Overview of newly-diagnosed glioma treatment

Since the introduction of Stupp's protocol for treating higher-grade IDH-mutant gliomas and glioblastoma to include radiotherapy and adjuvant chemotherapy, prognosis of glioma patients has only improved slightly: the median 5-year survival rate for newly-diagnosed glioma patients is only ~36% [10]. That number is as low as 7% for glioblastoma, IDH-wild-type patients [3]. Specifically, OS for patients with IDH-wild-type glioblastoma is 12-21 months; for patients with IDH-mutant, 1p/19q non-codeletion tumors is 7-8 years; and for patients with IDH-mutant, 1p/19q-codeletion oligodendrogliomas is 13-14 years after chemotherapy and RT [11]. However, improving OS and progression free survival (PFS) should not be the only goal. Since most newly diagnosed glioma patients, especially those with IDH-wildtype, grade IV glioblastoma, will eventually progress, it is equally important to improve patients' health-related quality of life (HRQoL). That means, the treatment strategy should be aggressive enough to kill tumor cells, while minimizing normal brain tissue function and neurological and cognitive impairment.

Treatment for all glioma types begin with maximal-safe surgical resection of the lesion, as more aggressive resection of tumor tissues is associated with higher OS, even for patients with

lower grade glioma [31]. After that, the course of treatment diverges for different patients depending on the molecular subtype and grade of the glioma. Higher grade patients, including those with glioblastoma, usually receive standard of care (SOC) treatment course following Stupp's protocol, which includes surgical resection, followed by RT and chemotherapy [2], [32]. This will be explored more in the next section. For low grade glioma patients, SOC treatment can be given, though some studies have questioned the high dose of RT for Grade 2 slow-growing gliomas [33]–[35]. In some cases, especially for younger patients (< 40 years old) with grade 2 glioma, low residual lesion after resection, and no tumor-related neurological symptoms, a "watch and wait" approach with regular MRI follow-up is recommended.

2.3.2. Treatment for newly-diagnosed glioblastoma patients

The treatment for newly-diagnosed glioblastoma patients begin with maximal safe surgical resection. The goal of the surgery is to safely remove as much as the tumor possible without affecting the patient's health and neurological state. A gross total resection (GTR) refers to the removal of >75% of the contrast-enhancing tumor observed in the T1-weighted post-Gd MRI image. However, surgery alone usually is not enough as microscopic tumor cells infiltrate well into the surrounding non-enhancing lesion and even into normal brain. While in many cases, it may be tempted to completely remove the whole tumor via surgery, doing so can easily cause damage to the patients' neurology state, depending on the tumor's location in the brain, worsening their quality of life [32], [36]. Because of this, surgery is usually followed by additional treatment in order to treat residual tumor cells.

After surgery, patients typically receive external beam radiotherapy (RT) (a total dose of 60 Gy in 2 Gy fractions over a course of 6 weeks), in conjunction with daily temozolomide chemotherapy (TMZ) (75 mg/m2), and six cycles of maintenance adjuvant TMZ chemotherapy (total 150-200 mg/m2). External beam radiotherapy works by directing electromagnetic (x-ray) or photon beam generated externally to the tumorous tissues [37]. The high energy particles cause

tumor cells death by either direct route (by directly ironizing the cells' DNA), or indirect route (by generating radiation-induced free radicals that cause double stranded DNA breaks) [38], [39]. This interaction is summarized in figure 2.3. The goal of radiotherapy is to control tumor growth locally while ensuring relatively low level of toxicity, since there is always a small risk of brain cell death due to radiation necrosis. But overall, RT is relatively safe, and patients receiving RT shows an improvement in OS compared to those without (for example, 29 weeks vs. 17 weeks for patients older than 70 [40]).

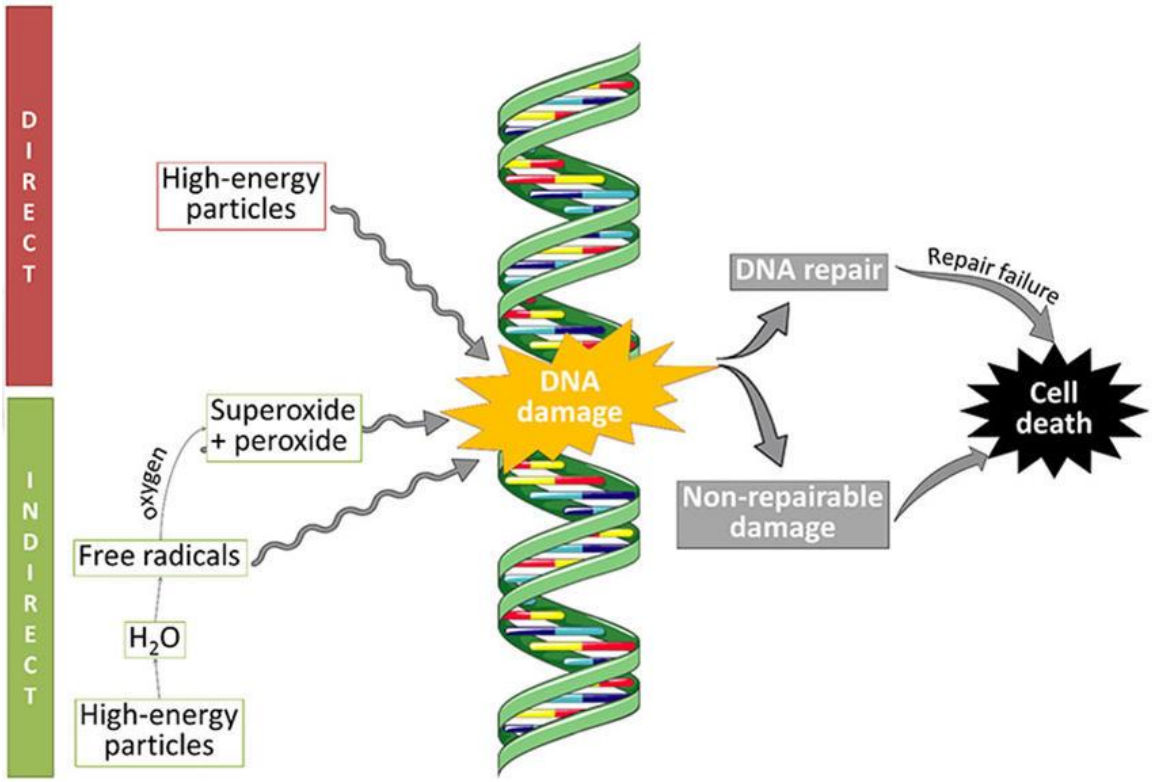

Figure 2.3. The biology of radiotherapy
High energy particles generated from external sources cause DNA damage and cell death either directly or indirectly (by generating free radicals). Figure adapted from Przystupski, 2019 [38]

Chemotherapy is the other component of SOC treatment. TMZ, a DNA alkylating drug that can cross the blood brain barrier, is usually the most widely used therapy in the brain [32]. Once entering the cell membrane, TMZ is spontaneously converted to monomethyl triazene 5-(3-methyltriazene-1-yl)-imidazole-4-carboxamide (MTIC), which then methylates cell's DNA,

resulting in cell damage, apoptosis, or limited cell growth and proliferation [41]. Many studies have shown that patients treated with TMZ in conjunction with RT have a higher OS compared to patients treated with only RT (14.6 months vs. 12.1 months) [42]. The dose for TMZ is 75 mg/m^2 daily in concurrent with RT, and six cycles of maintenance adjuvant TMZ (total 150-200 mg/m2).

Besides the three major SOC treatments described above, several other treatments are being studied vigorously. Tumor-Treating Fields (TTFs) is another FDA-approved treatment for GBM. This treatment works by applying alternating electric field at 200 kHz and low intensity (~1-3 V/cm) directly to the patient's scalp and cause apoptosis in rapidly diving cells. The addition of TTF to SOC treatment has shown to improve OS from 16.0 months to 20.9 months, and PFS from 4.0 months to 6.7 months [3]. Anti-angiogenic drugs have also been studied as newly-diagnosed GBM treatment. Bevacizumab is one of those drugs, as it targets vascular endothelial growth factor (VEGF), an important antigen in angiogenesis and neovascularization that is over-expressed in GBM. However, the use of bevacizumab only shows an improvement in PFS (by 3 months) but not OS [43], [44], likely because it is normalizing leaky tumor vascular by repairing the blood brain barrier, preventing the classic markers of tumor progression to be shown in MR images. Enzastaurin is another anti-angiogenic drug that target protein kinase C β. Although certain patients in the phase 2 trial exhibited a distinct OS benefit, phase 3 clinical trials did not reveal a significant improvement in OS or PFS when using the drug [45].

Despite decades of clinical trials incorporating novel systemic agents and more aggressive surgical approaches, only minimal improvements in outcome have been achieved, and SOC treatment remain largely the same since 2005. This is mainly due to the difficulty in identifying and targeting the highly malignant tumors to the full extent while sparing normal brain tissue. In the next section, we will focus on advances in RT treatment planning and ways to improve it.

2.3.3. Radiotherapy treatment planning and challenges

RT planning protocols involve the isotropic geometric expansion of tumors seen in post-contrast T1-weighted and T2-weighted FLAIR images. This clinical target volume (CTV) includes the gross tumor volume (GTV = resection cavity plus residual CEL seen in MRI), plus a margin for subclinical disease spread, typically an isotropic 1-2 cm expansion of the GTV. A third volume, the planning target volume (PTV) is the dose-delivery volume, in which a margin (~0.3-0.5 cm) is added to the CTV in order to account for uncertainties of treatment delivery and/or avoid critical brain structures. The goal is to make sure that the highest prescribed dose gets delivered uniformly across the entire CTV.

One of the biggest challenges of RT treatment is the dosimetry planning. A high dose (>60Gy) of RT is usually delivered to the entire CTV, because the majority of progression occurs locally within 1-2 cm of the original lesion (thus the isotropic 1-2 cm expansion in CTV planning). However, distant progression beyond the high RT dose area can be found in 10-37% of GBM patients [46]–[49]. This percentage is even higher when looking at hypo-fractionated RT, with up to 47% of patients progressing beyond the CTV area [50], [51]. This indicates that current approach to generate CTV can undertreat subclinical tumor cells.

On the other hand, radiation can cause irreversible damage to normal brain tissue due to an increase in toxicity. Brain irradiation may cause long term cognitive impairment, regardless of OS and PFS [52]. While RT is a relatively safe treatment, many studies have shown that changes to RT planning to improve RT treatment volume, target aggressive tumors, and reduce healthy brain tissue irradiation can have a positive effect on OS, PFS, and quality of life. An example of such method is adaptive RT, where during RT the subsequent RT treatment course can be modified based on changes in tumor size and location as seen on MRI. This method can effectively reduce the totally irradiated volume and was shown to increase OS compared to SOC (28.9 months vs. 12.1 months) [53]. Changes in the CTV to intentionally exclude some T2-FLAIR hyperintense areas as part of GTV showed no significant change in either OS or PFS while

reducing brain irradiation [54]. Furthermore, tumor progression has been shown to spread differently depending on the patient's initial diagnosis [55], [56]. Thus, there is a need for a better way to define target volume for RT treatment for patients with GBM that is highly personalized, to target the true extent of infiltrating tumor while minimizing treatment-related toxicity leading to cognitive decline by sparing healthy brain tissue. This will be discussed further in Chapter 7.

2.3.4. Monitoring for tumor progression for newly-diagnosed GBM patients

All newly-diagnosed GBM (and the majority of lower grade) patients will eventually progress despite vigorous treatment. It is therefore important to monitor for early signs of progression to better manage and provide additional treatment for these patients. Following the course of SOC treatment, follow-up MRI scans are routinely prescribed, at the frequency depending on the initial diagnosis, with higher-grade tumors requiring more frequent visits as shown in figure 2.4. For GBM patients, MRI is usually performed every ~2 months.

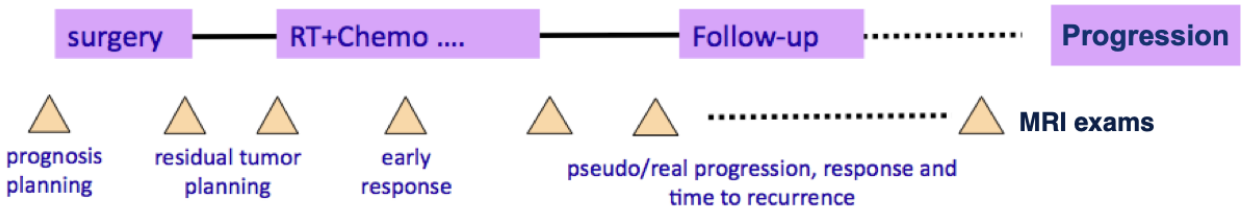

Figure 2.4. Schema for monitoring of tumor progression in high-grade gliomas
Following SOC treatment, follow-up MRI are performed every 2 months to monitor for tumor progression.

Chapter 3. Magnetic resonance imaging in brain tumors

3.1. Overview of MRI

3.1.1. Physics of MRI

MRI was developed to utilize the magnetic properties of many atoms such as ^{1}H or ^{13}C. Subatomic particles within an atom's nucleus, i.e. protons and neutrons, spins on their own axes. For a nucleus with an even number of protons and neutrons, these spins cancel out giving no net spin. However, in nucleus with an odd number of protons and neutrons like ^{1}H or ^{13}C, there is a non-zero net overall spin I, called nuclear spin or precession, which can be excited in the presence of an external strong magnetic field, giving these atoms magnetic properties [57], [58]. These atoms are also abundant in organic specimens, ^{1}H in water and fat and ^{13}C in organic molecules, making them suitable candidates to study changes to the human body and diseases such as cancers. This work will focus on ^{1}H MRI.

^{1}H nucleus possesses ½ spin and given the equation to calculate total spin state/orientation as $2I + 1$, it should have two spin states with different energy levels when an external magnetic field B_0, usually at 1.5 or 3 Tesla, is applied in the z-direction: a low energy state when the nucleus is parallel to the B_0 direction, and a high energy state when the nucleus is anti-parallel to the B_0 direction. After some time, equilibrium will be reached, and many nuclei will switch state to align in the direction of B_0 since a lower energy state is typically preferred. This created a net magnetization M_z in the z-direction, or the same direction of B_0, called longitudinal magnetization. Overall, the atoms will spin along the same direction of B_0 at the frequency given by Larmor's equation:

$$f_0 = \gamma B_0 \tag{3.1}$$

where f_0 is the precession frequency, B_0 is the strength of the magnetic field, and γ is the gyromagnetic ratio, which is a fixed constant for each type of atom. In the case for ^{1}H MRI, γ = 42.58 MHz/T. However, at this state, the system is at equilibrium and no signals are detected.

In order to generate signals, a 2nd smaller radiofrequency magnetic field B_1 is applied perpendicular to B_0 at the same Larmor frequency over time t. This excites the protons causing them to synchronize and rotate at frequency $f_1 = \gamma B_1$ following Larmor's equation, which then changes the net magnetization M_0 to spiral away from the z-direction parallel to B_0 onto the x-y plane. This is called transverse magnetization, denoted as M_{xy}. Flip angle is the angle in which M_0 is displaced from M_z as the result of M_{xy}, and common flip angles used are 90° and 180°, which are dependent on the strength of B_1 as well as time t. When B_1 is removed, the ^{1}H nucleus dephase and return to the original orientation along the z-axis parallel to B_0 again to be at the lower energy state. This looks like a spiraling of the magnetic vector along the z-axis (Figure 3.1-d), which in turn generate a damped sinusoidal electric signal, called the free induction decay (FID) that can be detect by the receiver coil of the MRI machine. That is the basis of how MRI signals are generated. This is summarized in Figure 3.1.

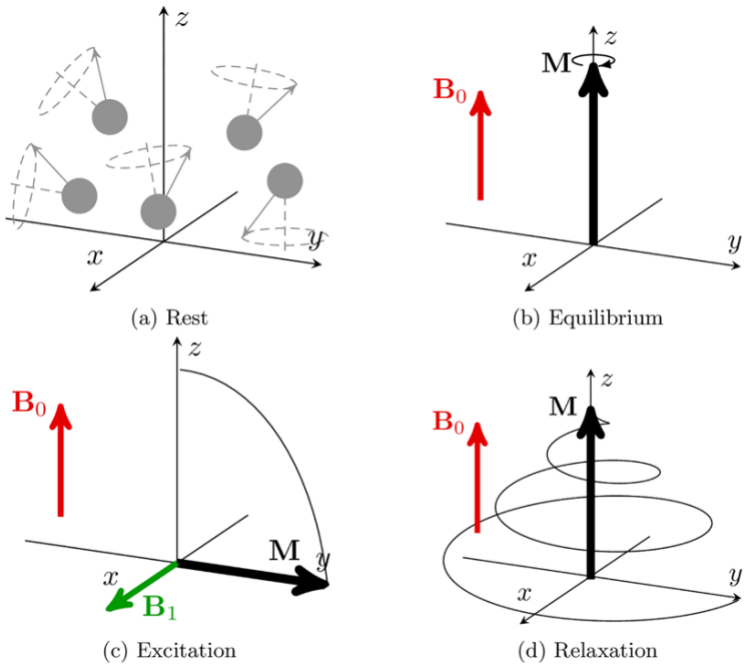

Figure 3.1. Schema of basic MRI principles
Figure shows the net magnetization of atoms at rest (a), during equilibrium in the presence of magnetic field B_0 (b), during excitation stage when introducing perpendicular magnetic field B_1 (c), and during relaxation after B_1 was turned off. Figure adapted from Thomas, 2021 [59]

3.1.2. T1 and T2 relaxation

MRI contrast is mostly driven by signal relaxation. Immediately following excitation of atoms by B_1 pulse, both longitudinal and traverse relaxation begin to happen to restore net magnetization to equilibrium state M_0. On one hand, at this stage, M_z is at 0 and is being repolarized by B_0 to reach equilibrium and return to its maximum level. This happens through a process called "spin-lattice relaxation" as energy is released back to the lattice. The time it takes to achieve 63% of M_z is called T1 relaxation time, or the signal recovery time. This relationship is depicted as:

$$M_z(t) = M_0(1 - e^{-t/T1}) \tag{3.2}$$

On the other hand, M_{xy} is at its peak and wants to return to 0 at equilibrium. This process is driven mainly by spin-spin interaction of the atoms. The time it takes for signal to decay to 37% of its original signal is called T2 relaxation time, or the signal persistence time. The decay follows:

$$M_{xy}(t) = M_0(e^{-t/T1}) \tag{3.3}$$

T1 and T2 are intrinsic properties of given tissue types based on their composition. External factors that result in magnetic field inhomogeneity can also further reduce the decay time. When this happens, the relaxation time is referred to as T2*. A 180° refocusing pulse is typically used to remove these magnetic field inhomogeneities, resulting in the formation of a spin echo that when reconstructed produces a T2-weighted image. Overall, it is generally expected that T1 > T2 > T2*. Figure *3.2* shows relaxation times for structures within the head.

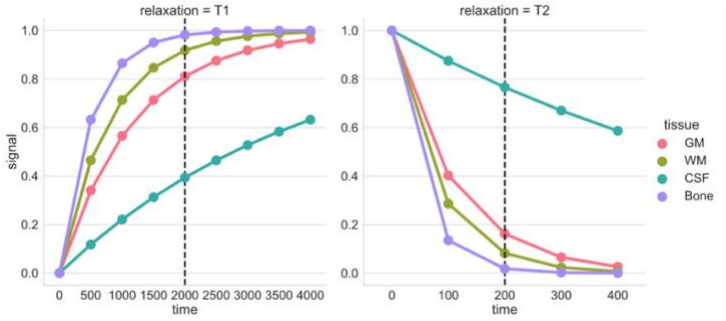

Figure 3.2. Examples of tissues and their relation time.

Figure adapted from Bhagwat, 2021 [60]

Factors that affect T1 and T2 relaxation time include tissue pathology, such as fluid. content, inflammation, cell density, necrosis, cell proliferation, etc., making MRI an important instrument in studying brain tumors. In practice, T1 and T2 are rarely directly measured, but instead we use them by varying pulse sequence timing and intensity to generate different type of images. Example pulse sequences include spin echo, gradient echo, inversion recovery, and echo-planar imaging, which are applied in various combinations often with specialized radio-frequency pulses to create diffusion-weighted, perfusion-weighted, functional, and susceptibility-weighted images, among a host of others. In the next few sections, I will describe the principles of three different types of MRI images used in my research: conventional anatomical images, diffusion-weighted images, and magnetic resonance spectroscopy imaging.

3.2. Conventional anatomical MR imaging

3.2.1. Principles of anatomical MRI to generate image contrast

There is no formal definition of "conventional anatomical MRI", however it can be understood as images generated from one of three basic pulse sequences: spin echo (SE), inversion recovery (IR), or gradient echo (GE) (Figure *3.3*). The most studied anatomical image contrasts in patients with brain tumors are T1-weighted from a GE sequence, T2-weighted from SE sequence, and T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) from IR sequence, and I will be focusing on these.

In SE sequences, a 90^{o} RF pulse (described in previous section) is followed by a 180^{o} RF pulse at time TE/2 to produce an echo. The goal of SE is to measure the T2 relaxation time instead of T2* by adding a 180 refocusing pulse to negate any dephasing due to inhomogeneities, thus disregarding the effect of any external factors that can affect the signal. An example is shown in figure 3.3. The strength of MRI signal thus equates to:

$$S = K(1 - e^{-\frac{TR}{T_1}})(e^{-\frac{TE}{T_2}}) \quad (3.4)$$

where K is a constant proportional to the proton density of the tissue; TR is the repetition time, or the amount of time between consecutive 90º pulse sequences applied; and TE is the echo time, or the time between the application of the RF pulse and the reception of the induced echo signal.

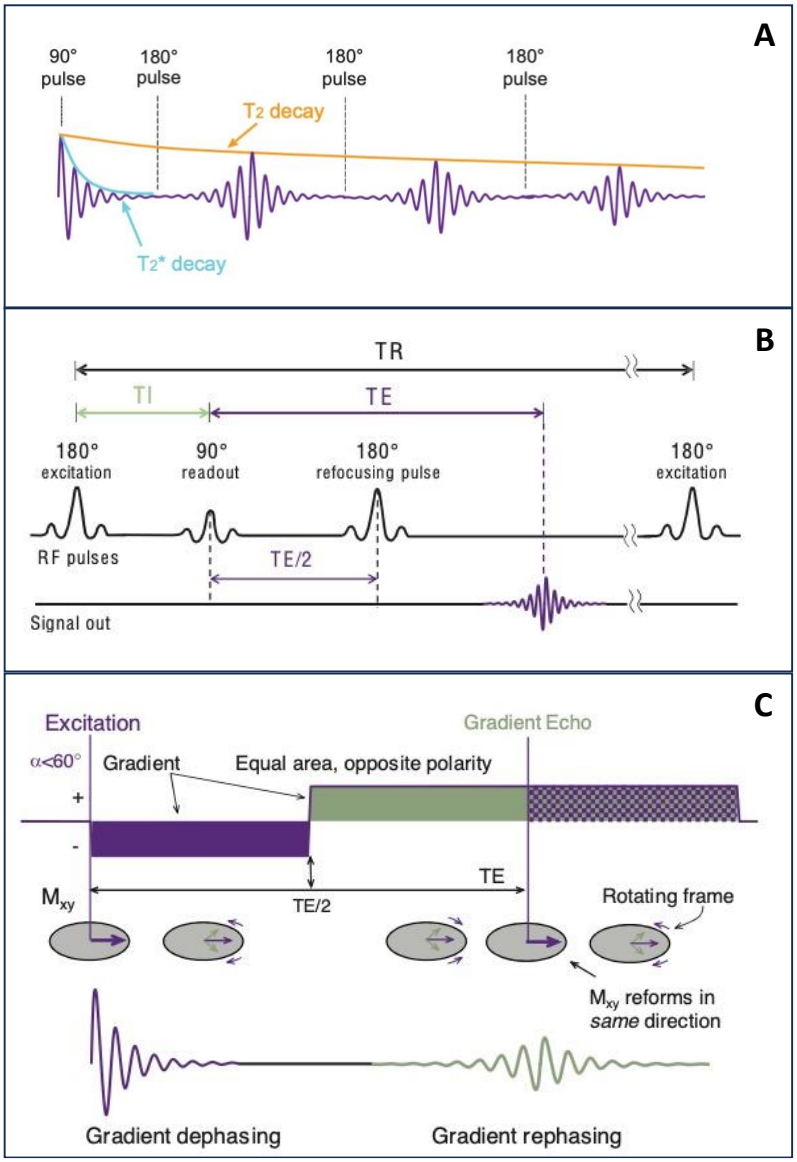

Figure 3.3. Example diagrams of common pulse sequences
A) Spin echo: T2 decay can be determined from the sequence, canceling out the effect of external factors; B) Inversed recovery: the initial 180º inverts magnetic field in the longitudinally, doubling the amplitude of longitudinal recovery, thus emphasizing T1 relaxation time; C) Gradient echo: echo is formed by magnetic field gradient instead of the 180º RF pulse, and dephasing and rephasing of transverse magnetic field is done by applying equal and opposite polarity gradients. Figure adapted from Mahesh, 2013 [61]

By changing TE and TR, the image contrast can be T1-weighted, T2-weighted, or proton density weighted. To generate T1-weighted image, TR is kept short (~ 400-600 ms) to accentuate the T1 difference between the tissues, while TE is also kept short (~5-30 ms) to minimize T2 decay. This causes CSF to appear dark and tissues to appear brighter. Inversely, to generate T2-weighted image, TR (~2000-4000 ms) and TE (~60-150ms) are relatively long, capturing more T2 decay of tissues. Because of this, T2-weighted image generally have higher tissue contrast compared to either T1-weighted image or proton density weighted image. CSF will appear bright and tissues will appear darker in T2-weighted image [61]. A visualization of how varying TE and TR in T1 and T2-weighted images can affect signal intensity and tissue contrast is shown in figure 3.4. One way to enhance the contrast of abnormal tissues is to add a gadolinium-based (Gad) contrast agent while performing T1-weighted imaging to increase contrast of tumors, abscesses, inflammation, etc. Typically, Gad cannot cross the blood-brain barrier (BBB), except when BBB is compromised such as in the case of tumors. When accumulated, Gad shortens both T1 and T2 relaxation for these tissues, causing them to appear brighter in T1-weighted images [62].

Another commonly utilized anatomical MRI for brain tumors is a T2-weighted FLAIR sequence. In this sequence, an initial 180^{o} RF pulse is first applied to invert the signal. After a time TI, a 90^{o} RF pulse is applied, followed by another 180^{o} pulse at time TE/2 to generate a spin echo. TI is chosen such that signal from CSF and other water-rich tissues is nulled. This allows a clearer separation of edema or tumor from the ventricles, the boundary of which is difficult to visualize on T2-weighted images. As a result, CSF appears dark on T2-FLAIR images, but fluid-filled tissue, such as tumor, still appears bright as shown in figure 3.4-C. Additionally, T1-weighted FLAIR can also be obtained, and is normally good for distinguishing fat from tissues in the body.

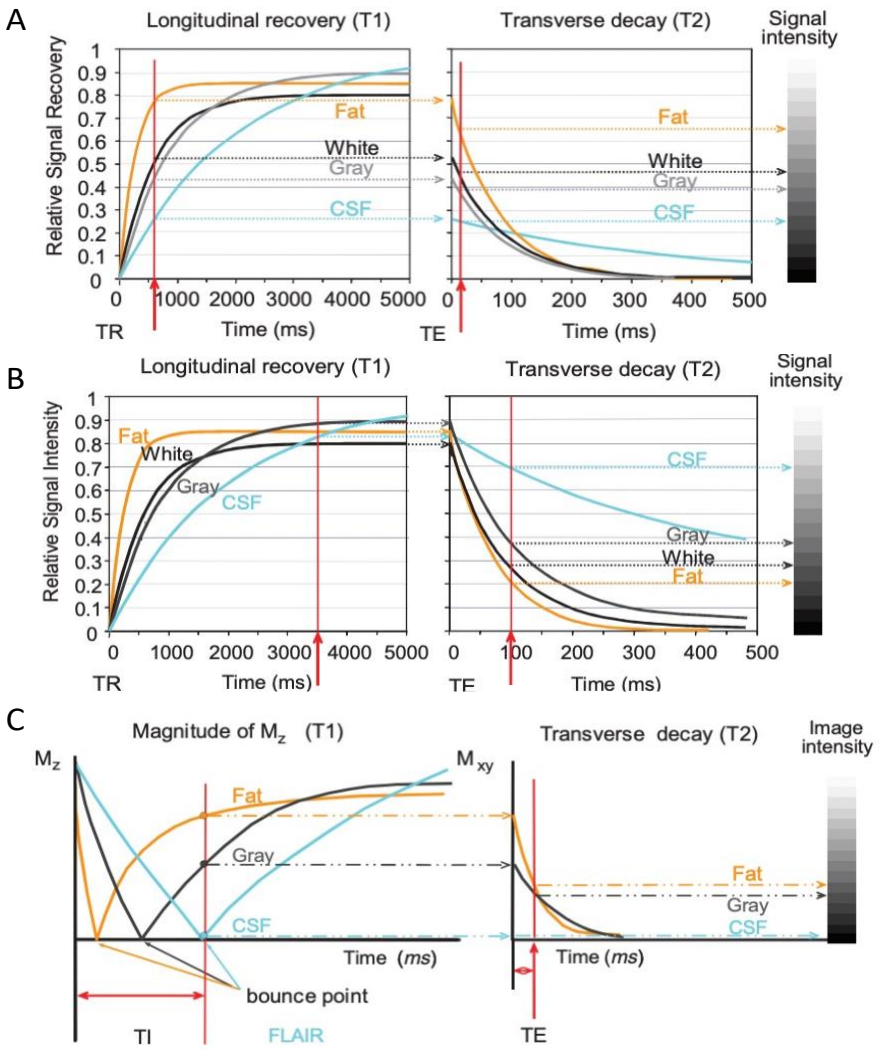

Figure 3.4. TI, TR, and TE affect signal intensity and tissue contrast
A) Short TR and short TE generate T1-weighted images; B) Long TR and long TE generate T2-weighted images; C) Inversion recovery time TI helps to null CSF signal. Figure adapted from Mahesh, 2013 [61]

Besides those mentioned previously, there are other ways to help amplify contrasts, such as fat suppression or post-acquisition processing, etc. All these techniques help in visualizing anatomic structures and diseases. Section 3.2.3 will discuss how anatomical MRI can aid in diagnosis, management, and treatment of glioma.

3.2.2. Signal localization to generate MR images

To determine and map the location of each signal to generate an image, varying linear magnetic field gradients need to be applied in all three directions. In theory, varying gradient causes proton to precess at different frequency depending on their spatial location along the gradient. This in turn will generate position-based variation of proton precessional frequency. Signal localization is usually done in three gradient application steps that cannot be done simultaneously: 1) frequency encode gradient; 2) phase encode gradient; 3) either a second phase encode (for 3D) or slice selecting gradient (for 2D). The slice selecting gradient is usually applied along the z-axis, resulting in a linear difference in resonance frequency across that axis. Each small range of frequencies correspond to a "slice", and protons within that slice can be excited separately. This helps determine the location of each signal in the z direction using the equation:

$$F_s = \gamma(B_o + zG_{ss}) = f_o + \gamma z G_{ss} \tag{3.5}$$

where F_s is the slice frequency at position z, f_o is the corresponding Larmor frequency and G_{ss} is the slice selecting gradient. Similarly, to map spatial location in the x-axis, a linear frequency-encoding gradient is applied along the x direction. This creates variation in resonance frequency along the x-axis following the equation:

$$f_x = f_o + \gamma x G_{fe} \tag{3.6}$$

where f_o is the main Larmor frequency and $\gamma x G_{fe}$ corresponds to the frequency offset due to gradient with respect to the location of the signal along the x-axis. Phase encoding works similar to frequency encoding, except gradient is applied in the y-axis and should be applied and then turned off after a time. After phase encoding gradient is turned off, this allows a phase shift between signals along the y-direction while keeping their frequency the same. Thus, determining the phase of spin vectors can reveal the location of signal in the y-direction. These gradients are typically applied multiple times to acquire multiple lines of "kspace" until enough spatial

frequencies are acquired to be able to resolve the object of interest at a high enough resolution (typically 256 phase encoded lines in y, each containing 256 encoded frequencies in x for the head). The final step is to perform Fourier transform to convert signal from the frequency domain or "kspace" into spatial domain or image space to generate an image.

3.2.3. Anatomical MRI in glioma diagnosis and management

Conventional anatomical MRI, such as T1-weighted (T1), T1-weighted post-Gd (T1C), T2-weighted (T2), T2-weighted FLAIR (FLAIR), are routinely acquired and used in clinical setting for both the diagnosis and management of patients with glioma. The advantages of these images include their high resolution, the ability of visualize all 3 planes without interference of bone, and their ease of use and interpretability. They can be used by radiologists to differential diagnose and visualize tumors: aggressive tumors appear brighter in T1C image due to the breakdown of blood-brain barrier, and the regions surrounding tumors (peritumoral), which is consisted of both infiltrative tumors and vasogenic edema, appear bright in T2 and FLAIR images due to an increase in water content [63]. Radiologist defined VASARI features derived from T1C and FLAIR images have been shown to correlate well with mutation status [64], change in cells metabolism [65], tumor subtype [66], and overall survival [66]. Anatomical MRI is also used by neurosurgeons to guide resection of tumors during surgery, radiation oncologists for RT planning, as well as by radiologists and neurooncologists to assess (but not early predict) both treatment response and tumor progression.

While useful, conventional anatomical MRI can be limited in its ability to describe structural, pathology, and metabolic changes of tumor cells. It does not have the ability to provide early/accurate diagnosis, to predict treatment response and outcome, or to thoroughly monitor tumors. For those purposes, additional physiologic and metabolic MR techniques should be included.

3.3. Diffusion-weighted Imaging (DWI)

3.3.1. Principles of Diffusion-weighted Imaging

DWI is another type of MRI sequence that generates signal contrast based on the difference in Brownian motion of tissues, i.e., the random movement and displacement of water molecules within and outside of brain tissue. This isotropic displacement is expressed in the equation:

$$s = (2DT_d)^{1/2} \tag{3.7}$$

where s is the displacement, D is the diffusion coefficient, and T_d is the diffusion time. D is an intrinsic property of molecules in a given environment. For example, free water in biological systems have a diffusion coefficient of about 3 $\mu m^2/ms$ [67]. However, diffusion of water molecules within tissue are hindered (such as water diffusion in extracellular spaces) and restricted (such as water diffusion within cells) depending on the size, shape, and properties of the specific tissue [68]. In this case, the diffusion coefficient is replaced by a different measurement, the apparent diffusion coefficient (ADC), to differentiate between this and diffusion of free water molecules. This restriction of diffusion can also be either isotropic or anisotropic (which means more restriction in a certain direction). For example, diffusion anisotropy happens within asymmetric tissues such as nerve cells or muscle bundles. More importantly, certain diseases can affect water diffusivity. Such is the case of brain tumors, where higher cell density can hinder the rate of water diffusion within the extracellular spaces (ECS), while the presence of edema can increase diffusion. Thus, modeling water diffusion can assist in disease monitoring. Fortunately, this can be done with MRI via a DWI sequence.

In a basic DWI sequence, a pulsed gradient spin echo is applied. This is similar to the T2-weighted spin echo sequence, with the addition of a pair of equal and opposite pulse gradients, called diffusion-sensitizing gradients, before and after the 180° echo pulse. The first pulse changes the phase shifts of each proton, while the second pulse acts to reverse this. However,

where there is water diffusion, this rephasing cannot be completed and MR signal is reduced as a result [69], [70]. The DWI sequence schema is depicted in figure 3.5.

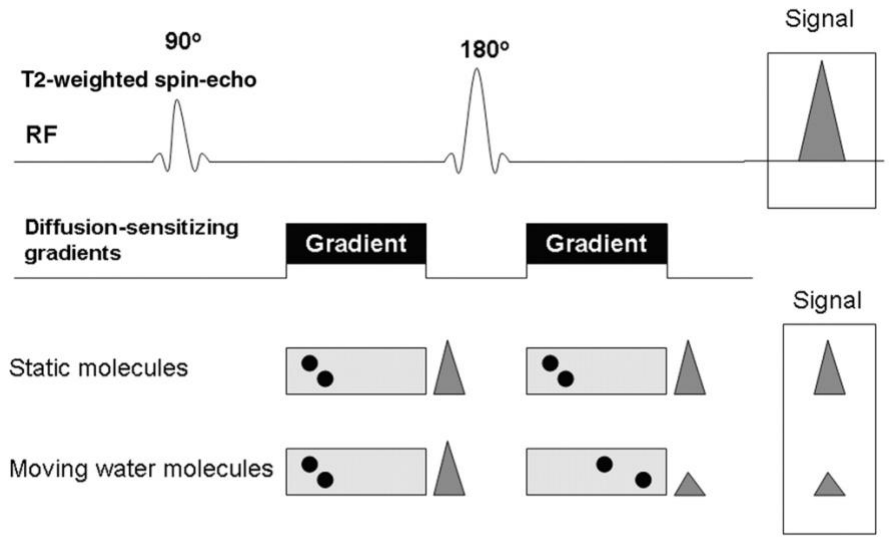

Figure 3.5. DWI sequence by applying pulsed gradient spin echo
MR signal is lost as the results of moving water molecules. Figure adapted from Koh, 2007 [69]

The sensitivity of DWI signals to water diffusion depends on the strength of gradient G, the time duration of each gradient application δ, and the time duration between the first and the second gradient Δ. This diffusion sensitivity can be altered by changing the b-value, as described in the equation:

$$b = \gamma^2 G^2 \delta^2 (\Delta - \delta/3) \quad (3.8)$$

As expected, higher b-value (> 500 s/mm^2) is usually required to detect diffusion over small distance, such is the case for many biological applications. However, higher b-value also decreases SNR causing the images to be uninterpretable. In brain imaging, DWI is usually done using at least two b-value, one at b = 0 /mm^2, and another b-value, usually chosen between 500-2000 s/mm^2, for the best quality images. Given the b value, the diffusion-weighted signal can be calculated using the equation:

$$S_b = S_0 e^{-bD} \quad (3.9)$$

where S_b is the signal at b = 0 s/mm^2, S_b is the DWI signal at a different b-value b, and D is the diffusion coefficient of free water. In biological setting, this becomes:

$$S_b = S_0 e^{-bADC} \tag{3.10}$$

As described previously in section 3.2.2, to generate DWI, diffusion-sensitizing gradients also need to be applied in all 3 directions. DWI signal is calculated separately for all 3 directions, and equation 3.9 becomes:

$$S_i = S_0 e^{-bADC_i} \tag{3.11}$$

Where S_i and ADC_i represents the DWI signal and the apparent diffusion coefficient for the ith-direction, respectively. For many biological applications, diffusion happens isotopically. For example, diffusion of some types of tumors can be isotropic assuming tumors cells grow randomly without a specific path. In these cases, we can combine the 3 directional signals to generate one single diffusion-weighted image, also known as the isotropic DWI, by computing their geometric mean:

$$S_{DWI} = \sqrt[3]{S_x S_y S_z} \tag{3.12}$$

In general, voxels with higher diffusion will appear darker. Confounding hyperintense signal on isotropic DWI can also be attributed to a dominant T2 effect, called the T2 shine-through effect. To mitigate the T2 shine-through effect, the ADC map is used instead. This map is generated using the equation, where areas of increased diffusion are now hyper- instead of hypo-intense:

$$ADC = -\frac{1}{b}\ln\left(\frac{S_{DWI}}{S_0}\right) \tag{3.13}$$

While the isotropic diffusion assumption works for many applications, most biological processes possess some level anisotropic diffusion. One prominent example is the brain's white matter, where diffusion mostly occurs in the direction along white matter tracts. In this case, diffusion cannot be adequately expressed as a single number, but instead is represented as a [3x3] matrix called the diffusion tensor:

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}; D_{isotropic} = \begin{bmatrix} D & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & D \end{bmatrix}$$

The isotropic diffusion coefficient can also be expressed in DTI format as shown above. Because $D_{yx} = D_{xy}$, $D_{xz} = D_{zx}$, $D_{yz} = D_{zy}$, the DTI matrix should have 6 unique coefficients. Thus, to estimate anisotropic diffusion, at least 6 gradient directions need to be applied. DTI can be generalized by using a different coordinate system, the diffusion ellipsoid, which can be represented by the 3 unit vectors called eigenvectors $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ with corresponding length called eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$. The eigenvectors are the three principle axes of the ellipsoid, and the eigenvalues are the diffusivity along those axes. This is shown in figure 3.6.

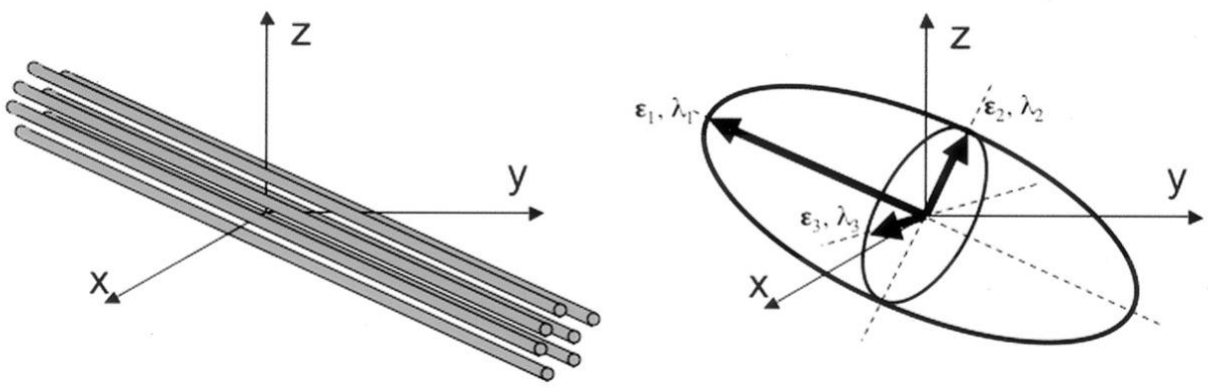

Figure 3.6. DTI represented by eigenvectors and eigenvalues
Figure adapted from Jellison, 2004 [71]

All eigenvalues and eigenvectors can be calculated from MRI signals. From those, we can estimate various diffusivity measurement and generate interpretable images: mean diffusivity (MD), which is synonymous to ADC, represents the average diffusivity independent of direction; fractional anisotropy (FA) represents the degree of directional dependency of diffusivity, or how much diffusivity is attributed in a specific direction. The equation used to map these two properties are:

$$MD = ADC = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (3.14)$$

$$FA = \sqrt{\frac{1}{2}\frac{(\lambda_1-\lambda_2)^2+(\lambda_1-\lambda_3)^2+(\lambda_2-\lambda_3)^2}{{\lambda_1}^2+{\lambda_2}^2+{\lambda_3}^2}} \quad (3.15)$$

In the ADC map, as described above, voxels with higher diffusivity will appear brighter, and voxels with lower diffusivity will appear darker. In the FA map, brighter areas are more anisotropic than darker areas. These maps have been used in many studies that evaluate patients with brain tumors [72]. Aggressive higher-grade tumors with high cellularity generally will have lower diffusivity and anisotropy, thus will appear darker in ADC map and FA map, while lower grade, non-enhancing tumors tend to have more fluid and thus appear bright on ADC. These relationships will be explored further in the next section.

3.3.2. Diffusion-weighted MRI in glioma diagnosis and management

DWI and its associated markers like ADC and FA can identify movement of water molecules within tissues. It has been shown that biological processes that increase the amount of free water in ECS result in an increase in ADC values, and inversely, processes that increase cell density and hinder water movement in ECS result in a decrease in ADC [73]. In the case of glioma, subclinical tumor cells can disrupt and infiltrate white matter tracts, resulting in an increase in edema and a decrease in directionality along white matter tracts. Both the ADC and FA maps can help identifying the presence of edema and tumor infiltration, which in turn can help map tumor progression and response to therapy. Particularly, a decrease in ADC and an increase in FA has been shown to correlate with an increase in proliferation (measured by KI-67) and cellularity in patients with glioblastoma [74]–[76]. Because of this, ADC and FA can be a useful marker in tumor diagnosis, specifically in determining tumor grade, as well as prognosis and predicting PFS and OS [74], [77]–[79]. Furthermore, ADC has also demonstrated promise in the classification of IDH mutations, 1p19q co-deletion, and MGMT promoter methylation [80]–[82].

Because ADC and FA can identify subclinical tumor cells before they appear abnormal on anatomical imaging, they have been used in identifying brain areas at risk for later progression

along with anatomical and metabolic MRI [9], demonstrating their promise in refining surgical and RT treatment planning. Advances in DTI and fibertracking also allows the visualization white matter tracts [83]. Because gliomas tend to grow along white matter tracts, DTI has allowed for identification of specific pathways of tumor growth and progression [84]. This has motivated our development of predictive tools that can be used to improve and personalize surgery and RT treatment.

3.4. Magnetic Resonance Spectroscopy Imaging (MRSI)

3.4.1. Principles of ^{1}H MRSI

^{1}H (protons) are present not only in water, but also other biological molecules. In previous sections, we discussed anatomical MRI, where it is not important to distinguish the source of protons molecules. In this section, we will discuss MRSI, an imaging technique that can provide a spectrum of brain metabolites spatially by evaluating signals from each metabolite generated due to their chemical shift from a standard frequency. This is done by first suppressing the overwhelming signals from both water and fat. Furthermore, voxel size for MRSI is typically much larger than anatomical MRI in order to accumulate enough SNR since the metabolites of interest are at a much lower concentration than water. MRSI has been used extensively in studying cellular metabolism, and in detecting changes in metabolism due to the presence of tumor. This will be discussed later in more detail.

In previous section, it was assumed that all protons' nuclei behave the same way when magnetic field B_0 is applied. However, in reality, protons from different molecules behave slightly differently based on their interactions with neighboring protons and electrons. This is due to electrons within the molecules generating a secondary magnetic field B_{sh}, shielding the protons from B_0. Effectively, the total magnetic field experience by the proton's nucleus is $B_{eff} = B_0 - B_{sh}$. The shielding factor σ can be expressed as:

$$\sigma = \frac{B_{sh}}{B_0} \tag{3.14}$$

By definition, σ also relates the chemical shift δ of the molecule. Because B_{sh} is many orders of magnitude less than B_0, chemical shift is typically expressed in ppm. In practice, however, the chemical shift is determined by using a reference compound, such as Tetramethylsilane (TMS) which have a chemical shift of 0:

$$\delta = \frac{f - f_{ref}}{f_{ref}} \tag{3.14}$$

3.4.2. ^{1}H MRSI acquisition

There are two main techniques used in human MR spectroscopy: single voxel spectroscopy (SVS) and multi-voxel chemical shift imaging (CSI) or MRSI. As the name implied, CSI/MRSI covers a much larger area and can be acquired at higher resolution, thus more suitable in studying brain tumors and their progression over time. CSI uses phase encoding gradients to further segment the excited volume into smaller voxels. Point Resolved Spectroscopy (PRESS) sequence is the most commonly used method for MRSI. It is done by applying three RF-pulses (90^o, 180^o, 180^o) in conjunction with the three orthogonal gradients (in z, y, x direction). CSI and PRESS allows for 3D localization of voxels, from which a free induction decay (FID) of various spectral frequencies can be acquired.

Two important steps in acquiring MRSI include water and fat suppression. The concentration of water is about 10,000 times that of other metabolites, and its signal can be too dominant to view metabolites of interest in a typical MRS spectrum. Water suppression is typically performed via a process called Chemical Shift Selective saturation (CHESS), where a 90^o pulse is applied at exactly water frequency to shift water into the xy-plane, where it is de-phased completely by an additional strong spoiler gradient. Similarly, there is a high concentration of lipid near the brain (in the scalp and bone marrow near the skull) that can either fold into the brain if the prescribed excitation volume is not large enough to encode this region, and/or result in ringing

artifacts affecting the overall signal (even in area far away from the skull). To suppress lipid artifacts, very selective saturation pulses (VSS) can be used, reducing signals coming from area near the skull. An unwanted side effect is that some brain voxels near VSS bands can have lower signals as the results. Figure 3.7 shows an example of water/lipid suppression, as well as an example of saturation bands place manually around the scalp when acquiring MRSI data.

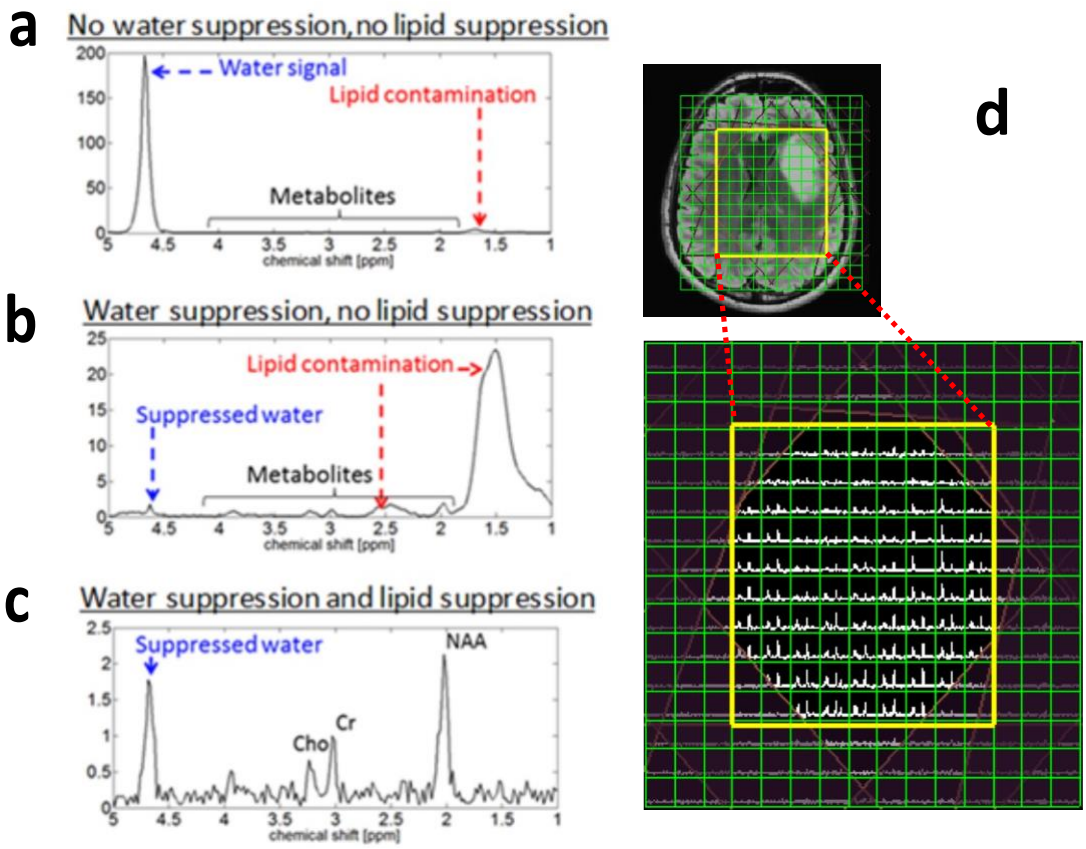

Figure 3.7. Water and lipid suppression allow better spectral resolution for other metabolites
a) MR spectrum with no water suppression and no lipid suppression; b) MR spectrum with water suppression but no lipid suppression; c) MR spectrum with both water and lipid suppressed; d) OVS band for lipid suppression. Figure a,b,c adapted from Kok, 2012 [85]

Similar to other MRI sequences, the choice of echo time (TE) is also an important factor in brain MRSI. Short TE (20-40ms) allows visualization of extra peaks such as myoinositol (MI), glutamine/glutamate (Glx), amino acids, etc. Using a long TE of 144ms results in the lactate peak to invert due to J-coupling, allowing differentiation of lactate and lipid when spectral editing is

applied [6], [7]. It also results in a smoother baseline, facilitating the quantification of other high SNR peaks such as NAA and choline [86].

3.4.3. MRS quantification and metabolite maps

While MRSI signals are acquired in the time-domain as FIDs for every phase encoded voxel in kspace, quantification of MRSI is usually done in frequency domain, which can be derived from time-domain signals by applying a Fourier Transform in the 4^{th} dimension after inverse Fourier transforming the three spatial dimensions [87]. Two example spectra are shown in figure 3.8, each collected using a different echo time. Frequency domain signals are complex numbers, which include a real and an imaginary component, but quantification of metabolites is usually performed on the real component. It has been found that the integrated amplitude (or the area under the curve) of the signal's real component for a given peak is directly proportional to the concentration of nuclei contributing to that peak (not concentration of metabolite). Thus, measuring this integration, either by calculating the peak's area under the curve or by modeling (e.g. LC model), can provide information about the peak's relative concentration [88].

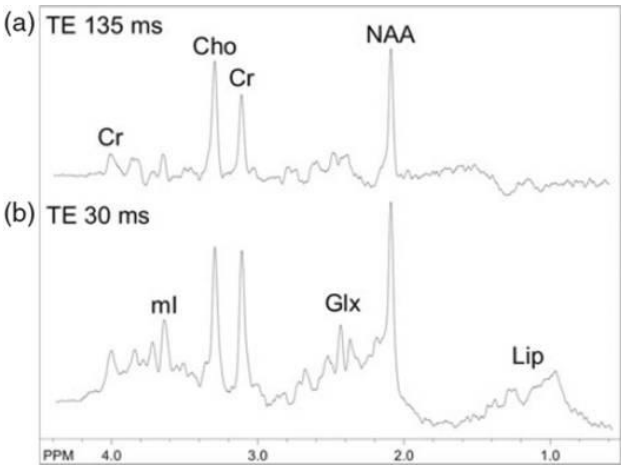

Figure 3.8. Sample MRSI spectra in frequency domain
Sample MRSI spectra in frequency domain and metabolites seen in a) long TE acquisition, and b) short TE acquisition.

Because of the constraint described previously in relating concentration of metabolites to the absolute peak characteristics, many studies choose to instead report and utilize the relative ratio between metabolites, which usually achieve higher statistical significance and offer additional benefit such as reducing sample size requirement [89]. Three commonly used ratios are the Cre/NAA, Cho/NAA, and Cho/Cre. Relative ratios are particularly useful in studying diseases when each metabolite is expected to shift in the opposite direction, e.g. Cho/NAA is used to study brain tumors since the presence of tumor means ↓NAA and ↑Cho simultaneously, which means ↑Cho/NAA. The choline-to-NAA index (CNI), which is an iterative residual of the deviation from choline to NAA in normal brain, has also been extensively shown to be a more robust metric from long TE spectra [7], [90], [91]. However, true concentration of metabolites is very difficult to measure correctly due to J-coupling, metabolites containing more than one group of ^{1}H nuclei, and overlapping peaks. This gave rise to a question: for diagnosis and monitoring of brain tumors, would using the full spectra providing more information than using peaks heights, ratios, or indices alone?

Each quantified metabolite can also be mapped spatially and displayed like an image (similar to anatomical MRI). This is done by incorporating spatial information collected during acquisition. Maps can be generated for metabolite peak heights, area under the curve, ratios, and indices.

3.4.4. ^{1}H MRSI in diagnosis and management of glioma

In long TE ^{1}H MRSI of the brain, the following peaks can be detected:

- *N-Acetylaspartate (NAA) at 2.01 ppm*: this includes mostly N-Acetylaspartate and a small trace of N-acetylaspartylglutamate (NAAG) near 2.04 ppm. NAA is an amino acid commonly found in the brain and neurons. It is a classic marker of neuron viability and healthy brain tissue. NAA level is lower in diseases that cause neuronal death or

dysfunction [92], such as in the case of brain tumors. As NAA originates in the brain, no NAA is found in metastases [93].

- *Choline (Cho) at 3.20 ppm*: this peak includes glycerophosphocholine (GPC), Phosphocholine (PC), and free choline. These metabolites are important for cell membrain systhesis and break-down, thus is often elevated during rapid proliferation of tumor cells. Cho peak is especially high in malignancy.
- *Creatine (Cre) at 3.03 ppm*: this peak includes both creatine and phosphocreatine. Both of these compounds involve the creatine kinase reaction during ATP generation, and thus play a large role in energy metabolism [86]. It is generated in the liver and migrates to the brain, so its level is relatively stable, and is often used as a reference metabolite, although Cre levels can fluctuate in certain types of glioma and with treatment [94].
- *Lactate (Lac) at 1.3 ppm*: this peak is usually inverted at TE = 144 ms. Lactate is usually not found in normal brain, but is detected in brain diseases that result in hypoxic conditions such as acute hypoxic [95] or glioma [96].
- *Lipid (Lip) at 0.9 ppm and 1.3 ppm*: this peak includes lipid from free fatty acids and tissue triglycerides. Lipid is associated with membrane break-down, and is often elevated in regions of hypoxia and necrosis [92], [97].

Other metabolites can also be found when MRSI is acquired at a lower TE, such as Glutamate/Glutamine (at 2.3 ppm) and Myo-inositol (at 3.6 ppm), as shown in figure 3.8b.

^{1}H MRSI can often compliment anatomical images in studying gliomas. It has been found that tumor usually grows beyond the hyperintense area of the T1-post-contrast and T2-weighted MR images [98], [99]. Furthermore, as discussed previously, hyper-intense regions on T2-weighted MR is also not specific to tumor. Because of this, anatomical MRI alone cannot detect the spread of glioma and diagnose the true heterogeneous nature of these lesions, and metabolic information from MRSI is needed. Overall, it has been found that more aggressive lesions usually

are associated with: 1) elevated Choline (Cho), with an increase in cell membrane turnover due to rapid growth of tumor tissues; and 2) lower N-acetylaspartate (NAA), signaling a decrease in neuronal viability due to the presence of aggressive tumors [92], [100]. The level of Creatine (Cre) can be elevated, associated with an increase in tissue energy metabolism, although Cre has also been shown to decrease in some studies [101]. The presence of a lactate peak, a marker for anaerobic metabolism, has also been observed in higher grade tumor and necrotic areas [92], [98], [102]. Cho/NAA >2 typically indicates metabolic abnormality in the growing lesions [103].

Many studies have demonstrated the benefit of using MRSI, particularly the Cho/NAA and Cho/Cre, in probing the underlying cellular metabolism associated with tumor growth [9], [104], [105] and poor survival [98], [106]; and more recently, identifying IDH and TERT promoter mutation status [107]–[109]. These studies show that 1H-MRS has the potential to be used in evaluating tumor heterogeneity. Furthermore, MRSI markers can also identify infiltrative tumor cells beyond the hyperintense T2 FLAIR [7], [9], [105], making it a good candidate to be used in RT treatment planning and predicting subsequent progression on anatomical imaging. The use of MRSI data, however, is still limited to metabolite peak area, height, ratios, and indices, though a few recent studies have found that using the entire MRSI spectrum is possible and can lead to better performance [110]. One disadvantage of using MRSI is the low resolution, but recent studies have also attempted to improve this with neural networks [111].

Chapter 4. Artificial intelligence and its application in brain tumor research

4.1. Introduction to artificial intelligence

Artificial intelligence (AI) refers to computer algorithms developed and operated by humans that can perform difficult tasks requiring intelligence [112]. AI research has been a vital part of modern medicine in the last decade, and radiology has been one of the earliest fields to adapt cutting edge AI models into its workflow. This is in part due to the remarkable advances in image recognition models, and an abundance amount of training data: unlike many other branches of medicine, the majority of data collected in radiology are images or other type of digital data, which can easily be fed into an AI algorithm [112]. The primary goal of AI in healthcare is to improve patient outcome and reduce costs by assisting healthcare providers in delivering more accurate and timely diagnosis and treatment. It is important to also understand that AI tools exists to assist physicians whenever needed, and never to replace the important work of physicians and radiologists [113].

Machine learning (ML), including its subset deep learning (DL) is a major branch of AI that can make complex decisions and perform tasks by learning from available data. The 4 main types of ML include: 1) supervised learning, where models can learn to map outputs from inputs, trained using labeled datasets; 2) unsupervised learning, where models can learn to discover patterns from a set of unlabeled data; 3) semi-supervised learning, where models learn from a small subset of labeled data; and 4) reinforcement learning, where a model acts as an agent that can make decision based on its previous experience to maximize the total rewards [114]. As a rule of thumb, most ML models benefit from having an abundance of high-quality data. In medicine, these data can include patient's demographic information, electronic medical/health records, images, lab results, genomics, and data from sensors and wearable devices [114]. From these data,

application of ML includes disease diagnosis, precision and personalized medicine, medical images analysis, predictive medicine, drug discovery, etc. [115].

4.2. Principles of supervised machine learning

This dissertation will focus on supervised learning methods for classification, regression, and image segmentation tasks. Given a dataset of n training samples in the form $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ where x_i is the input feature and y_i is its corresponding label, mathematically, a supervised learning model will learn a function f that can map output variables Y = $\{(y_1, y_2, \dots, y_n)\}$ from feature vectors X = $\{(x_1, x_2, \dots, x_n)\}$ following the equation:

$$Y = f(X) + \varepsilon \tag{4.1}$$

where ε is an error term with mean 0. In prediction task, while input X is always available, but Y cannot be obtained for all instances, we can get an estimate of Y by applying:

$$\hat{Y} = \hat{f}(X) \tag{4.2}$$

where $\hat{f}$ is a "blackbox" estimation of f and can be trained. Notice that the term ε is dropped from equation 4.2 because ε cannot be predicted using X, thus is called "irreducible error".

The goal of a supervised learning model, therefore, is to apply the most appropriate machine learning technique to provide the most accurate $\hat{f}$ so that $Y \approx \hat{f}(X)$. In addition, given a set of unseen test data X^* and Y^*, then $Y^* \approx \hat{f}(X^*)$. To find the best model $\hat{f}$ from a set of possible models $\hat{F}$, let's first define a loss for a given training sample (x_1, y_1) as $L(y_i, \hat{y}_i)$. The expected training loss of function $\hat{f}$, or the empirical risk of $\hat{f}$, is $R_{emp}(\hat{f})$, and can be estimated as

$$R_{emp}(\hat{f}) = \frac{1}{N}\sum_i L(y_i, \hat{y}_i) \tag{4.3}$$

This is sometimes called the objective function. The first principle of supervised ML is to choose $\hat{f}$ to minimize the empirical risk $R_{emp}(\hat{f})$. However, this can overfit to training data and cannot be generalized. To overcome this overfitting issue, we need to add a regularizer function $r(\hat{f})$ that

penalizes model complexity to simplify $\hat{f}$. This is called the structural risk minimization principle, and the full supervised ML optimization problem becomes:

$$\min L(\hat{f}) = \min R_{emp}(\hat{f}) + \lambda r(\hat{f}) \tag{4.4}$$

where λ is a set of hyperparameters that control bias-variance trade-off, and $L(\hat{f})$ is a combined loss function that both minimizes training error and penalizes models with high complexity. λ is typically learned using cross-validation methods [116]–[118].

In general, machine learning can be subdivided into traditional ML or DL, depending on the types of input features. With respect to medical imaging tasks, in traditional machine learning, there is a big focus on feature engineering, data pre-processing, and feature reduction methods to generate meaningful inputs to the models. On the other hand, in DL, because features are typically generated as part of the training, feature engineering is not as important. Historically, DL models typically require lots of data to generate a good performing model, and when there is an abundance of clean data, deep learning usually outperforms traditional ML. But recent advances in DL architectures and methods have allowed models to be trained accurately using much smaller datasets, making DL a more suitable choice for medical imaging tasks.

Some of the biggest obstacles to training ML models on medical imaging tasks are small and imbalanced datasets. For small datasets, data augmentation and transfer learning [119] are two commonly taken approaches. Data augmentation artificially generates training data via flipping, zooming, cropping, or by using Generative Adversarial Networks (GAN) [120]. Transfer learning helps simplify the training task by utilizing a pretrained network (of a different but similar task) before finetuning it to the actual task. For imbalanced datasets, we can choose to apply sample weighting [121], or resample data using sampling methods such as up-sampling, down-sampling, or Synthetic Minority Oversampling Technique (SMOTE) [122], etc.

Solving a ML problem involves the following steps: a) data preparation; b) selecting among different ML and DL models; c) selecting an appropriate loss function for the problem; d) fitting

and optimizing the model using cross-validation; and e) evaluating the model. Model evaluation is one of the most important aspects of AI research in healthcare that is currently not treated as such. Medical related tasks are typically very different than regular ML tasks, and we cannot always apply standard evaluation metrics for medical problems. For one, medical data is almost always highly imbalanced, typically with many negative samples and small number of positive samples. A good evaluation metric should not misinterpret when the data is imbalance and should be intuitive and specific to the task. Metrics for model evaluation can often be confused with the loss function. While sometimes they can be used interchangeably, in many cases they are different. A criterion for loss function is ease of model optimization, and that needs to be considered when designing an experiment. The next section will give an overview of several common ML and DL algorithms used in my research, as well as methods to evaluate ML models with a focus on MRI and brain tumors.

4.2.1. Classification

Classification is a field of supervised machine learning where the primary object is to generate discrete output labels based on input data, drawing from the knowledge acquired by the model during its training phase. Classification tasks can be binary (e.g., IDH mutant vs IDH wild-type), multiclass (can be ordinal, e.g., tumor grade prediction; or categorical, e.g., astrocytoma vs glioblastoma vs oligodendroglioma vs ependymoma), or multilabel (similar to multiclass problem, but more than one class can be assigned to each sample). A special case of the classification task, segmentation, will be discussed in section 4.2.3. In general, the objective of a classification problem is to find a model that can best separate the input data into two or more classes. Various ML model architectures can be employed for classification tasks such as Logistic Regression (LRs), Decision Tree (DTs), Support Vector Machines (SVMs), and Artificial Neural Network (ANNs), as well as various deep learning techniques.

Loss functions: The choice of loss function is another important factor to consider in ML and DL classification tasks. It depends on both the model architecture and specific problem. The most commonly used loss function for multi-class classification tasks in neural network architectures, for example, is the softmax cross-entropy loss function. Assuming C is the total number of classes, for every pair of input and output (x, y):

$$L_{CE} = -\sum_{i=0}^{C} y_i \log\left(P(y|x)_i\right) \tag{4.5}$$

with $P(y|x)$ the output of the activation function softmax, or the softmax probability of the model:

$$P(y|x) = \frac{e^{f_y(x)}}{\sum_c e^{f_c(x)}} \tag{4.6}$$

$f_y(x)$ is the model's decision function for output y, and $f_y(x)$ is the model's decision function for class c. Softmax is used to ensure the sum of model output for all classes is 1. The objective function needed to solve the problem becomes:

$$L = -\frac{1}{N}\sum_{i=1}^{N} log\ \frac{e^{f_y(x)}}{\sum_c e^{f_c(x)}} \tag{4.7}$$

In binary classification, the same cross-entropy loss can be applied, but a sigmoid activation function is typically used instead of softmax. For conventional machine learning models, other types of losses can be used, such as 0-1 loss, hinge loss, and ramp loss for SVM, logarithmic loss for logistic regression, etc. [123]

Model evaluation: For binary classification (and multi-class) tasks, it always begins with the confusion matrix (Figure 4.1). Actual ground truth and prediction results are compared, and sorted into one of the four categories: True Positive, False Positive, True Negative, and False Negative. Various evaluation metrics can be calculated using these 4 quantities, and the most common metrics are accuracy, sensitivity (recall), precision, specificity, and accuracy.

		Prediction		
		Positive	**Negative**	
Actual	**Positive**	True Positive (TP)	False Negative (FN)	**Sensitivity/Recall** $\frac{TP}{TP + FN}$
	Negative	False Positive (FP)	True Negative (TN)	**Specificity** $\frac{TN}{TN + FP}$
		Precision $\frac{TP}{TP + FP}$	**Negative Predictive Value** $\frac{TN}{TN + FN}$	**Accuracy** $\frac{TP + TN}{TP + TN + FP + FN}$

Figure 4.1. Confusion matrix for classification task

To figure out the best metrics for evaluation, we must assess how balanced the dataset is by looking at the data distribution. If the data is balanced, i.e., having approximately the same amount of data in each class, accuracy can be used as the metric. This is rarely the case in medical data, however. Therefore, different metrics are typically used depending on the objective of the task, e.g., balanced accuracy, sensitivity, specificity, precision, F1-score, etc. [124] It is always a good practice to present multiple metrics when comparing models. The receiver operating characteristic (ROC) curve and the Precision-Recall (PR) curve are also typically used to assess imbalanced data in binary classification task. The goal of these curves is to measure performance between models via the area under the curve (AUC) without assigning a decision threshold, thus is not suitable if the goal is to generate definitive prediction. It has been reported that PR curves are more informative than ROC curves given extremely imbalanced data [125].

4.2.2. Regression

Regression is another subclass of ML where the task is to predict a continuous output from the input (e.g., predicting the KI-67 from radiomic markers). Regression tasks in medical imaging are still very under-studied due to their high complexity in both training and model

evaluation. Regression models also tend to not generalize well even when achieving good performance in the training phase due high model variance (overfitting), or due to unseen test data being outside of the distribution of the training data. Distribution of the training data for the majority of regression models often does not follow a normal distribution and is often highly skewed (equivalent to the class imbalance issue seen in classification task). To deal with this problem, we can apply the same methods used for class imbalanced described previously. Another way to improve regression performance is to first train a classifier as a proxy task before utilizing its weight in the main regression task [126]. Just like classification, a regression problem can be solved using several ML models, such as Logistic Regression (LR), RandomForest Regression, Support Vector Regression (SVR), or deep learning methods (CNN, RNN, etc.).

Loss functions: The most commonly used loss functions for regression task are the Mean Squared Error (MSE), or the L2 loss, and Mean Absolute Error (MAE), or the L1 loss. For true output y_i and prediction $\hat{y}_i$:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{4.8}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{4.9}$$

Similar to the classification task, the choice of loss should reflect the goal of the specific task. MSE is much more sensitive to outliers, can converge more easily during training due to a decreasing gradient when the loss becomes smaller, and is suitable when the outliers are important to detect biologically (such is the case for many medical imaging tasks such as predicting KI-67). Other loss functions to consider includes the Huber loss (which combine both MSE and MAE by using a threshold), the Log Cosh loss, the Root Mean Squared Error, Mean Absolute Percentage Error (MAPE), etc. Manually designed loss can often be used although there is no guarantee the training will converge.

Model evaluation: Evaluation of a regression model usually involves visualizing the prediction by plotting the prediction and ground truth. To summarize model performance, MSE

and MAE are most often used to evaluate its accuracy, along with the coefficient of determination (R^2) and adjusted-R^2 values to assess how much variance in the outcome can be explained by model prediction. However, MAE and MSE are very unintuitive as their values can range from 0 to infinity, and there is usually no clear guideline of what a good MAE/MSE score is. Furthermore, unlike in the case of classification, it is often not recommended to compare MSE and MAE between models tested using different datasets. A percentage metric, such as Symmetric Mean Absolute Percentage Error (SMAPE), and the R^2 values are usually better for that purpose [127]:

$$SMAPE = \frac{100\%}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \tag{4.10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \; ; \; \bar{y} = \sum_{i=1}^{N} y_i \tag{4.11}$$

4.2.3. Semantic Segmentation

Semantic Segmentation is a subclass of classification task, where the goal is to assign each pixel (or voxel in the case of MRI) to a specific class label and reconstruct the image using these labeled representations. For example, in a brain tumor segmentation task, each voxel of the MR image can be classified as either normal brain tissue, contrast enhancing lesion, non-enhancing lesion, or necrosis. Semantic segmentation usually uses a convolution neural network (CNN) with an encoder/decoder architecture (U-net) where the encoder down-samples the original image and generates a low-resolution feature map that can be trained to classify different classes. The decoder then up-samples this feature map into a segmented image of the same resolution as the original image. A recent advance in this field includes the integration of vision transformer-base DL into standard U-net architecture that can be used for a segmentation task [128]–[130]. Overall, the performance of these models show a small improvement, but in some tasks, can still underperform compared to a standard CNN U-Net [131].

Loss function: The choice of loss function has been shown previously to greatly affect the performance of a segmentation task. If we look at a segmentation problem where the ground truth

is Y = $\{y_i\}$ and prediction S = $\{s_i\}$, given y_i^c as the indicator of whether the voxel i prediction for class c is correct, and s_i^c is the corresponding predicted probability of that, the loss function for segmentation tasks can be grouped into the following categories [132], [133]:

- *Distribution-based losses*: the goal of the training is to minimize the difference between two distributions. The most used loss function here is the cross-entropy (CE) loss, and most distribution-based losses are directly derived from CE loss, such as the focal loss, the TopK loss, etc.

$$L_{CE} = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_i^c log s_i^c \tag{4.12}$$

- *Region-based losses*: the goal of the training is to maximize the overlap of ground truth region Y and predicted region S. The most used region-based loss is the Dice loss, and other losses can be directly derived from Dice loss, such as the IoU loss, generalized Dice loss, Tversky loss, focal Tversky loss, etc. Dice loss is similar to the F1 score in a classification task, and usually penalizes false positives and false negatives equally, thus can be more sensitive to skewed and imbalance datasets compared to Tversky or focal Tversky.

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N}\sum_{c=1}^{C} y_i^c s_i^c}{\sum_{i=1}^{N}\sum_{c=1}^{C}(y_i^c)^2 + \sum_{i=1}^{N}\sum_{c=1}^{C}(s_i^c)^2} \tag{4.13}$$

- *Boundary-based losses*: the goal of training here is to minimize the distance between the boundary of Y and S, which can be good for imbalance datasets. The most used loss in this case is the Hausdorff Distance loss, though it has been previously reported that training with HD loss does not converge well due to its convex nature [133].

$$L_{HD} = \frac{1}{N}\sum_{i=1}^{N}[(s_i - y_i) \cdot ({HD_{Gi}}^2 + {HD_{Si}}^2)] \tag{4.14}$$

To pick the right loss function, it is important to understand the task's objective and the dataset distribution. In general, for tasks with highly imbalanced datasets, a region-based loss function typically will achieve better performance. For most medical applications, mis-diagnosis of true diseases is especially costly, therefore a loss that penalizes false negatives more than

false positives, such as the Tversky or Focal Tversky loss, is highly preferred. For example, in various segmentation tasks where the region of interest is very small, Focal Tversky loss outperforms all other loss functions by increasing model sensitivity [133], [134]. Finally, training can be performed by manually designing loss functions that combine one or more losses together, such as the BCE + Dice loss that combines both cross-entropy and dice losses into one.

Model evaluation: Similar to classification, it is extremely important to pick the right metrics for model evaluation based on the characteristics of the task. In most medical application, sensitivity needs to be high in order to prevent mis-diagnosis of true disease. Good metrics should also be intuitive as well.

4.3. Classical machine learning techniques

As discussed previously, classical machine learning algorithms heavily rely on feature engineering, the crafting of features that can be fed into a ML model. Domain knowledge is generally required to generate good features, although sometimes a variety of features can be selected later in a feature selection step. Typically, ML models work better if the number of training data is many orders of magnitude bigger than the number of features. Thus, data preprocessing should include some type of feature selection steps, typically done via statistical testing, or can be incorporated into the model as well. Some methods for feature engineering are listed in Heaton et al. [135]. Once features are generated, they can be fed into ML models for prediction tasks.

There are several ML models that one can choose from. The most commonly used supervised ML models are [136]:

- *Support vector machines (SVM)* [137]: This type of model generates one or more hyperplanes in a high dimensional space to best separate training data among different classes. It does so by maximizing the boundary distance between the two classes. SVM usually does not work well if there noise exists in the dataset, or if data of different classes are too close in high dimensional space.

- *Logistic regression* [138]: This is a probability-based model, meaning it will provide decision function using logistic function (sigmoid). This model assumes linearity between the inputs and the log odds, and no collinearity between different features.

- *Decision trees* (DT) [139]: DT classify data by using a hierarchy of decisions. Starting from the root node, DT includes a number of decision nodes, each representing a test of a feature. Nodes are then branched depending on the outcome (for example, nodes can be split based on thresholding a feature.) The criteria for splitting can be either the Gini impurity or the entropy. DTs typically have high interpretability, although can be too simple for many datasets.

- *Random forest* (RF) [140]: The goal of RF is to fit many individual DTs in parallel using different subsets of the original dataset (via bootstrap aggregating, or random sampling of data with replacement.) The final decision is made by using majority voting of all decision trees. Because of this, RF prevents overfitting problem and increases model generalization during testing. Another advantage of RF is that it can deal with colinear features well.

4.4. Deep learning techniques

Deep learning (DL) is a subfield of ML that solves complex tasks by using artificial neural networks that mimic the human brain. Many papers have reviewed DL architectures as well as their applications in the medical imaging domain [141], [142]. DL architectures are typically made up of an input layer, multiple interconnected hidden layers, and an output layer. The hidden layers basically perform non-linear transformation of input data into the output space via a series of weights, biases, and activation functions. Figure 4.2 below shows an example of how a simple neural network, a multilayer perceptron (MLP) with one hidden layer, works.

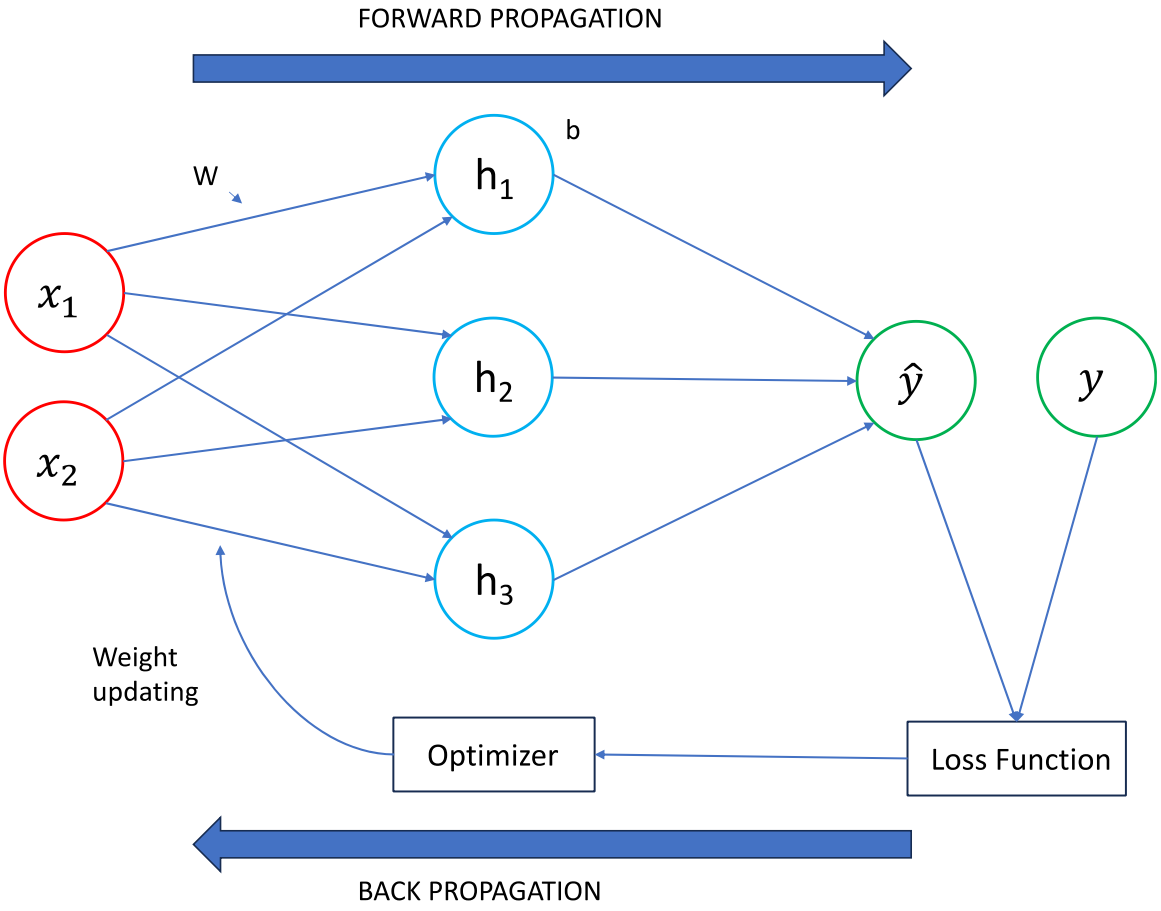

Figure 4.2. Simple multilayer perceptron model with one hidden layer
Layers are connected by a series of weight, and each neuron also has a bias term, all of which can be trained using back propagation.

A neural network typically includes a forward and a backward propagation. In the forward propagation, each layer is a function of all the outputs from the previous layers and their weights, as well as a bias, which is typically randomly assigned initially. For the previous simple MLP network with I input neurons and H hidden neurons:

Hidden layers: $$h_j = \sigma(\sum_{i=1}^{I} w_{ij}^1 x_i + b_j) \quad (4.15)$$

Output layer: $$o_j = f(\sum_{j=1}^{H} w_{jk}^2 h_j + b_k) \quad (4.16)$$

Both functions $\sigma(\cdot)$ and $f(\cdot)$ are called the activation functions and are usually non-linear, such as a sigmoid function or the rectified linear unit (RELU), in order to allow complex networks to be solved. The choice of activation function is important to solve the neural network problem. Fortunately, for most neural network hidden layers, RELU and leaky RELU are always safe choices, as they don't experience the vanishing gradient problem as in a sigmoid function, which can affect training.

RELU: $$f(x) = max(0, x) \quad (4.17)$$

Hidden layers: $$f(x) = max(\alpha x, x) \quad (4.18)$$

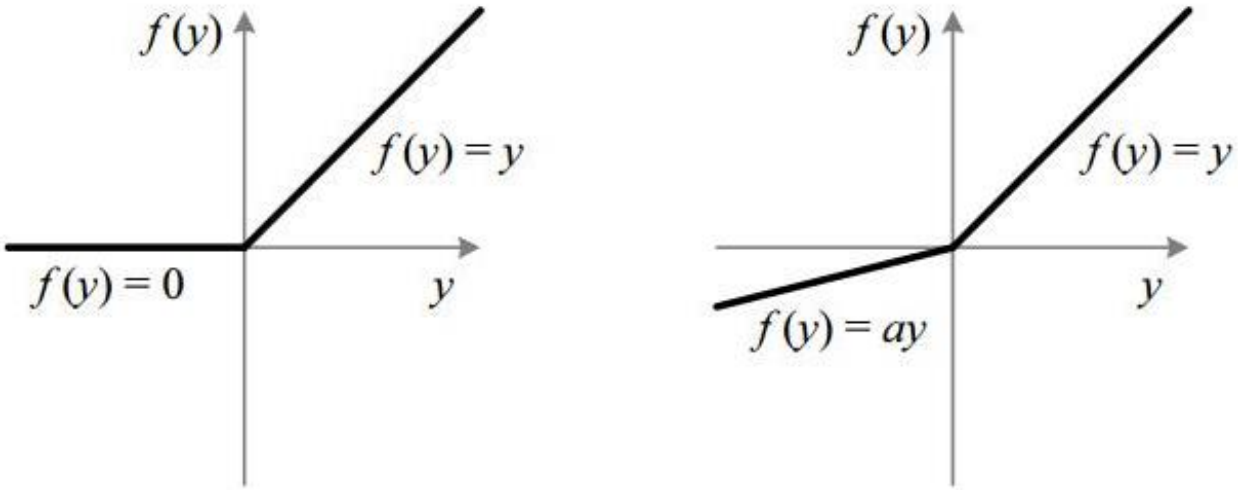

Figure 4.3. RELU examples
RELU function on the left and leaky RELU function on the right are popular activation function used for hidden layers.

The next step of solving a neural network is to calculate the loss between the predicted output and the real output using loss function described in previous section. The final goal is to readjust all weights and biases in order to minimize this loss. This is done via back propagation. The network will first calculate the local gradient for each weight and bias, which is the partial derivative of the lost function L with respect to weight w and bias b, i.e. $\frac{\partial L}{\partial w}$ and $\frac{\partial L}{\partial b}$. To train the neural network, an optimization function such as gradient descent (GD) will be used to adjust weights and biases:

$$w = w - \eta \frac{\partial L}{\partial w} \tag{4.19}$$

$$b = b - \eta \frac{\partial L}{\partial b} \tag{4.20}$$

where η is the learning rate. The choice of optimization function as well as the learning rate can affect how fast and how well the network converges, thus can also affect its model performance. Example optimization functions are stochastic gradient descent (SGD), RMSProp, Adaptive Moment Estimation (Adam), etc.

4.4.1. Convolution Neural Networks and UNets in semantic segmentation

Computer vision tasks such as classifying or segmenting an MRI image can be done via MLP. However, it would probably require an unlimited amount of data, as each 3D image can contain up to millions of voxels. For this reason, CNNs were developed as an extension to the MLP and is a much better option for those tasks. A CNN improves upon an MLP by reducing the total number of trainable parameters and retaining important spatial information of an image via hidden layers, which are consisted of a convolution layer, pooling layer, and more (e.g. batch normalization, activation, etc.).

In a CNN architecture, the goal of convolutional layers is to extract features of the input image tensor. This is done by applying a filter across the input tensor and generating a new output tensor by summing the element-wise product between the filter and each small patch of the input

image. This operation is then repeated K number of times using K different filters, with K typically being pre-defined. For most medical imaging application, each filter typically has size 3x3x3, 5x5x5, or 7x7x7 in the case of a 3D CNN, and 3x3, 5x5, or 7x7 in the case of a 2D CNN. All the elements in these filters are trainable weights and can be updated via backpropagation. The new output image is generated using a non-linear activation function such as RELU or leaky RELU to produce a feature map. Besides convolution layers, a CNN architecture also uses pooling layers immediately following convolution layers to down-sample the size of convoluted feature map, thus reduce the number of trainable parameters. This down-sampling operation of pooling layers also helps to further increase the receptive field of subsequent convolution layers. These feature maps output from the convolution or convolution-pooling block are then concatenated and passed on as input to the next hidden layer. Other operation such as batch normalization and drop-out can also be included to improve performance.

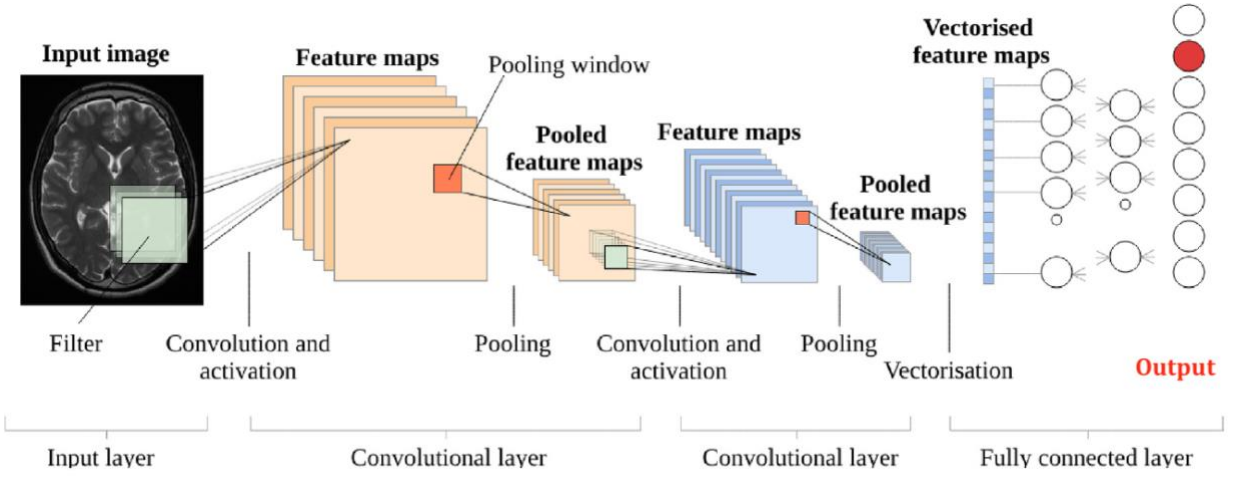

Figure 4.4. Example of a CNN for classification
A CNN consisting of 2 convolution layers, each followed by a pooling layer to extract features, before flattening and performing the classification task via fully connected layers.

A typical CNN will consist of multiple convolution layers (figure 4.4); although a very deep CNN model means a higher number of weights to train, requiring more data and computing resources. The equation for the feature map in the n^{th} layer using a k^{th} filter is:

$$X_k{}^n = \sigma(w_k{}^n * X^{n-1} + b_k{}^n) \tag{4.20}$$

where σ is a non-linear activation function as described previously. In a typical classification task, features from the last convolution layer will be flattened, and a final fully connected layer (similar to hidden layers of MLP) will be added to perform the classification task. One other thing to consider is the choice of 2D vs. 3D CNN models. As most medical images are 3D volumes, training using 3D CNNs have been proven to be superior because they incorporate the spatial relationship of the 3rd dimension, but computational demands are also higher [143].

Recent advances in CNN development have allowed CNNs to achieve state of the art performance in various tasks with advanced model architectures such as VGGNet [144], Inception [145], [146], and ResNet [147]. These advanced models allow deeper networks to be trained without reproducing the vanishing gradient issue discussed earlier. But what about using CNNs for segmentation tasks? How do we generate a full-resolution segmentation map from low resolution convolutional feature maps?

Many models have been proposed and used for segmentation with both 3D and 2D CNN. Most of these models utilize the encoder-decoder architecture in order to achieve full resolution of the output images. The encoder consists of a series of convolution and pooling layers that extract features just like in classification task. Then, the decoder applies a series of up-sampling steps to the features generated to produce a segmentation map of the original resolution. In each up-sampling step, the upsampled output is concatenated with corresponding features from encoder steps [148]. A U-Net improves upon this schema by adding several additional key features: the U-Net architecture has a symmetrical encoder and decoder, and every layer of the encoder is connected to the corresponding layers of the decoder via skip connections. U-Nets achieve dimension recovery by first using a 2x2 up-convolution layer before concatenating the output of that with the corresponding layers of the encoder. This allows for high level spatial details of the image to be preserved and used in the decoder stage. The result is a high resolution and detailed segmentation map [149]. This is particularly crucial for medical image segmentation, as

the spatial location of fine details is needed for most tasks. An example U-Net CNN is shown in figure 4.5. for brain tumor segmentation.

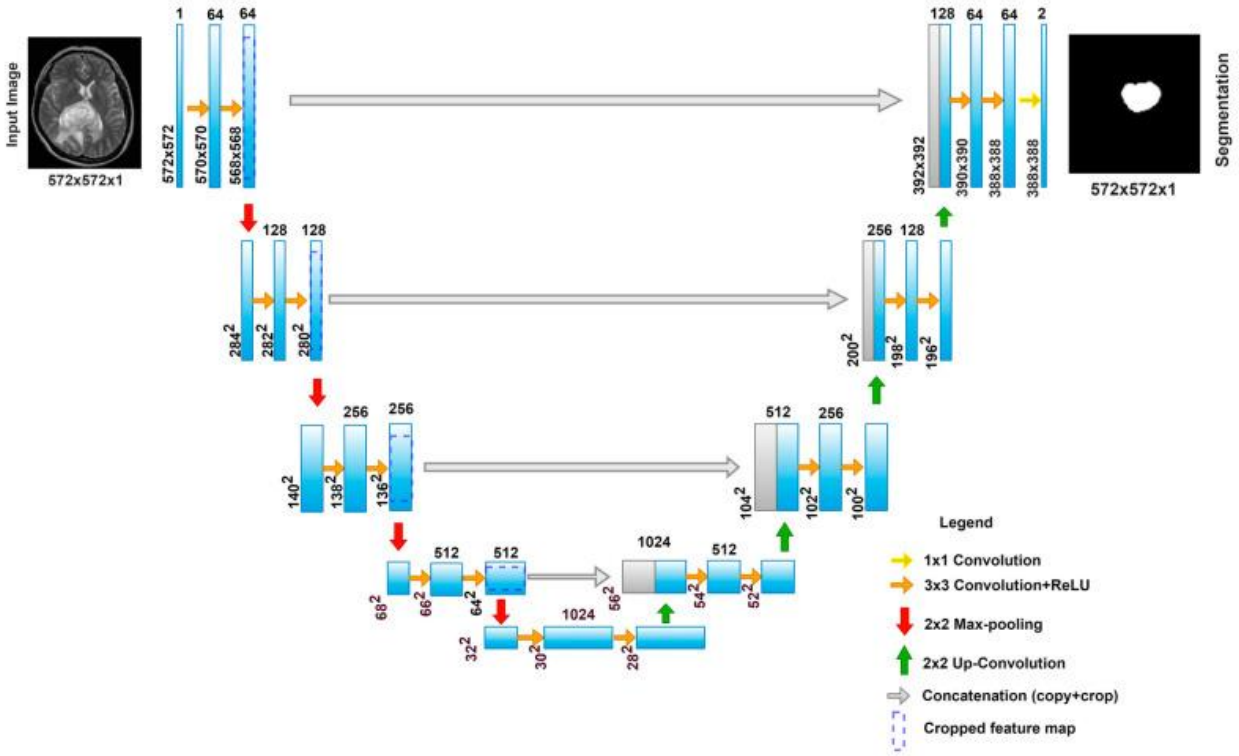

Figure 4.5. Application of U-Net architecture for brain tumor segmentation
Figure adapted from Yousef et al., 2023 [150]

Many other models have improved upon the U-Net architecture specifically for medical image segmentation, such as Residual U-Net [151], Attention U-Net [152], UNET++ [153], etc. That said, a well-trained standard U-Net has been shown to be capable of achieving better performance than other complicated architectures for brain tumor segmentation tasks [150], [154].

4.4.2. Recurrent Neural Networks and Long Short-Term Memory Network

Feed-forward NN and typical CNNs are great for a variety of tasks and applications. However, one limitation of these networks is their inability to interpret sequential data (such as time series or a sentence) without an attention layer. A different class of NN architecture is more suitable for that, called recurrent neural network (RNN). It is said to have “memory” and can

remember input from previous nodes within the sequence. It does this via a feedback loop in its cell, allowing output of a layer to be recycled within a layer (as opposed to feedforward NN where output can only be passed onto the next layer). Mathematically, the expression of the RNN cells is depicted as:

$$h_t = \sigma(w_h h_{t-1} + w_x x_t + b) \tag{4.21}$$

where x_t and h_t represent the input and output of the current node at time t, and h_{t-1} represents the output of node at time t-1 [155].

LSTM is a special case of RNN that can also remember long term memory via a set of gates, which basically are just a way to instruct what information to remember. This is done via a sigmoid layer, outputting a number from 0 to 1 depending on how much the network should remember. Different variants of LSTM will contain different types of gates, which help to solve the vanishing or exploding gradient issue seen in earlier RNN models [155], [156]. Figure 4.6 shows a typical RNN cell and LSTM cell.

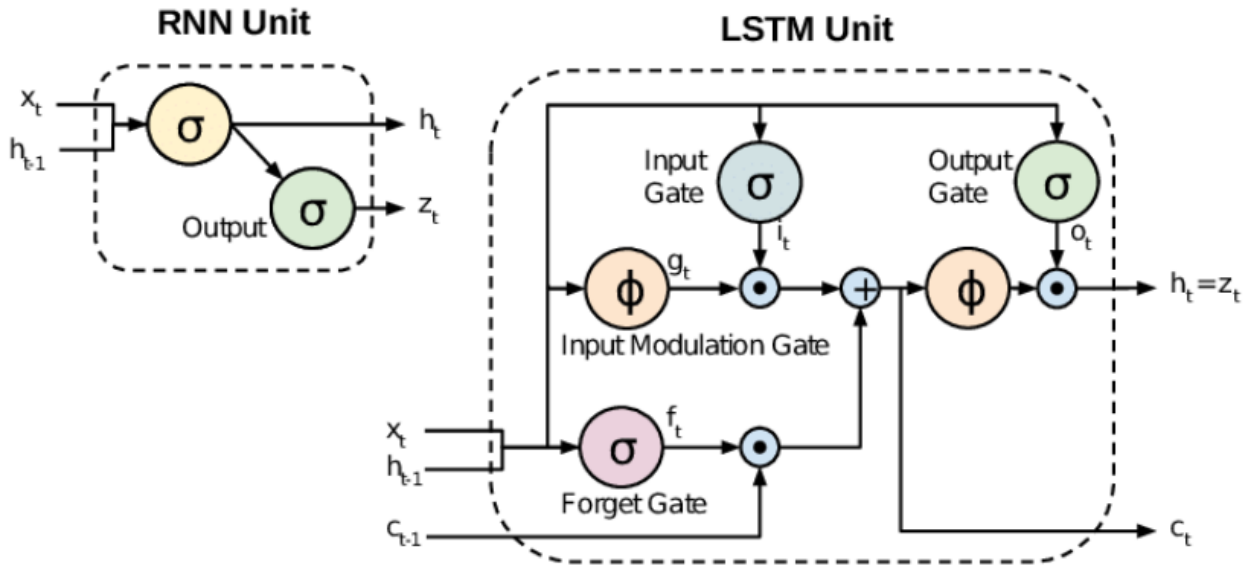

Figure 4.6. Examples of standard RNN and LSTM unit cells
The LSTM cell has a forget gate, input gate, and output gate. Figure adapted from Donahue et al. [156]

RNN and LSTM have found their way into tasks that involve sequential data classification, including spectral data. Many studies have shown that using LSTM based models on spectral data can achieve better or comparable performance compared to much deeper CNN models, with an added benefit of long term memory [157]–[159]. LSTM can also be used in conjunction with CNN feature extractors in a larger network that combines the benefit of both NN [156].

Chapter 5. Machine learning (ML) for predicting voxel-wise histopathology of tumor cells in newly diagnosed glioblastoma patients using and Proton Magnetic Resonance Spectroscopy (1H-MRSI)

5.1. Introduction

Gliomas are highly infiltrative, heterogenous brain tumors with poorly defined margin, and varying overall survival based on molecular subtype and grade [1], [14]. Currently, the gold standard for the diagnosis of tumors includes the histopathological and molecular evaluation of tissue samples randomly taken from the patients during surgery. Histopathological metric such as the % Ki-67 (or % MIB-1) and cellularity measured from these tissue samples have been used extensively as markers for tumor cell proliferation. However, these samples contribute to a very small fraction of tumor tissue, neglecting to capture the statistical distribution of biological properties that would allow for accurate characterization of tumor biology throughout the whole lesion [29], [30]. Furthermore, due to the biological heterogeneity of the tumor tissues, it is also challenging to identify the most malignant area of the lesion to obtain tissue samples from to make an accurate diagnosis. As the result, this can lead to poor outcomes, partially due to the difficulty in defining and treating the full extent of these infiltrative tumors. Having a spatial map of tumor histopathology and aggressiveness based on underlying tumor metabolism can mitigate this issue by: 1) help guiding the selection of tissue samples for more accurate diagnosis of heterogenous lesions, 2) potentially increasing the extent of resection of these highly infiltrative lesions, and 3) non-invasively characterizing tumor that remains after surgery to inform subsequent treatment.

Previous biopsy studies have shown that tumor usually grows beyond what was defined by T1-post-contrast and T2-weighted MR images [98], [99]. Therefore, anatomical images alone are not reliable in determining the spread of glioma and diagnosing the true heterogeneous nature of the lesions. More recent studies have demonstrated the benefit of using of Proton Magnetic Resonance Spectroscopy (1H-MRS) and the derived Choline-to-NAA index (CNI) in probing the

underlying cellular metabolism associated with tumor growth, tumor progression, and poor survival [78], [98], [106]; and more recently, identifying IDH and TERT promoter mutation status using the entire spectrum [107], [108], [160]. Specifically, poor tumor outcome and more aggressive lesions usually associate with 1) elevated Choline (Cho), signaling an increase in cell membrane turnover due to rapid growth of tumor tissues; and 2) lowered N-acetylaspartate (NAA), signaling a decrease in neuron viability. Creatine level is stable but can be lowered by about [161]15-40% in some gliomas. The presence of Lactate peak (Lac), which is a marker for anaerobic metabolism, were also seen in higher grade tumor and necrosis area [92], [98], [102]. These studies show that 1H-MRS has the potential to be used in determining tumor heterogeneity and identify subclinical tumors.

Predicting and generating spatial map of tumor pathology and have been tried before with various level of success. For example, Li et al. achieved a high AUC of 0.788 in classifying clinically high and low KI-67 using anatomical and diffusion images of 263 patients [162]; Gates et al. was able to generate spatial map for KI-67 using anatomical, diffusion, and perfusion images of 23 patients [163]. While promising, many of these papers either have few numbers of patients (less than 30), which can limit the number of biopsies with high KI-67 and skew the true distribution of pathological values, or contains incorrect assumption that one single KI-67 score can represent the KI-67 for the entire tumor regions of the patient, even though we know that all pathological parameters should be heterogenous even within the same patient.

Machine learning (ML) and deep learning (DL) technologies have been one of the most exciting and important development in radiology research. From image reconstruction, outcome prediction, to image segmentation, et cetera, ML allows MR images to be acquired and interpreted correctly and quickly, assisting physicians and radiologists in various stages of patients care. However, most ML research using MRSI data uses individual metabolites peak height, or map of metabolites, which heavily relies on the accuracy of post-processing techniques, while failing to utilize the information-rich 1H-MRSI full spectrum, such as the peak width, minor peaks, and

relationship of different peaks in spectrum (ratio, etc.). More recently, there have been a bigger focus on applying deep learning technique to analyze the full 1H-MRSI spectrum, performing various tasks such as spectral quality assurance [164], [165], estimating metabolite concentrations and uncertainty [164], [166]–[168], and classification [110], [169], [170]. Many of these papers show results exceeding the conventional method, while also improving processing speed. However, there have not been any attempt in applying deep learning techniques in regression tasks, particularly in predicting the histopathology of the patients, mostly due to limited amount of data, as well as the complexity of the tasks involved. In this study, we want to investigate the use of various deep learning techniques (1D-CNN and bidirectional LSTM), and imbalance techniques (sample weighting and 2-stages training) to predict the voxel-wise histopathology of lesion. The goal of this study was to develop a deep learning model that utilizes the wealth of information contained in the entire spectrum to predict voxel-wise histopathology of tumor cells, including tumor cellularity, mitotic activity (Ki67), and a new composite tumor aggressiveness index (CTAI; defined as the sum of normalized cellularity and Ki67) using tissue samples with spatially mapped coordinates on 3D 1H-MRSI. Our dataset is much larger than most previous studies (includes 607 tissues samples from 281 newly diagnosed glioma patients), and multiple tissues samples were collected from each patient, preserving the heterogenous nature of tissue samples within a patient.

5.2. Methods

5.2.1. Patients Characteristics:

A total of 397 newly diagnosed and clinically confirmed glioma patients between 2007 and 2018 were enrolled in this retrospective study. Patients were excluded if their preoperative MRI acquisition did not include the 3D ^{1}H MRSI. For each patient, about 1-5 tissue samples were collected. Of all 1159 tissues samples obtained from 397 patients, samples were only used in this study if: (i) it has one or more histological outcome, including Ki-67 and cellularity; (ii) it comes from tumor regions that is not necrosis; (iii) the patient receives 3D ^{1}H MRSI, and the tissue sample is within the PRESS box but not covered by the saturation bands. This results in a total of 607 tissues samples from 281 patients. The demographic and clinical characteristics of all patients, and the pathology distribution of tissues used in this study were shown in Table 1.

Table 5.1. Demographics & clinical characteristics of patients and tissue sample pathology used in this study

Patient Demographic		**Patients (n)**	**Patients (%)**
Total		**281**	
Sex	Female	115	41%
	Male	166	59%
Clinical Diagnosis			
WHO Grade	Grade II	99	35%
	Grade III	53	19%
	Grade IV	129	46%
Mutation status	IDH-wildtype	121	43%
	IDH-mutation + 1p19q intact	95	34%
	IDH-mutation + 1p19q-codeletion	65	23%

Tissue pathology		Tissue samples (n)	Tissue samples (%)
Good quality tissue		**607**	
Ki-67 (%)	**Total**	**549**	**100%**
	Low (< 10%)	360	65%
	Average (10% -25%)	114	21%
	High (> 25%)	75	14%
Cellularity (Cells/mm^2)	**Total**	**435**	**100%**
	Low (< 1500)	296	68%
	Average (1500 – 2500)	97	22%
	High (> 2500)	42	10%
Tumor Score	**Total**	**501**	**100%**
	0	37	7%
	1	67	13%
	2	199	40%
	3	198	40%

5.2.2. MR Acquisition:

The 3D 1H MRSI was acquired using point-resolved spectroscopic selection (PRESS) for volume localization and very selective saturation (VSS) pulses for lipid signal suppression (excited volume = 80 × 80 × 40 mm, TR = 1100–1250 ms, TE = 144 ms, overpress factor = 1.5 if lactate edited, otherwise 1.2, field of view = 16 × 16 × 16 or 18 × 18 × 16 cm, nominal voxel size = 1 × 1 × 1 cm), flyback echo-planar readout gradient in the superior–inferior direction, 988 Hz sweep width and 712 dwell points. A dual-cycle lactate-edited sequence [6] was used for 202 patients (440 samples, 11 min), while a standard single-cycle sequence was used for the remaining 79 patients (167 samples, 6 min).

5.2.3. Spectroscopy Data Processing of Tissue Samples:

To generate the 1D ^{1}H-MRSI spectrum centered at the location of each tissue sample, we first shifted the 3D spectral arrays shifted in k-space and reconstruct a spectral voxel at the center coordinates of each tissue sample location, and then perform phase correction to account for flyback echo-plannar gradient [90], [91], [171]. We also performed residual lipid and water removal and motion correction on the spectra. This process is done using software developed at UCSF. The sum and difference spectra of lactate edited sequence were sum and divide by 2 to match with data from single-cycle sequence.

Additional preprocessing of ^{1}D 1H-MRSI spectrum includes Gaussian process filtering smoothing and normalization by the average NAA of the normal appearing brain for each patient. We also calculate features for Machine learning model by measuring the metabolite peak height and area, including Choline, Creatine, NAA, Lipid, and Lactate, and normalized them by the median value of the corresponding metabolite in normal appearing brain. The indices choline-to-NAA index (CNI), choline-to-creatine (CCrI) and creatine-to-NAA (CrNI) were also calculated.

5.2.4. Histopathological Assessment of Tissue Samples:

At least 4 different tissue samples within the hyperintense region of the T2 FLAIR image and outside of the tumor cavity were planned preoperatively, with at least 1 cm apart between samples. An intraoperative navigation system (BrainLab or Surgical Stealth) was used during surgery to guide the resection of tissue samples, and the precise location of the tissues were recorded. Samples were then fixed in formalin and embedded in paraffin (Barajas). Hematoxylin and Eosin (H&E)-stained tissue samples were then sent to a board-certified pathologist for evaluation.

For each tissue sample, a maximal labeling index for %MIB-1-positive nuclei (or Ki-67 score), and total cellularity have been calculated from at least three fields and >1000 cells. A tumor score (TS; 0-3) has been assigned based on the contribution of tumor to total cellularity, with microvascular hyperplasia, necrosis, and gliosis being quantified as described previously. The primary pathology outcome of interest is the composite tumor aggressiveness index, or the CTAI score, calculated as following:

$$\mathrm{CTAI} = \log \frac{\mathrm{n(Ki\text{-}67)} + \mathrm{n(Cell)}}{1/\mathrm{Tumor\ Score}}$$

During training and testing, due to the heavily skewed distribution of the tissues' Ki-67 score and total cellularity, we perform logarithmic transformation on the Ki-67 score and square-root transformation on the cellularity.

5.2.5. Statistical Analysis and Machine Learning for Baseline Model

Each dataset (for each histopathological target) were split into 75% training data and 25% test data. Test samples were identical for both machine learning and deep learning methods. Only training dataset were used to perform statistical analysis. We calculated the correlation between each ^{1}H-MRS parameter and the 3 histopathological prediction targets of the training set using the Kendall's Tau Correlation coefficient, ensuring only one sample was assessed per patient.

Cut off p-values were chosen at 0.05, 0.01, and 0.001. We trained a Linear Regression (LR) model on the CNI to predict the KI-67, cellularity, and CTAI. This model was then used to establish baseline performance for both machine learning and deep learning models. We then trained a random forest regression (RFR) model using all individual metabolites and indices to predict Ki-67, cellularity, and CTAI. For each training step, we optimized the RFR model by performing 4-fold cross validation of the entire training dataset, using the negative mean squared error as the evaluation metric. Once the model was optimized with specific parameters for each prediction task, we refitted the model on all training samples and applied the trained model on the hold-out test set.

5.2.6. Deep Learning

Two types of neural network architectures were used to classify the spectra. We first trained a 1D-CNN based model architecture with 4 convolution layers, each with kernel size of [7, 3, 3, 3], and number of filters of [64, 64, 128, 256] respectively, following by ReLU activation, as well as a max pooling layer with pool size of 2 (but only for the first two convolution layers). For the second model architecture, the 1D-CNN-BiLSTM model, we utilize the same CNN architecture described previously, but added a bidirectional LSTM layer after the final convolution layer, but prior to flattening features and the fully connected layer. We trained and optimized the model using the MSE loss, as it provided better generalization than MAE loss.

Additionally, we explored the use of two separate methods (Figure 5.1B) to deal with the imbalanced dataset (due to histopathology outputs skewing left). The first method is sample weighting, where training samples are assigned weight using an inverse probability weighting technique. Samples with high probability distribution will weight much less than samples with low probability distribution. The second method is two-step training, where we explored the use of transfer learning by first training a classification model to predict the ordinal target, created by

binning each target into 3 groups (low, middle, and high, figure 5.1C), and then finetuning using the original target.

5.2.7. Model Evaluation

We evaluated the CNI linear regression model, the RFR model, and the deep learning model performance in predicting log(Ki-67), sqrt(cellularity), and CTAI using the standard metrics including mean squared error (MSE), mean absolute error (MAE), and the R^2. In order to further correctly evaluate model performance, we calculated three additional MAE metrics for each model: the MAE at predicting low, middle, and high pathological indices. We did this by first splitting each pathological index (Ki-67, cellularity, and CTAI) into three separate regions of low, middle, and high values (similar to how splitting was done for the two-step training method).

5.2.8. Gradient-weighted Regression Activation Mapping (Grad-RAM)

The Grad-RAM (similar to Grad-CAM) heatmap for each convolution layer was generated using the weighted sum of all feature maps within that layer. Since our CNN-BiLSTM network had 4 convolution layers, and we combined the RAM results of all convolution layers using the weighting 0.5*conv1 + 2*conv2 + 3*conv3 + 4*conv4 instead of just using the heatmap from the final convolution layer. This is so the Grad-RAM map can better reflect the finer details of the ^{1}H-MRSI.

5.2.9. Spatial map generation

For each histopathology index, we use the best performing model to generate spatial maps of histopathology for patients in the test set. To do this, we divide the PRESS box into 5x5x5 voxels, and obtain the full ^{1}H-MRSI for each voxel using their center location. We iterate and apply

the model on each voxel to generate a heatmap, that is then visualized in color with Slicer4 [172] software.

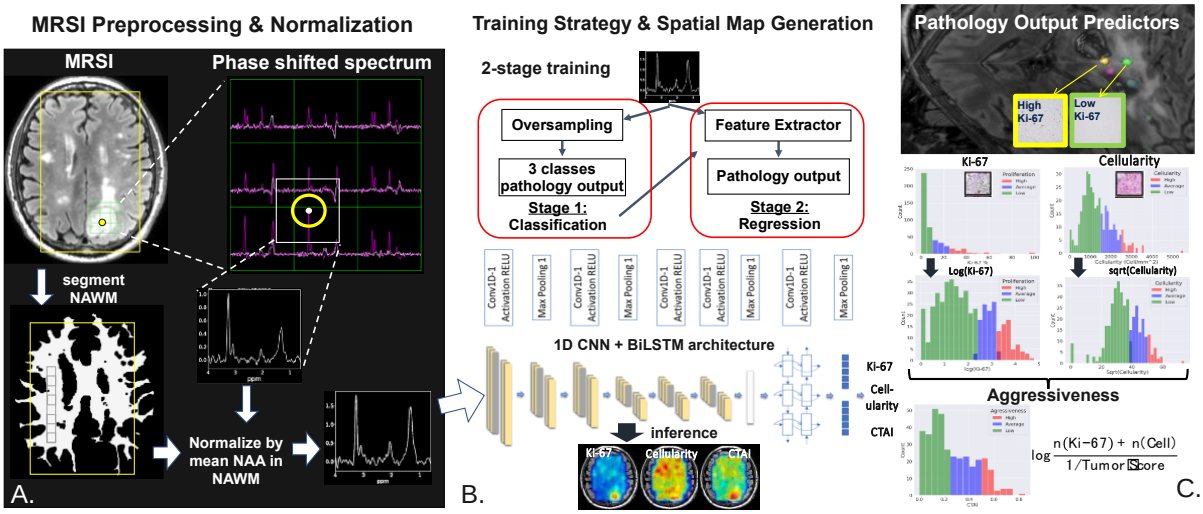

Figure 5.1. Images and Data Processing Schema
(A) To generate a single spectrum centered at the location of each tissue sample, 3D spectral arrays were first shifted in k-space to reconstruct a spectral voxel on the center coordinates of each tissue sample location. 1H-MRSI data were normalized by the mean NAA in normal-appearing-white-matter (NAWM). (B) Deep learning model architecture and the 2-stage training schema to improve performance, by first training a classifier prior to the regression task. Spatial maps are then generated for each test patients. (C). Log transformation was applied for Ki-67 and square-root transformation was applied for the cellularity. CTAI was calculated using Ki-67, cellularity, and the tumor score for each biopsy.

5.3. Results

5.3.1. Significant correlations between metabolites and pathology

The results of Kendall's Tau correlation between the metabolite measures and each histopathological metric are summarized in figure 5.2. CNI, CCrI, and nCho were significantly positively correlated to all 3 pathological indices. nLac was only significantly correlated to Ki-67 ($p < 0.01$), while nNAA was negatively correlated with both cellularity and CTAI values ($p < 0.001$).

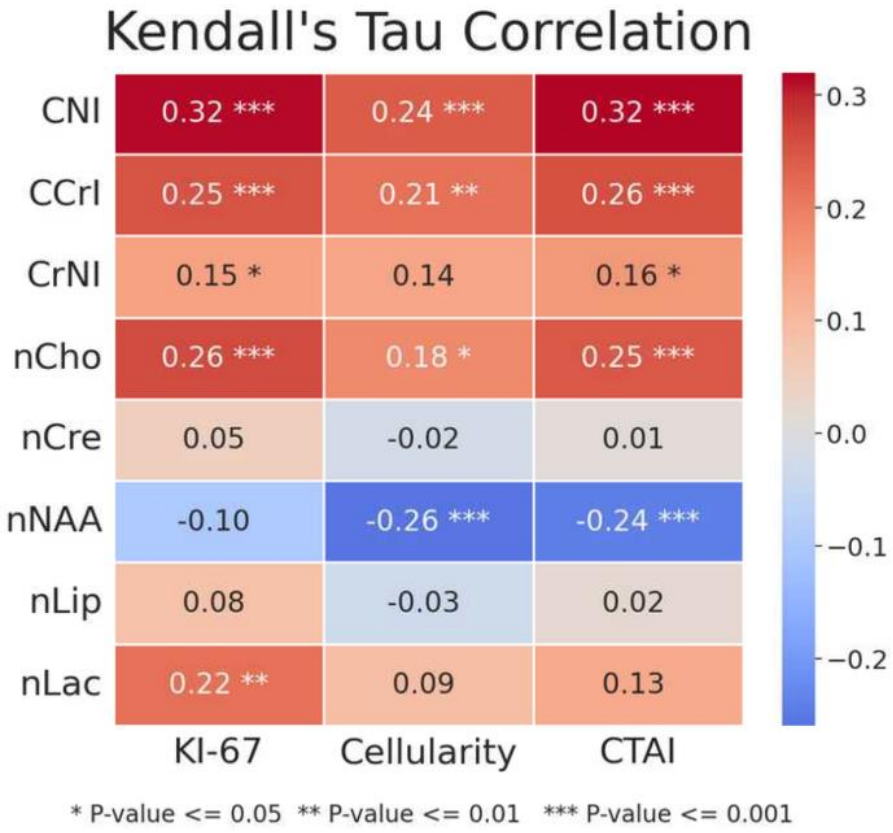

Figure 5.2. Kendall's Tau correlation between MRSI markers and tissue pathology
nCho, CNI and CCrI were strongly positively correlated with all pathology measurements, while was nNAA strongly negatively correlated to only cellularity and CTAI, and nLac only correlated with Ki-67.

5.3.2. Ki-67 prediction

Figure 4 shows example spectra with low and high Ki-67 values. High Ki-67 samples typically have high choline and creatine peaks, and reduced NAA. Table 5.2 summarizes the results for Ki-67 prediction. The best deep learning models were the 1D-CNN-BiLSTM trained with no sample weighting or 2-stage training and 1D-CNN-BiLSTM trained with 2-stage training but no sample weighting, both outperforming both the CNI linear regression baseline model and the Random Forest regressor model in mean-squared-error and mean-absolute-error. The actual Ki-67 vs. prediction plot for the deep learning models had an R^2 value of 0.25, much higher than the R^2 value of 0.01 in the Random Forest model.

Table 5.2. Summary of regression results for all models in predicting log(KI-67)

Model	MSE	MAE	MAE low	MAE mid	MAE high
Predicting log(Ki-67)					
CNI Linear Regression	1.32	0.93	0.63	1.03	1.85
RF Regressor model	1.29	0.95	0.71	0.95	1.83
1D-CNN	1.16	0.90	0.72	0.85	1.61
1D-CNN-BiLSTM	1.02	0.76	0.52	0.69	1.41
1D-CNN-BiLSTM with sample weighting	1.16	0.87	0.82	0.65	1.25
1D-CNN-BiLSTM with 2-stage training	0.97	0.79	0.57	0.65	1.34
1D-CNN-BiLSTM with both sample weighting and 2-stage training	1.01	0.81	0.64	0.68	1.35

5.3.3. Cellularity prediction

The middle row of figure 5.3 shows representative spectra with low and high cellularity values. Because Ki-67 and cellularity are well-correlated, overall similar spectral appearance was observed for tissue samples that had high Ki-67 and high cellularity compared to low Ki-67 and cellularity values: tissue samples with high cellularity also had elevated choline and creatine peaks with reduced NAA. Table 5.3 summarizes the results for cellularity prediction. The best deep learning model was the 1D-CNN-BiLSTM trained with 2-stage training, which outperformed both the CNI linear regression baseline model and the Random Forest regressor model with the lowest mean-squared-error (MSE = 82.89) and mean-absolute-error (MAE = 6.15). The actual

cellularity vs. prediction plot has an R^2 value of 0.26, significantly higher than the R^2 value of 0.03 in the Random Forest model.

Table 5.3. Summary of regression results for all models in predicting sqrt(Cellularity)

MODEL	MSE	MAE	MAE low	MAE mid	MAE high
Predicting Sqrt(Cellularity)					
CNI Linear Regression	113.5	7.90	6.59	8.61	17.26
RF Regressor model	115.26	7.99	6.32	9.62	16.94
1D-CNN	111.12	7.53	5.47	9.65	17.33
1D-CNN-BiLSTM	113.78	7.89	6.11	9.55	14.75
1D-CNN-BiLSTM with sample weighting	100.68	7.05	5.22	10.17	13.77
1D-CNN-BiLSTM with 2-stage training	82.89	6.15	5.53	5.41	15.12
1D-CNN-BiLSTM with both sample weighting and 2-stage training	102.77	7.20	5.72	10.92	14.01

5.3.4. CTAI prediction

The bottom row of Figure 5.3 shows the samples of spectra with low and high CTAI values. Because CTAI is a direct calculation from Ki-67 and cellularity, spectra at the location of tissue samples also have similar characteristics to both other pathology metrics. Table 5.4 summarizes results for CTAI prediction. The best deep learning model was the 1D-CNN-BiLSTM trained with both sample weighting and 2-stage training, which outperformed both the CNI linear regression baseline model and the Random Forest regressor model with the lowest mean-squared-error (MSE = 0.042), mean-absolute-error (MAE = 0.159). Training a classification model before

finetuning the regression model allowed the model weights to be initialized correctly, thus improving performance in this case. The actual vs. predicted cellularity had an R^2 value of 0.29, significantly higher than the R^2 value of 0.036 in the Random Forest model. Furthermore, the slope of the plot was 0.95, very close to 1, suggesting a strong positive correlation between our prediction and the actual CTAI values, which was not observed in the case for either Ki-67 or cellularity.

Table 5.4. Summary of Regression Results for all models in predicting CTAI

Model	MSE	MAE	MAE low	MAE mid	MAE high
Predicting CTAI					
CNI Linear Regression	0.049	0.184	0.179	0.140	0.418
RF Regressor model	0.047	0.175	0.181	0.118	0.385
1D-CNN	0.046	0.171	0.121	0.148	0.425
1D-CNN-BiLSTM	0.042	0.165	0.144	0.117	0.313
1D-CNN-BiLSTM with sample weighting	0.045	0.174	0.156	0.131	0.349
1D-CNN-BiLSTM with 2-stage training	0.042	0.161	0.137	0.120	0.325
1D-CNN-BiLSTM with both sample weighting and 2-stage training	0.042	0.159	0.137	0.121	0.319

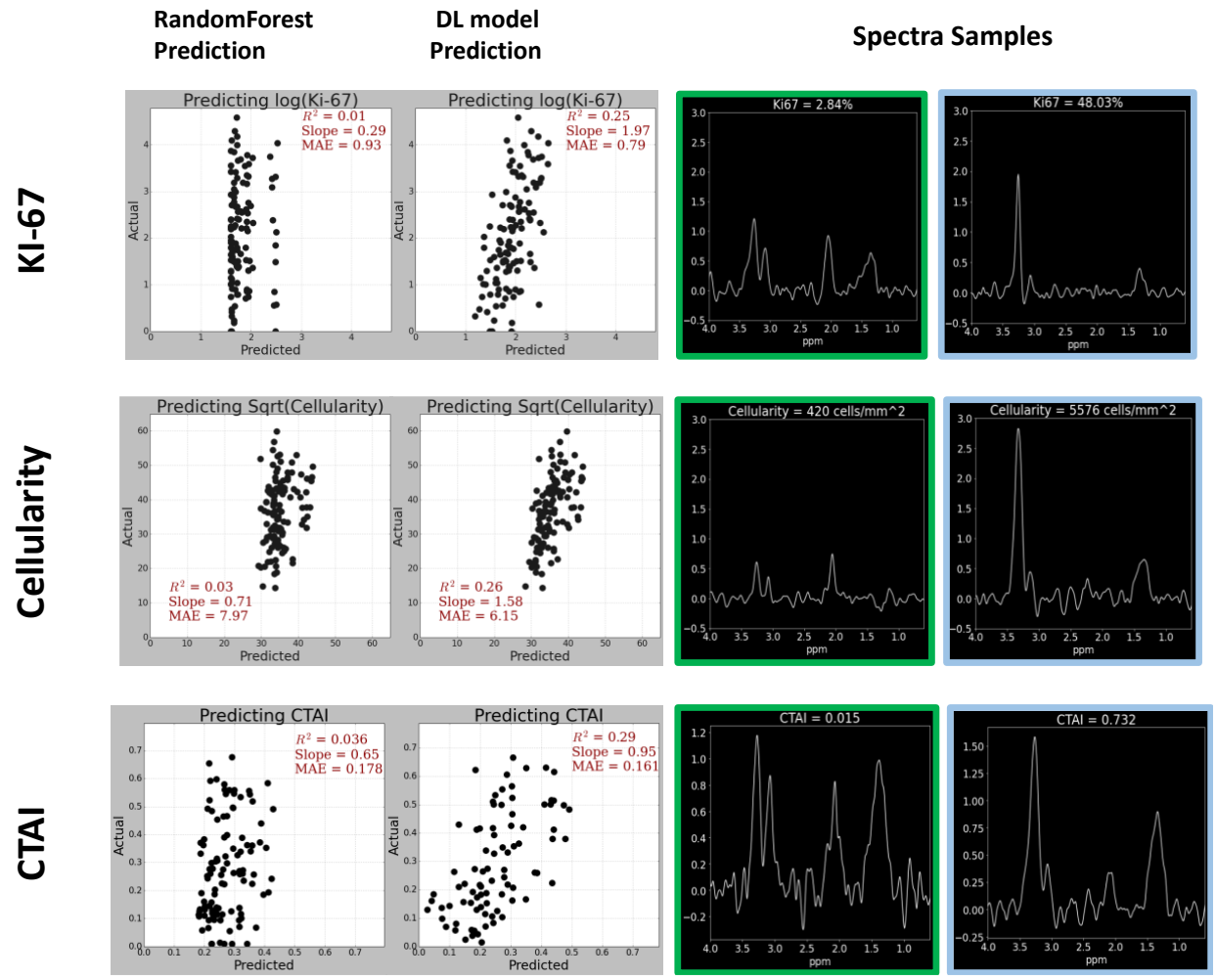

Figure 5.3. Main insights from the regression results
Deep learning models outperform Random Forest models in every instance, generating a prediction with much higher R^2 value. The CTAI model were the best with slope very close to 1. Green plots on the right show the full 1H-MRSI spectra of samples with low Ki-67, Cellularity, and CTAI respectively, while blue plots for high. In general, samples with high histopathology measurement tends to have high Choline and Creatine peaks; low NAA peak; or prominent Lactate peak.

5.3.5. Gradcam

Figure 5.4 shows Gradcam results highlighting where the model was able to correctly weight regions of high importance in the spectrum (near choline for highly aggressive samples, while NAA and creatine are highlighted for less aggressive samples). In samples where the model failed to correctly identify the true label, the network more heavily weighted regions of noise instead of focusing on areas of relevant metabolites.

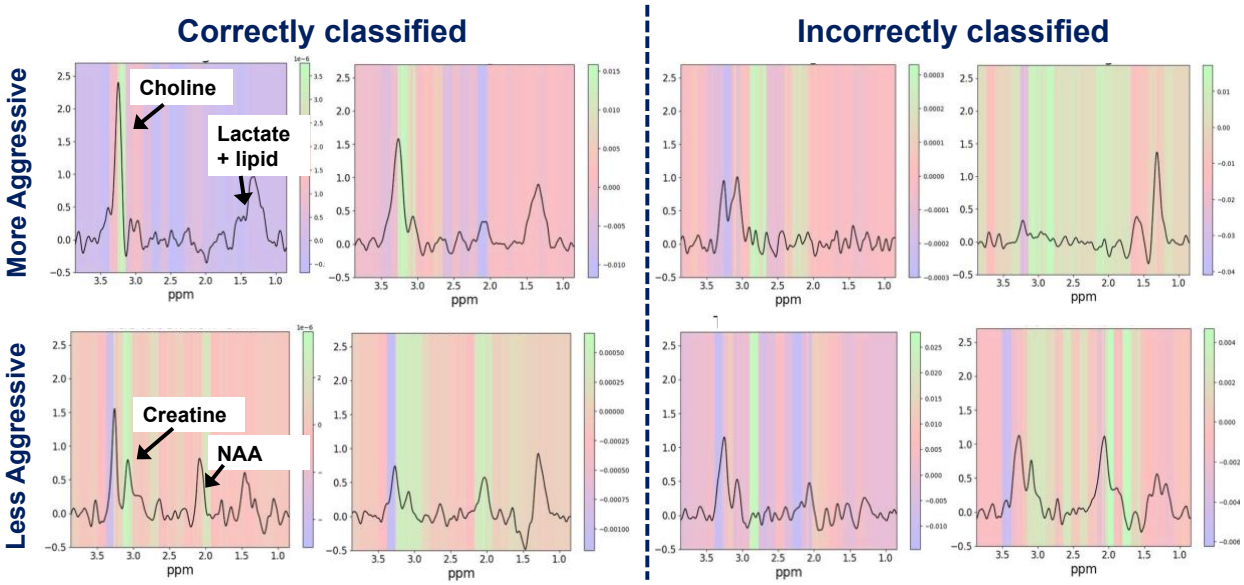

Figure 5.4. GradCam results
The model was able to correctly weight regions of high importance in the spectrum (near choline for highly aggressive samples, while NAA and creatine are highlighted for less aggressive samples). In samples where the model failed to correctly identify the true label, the weighting was not distributed correctly, with the network incorrectly focusing on regions of noise instead of metabolites.

5.3.6. Spatial map results

Figure 5.5 shows example spatial maps generated using the best deep learning model for 2 different patients. Tissue samples with higher proliferation, cellularity, and aggressiveness (fig 5.5b) compared to lower values in figure 5.5a. Overall, the spatial maps show areas of high Ki-67 and CTAI concentrated within tumor area, especially highlighting subregions of the contrast enhanced lesion that are more aggressive in the example in figure 5.5b. The predicted histopathology values at the location of the tissue samples within the spatial map are also similar to the actual histopathological metrics quantified from the tissue samples. Even though the overall range of predictions of our model is smaller, the resulting maps still capture relative differences and spatial heterogeneity within the lesion. Interestingly, the cellularity maps also highlight regions of abnormality beyond the anatomical tumor lesions.

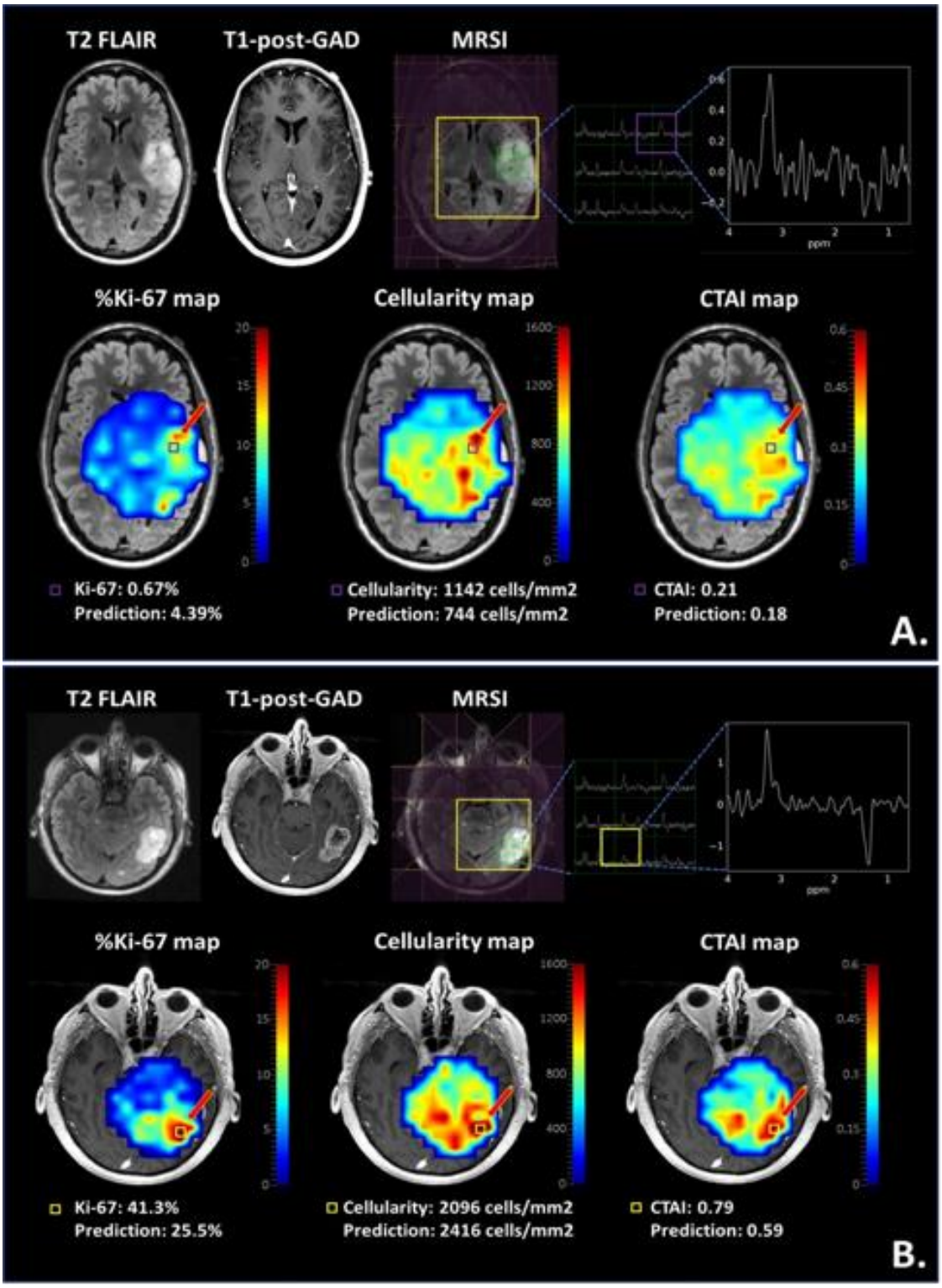

Figure 5.5. Spatial map results.
Spatial Map generation for two sample patients in (A) and (B) respectively. Overall, the model were able to correctly identify regions of high KI-67, Cellularity, and CTAI score as those within the contrast-enhanced and T2-lesions.

5.4. Discussion

Glioma is one of the most aggressive types of tumors, and outcomes for the most severe grade of glioma have not improved much in the last 30 years. Although obtaining tissue samples for histopathological and molecular analysis during surgical resection is the routine practice for generating a diagnosis, these samples are unfortunately only a small fraction of heterogeneous tumor tissue and do not capture the statistical distribution of biological properties that would allow for accurate characterization of tumor biology throughout the whole lesion, especially in what remains unresected. In this study, we have developed and tested a machine learning based tool using the 1D spectroscopy data that can reliably generate spatial maps of three different histopathological indices for every patient: the Ki-67, cellularity, and the newly derived CTAI. All of our models were built using 607 tissue samples from a large cohort of 281 patients newly diagnosed with glioma. To our knowledge, this is the first study to: a) use deep learning on 1D spectroscopy data to predict and generate the spatial brain maps for multiple histopathology indices; b) develop and explore a new histopathology index (CTAI) that combines the characteristics of both KI-67 and cellularity; c) incorporate multiple methods for training with imbalanced dataset, including the 2-stage-training and sample weighting; and d) visualize model performance using Gradient-weighted Regression Activation Mapping (Grad-RAM).

Our statistical analysis of the correlation between different MRSI markers and each pathology index shows strong agreement with what has been reported in the literature, with elevated choline being strong predictor for Ki-67, cellularity, and CTAI; lowered NAA being a strong predictor for cellularity; and elevated lactate being strong predictor for Ki-67. Additionally, increasing CNI and CCrI are good predictors for all 3 pathological indices. Spectra samples from figure 5.3 reflects this phenomenon, where a high choline peak and an abnormally low NAA peak are often found for tissues with both high Ki-67 and high cellularity.

Our machine learning and deep learning results showed that, for each of the 3 histopathology indices, the deep learning method using the full spectroscopy spectrum

outperformed both the conventional approach of linear regression over CNI data and the machine learning method using only metabolite peak heights and indices. This is expected, as the deep learning method utilized the full spectrum instead of just relying on just 1 or 2 quantified features, and also minimizing any errors due to processing. Furthermore, the 1D-CNN-BiLSTM architecture has the advantage of combining a CNN with Bidirectional Long-Short Term Memory to produce a model that can simultaneously capture abstract features and perform time series analysis, suitable for analyzing the full spectrum. We also observed that applying 2-stage training and sample weighting improved performance for prediction of all 3 histopathology indices, especially in lowering the MSE. Further evaluation by breaking down the MAE into MAE high, MAE mid, and MAE low that represented the MAE for tissue samples with high, mid, and low histopathology values helps explain this phenomenon. It appeared that applying 2-stage training and sample weighting particularly lowered the MAE for samples with high histopathology indices. As shown in the histopathology histogram in figure 5.1, these samples were less frequent, which created a highly imbalanced dataset. As a result, machine learning and the native deep learning model seemed to get stuck at local minima and unable to converge. This phenomenon was also reflected in the plot for the random forest prediction, where most predictions occurred near the median values of each histopathology index (figure 5.3). In the deep learning prediction plot, the deep learning model with the help of sample weighting and 2-stage training allowed predicted values to spread out more and better correlate with the true values. We hypothesized that, in the case of 2-stage training, training a classification model before finetuning the regression model would allow the model weights to be initialized correctly, thus improving performance. Similarly, sample weighting by inverse probability weighting allowed the model to focus on samples with high Ki-67, cellularity, and CTAI, thus improving performance in those regions.

We used Grad-Ram results (figure 5.4) to show both visualization of the results, as well as to provide some interpretability to the blackbox deep learning model. The regression activation map shows regions of the spectrum with high importance in determining the final prediction of the

model. We observed that for samples that were “correctly classified” (samples with % MAE < 5%), the models look at the correct area of the spectrum: higher weight is often assigned near the Choline peak, Creatine peak, or NAA peak. On the other hand, the model seems to be more confused when looking at spectra that is incorrectly classified, assigning higher weight to irrelevant areas of noise, and/or the spectra possesses unusual characteristics, such as a very high lactate/lipid peaks due to insufficient lipid suppression. We hypothesize that this could be the results of under-training due to insufficiently clean training data, or that the ^{1}H-MRSI is still a little noisy which can hinder model’s performance. Overall, the Grad-RAM results are promising, and with more cleaned training data, better performance can be achieved by the models.

We demonstrated how our results can potentially benefit clinical workflow by generating spatial maps for test patients. Overall, we observed that the spatial maps showed highest Ki-67 and CTAI concentrated within the area of CEL. The predicted histopathology values of tissue samples obtained from the spatial map were correlated with the actual histopathological values. Although the model under-predicted the actual values, the spatial maps still captured relative differences and spatial heterogeneity within the lesion. This can have major impact in determining which tissue samples to collect during diagnosis to best represent the full aggressiveness of the tumor, instead of relying on just the CNI value.

The CTAI score was developed in order to combine the characteristics of Ki-67, cellularity, and tumor score into one single index. Our plot from figure 5.3 shows that CTAI had better correlation with imaging markers, and potentially can reflect the aggressiveness of the tumor more than the individual Ki-67 or cellularity metrics. CTAI is designed to have a smaller range, mostly between 0 and 1, and its distribution is less skewed compared to Ki-67. As a result, learning CTAI was easier. The spatial map of CTAI also showed much less noise compared to either the Ki-67 or cellularity maps. That being said, CTAI needs to be further studied in order to establish its clinical implication.

Our study shows promise in using the full ^{1}H-MRSI spectra to predict tumor histopathology. Our dataset is one of the largest among all institutions that do similar analysis, with a larger percentage of tissues having high Ki-67 and cellularity. This is particularly important for all types of analysis and ML model development, as models can be trained using the fuller distribution of histopathology, which in turn will provide better generalization during inference. However, there are still several limitations to our study. First, while our dataset is one of the largest, it is still somewhat small for a regression task. This was reflected in the Grad-RAM results where models still sometimes focus more on noise and insignificant regions of the spectrum. This can be alleviated by more aggressive data augmentation to increase the dataset. We are also looking into pre-training a single network to first predict all metabolite levels. This is a more common and simpler task, and a massive amount of data can be used for training, since it does not require collection of tissue samples. Then, the whole network can be finetuned to predict histopathology. Another limitation of the study is the lack of other MRI markers, such as those in diffusion or perfusion data. This is by design, as we want to first look at the feasibility of using just the ^{1}H-MRSI data, before adding other MRI modalities. MRSI is quite unique in that the full spectrum can be used for training.

5.5. Conclusion

In conclusion, our work highlighted the use of ^{1}H-MRS data, both as individual metabolites and as the full 1D spectrum, to predict the tumor biology of tissue samples. Particularly, we developed a deep learning model that utilized the wealth of information contained in the entire spectrum to predict voxel-wise histopathology of tumor cells, including tumor cellularity, mitotic activity (Ki-67), and a new composite tumor aggressiveness index (CTAI; defined as the sum of normalized cellularity and Ki-67 times tumor score) using tissue samples with spatially mapped coordinates on 3D ^{1}H-MRSI. We showed that deep learning using the full 1D 1H-MRS data can improve model performance, especially when first training a classification model and finetuning

using a regression model. Even though metabolic data is the only MRI parametric used in this paper, which limits its power, we still achieve very promising results: our best model achieves an MAE within 10% of the range of the targets. Using the models developed, we were able to generate predictive maps for each of the histopathology metrics, which can be incorporated into not only diagnosis workflow, but also disease monitoring and treatment planning. In the future, we are looking into adding more patients to our dataset, using a different type of data normalization, and utilizing other additional MRI modalities to produce a fully working model.

Chapter 6. Defining radiation target volumes for glioblastoma (GBM) and predicting tumor recurrence with machine learning using pre-radiotherapy anatomical, diffusion & metabolic MRI

6.1. Introduction

GBM is a highly malignant, heterogenous, and invasive type of brain tumor, with a poor overall survival of 12-15 months [173]. Current standard of care (SOC) treatment of GBM begins with maximal safe surgical resection to remove the gross tumor volume. Because lesions are rarely fully removed by surgery, patients also typically receive external beam radiotherapy (RT) (a total dose of 60 Gy in 2 Gy fractions over a course of 6 weeks), daily temozolomide (TMZ) chemotherapy (75 mg/m^2 per day), and six cycles of maintenance adjuvant TMZ chemotherapy (total 150-200 mg/m^2). This protocol has been shown by Stupp et al. in 2005 to have a prolonged survival benefit. Since then, however, despite decades of clinical trials incorporating novel systemic agents and more aggressive surgical approaches, only minimal improvements in outcome have been achieved, mainly due to the difficulty in both identifying and targeting the highly malignant tumors to the full extent, while also sparing normal brain tissue to preserve brain function[174].

While recent advances in RT delivery can provide millimeter-scale precision and dose modulation, current RT treatment planning only utilizes post-contrast T1-weighted MRI and the hyper-intense signal from T2-weighted FLAIR MRI. The empirical RT dose is still often delivered uniformly as a 1-2 cm geometric expansion of the gross tumor volume (GTV) defined by lesions visible on these anatomical MRI images, without considering the spatial heterogeneity and infiltrative nature of this disease. This has the unintended consequences of undertreating subclinical disease, as well as unnecessarily irradiating normal brain tissue, adversely affecting clinical outcome, and increasing toxicity. While most of the tumor progression happens locally

within the 2 cm expansion of the hyperintense lesion from T2-weighted images, partially due to the consensus tendency to overtreat [175], tumor progression occurs beyond the high-dose treatment for about 10-37% of patients [46]–[49], [176]. At the same time, about 60% of irradiated tissue are normal-appearing brain tissue [2], causing neurotoxicity, which can negatively affect a patients' cognitive function, quality of life, and overall survival (OS) [177], [178]. To further over-complicate the matter, the introduction of anti-angiogenic treatment has altered the pattern of tumor progression, with non-enhanced tumor progression become more prevalent than previously observed. Thus, a better strategy to re-define RT target volumes by identifying regions with higher probability of tumor progression has great potential to improve outcomes as well as quality of life for patients with GBM.

Recent advances in diffusion-weighted and metabolic MRI have allowed precise voxel-level visualization and characterization of subclinical tumor tissue, and previous studies have shown that markers from diffusion and metabolic MRI can help identify voxels at risk for progression [8], [179]. However, these markers remained mostly unused in RT treatment planning outside of a few recent single arm phase II clinical trials [105], [180]–[184]. Subclinical tumor invasion, which causes an increase of edema and decrease in directionality along white matter tracts, can be reflected by an increase in apparent diffusion coefficient (ADC) and a decrease in fractional anisotropy (FA) using diffusion tensor imaging (DTI). DTI has also been shown to identify directionality of new tumor progression along white matter tracks [185]–[187]. Metabolite levels estimated using proton Magnetic Resonance Spectroscopy (^{1}H-MRS) and the derived Choline-to-NAA index (CNI) can help measure underlying cellular metabolism associated with infiltrative tumor [4], [5], hypoxia [6], as well as tumor growth and progression[7]–[9]. This is very promising, as several recent studies have begun to incorporate DTI tractography into presurgical assessment of patients with GBM [188], as well as CNI and Lactate-to-NAA ratio of 1H-MRSI data in simulating RT treatment plans, although in a very small patient cohort [105], [180], [181], [184]. Furthermore, our institution has previously reported that the combination of these advanced

imaging techniques can provide even more information into a tumor's malignant behavior, and can potentially predict tumor progression and assist with RT planning [9]. However, to our knowledge, despite the vigorous efforts to incorporate these advanced imaging techniques to study tumor progression and predict tumor behavior, only one study to date has successfully attempted an integrated voxel-level based approach to improve the clinical treatment volume (CTV) definition for precision RT treatment planning based on multi-parametric MR images on a large cohort. This retrospective study by Heo et al. [182] reported a sensitivity of 0.80 using diffusion and perfusion-weighted MRI markers from 88 early progression patients, but suffered from poor specificity (0.29), especially for those with later progression.

The growing use of machine and deep learning applied to brain tumor imaging applications has allowed tasks such as lesion segmentation, diagnosis of molecular, and outcome prediction, to be performed with high speed, accuracy, and precision. However, the use of AI in voxel-wise prediction of tumor progression remains understudied, due to the complexity of the problem, the lack of labeled patient data, and the difficulties in producing appropriate ground truth data that are correctly aligned to the input images. Interpretation of the results and evaluation of model performance is even further lacking. The goal of this study is to use multiparametric MRI at prior to RT along with machine and deep learning to predict regions of subsequent tumor progression for precision-based RT planning, and then compare the resulting predicted maps to the standard of care 2cm uniform expansions of anatomical lesion volumes for defining the clinical target volume in RT planning. The results presented offer several improvements over prior studies due to: 1) our inclusion of a larger number of patients who were treated with various types of therapy (SOC plus a subset with concomitant anti-angiogenic therapy) and experienced progression of both the contrast-enhancing lesion (CEL) and the less common non-enhancing lesion (NEL) due to the anti-angiogenic agents often obscuring the classic presence of contrast enhancement; and 2) specific methodological enhancements employed specifically for this problem. The latter includes: 1) performing temporal alignment by utilizing state-of-the-art alignment tools specifically

designed to account for tissue shift that occurs after surgical resection of a glioma [189]; 2) training a model that utilizes a 3D UNET deep learning architecture, allowing spatial resolution to be preserved and enabling us to directly generate a predicted target volume; 3) developing a novel approach for optimizing, evaluating, and visualizing model performance by incorporating new loss functions and evaluation metrics. We hypothesize that this comprehensive strategy will result in a more biologically-relevant definition of RT target volumes based on the true extent of infiltrating tumor in patients with GBM that will more closely cover the progressed lesion and minimize dose to healthy brain tissue.

6.2. Materials and Methods

6.2.1. Patient cohort

A total of 72 patients who were newly-diagnosed with primary GBM according to WHO 2014 criteria were included in this retrospective study. All patients received SOC treatment, including surgical resection followed by external beam radiotherapy (RT) (a total dose of 60 Gy in 2 Gy fractions over a course of 6 weeks), daily temozolomide chemotherapy (TMZ) (75 mg/m2), and six cycles of maintenance adjuvant TMZ chemotherapy (total 150-200 mg/m2). Of these 72 patients, 24 received no additional treatment, while the rest of the patients were treated with an additional anti-angiogenic agent: 26 with enzastaurin (250 mg daily), and 22 with erlotinib (150 mg/day continuously or 500 mg/ day continuously if on anti-epileptic drugs starting on day 1 of radiotherapy) and bevacizumab (10 mg/kg every 14 days starting in week 2 of radiotherapy). All patients gave informed consent according to established guidelines by the Institutional Review Board (IRB).

All patients received a baseline MRI scan (post-surgical resection but pre-radiotherapy and chemotherapy) that included at least pre- and post-contrast T1- and T2-FLAIR images, diffusion-weighted images, and MRSI. After the course of radiotherapy and chemotherapy,

patients were followed with multiple MRI scans about every two months (including at least pre- and post-contrast T1-weighted and T2-FLAIR imaging), until progression. To confirm the absence of pseudo-progression, the clinical histories of patients who either progressed within 12 weeks of the completion of radiotherapy or had a suspect scan followed by stable disease were centrally re- reviewed by a neuro-oncologist. If reoperation was performed, true progression in the location of recurrence was confirmed according to the recommendations of Wen et al [190]. Based on these criteria, none of the patients in this study exhibited pseudoprogression.

6.2.2. Image Acquisition

MR examinations were performed on a 3T GE Signa scanner using an eight-channel phased-array head coil. Standard anatomical imaging included T2-weighted FLAIR and 3D T1-weighted IR-SPGR imaging pre- and post- the injection of a gadolinium-based contrast agent. Diffusion-tensor images (DTI) were obtained with b=1000s/mm^2, 6- directional axial diffusion-weighted echo-planar imaging (EPI) sequence, and 4 b_0 excitations (TR/TE=1000/108ms, voxel size=1.7-2.0×1.7-2.0×2.0-3.0mm). Lactate-edited 3D ^{1}H-MRSI were acquired using point resolved spectroscopy (PRESS) volume localization and very selective saturation (VSS) bands to avoid chemical shift artifacts as well as to suppress residual lipid signals (excited volume =80×80×40mm, repetition time = 1100-1250ms, echo time = 144ms, overPRESS-factor = 1.5, nominal voxel size = 1×1×1cm, flyback echo-planar readout in SI, total acquisition time = 9.5 minutes, 988Hz sweep-width, and 712 dwell-points).

6.2.3. Pre-RT Exam Image Processing

From the pre-RT scans, relevant anatomical, diffusion, and metabolic maps were generated for all patients, described as follows. For DWI data, ADC and FA were calculated on a voxel basis by applying the FMRIB's Diffusion Toolkit [191]. To allow for cross-patient analysis,

the ADC and FA maps were then normalized to the mode of intensities in normal-appearing brain tissue (calculated using the entire brain volume after subtracting the CEL and NEL). For MRSI data, we applied previously described preprocessing methods to generate metabolite peak heights maps [6], [171] on a voxel-to-voxel basis from the 3D spectral data. Maps of NAA, Choline (Cho), Creatine (Cre), Lactate (Lac), and Lipid (Lip) were generated and normalized to the median peak height of normal-appearing brain voxels. The Choline-to-NAA index (CNI), Choline-to-Creatine index (CCrI), and Creatine-to-NAA index (CrNI) were also calculated as maps that reflect changes of the two relevant metabolites compared to normal brain [5].

All images from the pre-RT timepoint, including anatomical images (T2-weighted FLAIR, T1-weighted pre-GD) and physiological parametric maps from diffusion MRI were rigidly aligned to the post-Gd T1-weighted image using Slicer's BRAINSFit tool with B-spline warping [172], or FMRIB's FSL Linear Image Registration Tool (FLIRT) [192], [193], and then, along with metabolite maps, resampled to an isotropic 3x3x3 mm voxel resolution to account for inter-exam alignment error. Regions of interest (ROIs) included the CEL, NEL (defined as CEL subtracted from the T2L on T2-FLAIR images), and normal appearing voxels (NAV), defined as normal brain tissue from a skull-stripped brain mask obtained from HD-BET brain extraction tool [194] after subtraction of cavity, ventricles, and lesion ROIs. Voxels in the resection cavity region were excluded from all ROIs. CEL and NEL ROIs were semi-automatically defined on the pre- and post- contrast T1-weighted images (CEL) and T2-weighted FLAIR images (T2-hyperintense lesion).

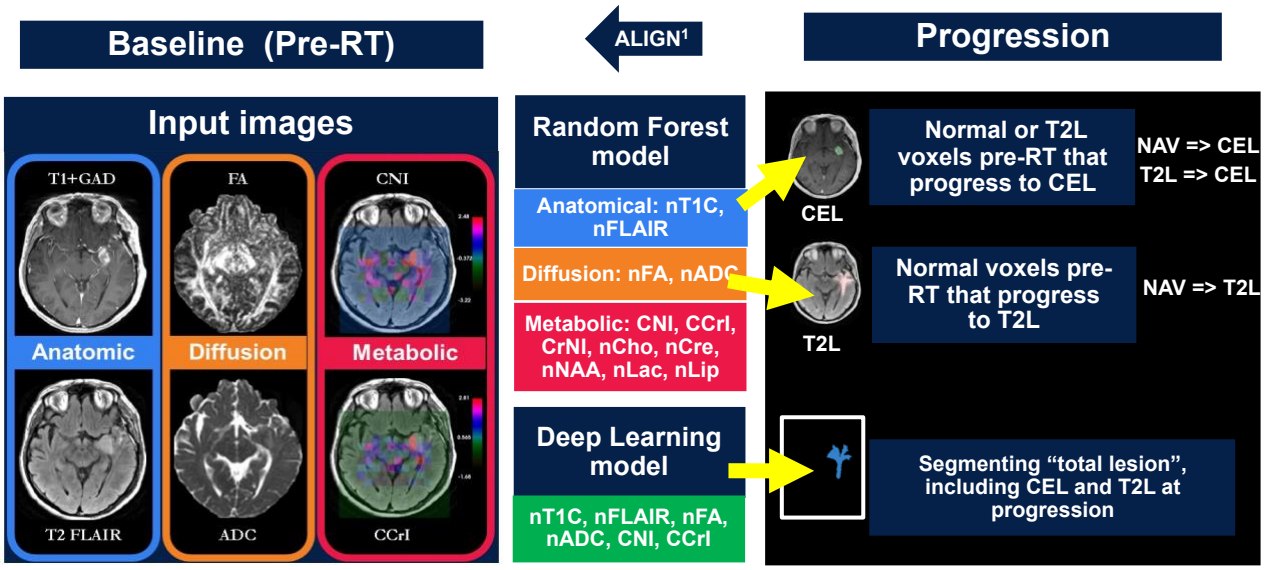

Figure 6.1. Study Schema
Images from progression scans were aligned to the baseline scan. Multi-parametric MRI from the baseline scan were used as input for both random forest and deep learning model (left). Two random forest models were trained, one to predict CEL progression and one to predict T2 lesion progression. The deep learning model was trained to segment the entire lesion at from both the pre-RT and progression time points combined in order to define the target volume for radiotherapy.

6.2.4. Inter-Exam Image Registration and Progression Exam Image Processing

We tested three different methods to register the images from the progression time point to the images from the preRT time point: 1) direct nonrigid registration using T1-weighted pre-contrast images; 2) multistep registration, where T1-weighted pre-contrast images of intermediate time points (i.e., 1 month or 2 months follow-up after the start of radiotherapy) are first nonrigidly aligned to the pre-RT time point, before being aligned rigidly to the progression time point; and 3) a deep learning method specifically trained on serial post-resection glioma data with tissue shift described by Mok et al. [189], where T1-weighted post-contrast and T2-weighted FLAIR images from both the pre-RT and progression time points were used as inputs. Visual analysis showed that the deep learning registration method outperformed the other two methods, and thus were used for all patients in this analysis. All images and ROIs from the progression scan were then transformed using the transformation matrix produced by the deep learning registration method,

providing a voxel-to-voxel mapping from the progression to the baseline scan. All images (from both the pre-RT and progression time point) were then resampled to 3mm x 3mm x 3mm to mitigate any errors caused by the alignment process.

6.2.5. Voxel Classifications and Statistical Analysis

Using the lesion ROIs that were manually segmented, voxels from both pre-RT and progression scans were classified as either CEL, NEL, or NAV and grouped into the following categories: 1) stable NAV (NAV→NAV), 2) progressed NAV (NAV→NEL, NAV→CEL), 3) stable NEL (NEL→NEL), 4) progressed NEL (NEL→CEL), 5) stable CEL (CEL→CEL). There were also a small number of voxels that were lesion at baseline and became NAV at time of progression, due to either treatment or misalignment issues that were excluded from this analysis.

To assess if there was a significant difference between stable NAV (NAV→NAV) voxels and progressed voxels (NAV→NEL, NAV→CEL, and NEL→CEL), we calculated the median value of each parameter map for each category per patient, ensuring that there were at least 5 voxels for each group. Only voxels within a 4cm expansion of the pre-RT lesion were evaluated to reduce class imbalance. A Mann-Whitney-U signed rank test was applied to compare between normal and progressed voxel groups. Significant levels were selected at p-value = 0.05, 0.01, 0.001, and 1e-4.

6.2.6. Machine Learning

To perform voxel-wise prediction of progression, we trained and tested two separate random forest models to classify: 1) stable NAV (NAV→NAV) vs. NEL progression (NAV→NEL); and 2) stable NAV (NAV→NAV) vs. CEL progression (NAV→CEL, NAV→CEL). The inputs of each model included the following: 1) normalized maps of T1C, T1, and T2-FLAIR from anatomical MRI; 2) normalized-ADC and normalized-FA from diffusion-weighted MRI; 3)

normalized metabolite maps of Cho, Cre, NAA, Lac, Lip; and 4) CNI, CCrI, CrNI from MRSI. Each model was trained and tuned to maximize the area under the curve (AUC) of the receiver operating characteristic (ROC) curve using 5-fold cross-validation (CV), and the CV results were recorded and averaged. CV were done using a patient-wise stratified splitting method with train/test ratio of 70/30%, where for each fold in the CV split we ensured that voxels from the same patient couldnot co-exist in both training and test folds. The average ROC-AUC score was then used to compare between models. To determine if time-to-progression played a role in our voxel-wise progression predictions, we split our patients based on median progression time (11 months) into an early progression group (patients who progressed before 11 months), and late progression group (patients who progressed after 11 months) and repeated the above random forest model training for each sub-group.

6.2.7. Deep learning

A deep learning segmentation-based model was also trained to generate hypothetical target volumes. Since clinical target volumes need to include both current tumor voxels in addition to our proposed current normal-appearing voxels based on anatomical imaging that are of highest risk to progress, we designed the deep learning task to segment a composite lesion mask of the NEL and CEL from both pre-RT and progression time points using only images from the pre-RT MRI scan. This segmentation task was performed using a 3D 4-staged U-Net architecture like the ones described in Cicek et al.[195] and Henry et al.[196] There were 4 stages in the encoder part of the network, and the numbers of filters were 48, 48, 96, and 192 for each stage, respectively. Convolution layers at each stage were always followed by a group-normalization layer (instead of batch normalization), to keep the batch size small, which has been found to result in better performance for medical image-based tasks[196], [197], and RELU activation layers. Down-sampling was performed using Max-pooling layers of size 2x2x2 and stride 2. The decoder part was symmetrical to the encoder part, and up-sampling between stages was performed using

trilinear interpolation. Shortcut connections were added between the encoder and decoder of the same stage via concatenation. Finally, a sigmoid activation layer was added before the output layer, which is a combined segmented tumor mask. Figure 6.2 depicts the model architecture used in our deep learning task. Models were optimized using Ranger optimizer introduced by Wright et al. [198], which combines both the Rectified Adam [199] and Lookahead [200] optimizers into one. The inputs of the deep learning models were chosen among the following images: maps of ADC, FA, CNI, CCrI, T1C, and FLAIR at the pre-RT time point. CNI and CCrI were chosen to represent MRSI modality due to their higher significant level in predicting progressed voxels in both statistical analysis and machine learning steps. All models were trained using 58 patients and tested using the remaining 14 patients. During training, patients were further split into training set (46 patients) and validation set (12 patients). To ensure the consistency needed for model evaluation, we kept both the validation and test sets the same for all models.

Loss function study: To study how different loss functions affect the model performance, we trained and tested the models using the following losses: Dice loss, Tversky loss (with varying α and β values), focal Tversky loss (with varying α, β, and γ values). In addition, we designed a loss function based on Tversky loss, but dynamically varying α and β values for each patient depending on the size of the patient's tumor called the Individualized Progression Coverage Coefficient (PCC) calculated using the following equation:

$$PCC = \frac{TP}{TP + \alpha FP + \beta FN} \tag{6.1}$$

where $\beta = \frac{1}{f+1}$, α = 1- β , f = $\frac{n_{lesion\_voxels}}{n_{brain\_voxels}}$. Our rationale was that patients with smaller lesions equate to more highly imbalance dataset and may require a higher β value to reduce false negatives and improve sensitivity. Thus, varying α and β based on lesion size would provide a way to account for this imbalance and determine more-optimal thresholds for tolerance in the model, potentially improving its performance. A compound loss function created by adding a varying levels of binary cross-entropy (BCE) loss to the PCC loss was also evaluated. All models

were trained using images from 3 MRI modalities (i.e. the maps of ADC, FA, CNI, CCrI, T1C, and FLAIR) at the pre-RT time point, and models were optimized separately. Wilcoxon rank sum tests were used to statistically evaluate difference between models trained using different loss functions. Significant levels were selected at p-value = 0.05, 0.01, and 0.001.

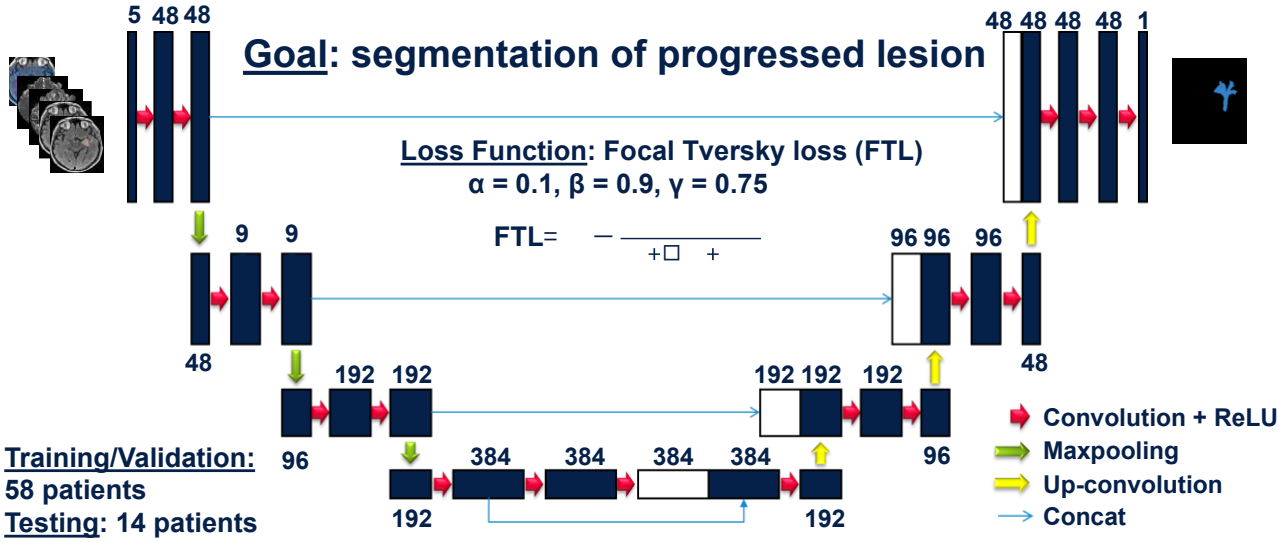

Figure 6.2. Deep learning model architecture
The chosen network used an encoder-decoder architecture, heavily inspired by the 3D U-Net architecture from Cicek et al. The encoder had four stages. Each stage consisted of two 3x3x3 convolutions. The decoder part of the network was almost symmetrical to the encoder. Shortcut connections between encoder and decoder of the same stage were performed by concatenation. After experimenting with several different options, the best loss function observed for this task was the Focal Tversky loss as shown.

MRI modality study: To study how using different image modality can affect the model performance, we trained and tested the models using the following groups of inputs: 1) only anatomic images (pre-RT T2-FLAIR and T1C); 2) anatomic + diffusion images (pre-RT T2-FLAIR, T1C, FA, and ADC); 3) anatomic + MRSI images (pre-RT T2-FLAIR, T1C, CNI, and CCrI); and 4) anatomic + diffusion + MRSI images (pre-RT FLAIR, T1C, FA, ADC, CNI, and CCrI). All models were trained using the PCC + BCE loss, and models were optimized separately. Wilcoxon rank

sum tests were used to statistically evaluate difference between models trained using different loss functions. Significant levels were selected at p-value = 0.05, 0.01, and 0.001 respectively.

6.2.8. Model Evaluation

To evaluate the performance of our hypothetical target volume (HTV) generated using the deep learning method and multi-modal MRI, we compared standard evaluation metrics between our HTV with other CTVs, namely 1) a HTV that only treat the tumor presented at time pre-RT; 2) a HTV similar to the SOC RT CTV, with GTV including the combined CEL and NEL at pre-RT + a 2 cm expansion of GTV; and 3) a HTV generated using the same deep learning method, but trained only using the anatomical images. Evaluation metrics included sensitivity, specificity, Dice coefficient, Tversky coefficient with α = 0.05 and β = 0.95, and the newly-derived individualized PCC described above.

6.3. Results

6.3.1. Patient characteristics

Patient characteristics are shown in Table 1. The full patient's cohort median PFS and OS are 7.0 and 17.6 months, respectively. There was a significant difference ($p = 0.01$, Mann-Whitney U test) in OS between SOC+Bevacizumab cohort (median = 20.3 years) and SOC cohort (14.6), but no significant difference was found in OS between SOC+Bevacizumab and SOC+Enzastaurin ($p = 0.14$), or SOC+Enzastaurin and SOC ($p = 0.22$). However, there was a highly significant difference in PFS between SOC+Bevacizumab and SOC+Enzastaurin ($p = 0.002$) and SOC+Bevacizumab and SOC ($p < 0.0001$). PFS of SOC+Enzastaurin patients was also significantly longer than the SOCcohort (7.1 vs 4.7, respectively; $p = 0.02$).

Table 6.1. Patients characteristics
Of 72 patients used in this study, 25 received SOC treatment, 23 received SOC + Bevacizumab, and 24 received SOC + Enzastaurin

Patient cohort	All patients	SOC	SOC+Bevacizumab	SOC+Enzastaurin
N	72	25	23	24
Age (year)				
Median	53	53	54	57
Range	[25 - 77]	[27 - 77]	[28 - 75]	[25 - 70]
OS (months)				
Average	23.4 $\pm$ 24.3	18.0 $\pm$ 11.6	40.1 $\pm$ 43.6	19.7 $\pm$ 10.3
Median	17.6	14.6	20.3	17.9
Range	[5.8 – 139.7]	[5.9 – 63.9]	[5.8 – 139.7]	[8.7 – 54.0]
PFS (months)				
Average	8.2 $\pm$ 5.7	5.2 $\pm$ 3.3	13.0 $\pm$ 5.9	8.4 $\pm$ 6.1
Median	7.0	4.7	12.0	7.1
Range	[0.8 – 25.9]	[0.8 – 11.7]	[1.9 – 25.9]	[1.8 – 23.0]

6.3.2. Statistical analysis

Figure 6.3 shows the statistical analysis of all voxels within a 4cm boundary of the original lesion. Statistically significant differences ($p \leq .0001$) between stable NAV voxels vs. both NAV→CEL and NEL→CEL voxels were observed using the median nFLAIR, nADC, nFA, CNI, and CCr. Highly statistically significant differences between the stable NAV voxels vs. NAV→NEL voxels were observed using the median nFLAIR, CNI, and CCrI ($p<0.0001$) while increases in median nADC also reached statistical significance of $p<0.05$.

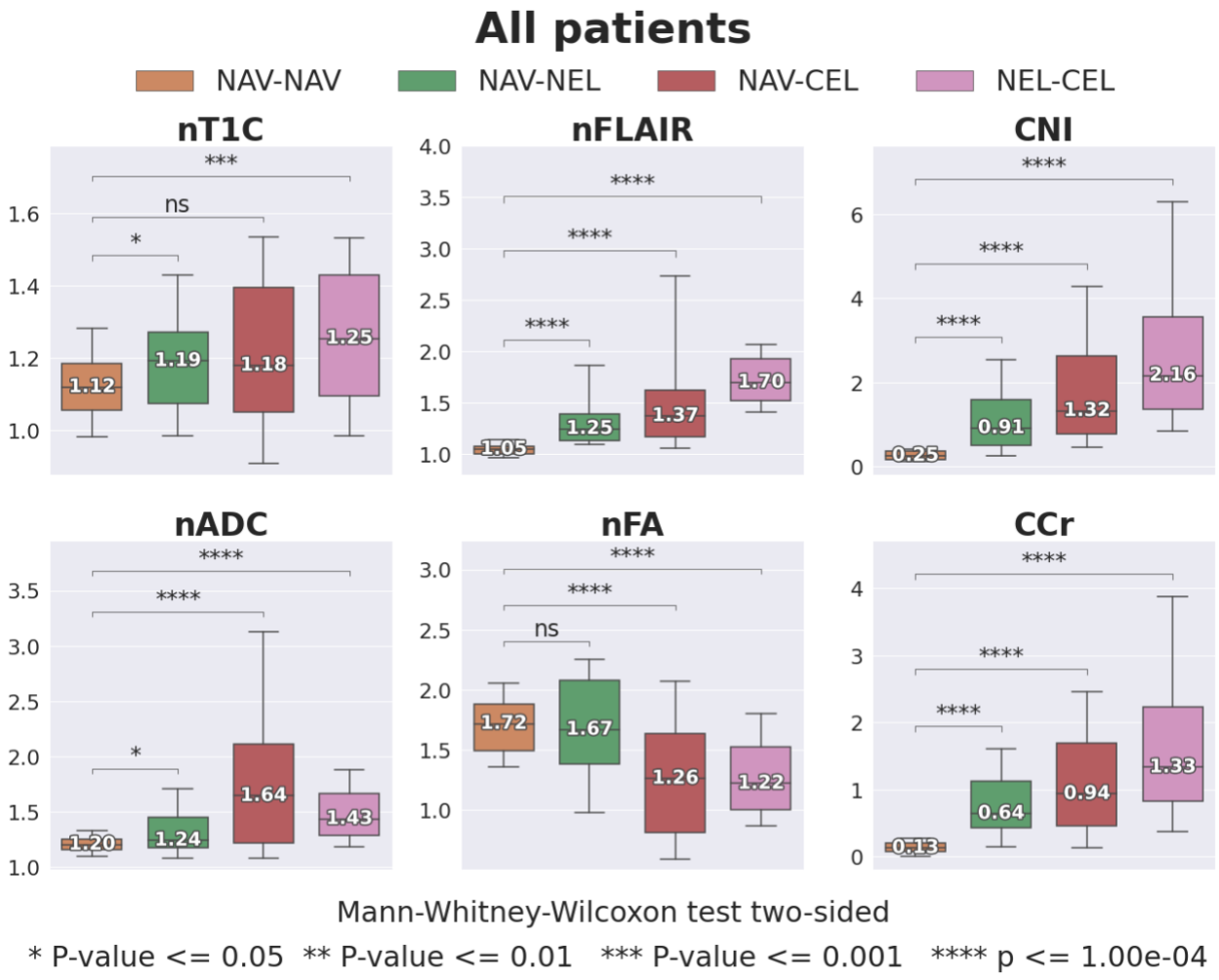

Figure 6.3. Region level analyses
Statistical comparison of MRI parameters between stable voxels (NAV→NAV) progressed voxels (NAV→NEL, NAV→CEL, NEL→CEL)

6.3.3. Machine learning

Figure 6.4 shows the 5-fold CV ROC curves for voxel-based predictions of CEL (top row) and NEL (bottom row) progression. The best random forest model to predict stable NAV vs. NAV-CEL and NEL-CEL progression voxels achieved a mean AUC of 0.88 when combining all patients. Interestingly, when splitting patients by median time to progression (t = 11 months), the CEL-progression model performed very well for patients who progressed before 11 months (AUC = 0.94) but performed poorly (AUC=0.74) for patients who progressed later. Similarly, the best random forest model to classify stable (NAV→NAV) vs. non-enhancing (NAV→NEL) progression achieved a mean AUC of 0.81 when combining all patients. Again, when splitting patients by median time to progression, the non-enhancing progression model performed better for patients who progressed before 11 months (AUC = 0.83) than for patients who progressed later (AUC = 0.73), although the difference was not as drastic as the CEL-progression model.

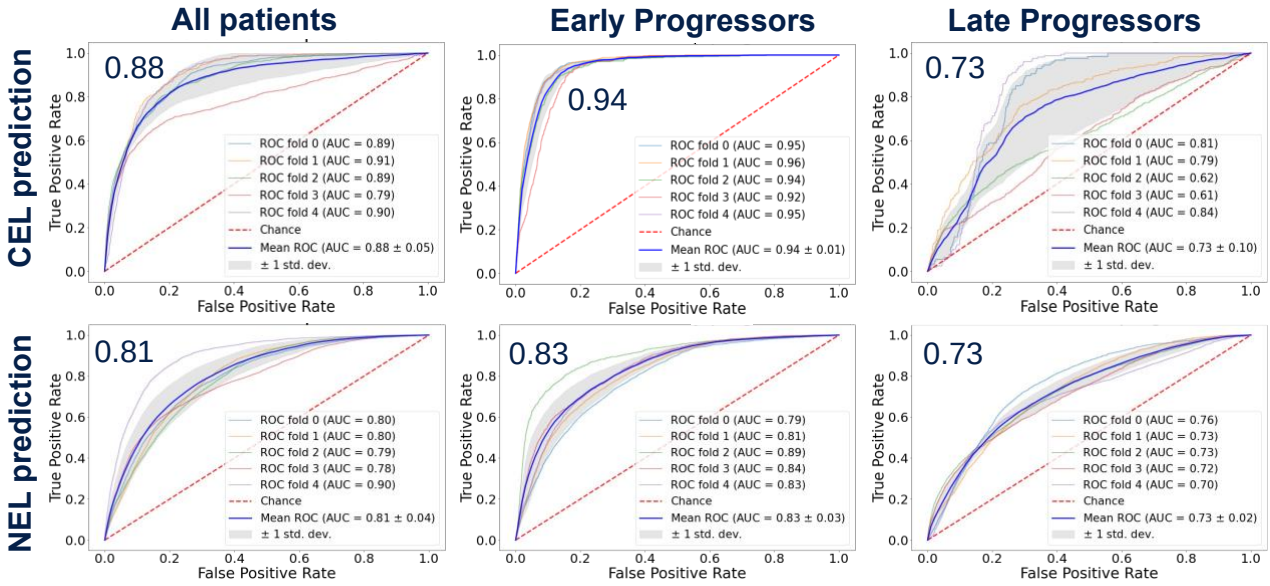

Figure 6.4. 5 fold cross-validation ROC plots and AUC results
Random-Forest models were able to predict subsequent contrast-enhancing lesions (CEL – top row) and non-enhancing lesions (NEL – bottom row) progression. Highest AUC was achieved for predicting early progressors (t < 11 months), especially for CEL recurrence.

Most important features for the CEL-progression model for all patients are the CNI, CCrI, nLipid, and the nFLAIR, and the most important features for the NEL-progression model are the CCrI, CNI, and nFLAIR.

6.3.4. Deep learning model optimization

Figure 6.5 shows how various loss functions affect model performance. As shown in the validation performance during training, using the combination loss (PCC+BCE or PCC+0.5BCE) achieve the best performance, where the model was able to converge quickly, followed by the model using only PCC loss. When observing model performance on the test set, we found these loss functions to have significantly higher sensitivity at the expense of a small drop in specificity. Tversky and PCC scores for these models were also high. In contrast, models trained using dice loss and Tversky loss with higher α values (less than 0.1) were observed to be undertrained. While specificity can be slightly higher for these models, sensitivity was much lower than desired for a medical imaging segmentation task.

Figure 6.6 shows a similar analysis, but we compared models trained using different combination of MRI modalities. The validation curves in (A) show a less drastic difference compared to figure 6.5, but overall, we observed that the model using anatomical + diffusion + MRSI achieved the most sufficient learning. This was further validated during inference, where this model also achieved the best sensitivity and PCC score, while keeping the specificity level relatively high. Furthermore, there was also less variation between patients compared to other models. Models trained without MRSI (using only anatomic images and anatomic + diffusion images) did not achieve optimal performance during training, and therefore achieved much lower sensitivity during testing, as well as higher variation between patients.

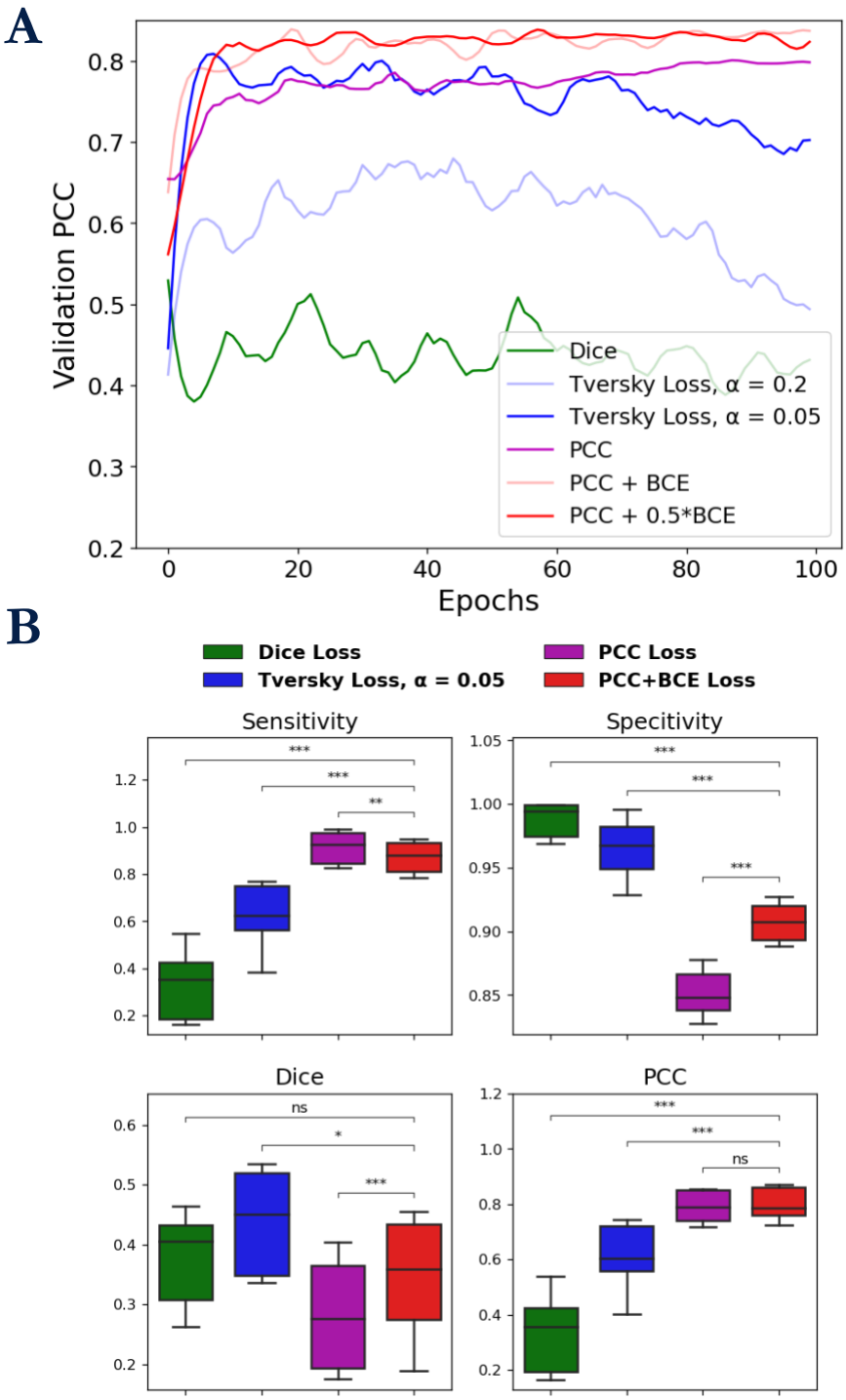

Figure 6.5. Comparing performance of models with different loss functions
A) The PCC score during training of all models for patients in validation set; and B) The sensitivity, specificity, Dice, and PCC of all models for patients in the test set. Wilcoxon rank sum test was used, and significant level was defined as (*, **, *** for p-value < 0.05, 0.01, and 0.001 respectively). All models were trained and optimized separately using anatomic + diffusion + MRSI input. Models trained using combination loss function PCC + BCE achieved much better performance than any other loss functions.

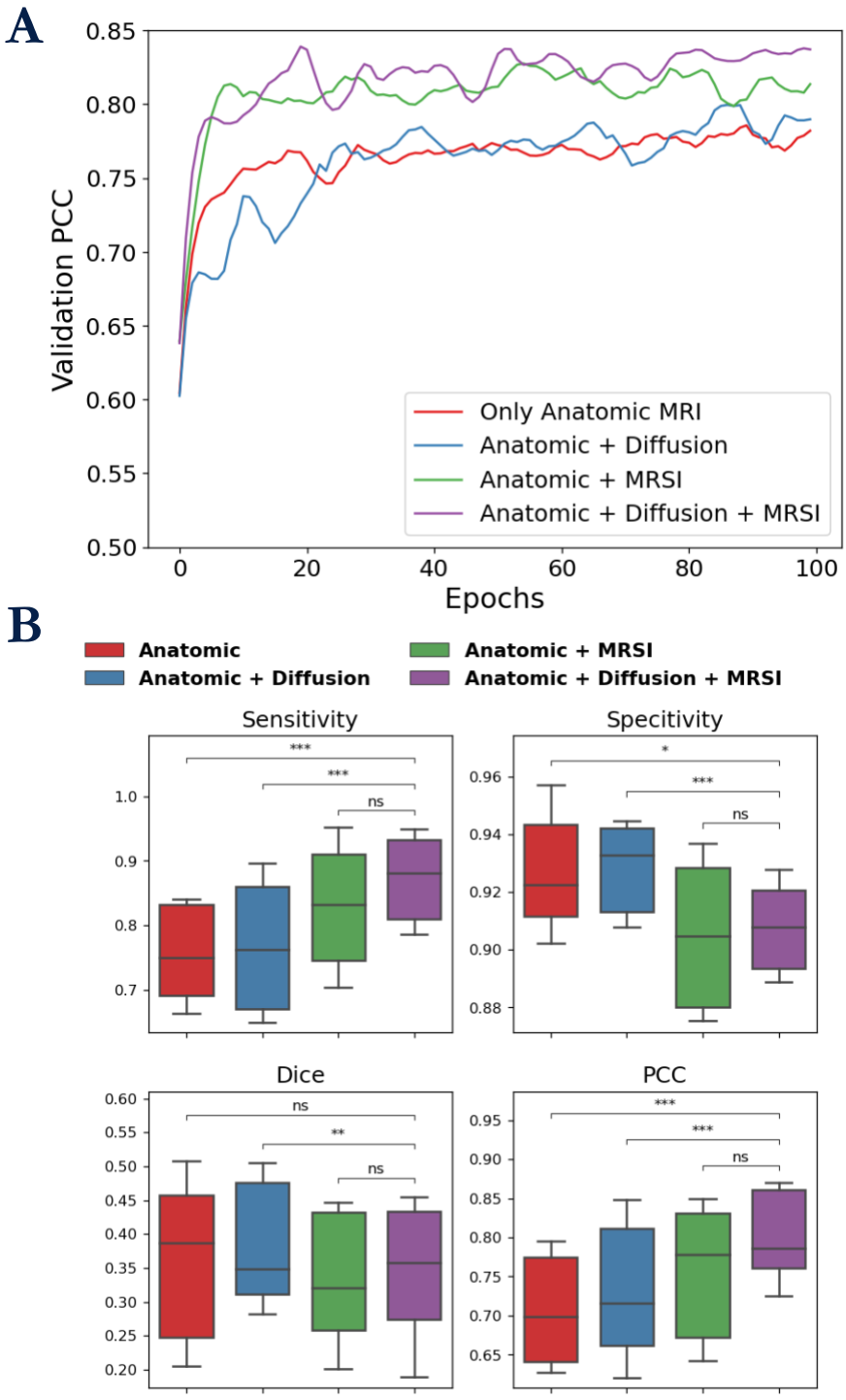

Figure 6.6. Comparing performance of models with different MRI input modalities
A) The PCC score during training of all models for patients in validation set; and B) The sensitivity, specificity, Dice, and PCC of all models for patients in the test set. Wilcoxon rank sum test was used, and significant level was defined as (*, **, *** for p-value < 0.05, 0.01, and 0.001 respectively). All models were trained and optimized separately using the PCC + BCE loss function. Model trained using all MRI modality (Anatomic + Diffusion + MRSI) achieve significantly higher PCC by improving the sensitivity of the model.

6.3.5. Model comparison among different treatment plans

Our best performing model gathered from previous experiments was found to use anatomic + diffusion + MRSI as input, and trained using PCC + BCE loss function, with initial learning rate at 5e-5, optimized using Ranger optimizer, and converged at epoch 135. To further evaluate the HTV generated using this model, we compared its performance against a HTV of only the pre-RT lesion, an SOC 2cm-CTV (2cm expansion of pre-RT lesion), and a HTV generated model trained using only anatomical images. Table 6.2 summarizes the average sensitivity, specificity, Dice score, Tversky coefficient, and PCC results of all HTV for patients in the test set, and Table 6.3 shows the individual patient comparisons between our best HTV and the SOC 2cm-CTV. As expected, the HTV that included only the pre-RT lesion had the lowest sensitivity (0.35 ± 0.18), severely undertreating the lesions. On the other hand, the SOC 2cm-CTV achieved high sensitivity (0.83 ± 0.14) but also the lowest specificity (0.87 ± 0.08), overtreating the normal-appearing brain. The HTV generated by our deep learning model outperformed the 2cm-CTV in covering the progressed lesion, with the highest sensitivity (0.87 ± 0.10), higher specificity than the 2cm-CTV (0.91 ± 0.02), and most importantly, higher PCC (0.80 ± 0.09). This means sparing more normal brain tissue compared to the 2cm-CTV, while targeting more subclinical subclinical disease in the location of future progression. Furthermore, in patient-wise comparisons, our best deep learning HTV also showed less variation in performance between patients, with lower standard deviation in all metrics calculated compared to the 2cm-CTV. Improvement in sensitivity and PCC values were observed in 9/14 patients, while improvement in specificity was observed in 7/14 patients.

Visual comparison of resulting target volumes in 2 example patients are shown in Figure 6.8. For the first patient (A), which is the same as patient #8 in Table 6.3, the HTV from our multimodality MRI model performed the best with highest sensitivity, Tversky score, and PCC compared to other plans, and specificity also higher than the 2cm-CTV. Visually, the HTV covers

the entire tumor area of progression scan with a small area of buffer, even though the original tumor at pre-RT was relatively small. The 2cm-CTV missed lesion voxels, toward the right ventricle, lowering its sensitivity. For the second worst performing patient (B), sensitivity is quite low across all HTVs, suggesting an unpredictable progression path. Of all patients in Table 6.3, this is the only patient where our model resulted in slightly lower sensitivity and specificity compared to the 2cm-CTV, which leads to lower Tversky and PCC scores as well. Despite the lower performance metrics, we have observed several slices where our HTV correctly identified the direction of progression (red arrow).

Table 6.2. Comparison between deep learning performance and other hypothetical treatment plans
The 2cm-CTV treatment plan achieved high sensitivity but the lowest specificity, overtreating normal-appearing-brain. Deep learning model outperformed all other hypothetical treatment plans in sensitivity and PCC and has higher specificity and dice than the standard of care treatment.

Methods	Sensitivity	Specificity	Dice	PCC	Tversky ($\alpha = 0.03$)
Only treating pre-RT lesion	0.35 ± 0.18	0.98 ± 0.02	0.38 ± 0.12	0.35 ± 0.17	0.35 ± 0.17
Standard of care 2cm-CTV	0.83 ± 0.14	0.87 ± 0.08	0.33 ± 0.15	0.75 ± 0.12	0.74 ± 0.13
Anatomic DL model	0.79 ± 0.13	0.93 ± 0.02	0.38 ± 0.14	0.74 ± 0.12	0.73 ± 0.13
Multi-modal DL model	0.87 ± 0.10	0.91 ± 0.02	0.36 ± 0.14	0.80 ± 0.09	0.78 ± 0.10

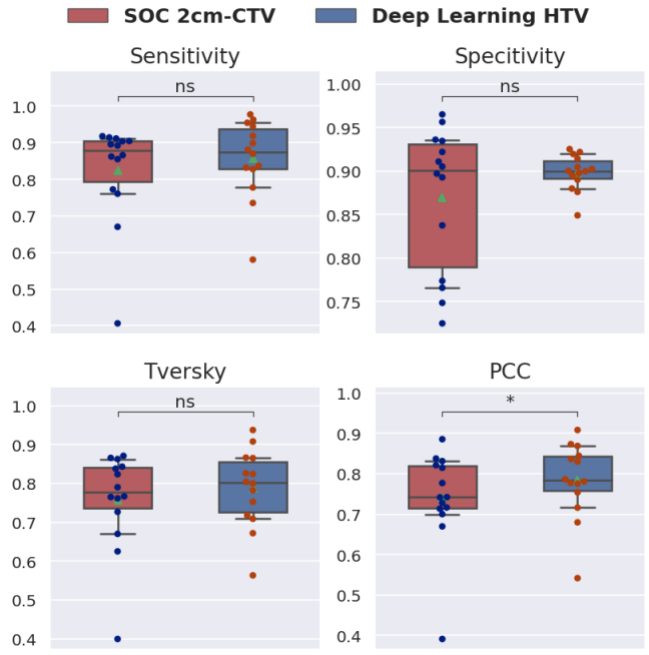

Figure 6.7. Comparing performance between SOC 2cm-CTV and our best HTV
Our model showed a higher mean sensitivity, specificity, Tversky, and PCC, although only the PCC shows significant improvement (p = 0.04, Wilcoxon signed-rank test

Table 6.3. Patient-wise comparison between deep learning performance and SOC
Of 14 patients used in our test set, we observed an improvement in sensitivity in 9/14 patients, and improvement in specificity in 7/14 patients. PCC was also improved for 9/14 patients (in red).

	SOC – 2cm expansion treatment			***DEEP LEARNING TREATMENT***		
Patient ID	*Sensitivity*	*Specificity*	*PCC*	*Sensitivity*	*Specificity*	*PCC*
1	*0.902*	*0.848*	*0.787*	*0.832*	*0.922*	*0.776*
2	*0.924*	*0.966*	*0.895*	*0.963*	*0.890*	*0.869*
3	*0.782*	*0.921*	*0.726*	*0.919*	*0.900*	*0.836*
4	*0.770*	*0.903*	*0.710*	*0.837*	*0.914*	*0.777*
5	*0.928*	*0.735*	*0.738*	*0.870*	*0.880*	*0.781*
6	*0.872*	*0.784*	*0.723*	*0.899*	*0.849*	*0.786*
7	*0.826*	*0.917*	*0.770*	*0.803*	*0.917*	*0.766*
8	*0.416*	*0.932*	*0.401*	*0.580*	*0.904*	*0.542*
9	*0.914*	*0.915*	*0.848*	*0.977*	*0.919*	*0.908*
10	*0.865*	*0.946*	*0.825*	*0.945*	*0.876*	*0.844*
11	*0.680*	*0.975*	*0.680*	*0.881*	*0.925*	*0.830*
12	*0.921*	*0.758*	*0.752*	*0.778*	*0.897*	*0.716*
13	*0.915*	*0.907*	*0.841*	*0.954*	*0.902*	*0.873*
14	*0.906*	*0.776*	*0.752*	*0.735*	*0.895*	*0.680*

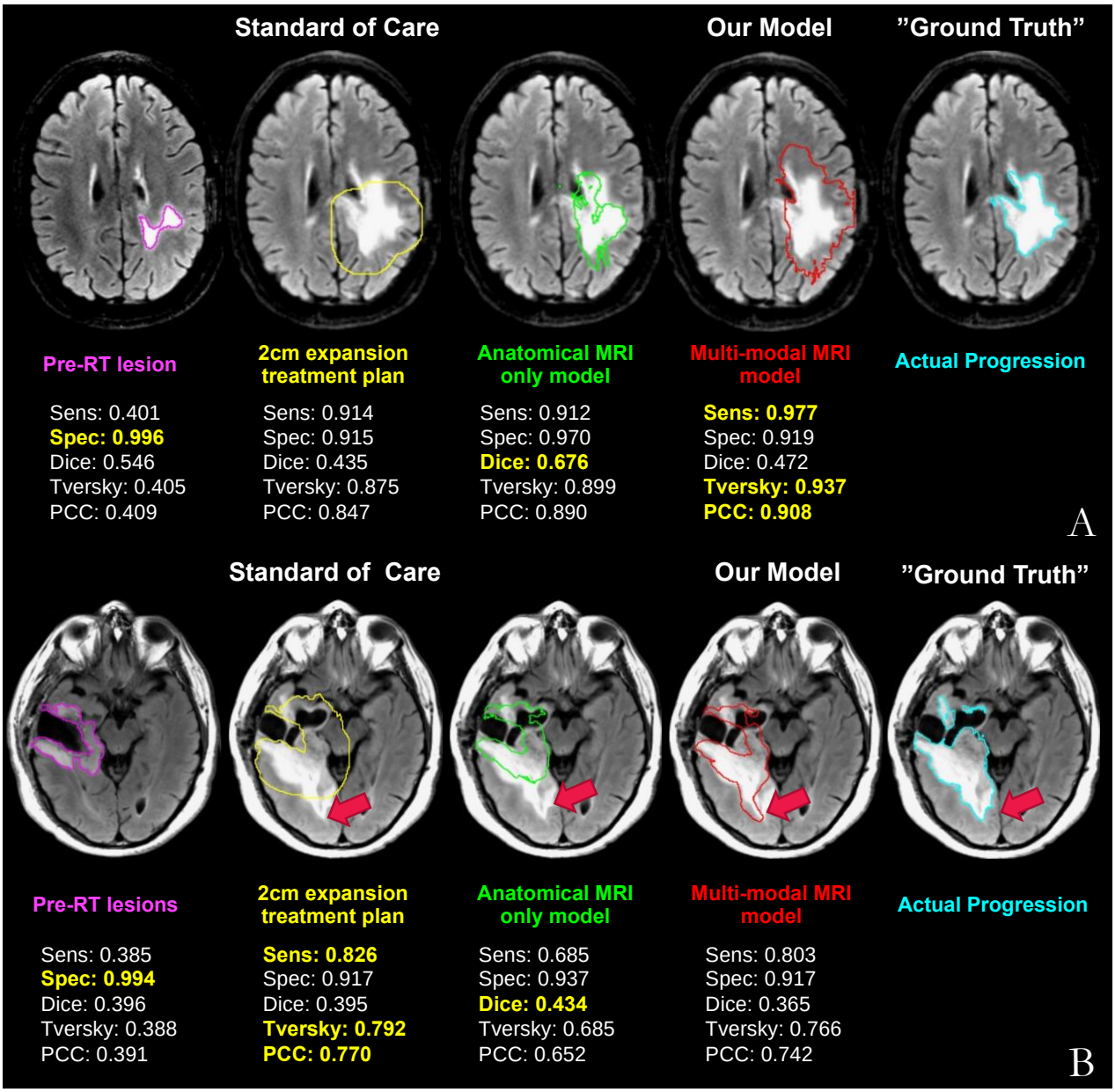

Figure 6.8. Visual assessment of all HTVs for 2 different example patients
The patient in (A) is the same as patient #8 of Table 6.3. This is a patient where both sensitivity and specificity were higher in our HTV compared to the 2cm-CTV, resulting a higher Tversky score and PCC score as well. On the other hand, the patient in (B) is the same as patient #6, and this is the only patient in the test set where both sensitivity and specificity were slightly lower in our HTV. Despite the reduced specificity, our HTV was able to detect the correct direction of tumor progression towards the posterior region (depicted with red arrow).

6.4. Discussion

Since the introduction of Stupp's protocol for treating glioblastoma to include radiotherapy and adjuvant chemotherapy, prognosis of glioma patients has only improved minimally despite decades of clinical trials incorporating new therapeutic agents. This is mostly because it is challenging to fully identify and target aggressive and subclinical tumor cells while preserving healthy brain tissue. Recent developments in the field of radiotherapy have allowed empirical doses to be delivered precisely to the planned CTV. However, current RT treatment is still only guided by anatomical T1-weighted and T2-weighted MRI with the highest dose delivered to the 2cm expansion of the gross tumor volume, failing to acknowledge the spatial heterogeneity and infiltrative nature of this disease. While tumor progression mostly occurs locally, previous studies have shown that tumor can spread beyond the CTV in about 10-37% of patients [46]–[49], [176], which is consistent with our dataset where about 28% of all patients with tumors progressed outside of the high-dose margin. At the same time, current CTV has low specificity, with about 60% of all radiated voxels being normal brain tissue [2], and overtreating normal brain tissue can lead to cognitive decline, reduced quality of life, and in extreme cases shortened overall survival [177].

Although recent advances in diffusion-weighted MRI and metabolic ^{1}H-MRSI [8] have allowed the visualization and detection of subclinical tumor cells in patients with GBM, they remain unused in clinical RT planning with the exception of a few research studies [105], [180]–[184]. We have previously shown [9] that abnormalities in ADC, FA, and CNI are all good indicators of future tumor progression, with higher ADC, lower FA, and higher CNI all denoting regions at higher risk of progression. In this study, we expanded on that knowledge, and aimed to provide a tool that can guide precision-based RT planning by utilizing multiparametric MRI at the pre-RT time point in conjunction with machine learning to predict regions of subsequent tumor progression. We first trained random forest models to predict the conversion of normal-appearing voxels to CEL and NEL progression for all patients and compared the results when splitting the cohort by treatment

as well as by time of progression. We then trained and applied a U-Net based deep learning model using the multi-parametric maps as inputs to generate a hypothetical treatment plan that included new areas of tumor progression on T2-FLAIR and post-contrast T1-weighted images and compared the predicted maps of test patients to SOC 2cm-CTV, which is a clinical 2cm uniform expansions of anatomical lesion volumes. To our knowledge, this is the first study to: a) apply deep learning to pre-RT multiparametric MRI images and predict regions of subsequent tumor progression; b) utilize cutting edge technology in inter-exam image registration; c) develop and test new loss functions in training and evaluation metrics to improve model prediction; and d) perform a comprehensive analysis on three different upfront treatment cohorts.

Our statistical analysis of diffusion and metabolic MRI markers have confirmed how diffusion-weighted MRI and MRSI at pre-RT can identify subclinical disease that appears normal on anatomical MRI, especially for CNI and CCrI maps. However, interestingly, we also observed a small but significant difference in the median normalized T2-FLAIR between stable NAV and progressed NAV voxels, suggesting subtle differences in normalized T2-FLAIR hyperintense signal (that visually appears normal when annotating the T2 lesion manually) can also signify abnormal tissue. The diffusion-weighted markers (nFA and nADC) were useful in identifying differences between stable normal voxels and voxels that progressed to become contrast enhancing (NAV→CEL and NEL→CEL), but not as useful between voxels that were remained normal and progressed via non-enhancing T2-FLAIR hyperintense signal (NAV→NEL). Most of these results agree with our previous findings by Anwar et al.[9].

Our random forest model further solidified our hypothesis that advanced MRI can help predict future progression, as we were able to achieve an average ROC-AUC score of 0.88 for the CEL-progression model and 0.81 for the NEL-progression model. The higher performance for CEL-progression prediction task was expected, as those voxels are more likely appear abnormal pre-RT. The most important features for the progression by CEL model were CNI, CCrI, nLipid, and nFLAIR, while the most important features for progression by NEL model are the CNI, CCrI,

and nFLAIR. This is also in agreement with our statistical analysis results. The fact that nT1C at the pre-RT time point did not contribute much in predicting tumor progression in our cohort suggested that it was the infiltrative tumor outside of the contrast-enhancement regions that drove subsequent progression, and can be better visualized using the T2-FLAIR signal, as well as choline-containing metrics in MRSI (CNI, CCrI). Interestingly, even though diffusion-weighted images were shown to be a significant predictor of both NEL and CEL progression, neither ADC nor FA were among the most important features for the random forest model. After further investigation, we hypothesized that the intrinsic lower SNR of diffusion-weighted images may hinder its usefulness when performing voxel-wise analysis. This was also observed in our deep learning results, where improvement in performance was minimal with the addition of diffusion-weighted images. We also observed in our random forest results that it was easier to predict both CEL and NEL progression for patients who progressed earlier compared to patients who progressed later. This is also in agreement with some previously published results, that MRI markers in the pre-RT scans better reflect early progression than late progression [182]. Furthermore, when anti-angiogenic drugs (Bevacizumab or Enzastaurin) are part of the treatment regimen, which was the case for 65% of patients, the classic MRI markers for progression can be delayed, inflating the false positive rate of our model.

Designing optimal loss functions for training, and metrics for evaluating the segmentation task was a difficult task. In a typical tumor segmentation task where the main goal is to produce an exact segmentation of the tumor in real time, both false positives and false negatives can be penalized quite equally, making Dice coefficient a suitable candidate for both loss function and evaluation metric. That is not the case in our task, however, since a low sensitivity of a model can often lead to undertreating the tumors. The key is to design a loss function and an evaluation metric that can better balance the sensitivity and specificity in a way that makes sense clinically. We did this by designing the loss function PCC based on the Tversky loss, in which the term α and β values can be dynamically altered based on the patient's tumor size. Patients with smaller

lesions have a higher class-imbalance, thus requiring higher β to further reduce false negatives and improve the sensitivity. In our training patient cohort, the average α is (0.031 ± 0.021) with a range of [0.003 - 0.093]. As a result, models trained using a PCC-based loss function outperformed Dice-based and simple Tversky-based loss functions (Figure 6.5), improving the sensitivity to 0.87, while keeping the specificity relatively high as well (at 0.91). The benefit of using Tversky scores (with α = 0.03) and PCCs as evaluation metrics is also demonstrated in Table 6.2, in comparison to the standard Dice metric. The HTV that only includes the pre-RT lesion had the highest Dice (due to an almost perfect specificity score), yet is unanimously considered the poorest model by any measure, since none of the infiltrating tumor cells will be irradiated at all. Its Tversky and PCC scores, however, are both low, reflecting its true performance. In contrast, the standard of care 2cm-CTV, which is a highly acceptable model, has a very poor Dice score, but relatively high Tversky (α = 0.03) and PCC. One question remains regarding Tversky-based loss and metrics: how to determine α and β to appropriately to best represent the segmentation task. We hypothesized that, since α and β play a role in correcting the intrinsic imbalanced class issue (with low number of positive-labeled voxels), the best α should be approximately the ratio between the number of lesion voxels over the total number of brain voxels. Our experiment (Figure 6.5) partially reflected this, since model performance is the best between α = 0.01-0.05, which is close to the median percentage of lesion voxels in our patient cohort at ~0.03. As for PCC, it goes another step further and allow personalization of the model, both during training and in evaluating the models. Clinically, PCC also makes the most sense for our task: patients with a large tumor can benefit from a little higher specificity as to prevent too much brain damage from radiation, while patient with small tumor can benefit from a higher sensitivity to ensure full coverage of progression. The most optimal sensitivity and specificity should depend on each specific patient, and we are looking into doing more experiments with PCC to show its utility.

Figure 6.6 shows our model performance when altering the input images. When training and optimizing models using the same PCC+BCE loss, the multimodal anatomic + diffusion + MRSI (ADM) model achieved a significantly better performance compared to both the anatomic (A) model and the anatomic + diffusion (AD) model, but only slightly better than the anatomic + MRSI (AM). This demonstrates the utility of using MRSI markers in predicting progression. Overall, our best model was trained using all 3 MRI modalities, with trending higher sensitivity (0.87 vs. 0.83, p-value = 0.22) and specificity (0.91 vs. 0.87, p-value = 0.31) compared to the 2cm-CTV, and a significantly higher PCC (0.80 vs. 0.75, p-value = 0.04). Table 6.3 shows that among the 14 patients in the test set, 9/14 showed an improvement in sensitivity and PCC, and 7/14 showed an improvement in specificity. Furthermore, the plot in figure 6.7 shows a small variance in specificity values among test patients, signifying that the level of overtreatment is fairly low among all patients. Figure 6.7 also showed two outliers with very low sensitivity even when using the 2cm-CTV. Upon further investigation, we determined that tumor recurrence occured at a different location than the original tumor, causing the SOC 2cm-CTV to completely miss it, while our model was able to recover some of this sensitivity. We can visualize how the model was able to improve performance in figure 6.8. It appears that, in both patients shown, the correct path of progression was identified by the model (as showed by the red arrow), even when the performance of our model is slightly lower compared to the SOC 2cm-CTV (patient B).

Though our results are promising, there are several limitations to our study. While the sensitivity and specificity of our model are higher than any other study, we still only achieved minimal improvement compared to the SOC 2cm-CTV. Even then, designing a personalized target volume using our model may still have a significant clinical and survival impact by targeting more subclinical disease while also sparing healthy brain tissue. With more patients, we are confident that our model performance can be further improved to a more clinically meaningful level. Although we have demonstrated the potential of utilizing both Tversky-based coefficients and PCCs as loss functions and evaluation metrics, their ultimate utility and clinical relevance

remains uncertain. Thus, more studies and analyses of these metrics is necessary to increase the confidence of using them. We also realized that the use of MRSI is a limiting factor as it 1) is not routinely performed in clinical practice, which would limit the widespread adoption of our model in the future, and 2) has low resolution with smaller spatial coverage than other MRI modalities, limiting its potential to only regions closest to the lesion site. But since most tumor progression occurs locally, the full extent of tumor progression can still be captured using MRSI for most patients.

6.5. Conclusions

Our study demonstrated the feasibility of using pre-treatment diffusion-weighted and ^{1}H-MRSI, along with machine learning and deep learning techniques, to predict future regions of tumor progression and generate hypothetical target volumes for RT treatment. In our random forest models, CEL and NEL progression was more challenging to predict with longer time from treatment, suggesting that time to progression should be added in subsequent modeling. For the deep learning segmentation task, we found that the best model was trained using a novel loss function that improved model performance by taking the size of the original tumor into consideration. Our deep learning model using multi-parametric MRI performed better than the current practice of a uniform 2cm expansion for RT treatment planning and no expansion, suggesting that multi-parametric MRI with deep learning has the potential to assist future RT treatment planning. We also explored multiple ways to evaluate model performance and found that both Tversky's metrics and our newly-developed individualized PCC metric were better options for this segmentation task than the conventional Dice score. Future studies will look into further investigating PCC as a metric for model evaluation in this and other segmentation-based and outcome prediction tasks.

Chapter 7. Early prediction of progression free survival (PFS) and overall survival (OS) in patients with glioblastoma using machine learning and multi-parametric MRI

7.1. Introduction

Glioblastoma (GBM) is a highly malignant and invasive brain tumor with poor median OS of approximately 12-15 months [173]. However, OS can range anywhere from 0.7 months to 10 years, with about 3-5% of patients who have received SOC treatment of RT and temozolomide alive after 3 years, and only ~0.71% alive after 10 years [201]. As even more variation in survival has been observed with the addition of investigational anti-angiogenic- and immunotherapies, early prediction of individual prognosis and OS of patients with GBM before the onset of therapy is imperative for physicians to determine the appropriate treatment strategy for an individual patient [91]. Previous studies have identified factors that can explain significant variations in OS and PFS for newly diagnosed GBM patients, including their age at diagnosis, the Karnofsky performance score, the molecular subtype of GBM and mutations such as MGMT promoter methylation, EGFR, etc., [201]–[204] as well as the tumor's histopathology such as Ki-67 [201]. However, the course of treatment can highly affect OS and PFS of GBM patients [202], [203], as well as the validity of PFS as a surrogate endpoint of OS, and tumors can undergo malignant transformation during progression [205]. Taken together, predicting the OS and PFS at the time of diagnosis remains a significant challenge for patients with glioma.

Recent studies have individually demonstrated the utility of histogram and radiomic metrics derived from metabolic, diffusion, and perfusion-weighted MRI at pre-, mid-, and post- RT time points as potential markers for predicting clinical outcome of patients with GBM [78], [91], [206], [207]. Larsson et al. have shown that temporal changes in perfusion parameters significantly correlated with survival time [208]. Nelson et al. have identified anatomic lesion volumes at post-RT scan, metabolic lesion volume at mid-RT and post-RT scan, and the levels of choline, lactate, and lipid in MRSI all correlated well with OS [91]. However, most radiomic and

large-scale studies still only rely on anatomical MRI during pre-treatment time point to infer correlation with OS and PFS, and to our best knowledge, no study has performed a comprehensive prediction of OS and PFS using temporal changes of multimodal MRI. In this study, we developed a machine learning model that incorporates multi-parametric metabolic and physiologic MRI parameters from pre- and/or mid- therapy to predict OS and PFS in patients with GBM treated with upfront radiation, anti-angiogenic-, and cytotoxic-chemotherapy.

7.2. Methods

7.2.1. Subjects

A total of sixty-three patients (median age of 53, range [25 – 75]) newly-diagnosed with GBM, and scanned after surgical resection but before the onset of subsequent therapy, were included in this retrospective study. 28 patients received SOC treatment plus Avastin and Tarceva (or Bevacizumab - ATT), while 35 patients received SOC treatment and Enzastaurin (Enza). All patients underwent pre-RT baseline MR Imaging, 48 of these patients also received a mid-RT scan 3-4 weeks from the start of RT.

7.2.2. Image Acquisition

MR examinations were performed on a 3T GE Signa scanner using an eight-channel phased-array head coil. Standard anatomical imaging included the T2-weighted FLAIR and the 3D T1-weighted IR-SPGR imaging, both pre- and post- the injection of a gadolinium-based contrast agent. Diffusion-tensor images (DTI) were obtained in the axial plane with b=1000 s/mm^2, and either 6 gradient directions and 4 excitations or 24 gradient directions and 1 excitation or b=2000 s/mm2 and 55 gradient directions [repetition time (TR)/echo time (TE) = 1000/108 milliseconds, voxel size = 1.7-2.0 × 1.7-2.0 × 2.0-3.0 mm]. DSC perfusion-weighted images were obtained following a 3-ml/s bolus injection of 0.1 mmol/kg body weight gadolinium diethyltriamine

pentaacetic acid using a series of T2*- weighted echo-planar images [TR/TE/flip angle = 1250-1500/35-54 milliseconds/30°-35°, 128 × 128 matrix, slice thickness = 3-5 mm, 7-24 slices with 60-80 time points] before, during, and after the arrival of the contrast agent bolus. The temporal resolution was between 1 and 1.5 seconds, with total acquisition time ranging from 1-2 min.

3D ^{1}H MRSI were acquired using point-resolved spectroscopic selection for volume localization and very selective saturation (VSS) pulses for lipid signal suppression [excited volume = 80 × 80 × 40 mm, TR = 1100-1250 ms, TE = 144 milliseconds, overpress factor = 1.5 if lactate edited, otherwise 1.2, field of view = 16×16×16 or 18×18×16 cm, nominal voxel size=1×1×1cm], flyback echo-planar readout gradient in the SI direction, 988 Hz sweep width and 712 dwell points. A dual-cycle lactate-edited sequence was used for 54 patients, while a standard single-cycle sequence [6] was used for the remaining 9 patients.

7.2.3. Processing

From both pre-RT and mid-RT scans, all relevant anatomical, diffusion, perfusion, and metabolic maps were extracted for all patients. All images, including anatomical images (T2-weighted FLAIR, T1-weighted pre-/post-contrast) and physiological parametric maps from diffusion MRI, and DSC-perfusion MRI were aligned to the post-Gd T1-weighted image using Slicer's BRAINSFit tool with B-spline warping [172], or FMRIB's FSL Linear Image Registration Tool (FLIRT) [192], [193]. Anatomical imaging was used to manually define the contrast-enhancing and T2-hyperintesnse lesions (CEL and T2L).

For DWI MRI data, ADC and FA were calculated on a voxel-by-voxel basis by applying the FMRIB's Diffusion Toolkit on DWI and DTI data [191]. From the DSC perfusion MRI data, the relative cerebral blood volume (rCBV) and the normalized peak height (nPH) maps were generated on a voxel-by-voxel basis using the modified gamma-variate function that also considers the leakage of the contrast agent [33], as well as nonparametric post-processing methods [209], [210]. From the lactate-edited MRSI sequence, we performed k-space Fourier

transforms to generate 3D spectral data, following by a phase correction step to account for flyback echo-planar gradient. Metabolite peaks height was then calculated on a voxel-to-voxel basis using the 3D spectral data. Maps of NAA, Choline (Cho), Creatine (Cre), Lactate (Lac), and Lipid (Lip) were generated and normalized to the median peak height of normal brain voxels. The Choline-to-NAA index (CNI), Choline-to-Creatine index (CCrI), and Creatine-to-NAA index (CrNI) were also calculated as maps that reflect changes of the two relevant metabolites compared to normal brain [6], [171].

Four different masks were generated semi-automatically using in-house software and used for subsequent analysis: 1) the CEL, 2) the T2L (which includes the CEL), 3) the union of the T2L mask and voxels with CNI<1, 4) the union of the T2L mask and voxels with CNI<2, and 5) the map of all supratentorial brain tissue (or the brain mask). Figure 7.1 summarizes the full study schema, including the visualization of each map used.

7.2.4. Analysis and Machine Learning

For each set of MRI images (pre-RT and mid-RT), histogram analysis of parameters, including median, mean, percentiles (25 and 75), sum, kurtosis, and skewness, was performed on each parametric map after the application of each of the 5 masks previously described in section 7.2.3. The CEL and T2L volumes at both time points and type of treatment, were also included as part of our feature space.

Two separate binary outcomes were simultaneously analyzed: 1) high PFS vs. low PFS, with PFS split based on the median PFS of the entire patient population at 45 weeks; and 2) high OS vs. low OS, with OS also split based on the median OS of the entire patient population at 76 weeks. Patients were split into training data (44 patients) and test data (19 patients). A Mann–Whitney U test was performed on only the training data to determine relevant features to include in our machine learning model, with a p-value ≤ 0.1 used as the threshold for a feature to be selected.

Machine Learning: For each mask, a random forest model was trained using relevant features determined from the Mann-Whitney U test. Hyper-parameters were tuned by performing 3-fold cross-validation on all imaging parameters from 44 patients, and the CV results were recorded and averaged. The best model was then used to test the remaining 19 patients in our test set. Each model was fit to maximize the area under the receiver operating characteristic curve (AUC-ROC), and the AUC-ROC was also chosen as the main model evaluation metric.

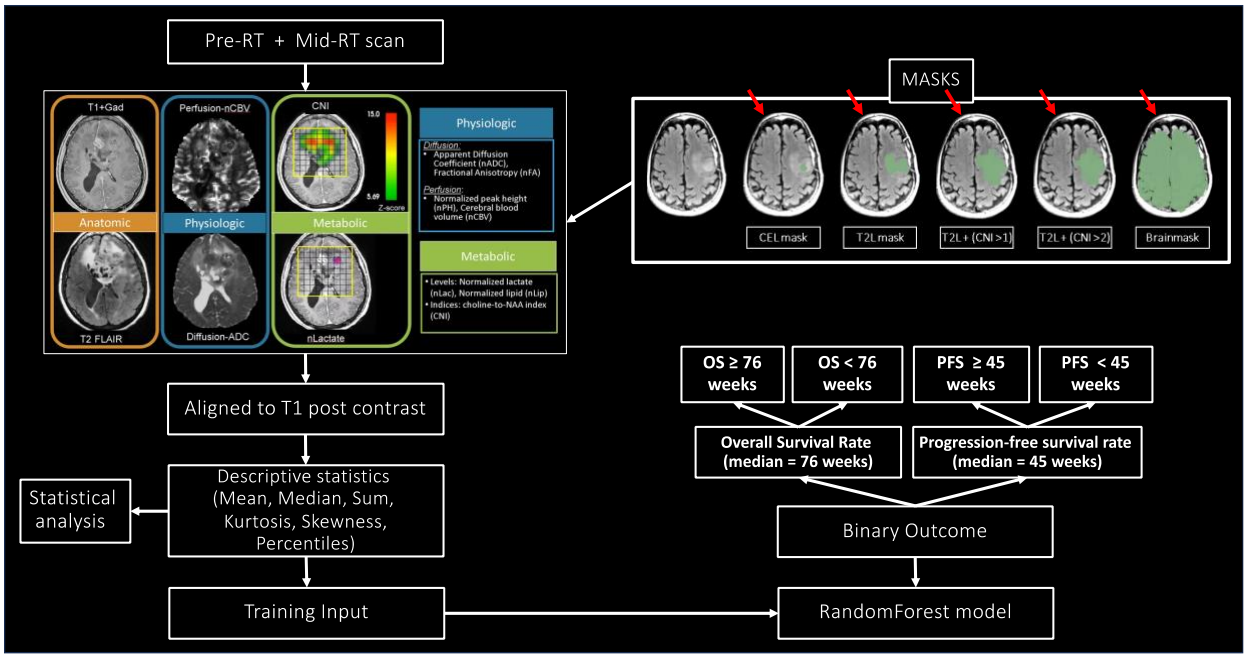

Figure 7.1. Study schema
From the pre-RT and mid-RT scan, relevant diffusion, perfusion, and metabolic maps were extracted and aligned for all patients,. Five different types of masks were generated for the analysis: 1) the CEL mask, 2) the T2L mask, 3) the union of the T2L and voxels with CNI>1, 4) the union of the T2L and voxels with CNI>2, and 5) supratentorial brain tissue mask. Once each mask was applied, relevant descriptive statistics were derived from each parametric map. Two separate binary outcomes were split based on median OS of our cohort (76 weeks) and median PFS (45 weeks).

7.3. Results

7.3.1. Patient characteristics

The characteristics of our patient cohort is shown in Table 7.1. The median PFS and OS for all patients in our dataset was 45 weeks and 76 weeks, respectively. While there was no significant difference in the OS between our two cohorts SOC+Bevacizumab (81.3 weeks) and SOC+Enzastaurin (73.2 weeks), there was a significant difference in PFS (50.8 weeks vs. 31.1 weeks; $p = 0.02$). This difference reflects the ability of bevacizumab as an anti-angiogenic agent to normalize the vasculature further than Enzastaurin, making T1 post-contrast imaging an ineffective marker for assessing progression given the similar OS between the two groups.

Table 7.1. Patients characteristics
Of 63 patients used in this study, 28 received SOC + Bevacizumab treatment, and 35 received SOC + Enzastaurin treatment.

Patient cohort	All patients	SOC + Bevacizumab	SOC + Enzastaurin
N (total)	63	28	35
Pre-RT ($N_{pre\text{-}RT}$)	63	28	35
Mid-RT ($N_{mid\text{-}RT}$)	48	24	24
Age (year)	53	51	56
Median OS (weeks)	75.0	81.3	73.2
Median PFS (weeks)	45.0	50.8	31.1

7.3.2. Statistical Analysis

In figure 7.2, we show the pre-RT T2-FLAIR, nCBV, CNI, and nLac images of two different patients at the same scale. Patient B has a CEL volume of 28.4 cm^3, progressed at 84 weeks, and died at 146 weeks. Patient A has a smaller CEL volume of 10.9 cm^3 (and a smaller T2L volume as well), but progressed much sooner at 10 weeks, and had a much shorter OS of 47 weeks. Patient A also had visibly elevated CNI and rCBV near the edge of the T2L, suggesting a more infiltrative lesion, resulting in a smaller PFS and OS.

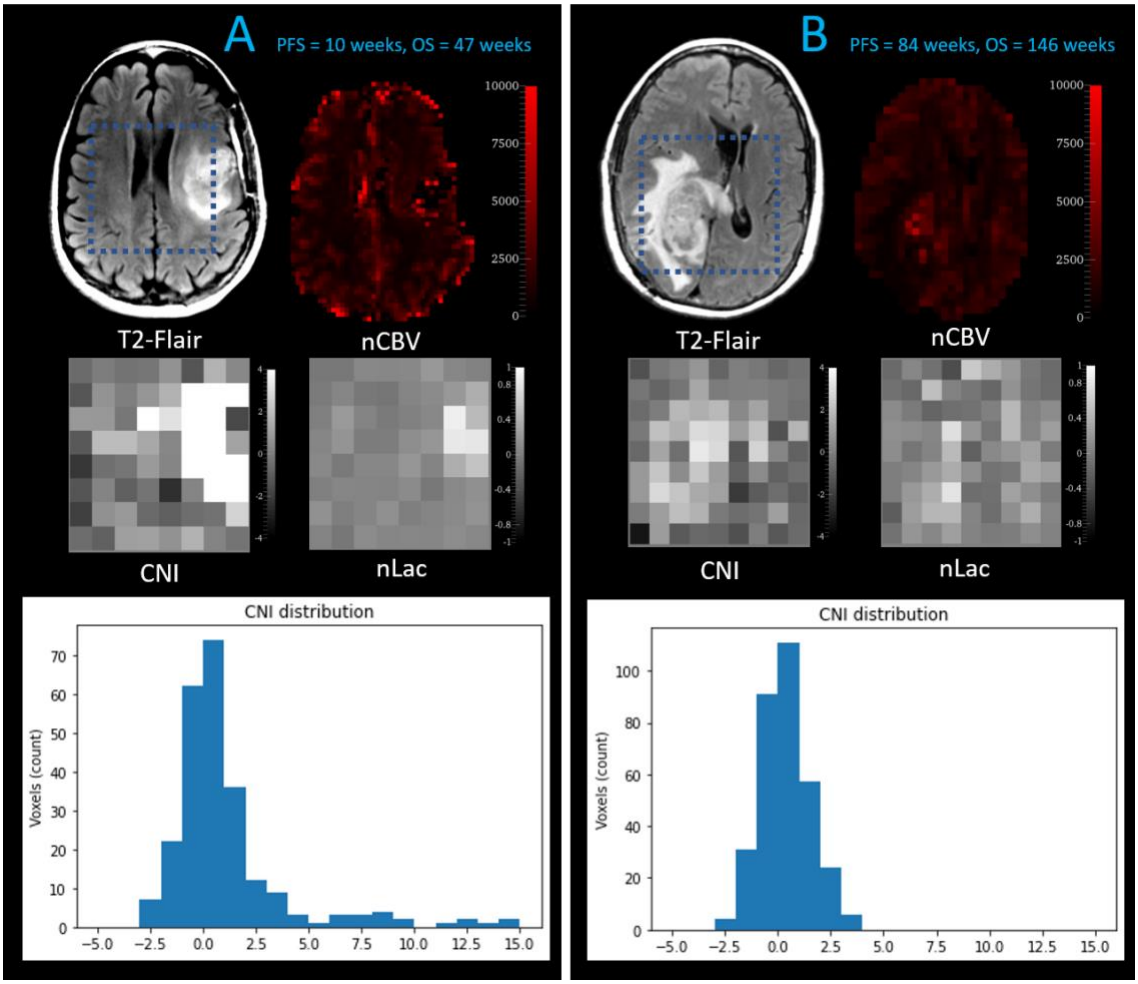

Figure 7.2. Example T2-FLAIR, nCBV, CNI, and nLac images at pre-RT scan
Example T2-FLAIR, nCBV, CNI, and nLac images at pre-RT scan for A) patient progressed at week 10; and Panel B/ patient progressed at week 84. Color bar is the same for both patients. Noted that patient A has a CEL volume of 10.9 cm^3, while patient B has a larger CEL volume of 28.4 cm^3 at time t1. Bottom: Histogram of CNI values for each patient.

Figure 7.3. shows sample plots of the median of all images using the [T2all + CNI>1] mask, and the p-value calculated from performing a Mann-Whitney U test on the training set. For each image, the left plot represents difference between patients with PFS > 45 weeks (blue) and patients with PFS < 45 weeks (orange) at pre-RT, while the right is at mid-RT. The only significant parameters here (p-value less than 0.05, depicted with '*') are the CEL volume at mid-RT and the median CNI at pre-RT. This shows that both pre-RT and mid-RT parameters are useful in predicting OS and PFS.

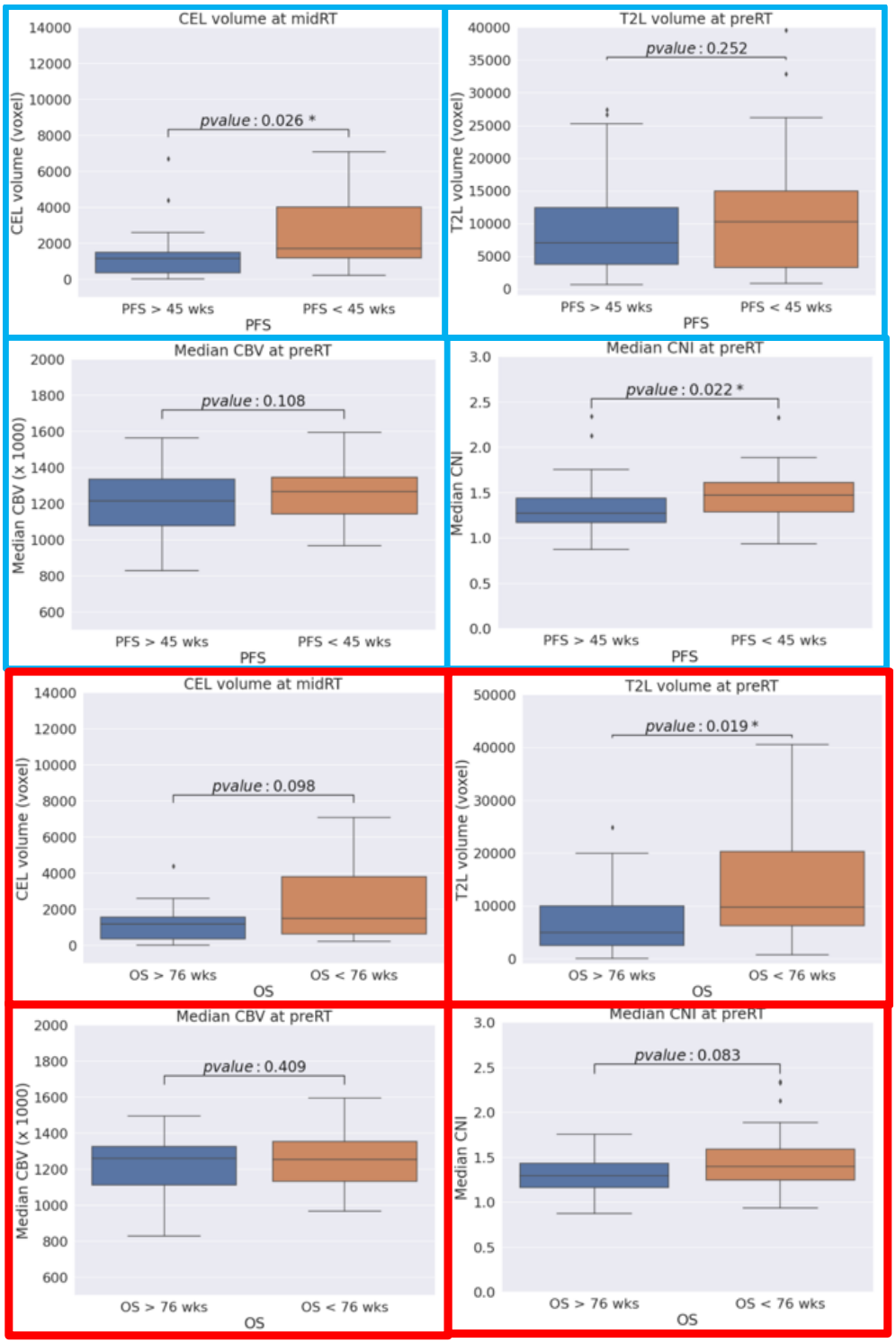

Figure 7.3. Statistical analysis results
Statistical analysis using [T2L + CNI > 1] mask, median values, and Mann-Whitney U test. Significant predictors of PFS include the pre-RT CNI and mid-RT CEL volume, and significant predictors of OS include the pre-RT T2L volume.

Table 7.2 and Table 7.3 provided the full list of all descriptive statistics parameters that can significantly explain the difference between high vs. low PFS and OS respectively (with $p < 0.05$ using Mann-Whitney U test), for each of the 5 brain masks. This again showed that parameters from both pre-RT and mid-RT time point were useful in predicting PFS. Diffusion maps (ADC and FA) and MRSI maps (CNI, lactate, and lipid) were consistently the most significant predictors for both the PFS and OS tasks, while perfusion maps did not contribute significantly for either task.

Table 7.2. Significant parameters for PFS prediction
List of all descriptive statistics parameters with $p < 0.05$ for PFS prediction (PFS < 45 weeks vs. PFS > 45 weeks) using Mann-Whitney U test for each of the 5 masks.

CEL MASK		**[T2ALL + CNI>1] MASK**		**BRAIN MASK**	
Descriptive Stats	**P-value**	**Descriptive Stats**	**P-value**	**Descriptive Stats**	**P-value**
pre-RT summed FA	0.014	mid-RT median FA	0.006	mid-RT skew CNI	0.013
mid-RT summed FA	0.027	mid-RT median CNI	0.012	mid-RT mean CNI	0.013
mid-RT CEL volume	0.046	mid-RT 25-prctile FA	0.013	mid-RT 25-prctile LAC	0.014
pre-RT CEL volume	0.047	mid-RT 95-prctile CNI	0.019	mid-RT 95-prctile CNI	0.019
		mid-RT 75-prctile CNI	0.021	mid-RT summed CNI	0.022
		mid-RT mean FA	0.022	mid-RT skew LAC	0.023
		mid-RT 25-prctile CNI	0.023	mid-RT stdev CNI	0.035
		mid-RT mean CNI	0.023	mid-RT 25-prctile LIP	0.038
		mid-RT 25-prctile LAC	0.025	mid-RT 5-prctile LIP	0.038
		mid-RT 5-prctile LIP	0.032	mid-RT 5-prctile LAC	0.048
		mid-RT 75-prctile FA	0.036		
		mid-RT summed CNI	0.037		
		mid-RT stdev CNI	0.037		
		pre-RT kur ADC	0.042		
		mid-RT skew LIP	0.045		
T2ALL MASK		**[T2ALL + CNI>2] MASK**			
Descriptive Stats	**P-value**	**Descriptive Stats**	**P-value**		
mid-RT 95-prctile FA	0.015	mid-RT summed LIP	0.007		
mid-RT summed CNI	0.015	mid-RT mean LIP	0.013		
mid-RT stdev FA	0.017	mid-RT median LIP	0.015		
mid-RT 25-prctile CNI	0.02	mid-RT 75-prctile CNI	0.021		
pre-RT 95-prctile CNI	0.025	mid-RT summed CNI	0.027		
pre-RT 75-prctile CNI	0.025	mid-RT median FA	0.029		
mid-RT summed LIP	0.032	pre-RT skew T1C	0.032		
mid-RT 25-prctile LIP	0.038	mid-RT summed LAC	0.033		
mid-RT mean CNI	0.041	mid-RT 25-prctile LIP	0.035		
mid-RT median CNI	0.041	mid-RT 5-prctile LIP	0.038		
pre-RT stdev LAC	0.041	mid-RT stdev CNI	0.038		
mid-RT 5-prctile CNI	0.041	mid-RT mean FA	0.039		
pre-RT summed FA	0.045	mid-RT 25-prctile FA	0.039		
pre-RT mean CNI	0.048	mid-RT 95-prctile CNI	0.041		
		mid-RT 75-prctile FA	0.041		
		mid-RT skew LAC	0.047		
		pre-RT kur T1C	0.05		
		mid-RT mean CNI	0.05		

Table 7.3. Significant parameters for OS prediction
List of all descriptive statistics parameters with p-val < 0.05 for OS prediction (OS < 75 weeks vs. OS > 76 weeks) using Mann-Whitney U test for each of the 5 masks

CEL MASK		**[T2ALL + CNI>1] MASK**		**BRAIN MASK**	
Descriptive Stats	**P-value**	**Descriptive Stats**	**P-value**	**Descriptive Stats**	**P-value**
pre-RT summed FA	0.007	pre-RT summed FA	0.017	pre-RT mean LAC	0.014
mid-RT summed CNI	0.013	pre-RT mask volume	0.029	pre-RT median CNI	0.016
mid-RT 5-prctile LAC	0.016	pre-RT skew T2FLA	0.032	pre-RT 75-prctile CNI	0.021
mid-RT CEL volume	0.021	pre-RT 25-prctile FA	0.043	pre-RT summed CNI	0.024
pre-RT summed ADC	0.048	mid-RT 25-prctile FA	0.044	pre-RT 25-prctile CBV	0.025
				pre-RT kur LAC	0.026
T2ALL MASK		**[T2ALL + CNI>2] MASK**		pre-RT 25-prctile CNI	0.027
Descriptive Stats	**P-value**	**Descriptive Stats**	**P-value**	pre-RT mean CNI	0.029
pre-RT summed FA	0.005	pre-RT summed FA	0.011	mid-RT median CNI	0.031
mid-RT stdev FA	0.006	pre-RT skew LIP	0.016	pre-RT 5-prctile CNI	0.033
pre-RT median CNI	0.009	mid-RT summed FA	0.018	mid-RT mean CNI	0.034
pre-RT 95-prctile CNI	0.01	pre-RT skew T2FLA	0.022	pre-RT 95-prctile CNI	0.043
pre-RT mean CNI	0.016	pre-RT 25-prctile CBV	0.027	mid-RT summed CNI	0.049
pre-RT summed ADC	0.017	mid-RT mask volume	0.028		
mid-RT summed CNI	0.02	mid-RT summed CNI	0.032		
pre-RT stdev CNI	0.022	pre-RT summed ADC	0.034		
pre-RT T2L volume	0.023	mid-RT summed ADC	0.035		
mid-RT summed FA	0.023	pre-RT 5-prctile FA	0.036		
pre-RT summed CNI	0.026	pre-RT skew LAC	0.046		
mid-RT 95-prctile FA	0.027	mid-RT summed T1C	0.047		
pre-RT 75-prctile LAC	0.029	pre-RT 75-prctile LAC	0.049		
pre-RT 75-prctile CNI	0.031	pre-RT skew LAC	0.046		
pre-RT summed T1C	0.034	mid-RT summed T1C	0.047		
pre-RT 25-prctile CBV	0.035	pre-RT 75-prctile LAC	0.049		
pre-RT skew LIP	0.035	pre-RT skew LAC	0.046		
mid-RT summed T1C	0.035	mid-RT summed T1C	0.047		
pre-RT 25-prctile PH	0.037	pre-RT 75-prctile LAC	0.049		
pre-RT 25-prctile CNI	0.045	pre-RT skew LAC	0.046		
		mid-RT summed T1C	0.047		
		pre-RT 75-prctile LAC	0.049		
		pre-RT 75-prctile LAC	0.049		
		pre-RT skew LAC	0.046		
		mid-RT summed T1C	0.047		
		pre-RT 75-prctile LAC	0.049		
		pre-RT skew LAC	0.046		
		mid-RT summed T1C	0.047		
		pre-RT 75-prctile LAC	0.049		
		pre-RT 75-prctile LAC	0.049		

7.3.3. Machine Learning

Tables 7.4 and 7.5 show the performance of the Random Forest model in predicting early PFS (PFS < 45 weeks) and early OS (OS < 76 weeks) for each mask using just pre-RT images (first row), and both pre-RT and mid-RT images (second row). Overall, it was easier to predict PFS compared to OS.

Early PFS prediction (PFS < 45 weeks): We observed that using the brain mask on both pre-RT and mid-RT images resulted in the best ROC score of .802. Larger masks tended to retain more information, and parameters such as skewness and kurtosis provided information about outliers and symmetry. Surprisingly, the smallest CEL mask gave the second-best performance, although the prediction relied more heavily on summed voxel metrics. It is important to note that the patients' age, the CEL volume at mid-RT, as well as the type of treatment (ATT vs. ENZA) were among the most important features for most models in predicting PFS. Leaving out treatment reduced our model's performance by 0.3 on average, suggesting that the type of treatment plays a role in predicting PFS. This agrees with our previous hypothesis that anti-angiogenic drugs delayed and altered the classic blood-brain-barrier breakdown markers for progression, and further suggests that each different type of anti-angiogenic drugs acts differently. As reported in our statistical analysis, diffusion markers (FA and ADC), and MRSI markers (CNI, lactate, and lipid) were among the best predictors for PFS, although diffusion metrics were only relevant within the CEL.

Early OS prediction (OS < 76 weeks): Compared to early PFS prediction, the type of treatment and CEL volume were not as relevant when predicting OS. Instead, patient age and the T2L volume at pre-RT were the most predictive parameters. And as mentioned previously, the majority of OS predictors (CNI and lactate) were from the pre-RT time point, although the best performing model only achieved an AUC of .725 (using the brain mask and only images from the pre-RT time point).

Table 7.4. Performance of the Random Forest model in predicting PFS
Performance of the Random Forest model in predicting PFS for each mask. Using the entire brain mask on both pre-RT and mid-RT images gives the best ROC score of .802. Overall, better performance is observed using larger masks and images at both timepoint.

PFS < 45 weeks	CEL mask	T2all mask	(T2all + CNI>1) mask	(T2all + CNI>2) mask	Brain mask
Only Pre-RT images	.695	.736	.776	.745	.759
Both Pre-RT and Mid-RT images	.714	.723	.781	.753	.802
Top 5 features (Gini-importance)	1. ADC sum mid-RT	Treatment	Treatment	Treatment	Skewness CNI pre-RT
	2. Age	Age	CEL volume mid-RT	CEL volume mid-RT	CEL volume mid-RT
	3. Treatment	Stdev lactate pre-RT	Median CNI pre-RT	Median lipid pre-RT	25 prc CNI pre-RT
	4. FA sum pre-RT	Stdev CNI pre-RT	Median lipid pre-RT	Median CNI pre-RT	Skewness ADC pre-RT
	5. CEL volume mid-RT	CEL volume mid-RT	Median CBV pre-RT	Age	Age

Table 7.5. Performance of the Random Forest model in predicting OS
Performance of the Random Forest model in predicting OS rate for each mask. Overall, it is harder to predict survival than PFS, most likely due to variations in therapy after progression. Using the brain mask on the pre-RT images gives the best ROC score of .725.

OS < 76 weeks	CEL mask	T2all mask	(T2all + CNI>1) mask	(T2all + CNI>2) mask	Brain mask
Only Pre-RT images	.529	.675	.638	.635	.725
Both Pre-RT and Mid-RT images	.641	.689	.652	.661	.710
Top 5 features (Gini-importance)	1. Median CNI pre-RT	Age	Treatment	Age	Age
	2. Age	Sum lactate mid-RT	Age	Treatment	Kurtosis CNI pre-RT
	3. Sum ADC mid-RT	Median CNI pre-RT	Kurtosis ADC pre-RT	T2L volume pre-RT	Median lactate pre-RT
	4. Sum FA pre-RT	Sum FA pre-RT	Kurtosis Lactate pre-RT	Kurtosis CNI pre-RT	T2L volume pre-RT
	5. T2L volume pre-RT	T2L volume pre-RT	T2L volume pre-RT	Kurtosis ADC pre-RT	Median CNI pre-RT

7.4. Discussion

GBM is the most common adult primary malignant brain tumor with poor median survival of about 12-15 months for patients receiving standard of care treatment [173] that can range from as little as 0.7 months to as long as 10 years [201]. Given this wide variation in OS, it is especially critical to identify early on during the course of therapy imaging markers that result in a worse outcome so that adjustments can be made to treatment strategies accordingly in order to personalize patient management [91]. As the revised Response Assessment in Neuro-Oncology (RANO) 2.0 criteria has been recently modified to use the post-RT MRI as a landmark from which to base subsequent tumor progression instead of the postsurgical MRI time point [211], it is even more crucial to explore the temporal changes in imaging features during the course of upfront radiotherapy treatment and investigate how therapeutic alterations on imaging relate to outcome measures.

In prior studies conducted at our institution, we have found that metabolic metrics from ^{1}H MRSI, along with lesion volumes post-RT and metabolic lesion volume at both mid-RT and post-RT, were significantly associated with the OS [212]. In addition, we were able to independently correlate diffusion and MRSI metrics with higher risk of tumor progression for patients treated with SOC + bevacizumab [78], [101] and for patients treated with SOC + enzastautin [91], [106], [207]. In this study, we delved deeper and performed a more comprehensive analysis by: 1) jointly analyzing anatomical MRI, DWI, DSC perfusion, and ^{1}H MRSI in our analysis; 2) expanding the feature space by incorporating more descriptive statistics metrics such as skewness, percentiles, kurtosis, etc.; 3) expanding our ROIs beyond the CEL and NEL; and 4) applying machine learning to improve prediction performance over more basic univariate and multivariate statistical analyses. Another advantage of our study is the rare addition of upfront treatment along with SOC RT and temozolomide used in our patient's cohort which allowed us to identify therapy-specific alterations that serve as markers of PFS and OS.

In both of our univariate analysis and our machine learning model, we found that patient age at the time of diagnosis, tumor volume, metabolic tumor volume, anatomic metrics (T2 FLA and T1C), diffusion metrics (FA and ADC), and MRSI metrics (CNI, lactate/lipid) were the most frequent significant predictors of both PFS and OS. This agrees with the results from our previous study where both anatomic and diffusion parameters were significantly associated with overall survival, and that the type of treatment does affect this association [212]. Using these parameters as features, our best random forest model achieved an AUC-ROC of 0.802 in predicting early PFS, and 0.725 in predicting early OS. Overall, it is easier to predict PFS output as compared to OS. This makes sense because the assumption that a patient's survival being fully dependent on the tumor's progression might not be entirely correct. Furthermore, the anti-angiogenic drugs were hypothesized to play a role in delaying or altering the radiomic markers for progression, making the pre-RT markers less reliable for prediction. Since pre-RT markers were predominantly used to predict OS (Table 7.3), it is understandable why OS prediction performance is lower compared to PFS (which relies more on mid-RT markers – Table 7.2).

Interestingly, incorporating mid-RT markers lowered the model performance for the OS model, while improving performance for the PFS model. We believe that both observations can also be explained by the effect of anti-angiogenic drugs. Theoretically, anti-angiogenic drugs can take away classic radiomic markers for progression but does not cause a big improvement in overall survival (thus pre-RT MRI might not perform as well in predicting PFS but does well in predicting OS). This was shown in chapter 6, where we observed that the addition of both enzastaurin and bevacizumab was significantly associated with longer PFS, while only Bevacizumab was associated with prolonged OS in our cohort (but to a much lesser degree of significance compared to PFS). We postulate that this may be due to the fact that both mid-RT imaging features and the PFS time points are more similarly affected by anti-angiogenic drugs, potentially revealing common features from mid-RT can help explain the resulting PFS, but not OS.

This study also found that using a mask that extends beyond the conventional anatomical (T2) lesions improved predictive performance. With smaller masks, the sum and count parameters tended to improve performance, while with larger masks, the median and kurtosis parameters were better features. This is not surprising given that smaller masks result in fewer voxels on which to perform histogram analyses, resulting in parameters that describe the data distribution (such as median, percentile, or kurtosis) not actually reflecting the true distribution, but also highlights the need for identifying subclinical infiltrating tumor cells that extend beyond the T2L and impact survival. This finding is highly significant for the design of future imaging studies to look beyond the CEL and T2L as lesion boundaries and include more descriptive statistical features than just median values.

Although we reported some promising results, our study still had several limitations. First, it did not include analysis of the post-RT MRI (at 8 weeks after the start of RT). While including imaging data from a mid-RT time point improved performance for the PFS model, it is likely that including imaging features from the post-RT scan will further improve prediction accuracy. Another limitation of our study is the lack of features that reflect a temporal imaging change (such as [Mid-RT median CNI] – [Pre-RT median CNI]). While this was intentional, as we believed the random forest model could still detect these alterations, including those features might also improve our model performance. Finally, our patient cohort, while much larger than many previous studies, only included 63 patients, who received additional anti-angiogenic treatment. Future studies will include additional patients who underwent only SOC treatment, both to increase the total dataset and to further confirm our findings with regards to the effect of anti-angiogenic drugs. We are also looking to develop models to predict outcome as a continuous output, incorporate feature reduction methods, and evaluate reproducibility of results with other machine learning models.

7.5. Conclusion

Our work highlighted the benefit of using multi-parametric MRI images from multiple time points when predicting PFS and OS. We found that age, the tumor volume, diffusion (FA and ADC), and metabolic (CNI, lactate/lipid) metrics were the most frequent significant predictors of both PFS and OS. Interestingly, incorporating mid-RT metrics improved the prediction of PFS, but prediction for early OS was mostly driven by pre-RT imaging metrics, likely due to anti-angiogenic drugs might delaying classic imaging markers of progression. Using a mask that extended beyond the conventional anatomical T2L improved predictive performance. Future analyses will predict outcome as a continuous measure, incorporate feature reduction methods, and evaluate reproducibility of results with other machine learning models.

Chapter 8. Conclusions and future directions

8.1. Conclusions

This dissertation provided a deep dive into various methods that can help improve the diagnosis and management of newly-diagnosed glioma using artificial intelligence and multi-parametric MRI.

In Chapter 5, we highlighted the promise of 1H-MRS data, both as individual metabolite levels and as the full 1D spectrum, in predicting tumor biology of tissue samples obtained during surgery. Particularly, we developed a deep learning model that utilized the wealth of information contained in the entire spectrum to predict voxel-wise histopathology of tumor cells, including cellularity, mitotic activity (Ki-67), and a new composite tumor aggressiveness index (CTAI; defined as the sum of normalized cellularity and Ki-67 times tumor score) using tissue samples with spatially mapped coordinates on 3D 1H-MRSI. Using the models developed, we were able to generate predictive maps for each of the histopathology metrics, which can be incorporated into noninvasive diagnosis, but also disease monitoring when obtaining tissue is not feasible.

In Chapter 6, we demonstrated the feasibility of using pre-treatment diffusion-weighted imaging and ^{1}H-MRSI, along with machine learning and deep learning techniques, to predict future regions of tumor progression and generate a hypothetical target volume for RT treatment. In our random forest models, progression was more challenging to predict with longer time from treatment, suggesting that time to progression should be added in subsequent modeling. For the deep learning segmentation task, we found that the best model was trained using a loss function that took the size of the original tumor into consideration. Our deep learning model using multi-parametric MRI performed better than current practice of performing a uniform 2cm expansion for RT treatment planning and no expansion, suggesting that multi-parametric MRI with deep learning has the potential to assist future RT treatment planning. We also explored multiple ways to evaluate model performance and found that both Tversky's metrics and a newly-developed

Individualized Progression Coverage Coefficient metric are better options for assessing performance of this segmentation task than the conventional Dice score.

In Chapter 7, we highlighted the benefit of using multi-parametric MRI images from multiple time points (pre-RT and mid-RT) when predicting PFS and OS. We found that the patient's age at the time of diagnosis, the tumor volume, metrics from diffusion MRI (FA and ADC), and metrics from metabolic MRI (CNI, lactate/lipid) were the most frequent significant predictors of both PFS and OS. Interestingly, while incorporating mid-RT metrics improved the prediction of early PFS (PFS < 45 weeks), prediction of early OS (OS < 76 weeks) was mostly driven by pre-RT radiomic metrics, confirming our hypothesis that anti-angiogenic drugs might play a role in delaying classic radiomic predictors for progression. We also observed that using a mask that extends beyond the conventional anatomical (T2) lesions improved predictive performance.

8.2. Future directions

In Chapter 5, we focused mostly on the use of deep learning on the full MRSI spectrum to predict histopathology metrics. While our results are acceptable, this approach is still only limited to the use of a single MRI modality. It has been reported previously that physiologic MRI like DWI or DSC perfusion have promise for predicting Ki-67 and cellularity metrics. Therefore, in subsequent modeling, we are looking to incorporate these MRI modalities into our modeling. To do this, for each tissue sample, we will use deep learning methods to individually extract features from each MRI map, before concatenating them with our MRSI features generated in Chapter 5. We hope that this will improve our performance. Finally, we hypothesized that our CTAI metric can be more useful than both Ki-67 and cellularity, and we are looking to further explore the utility CTAI metric in outcome studies.

In Chapter 6, we developed the patient specific Progression Coverage Coefficient (PCC) and used it as both the loss function for our deep learning model as well as the evaluation metric. While the approach alleviates some of the class imbalance issues inherent in lesion segmentation

tasks judging from its performance as loss function, the use of PCC as an evaluation metric needs to be studied further. We hypothesize that the PCC makes sense clinically because patients with larger tumors benefit more from a slightly higher specificity (as to reduce normal brain irradiation), while patients with smaller tumors benefits more from a slightly higher sensitivity (as to ensure the radiation of the entire tumor). We also want to further improve our model with additional patients, especially models that do not require MRSI, as this data is not routinely collected clinically.

In Chapter 7, we want to further study the effects of anti-angiogenic therapies by including another cohort of patients who only received the standard of care treatment. This will allow us to determine whether a different set of imaging features are relevant in the absence of anti-angiogenic therapy. We also want to conduct the study using continuous PFS and OS in order to generate more useful future models. Lastly, we want to include other MRI time points into our models, including the post RT MRI (8 weeks into RT treatment), as well as additional follow-up MRI post treatment, but this would involve increasing datasets.

References

[1] Q. T. Ostrom *et al.*, "The epidemiology of glioma in adults: A state of the science review," *Neuro-Oncology*, vol. 16, no. 7. Oxford University Press, pp. 896–913, 2014. doi: 10.1093/neuonc/nou087.

[2] R. Stupp *et al.*, "Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma," *New England Journal of Medicine*, vol. 352, no. 10, pp. 987–996, Mar. 2005, doi: 10.1056/NEJMoa043330.

[3] R. Stupp *et al.*, "Effect of Tumor-Treating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Survival in Patients With Glioblastoma," *JAMA*, vol. 318, no. 23, p. 2306, Dec. 2017, doi: 10.1001/jama.2017.18718.

[4] T. R. McKnight, S. M. Noworolski, D. B. Vigneron, and S. J. Nelson Dr rer Nat, "An automated technique for the quantitative assessment of 3D-MRSI data from patients with glioma," *Journal of Magnetic Resonance Imaging*, vol. 13, no. 2, 2001, doi: 10.1002/1522-2586(200102)13:2<167::aid-jmri1026>3.3.co;2-b.

[5] T. R. McKnight *et al.*, "Histopathological validation of a three-dimensional magnetic resonance spectroscopy index as a predictor of tumor presence," *J Neurosurg*, vol. 97, no. 4, 2002, doi: 10.3171/jns.2002.97.4.0794.

[6] I. Park, A. P. Chen, M. L. Zierhut, E. Ozturk-Isik, D. B. Vigneron, and S. J. Nelson, "Implementation of 3 T lactate-edited 3D 1H MR spectroscopic imaging with flyback echo-planar readout for gliomas patients," *Ann Biomed Eng*, vol. 39, no. 1, pp. 193–204, Jan. 2011, doi: 10.1007/s10439-010-0128-x.

[7] S. J. Nelson *et al.*, “In vivo molecular imaging for planning radiation therapy of gliomas: An application of 1H MRSI,” *Journal of Magnetic Resonance Imaging*, vol. 16, no. 4, pp. 464–476, Oct. 2002, doi: 10.1002/jmri.10183.

[8] S. J. Nelson, “Assessment of therapeutic response and treatment planning for brain tumors using metabolic and physiological MRI,” *NMR Biomed*, vol. 24, no. 6, pp. 734–749, Jul. 2011, doi: 10.1002/NBM.1669.

[9] M. Anwar *et al.*, “Identifying Voxels at Risk for Progression in Glioblastoma Based on Dosimetry, Physiologic and Metabolic MRI,” *Radiat Res*, vol. 188, no. 3, p. 303, Jul. 2017, doi: 10.1667/RR14662.1.

[10] Q. T. Ostrom, G. Cioffi, K. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, “CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014–2018,” *Neuro Oncol*, vol. 23, no. Supplement_3, pp. iii1–iii105, Oct. 2021, doi: 10.1093/neuonc/noab200.

[11] L. R. Schaff and I. K. Mellinghoff, “Glioblastoma and Other Primary Brain Malignancies in Adults,” *JAMA*, vol. 329, no. 7, p. 574, Feb. 2023, doi: 10.1001/jama.2023.0023.

[12] D. N. Louis *et al.*, “The 2021 WHO Classification of Tumors of the Central Nervous System: a summary,” *Neuro Oncol*, vol. 23, no. 8, pp. 1231–1251, Aug. 2021, doi: 10.1093/neuonc/noab106.

[13] L. Valentinis *et al.*, “Headache attributed to intracranial tumours: A prospective cohort study,” *Cephalalgia*, vol. 30, no. 4, pp. 389–398, Apr. 2010, doi: 10.1111/j.1468-2982.2009.01970.x.

[14] Q. T. Ostrom *et al.*, “CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015–2019,” *Neuro Oncol*, vol. 24, no. Supplement_5, pp. v1–v95, Oct. 2022, doi: 10.1093/neuonc/noac202.

[15] H. Chen *et al.*, “Mutant IDH1 and seizures in patients with glioma,” *Neurology*, vol. 88, no. 19, pp. 1805–1813, May 2017, doi: 10.1212/WNL.0000000000003911.

[16] Benjamin M. Ellingson *et al.*, “Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials,” *Neuro Oncol*, Aug. 2015, doi: 10.1093/neuonc/nov095.

[17] J. E. Villanueva-Meyer, M. C. Mabray, and S. Cha, “Current Clinical Brain Tumor Imaging,” *Neurosurgery*, vol. 81, no. 3, pp. 397–415, Sep. 2017, doi: 10.1093/neuros/nyx103.

[18] H. K. Gumprecht, D. C. Widenka, and C. B. Lumenta, “Brain Lab VectorVision Neuronavigation System: Technology and Clinical Experiences in 131 Cases,” *Neurosurgery*, vol. 44, no. 1, pp. 97–104, Jan. 1999, doi: 10.1097/00006123-199901000-00056.

[19] S. H. Torp, O. Solheim, and A. J. Skjulsvik, “The WHO 2021 Classification of Central Nervous System tumours: a practical update on what neurosurgeons need to know—a minireview,” *Acta Neurochir (Wien)*, vol. 164, no. 9, pp. 2453–2464, Jul. 2022, doi: 10.1007/s00701-022-05301-y.

[20] American Society of Clinical Oncology, “Brain Tumor,” http://www.cancer.net/cancer-types/brain-tumor/view-all.

[21] R. F. Barajas *et al.*, “Regional variation in histopathologic features of tumor specimens from treatment-naive glioblastoma correlates with anatomic and physiologic MR

Imaging," *Neuro Oncol*, vol. 14, no. 7, pp. 942–954, Jul. 2012, doi: 10.1093/neuonc/nos128.

[22] O. Tynninen *et al.*, "MRI enhancement and microvascular density in gliomas. Correlation with tumor cell proliferation," *Invest Radiol*, vol. 34, no. 6, 1999, doi: 10.1097/00004424-199906000-00007.

[23] T. Scholzen and J. Gerdes, "The Ki-67 protein: From the known and the unknown," *J Cell Physiol*, vol. 182, no. 3, pp. 311–322, Mar. 2000, doi: 10.1002/(SICI)1097-4652(200003)182:3<311::AID-JCP1>3.0.CO;2-9.

[24] K. Uehara *et al.*, "Patterns of failure after multimodal treatments for high-grade glioma: effectiveness of MIB-1 labeling index.," *Radiat Oncol*, vol. 7, p. 104, Jun. 2012, doi: 10.1186/1748-717X-7-104.

[25] K. P. Padmavathy *et al.*, "Clinical Evaluation of Proline, Glutamic acid, and Leucine-Rich Protein 1 Expression in Astrocytomas and Correlations with the Proliferation Marker Ki-67," *Journal of Molecular Neuroscience*, vol. 71, no. 4, pp. 724–733, Apr. 2021, doi: 10.1007/s12031-020-01690-w.

[26] A. L. Johannessen and S. H. Torp, "The clinical value of Ki-67/MIB-1 labeling index in human astrocytomas," *Pathology & Oncology Research*, vol. 12, no. 3, pp. 143–147, Sep. 2006, doi: 10.1007/BF02893360.

[27] A. J. Skjulsvik, J. N. Mørk, M. O. Torp, and S. H. Torp, "Ki-67/MIB-1 immunostaining in a cohort of human gliomas.," *Int J Clin Exp Pathol*, vol. 7, no. 12, pp. 8905–10, 2014.

[28] S. S. Krishnan, S. Muthiah, S. Rao, S. S. Salem, V. C. Madabhushi, and A. Mahadevan, "Mindbomb Homolog-1 Index in the Prognosis of High-Grade Glioma and Its

Clinicopathological Correlation," *J Neurosci Rural Pract*, vol. 10, no. 02, pp. 185–193, Apr. 2019, doi: 10.4103/jnrp.jnrp_374_18.

[29] S. A. Bobholz *et al.*, "Radiomic features of multiparametric mri present stable associations with analogous histological features in patients with brain cancer," *Tomography*, vol. 6, no. 2, pp. 160–169, Jun. 2020, doi: 10.18383/j.tom.2019.00029.

[30] J. B. Elder *et al.*, "Histologic findings associated with laser interstitial thermotherapy for glioblastoma multiforme," *Diagn Pathol*, vol. 14, no. 1, Feb. 2019, doi: 10.1186/s13000-019-0794-4.

[31] P. Karschnia *et al.*, "Evidence-based recommendations on categories for extent of resection in diffuse glioma," *Eur J Cancer*, vol. 149, pp. 23–33, May 2021, doi: 10.1016/j.ejca.2021.03.002.

[32] A. Rodríguez-Camacho *et al.*, "Glioblastoma Treatment: State-of-the-Art and Future Perspectives," *Int J Mol Sci*, vol. 23, no. 13, p. 7207, Jun. 2022, doi: 10.3390/ijms23137207.

[33] J. Cluceru, "Toward improving the diagnosis of glioma," UC San Francisco, 2021. [Online]. Available: https://escholarship.org/uc/item/16c430mk

[34] M. van den Bent *et al.*, "Long-term efficacy of early versus delayed radiotherapy for low-grade astrocytoma and oligodendroglioma in adults: the EORTC 22845 randomised trial," *The Lancet*, vol. 366, no. 9490, pp. 985–990, Sep. 2005, doi: 10.1016/S0140-6736(05)67070-5.

[35] N. A. Oberheim Bush and S. Chang, "Treatment Strategies for Low-Grade Glioma in Adults," *J Oncol Pract*, vol. 12, no. 12, pp. 1235–1241, Dec. 2016, doi: 10.1200/JOP.2016.018622.

[36] M. Davis, “Glioblastoma: Overview of Disease and Treatment,” *Clin J Oncol Nurs*, vol. 20, no. 5, pp. S2–S8, Oct. 2016, doi: 10.1188/16.CJON.S1.2-8.

[37] T. M. Pisansky, “External-Beam Radiotherapy for Localized Prostate Cancer,” *New England Journal of Medicine*, vol. 355, no. 15, pp. 1583–1591, Oct. 2006, doi: 10.1056/NEJMct055263.

[38] D. Przystupski *et al.*, “The Cytoprotective Role of Antioxidants in Mammalian Cells Under Rapidly Varying UV Conditions During Stratospheric Balloon Campaign,” *Front Pharmacol*, vol. 10, Aug. 2019, doi: 10.3389/fphar.2019.00851.

[39] E. V. Maani and C. V. Maani, *Radiation Therapy*. 2023.

[40] F. Keime-Guibert *et al.*, “Radiotherapy for Glioblastoma in the Elderly,” *New England Journal of Medicine*, vol. 356, no. 15, pp. 1527–1535, Apr. 2007, doi: 10.1056/NEJMoa065901.

[41] A. Karachi, F. Dastmalchi, D. A. Mitchell, and M. Rahman, “Temozolomide for immunomodulation in the treatment of glioblastoma,” *Neuro Oncol*, vol. 20, no. 12, pp. 1566–1572, Nov. 2018, doi: 10.1093/neuonc/noy072.

[42] A. C. Tan, D. M. Ashley, G. Y. López, M. Malinzak, H. S. Friedman, and M. Khasraw, “Management of glioblastoma: State of the art and future directions,” *CA Cancer J Clin*, vol. 70, no. 4, pp. 299–312, Jul. 2020, doi: 10.3322/caac.21613.

[43] O. L. Chinot *et al.*, “Bevacizumab plus Radiotherapy–Temozolomide for Newly Diagnosed Glioblastoma,” *New England Journal of Medicine*, vol. 370, no. 8, pp. 709–722, Feb. 2014, doi: 10.1056/NEJMoa1308345.

[44] E. Le Rhun *et al.*, “Molecular targeted therapy of glioblastoma,” *Cancer Treat Rev*, vol. 80, p. 101896, Nov. 2019, doi: 10.1016/j.ctrv.2019.101896.

[45] W. Wick *et al.*, “Phase III study of enzastaurin compared with lomustine in the treatment of recurrent intracranial glioblastoma.,” *J Clin Oncol*, vol. 28, no. 7, pp. 1168–74, Mar. 2010, doi: 10.1200/JCO.2009.23.2595.

[46] S. H. Choi *et al.*, “Impact of Including Peritumoral Edema in Radiotherapy Target Volume on Patterns of Failure in Glioblastoma following Temozolomide-based Chemoradiotherapy,” *Sci Rep*, vol. 7, no. 1, p. 42148, Feb. 2017, doi: 10.1038/srep42148.

[47] G. Minniti *et al.*, “Patterns of failure and comparison of different target volume delineations in patients with glioblastoma treated with conformal radiotherapy plus concomitant and adjuvant temozolomide,” *Radiotherapy and Oncology*, vol. 97, no. 3, pp. 377–381, Dec. 2010, doi: 10.1016/j.radonc.2010.08.020.

[48] B. J. Gebhardt, M. C. Dobelbower, W. H. Ennis, A. K. Bag, J. M. Markert, and J. B. Fiveash, “Patterns of failure for glioblastoma multiforme following limited-margin radiation and concurrent temozolomide,” *Radiation Oncology*, vol. 9, no. 1, p. 130, Dec. 2014, doi: 10.1186/1748-717X-9-130.

[49] M. Rapp, J. Baernreuther, B. Turowski, H.-J. Steiger, M. Sabel, and M. A. Kamp, “Recurrence Pattern Analysis of Primary Glioblastoma,” *World Neurosurg*, vol. 103, pp. 733–740, Jul. 2017, doi: 10.1016/j.wneu.2017.04.053.

[50] M. M. Kim *et al.*, “Dose-intensified chemoradiation is associated with altered patterns of failure and favorable survival in patients with newly diagnosed glioblastoma,” *J Neurooncol*, vol. 143, no. 2, pp. 313–319, Jun. 2019, doi: 10.1007/s11060-019-03166-3.

[51] K. Reddy, L. E. Gaspar, B. D. Kavanagh, and C. Chen, “Hypofractionated intensity-modulated radiotherapy with temozolomide chemotherapy may alter the patterns of failure

in patients with glioblastoma multiforme," *J Med Imaging Radiat Oncol*, vol. 58, no. 6, pp. 714–721, Dec. 2014, doi: 10.1111/1754-9485.12185.

[52] T. A. Lawrie *et al.*, "Long-term neurocognitive and other side effects of radiotherapy, with or without chemotherapy, for glioma," *Cochrane Database of Systematic Reviews*, Aug. 2019, doi: 10.1002/14651858.CD013047.pub2.

[53] Z. VÉGVÁRY *et al.*, "Adaptive Radiotherapy for Glioblastoma Multiforme – The Impact on Disease Outcome," *Anticancer Res*, vol. 40, no. 8, pp. 4237–4244, Aug. 2020, doi: 10.21873/anticanres.14425.

[54] L. Zheng *et al.*, "The Definition and Delineation of the Target Area of Radiotherapy Based on the Recurrence Pattern of Glioblastoma After Temozolomide Chemoradiotherapy.," *Front Oncol*, vol. 10, p. 615368, 2020, doi: 10.3389/fonc.2020.615368.

[55] H. S. Phillips *et al.*, "Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis", doi: 10.1016/j.ccr.2006.02.019.

[56] P. Kleihues and H. Ohgaki, "Genetics of glioma progression and the definition of primary and secondary glioblastoma," *Brain Pathology*, vol. 7, no. 4, pp. 1131–1136, 1997, doi: 10.1111/J.1750-3639.1997.TB00993.X.

[57] R. W. Brown, Y. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging*. Wiley, 2014. doi: 10.1002/9781118633953.

[58] V. P. B. Grover, J. M. Tognarelli, M. M. E. Crossey, I. J. Cox, S. D. Taylor-Robinson, and M. J. W. McPhail, "Magnetic Resonance Imaging: Principles and Techniques: Lessons for

Clinicians," *J Clin Exp Hepatol*, vol. 5, no. 3, pp. 246–255, Sep. 2015, doi: 10.1016/j.jceh.2015.08.001.

[59] T. Puiseux, A. Sewonu, R. Moreno, S. Mendez, and F. Nicoud, "Numerical simulation of time-resolved 3D phase-contrast magnetic resonance imaging," *PLoS One*, vol. 16, no. 3, p. e0248816, Mar. 2021, doi: 10.1371/journal.pone.0248816.

[60] Nikhil Bhagwat and Erin Dickie, "Structural MRI (Pre)processing and Neuroimaging Analysis," https://carpentries-incubator.github.io/SDC-BIDS-sMRI/index.html.

[61] M. Mahesh, "The Essential Physics of Medical Imaging, Third Edition.," *Med Phys*, vol. 40, no. 7, Jul. 2013, doi: 10.1118/1.4811156.

[62] M. Rohrer, H. Bauer, J. Mintorovitch, M. Requardt, and H.-J. Weinmann, "Comparison of Magnetic Properties of MRI Contrast Media Solutions at Different Magnetic Field Strengths," *Invest Radiol*, vol. 40, no. 11, pp. 715–724, Nov. 2005, doi: 10.1097/01.rli.0000184756.66360.d3.

[63] N. Upadhyay and A. D. Waldman, "Conventional MRI evaluation of gliomas," *Br J Radiol*, vol. 84, no. special_issue_2, pp. S107–S111, Dec. 2011, doi: 10.1259/bjr/65711810.

[64] D. A. Gutman *et al.*, "MR Imaging Predictors of Molecular Profile and Survival: Multi-institutional Study of the TCGA Glioblastoma Data Set," *Radiology*, vol. 267, no. 2, pp. 560–569, May 2013, doi: 10.1148/radiol.13120118.

[65] R. R. Colen *et al.*, "Imaging genomic mapping of an invasive MRI phenotype predicts patient outcome and metabolic dysfunction: a TCGA glioma phenotype research group project," *BMC Med Genomics*, vol. 7, no. 1, p. 30, Dec. 2014, doi: 10.1186/1755-8794-7-30.

[66] D. Yang, G. Rao, J. Martinez, A. Veeraraghavan, and A. Rao, “Evaluation of tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma,” *Med Phys*, vol. 42, no. 11, pp. 6725–6735, Nov. 2015, doi: 10.1118/1.4934373.

[67] D. Le Bihan, “Looking into the functional architecture of the brain with diffusion MRI,” *Nat Rev Neurosci*, vol. 4, no. 6, pp. 469–480, Jun. 2003, doi: 10.1038/nrn1119.

[68] N. S. White *et al.*, “Diffusion-weighted imaging in cancer: physical foundations and applications of restriction spectrum imaging.,” *Cancer Res*, vol. 74, no. 17, pp. 4638–52, Sep. 2014, doi: 10.1158/0008-5472.CAN-13-3534.

[69] D.-M. Koh and D. J. Collins, “Diffusion-Weighted MRI in the Body: Applications and Challenges in Oncology,” *American Journal of Roentgenology*, vol. 188, no. 6, pp. 1622–1635, Jun. 2007, doi: 10.2214/AJR.06.1403.

[70] E. M. Charles-Edwards and N. M. deSouza, “Diffusion-weighted magnetic resonance imaging and its application to cancer.,” *Cancer Imaging*, vol. 6, no. 1, pp. 135–43, Sep. 2006, doi: 10.1102/1470-7330.2006.0021.

[71] B. J. Jellison, A. S. Field, J. Medow, M. Lazar, M. S. Salamat, and A. L. Alexander, “Diffusion tensor imaging of cerebral white matter: a pictorial review of physics, fiber tract anatomy, and tumor imaging patterns.,” *AJNR Am J Neuroradiol*, vol. 25, no. 3, pp. 356–69, Mar. 2004.

[72] S. E. Maier, Y. Sun, and R. V Mulkern, “Diffusion imaging of brain tumors”, doi: 10.1002/nbm.1544.

[73] T. Sugahara *et al.*, “Usefulness of diffusion-weighted MRI with echo-planar technique in the evaluation of cellularity in gliomas,” *Journal of Magnetic Resonance Imaging*, vol. 9,

no. 1, pp. 53–60, Jan. 1999, doi: 10.1002/(SICI)1522-2586(199901)9:1<53::AID-JMRI7>3.0.CO;2-2.

[74] L. Chen *et al.*, "The Correlation between Apparent Diffusion Coefficient and Tumor Cellularity in Patients: A Meta-Analysis," *PLoS One*, vol. 8, no. 11, p. e79008, Nov. 2013, doi: 10.1371/journal.pone.0079008.

[75] G. A. Gihr *et al.*, "Diffusion Profiling via a Histogram Approach Distinguishes Low-grade from High-grade Meningiomas, Can Reflect the Respective Proliferative Potential and Progesterone Receptor Status," *Mol Imaging Biol*, vol. 20, no. 4, pp. 632–640, Aug. 2018, doi: 10.1007/s11307-018-1166-2.

[76] A. Surov, H. J. Meyer, and A. Wienke, "Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: a meta-analysis," *Oncotarget*, vol. 8, no. 35, pp. 59492–59499, Aug. 2017, doi: 10.18632/oncotarget.17752.

[77] W. B. Pope *et al.*, "Apparent Diffusion Coefficient Histogram Analysis Stratifies Progression-Free Survival in Newly Diagnosed Bevacizumab-Treated Glioblastoma," *American Journal of Neuroradiology*, vol. 32, no. 5, pp. 882–889, May 2011, doi: 10.3174/ajnr.A2385.

[78] Q. Wen *et al.*, "Comparison of ADC metrics and their association with outcome for patients with newly diagnosed glioblastoma being treated with radiation therapy, temozolomide, erlotinib and bevacizumab," *J Neurooncol*, vol. 121, no. 2, pp. 331–339, Jan. 2015, doi: 10.1007/s11060-014-1636-6.

[79] W. Chang *et al.*, "Diffusion MRI Characteristics after Concurrent Radiochemotherapy Predicts Progression-Free and Overall Survival in Newly Diagnosed Glioblastoma," *Tomography*, vol. 1, no. 1, pp. 37–43, Sep. 2015, doi: 10.18383/j.tom.2015.00115.

[80] G. Gihr *et al.*, “Diffusion weighted imaging in high-grade gliomas: A histogram-based analysis of apparent diffusion coefficient profile.,” *PLoS One*, vol. 16, no. 4, p. e0249878, 2021, doi: 10.1371/journal.pone.0249878.

[81] Y. W. Park *et al.*, “Prediction of *IDH1* -Mutation and 1p/19q-Codeletion Status Using Preoperative MR Imaging Phenotypes in Lower Grade Gliomas,” *American Journal of Neuroradiology*, vol. 39, no. 1, pp. 37–42, Jan. 2018, doi: 10.3174/ajnr.A5421.

[82] P. Eichinger *et al.*, “Diffusion tensor image features predict IDH genotype in newly diagnosed WHO grade II/III gliomas,” *Sci Rep*, vol. 7, no. 1, p. 13396, Oct. 2017, doi: 10.1038/s41598-017-13679-4.

[83] B. Jeurissen, M. Descoteaux, S. Mori, and A. Leemans, “Diffusion MRI fiber tractography of the brain,” *NMR in Biomedicine*, vol. 32, no. 4. 2019. doi: 10.1002/nbm.3785.

[84] N. J. Mickevicius *et al.*, “Location of brain tumor intersecting white matter tracts predicts patient prognosis,” *J Neurooncol*, vol. 125, no. 2, pp. 393–400, Nov. 2015, doi: 10.1007/s11060-015-1928-5.

[85] T. Kok, “Magnetic Resonance Spectroscopic Imaging with 2D Spectroscopy for the Detection of Brain Metabolites ,” MASSACHUSETTS INTITUTE OF TECHNOLOGY, 2012.

[86] H. Zhu and P. B. Barker, “MR spectroscopy and spectroscopic imaging of the brain.,” *Methods Mol Biol*, vol. 711, pp. 203–26, 2011, doi: 10.1007/978-1-61737-992-5_9.

[87] J. R. Alger, “Quantitative Proton Magnetic Resonance Spectroscopy and Spectroscopic Imaging of the Brain,” *Topics in Magnetic Resonance Imaging*, vol. 21, no. 2, pp. 115–128, Apr. 2010, doi: 10.1097/RMR.0b013e31821e568f.

[88] Š. Mierisová and M. Ala-Korpela, “MR spectroscopy quantitation: a review of frequency domain methods,” *NMR Biomed*, vol. 14, no. 4, pp. 247–259, Jun. 2001, doi: 10.1002/nbm.697.

[89] S. E. Hoch, I. I. Kirov, and A. Tal, “When are metabolic ratios superior to absolute quantification? A statistical analysis,” *NMR Biomed*, vol. 30, no. 7, 2017, doi: 10.1002/nbm.3710.

[90] Y. Li *et al.*, “Survival analysis in patients with newly diagnosed glioblastoma using pre- and postradiotherapy MR spectroscopic imaging,” *Neuro Oncol*, vol. 15, no. 5, 2013, doi: 10.1093/neuonc/nos334.

[91] S. J. Nelson *et al.*, “Association of early changes in 1H MRSI parameters with survival for patients with newly diagnosed glioblastoma receiving a multimodality treatment regimen,” *Neuro Oncol*, vol. 19, no. 3, pp. 430–439, 2017, doi: 10.1093/neuonc/now159.

[92] A. Horská and P. B. Barker, “Imaging of brain tumors: MR spectroscopy and metabolic imaging.,” *Neuroimaging Clin N Am*, vol. 20, no. 3, pp. 293–310, Aug. 2010, doi: 10.1016/j.nic.2010.04.003.

[93] S. Delorme, “Applications of MRS in the evaluation of focal malignant brain lesions,” *Cancer Imaging*, vol. 6, no. 1, pp. 95–99, 2006, doi: 10.1102/1470-7330.2006.0015.

[94] D. Galanaud *et al.*, “Use of proton magnetic resonance spectroscopy of the brain to differentiate gliomatosis cerebri from low-grade glioma,” *J Neurosurg*, vol. 98, no. 2, pp. 269–276, Feb. 2003, doi: 10.3171/jns.2003.98.2.0269.

[95] J. Penrice *et al.*, “Proton Magnetic Resonance Spectroscopy of the Brain in Normal Preterm and Term Infants, and Early Changes after Perinatal Hypoxia-Ischemia,” *Pediatr Res*, vol. 40, no. 1, pp. 6–14, Jul. 1996, doi: 10.1203/00006450-199607000-00002.

[96] J. R. Alger *et al.*, “Metabolism of human gliomas: assessment with H-1 MR spectroscopy and F-18 fluorodeoxyglucose PET.,” *Radiology*, vol. 177, no. 3, pp. 633–641, Dec. 1990, doi: 10.1148/radiology.177.3.2243962.

[97] E. J. Delikatny, S. Chawla, D.-J. Leung, and H. Poptani, “MR-visible lipids and the tumor microenvironment.,” *NMR Biomed*, vol. 24, no. 6, pp. 592–611, Jul. 2011, doi: 10.1002/nbm.1661.

[98] J. Guo *et al.*, “The relationship between cho/naa and glioma metabolism: Implementation for margin delineation of cerebral gliomas,” *Acta Neurochir (Wien)*, vol. 154, no. 8, pp. 1361–1370, Aug. 2012, doi: 10.1007/s00701-012-1418-x.

[99] A. Stadlbauer, M. Buchfelder, M. Doelken, T. Hammen, and O. Ganslandt, “Magnetic Resonance Spectroscopic Imaging for Visualization of the Infiltration Zone of Glioma,” *Cent Eur Neurosurg*, vol. 72, no. 02, pp. 63–69, May 2011, doi: 10.1055/s-0030-1253410.

[100] J. A. Osorio *et al.*, “3D 1H MRSI of brain tumors at 3.0 tesla using an eight-channel phased-array head coil,” *Journal of Magnetic Resonance Imaging*, vol. 26, no. 1, pp. 23–30, Jul. 2007, doi: 10.1002/jmri.20970.

[101] S. J. Nelson *et al.*, “Serial analysis of 3D H-1 MRSI for patients with newly diagnosed GBM treated with combination therapy that includes bevacizumab,” *J Neurooncol*, vol. 130, no. 1, pp. 171–179, Oct. 2016, doi: 10.1007/s11060-016-2229-3.

[102] T. L. Luks *et al.*, “Relationship of In Vivo MR Parameters to Histopathological and Molecular Characteristics of Newly Diagnosed, Nonenhancing Lower-Grade Gliomas,” *Transl Oncol*, vol. 11, no. 4, pp. 941–949, Aug. 2018, doi: 10.1016/j.tranon.2018.05.005.

[103] M. Muruganandham *et al.*, “3-Dimensional Magnetic Resonance Spectroscopic Imaging at 3 Tesla for Early Response Assessment of Glioblastoma Patients During External Beam

Radiation Therapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 90, no. 1, pp. 181–189, Sep. 2014, doi: 10.1016/j.ijrobp.2014.05.014.

[104] E. Matsusue, J. R. Fink, J. K. Rockhill, T. Ogawa, and K. R. Maravilla, "Distinction between glioma progression and post-radiation change by combined physiologic MR imaging," *Neuroradiology*, vol. 52, no. 4, pp. 297–306, Apr. 2010, doi: 10.1007/s00234-009-0613-9.

[105] J. S. Cordova *et al.*, "Simulating the Effect of Spectroscopic MRI as a Metric for Radiation Therapy Planning in Patients with Glioblastoma," *Tomography*, vol. 2, no. 4, pp. 366–373, Dec. 2016, doi: 10.18383/j.tom.2016.00187.

[106] E. Essock-Burns *et al.*, "Assessment of perfusion MRI-derived parameters in evaluating and predicting response to antiangiogenic therapy in patients with newly diagnosed glioblastoma," *Neuro Oncol*, vol. 13, no. 1, pp. 119–131, Jan. 2011, doi: 10.1093/neuonc/noq143.

[107] J. E. Eckel-Passow *et al.*, " Glioma Groups Based on 1p/19q, IDH , and TERT Promoter Mutations in Tumors ," *New England Journal of Medicine*, vol. 372, no. 26, pp. 2499–2508, Jun. 2015, doi: 10.1056/nejmoa1407279.

[108] R. Ogura *et al.*, "Immunohistochemical profiles of IDH1, MGMT and P53: Practical significance for prognostication of patients with diffuse gliomas," *Neuropathology*, vol. 35, no. 4, pp. 324–335, Aug. 2015, doi: 10.1111/neup.12196.

[109] L. E. Jalbert *et al.*, "Metabolic profiling of IDH mutation and malignant progression in infiltrating glioma," *Sci Rep*, vol. 7, Mar. 2017, doi: 10.1038/srep44792.

[110] N. Dikaios, “Deep learning magnetic resonance spectroscopy fingerprints of brain tumours using quantum mechanically synthesised data,” *NMR Biomed*, vol. 34, no. 4, Apr. 2021, doi: 10.1002/nbm.4479.

[111] Z. Iqbal, D. Nguyen, G. Hangel, S. Motyka, W. Bogner, and S. Jiang, “Super-Resolution 1H Magnetic Resonance Spectroscopic Imaging Utilizing Deep Learning,” *Front Oncol*, vol. 9, Oct. 2019, doi: 10.3389/fonc.2019.01010.

[112] M. A. Mazurowski, M. Buda, A. Saha, and M. R. Bashir, “Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI,” *Journal of Magnetic Resonance Imaging*, vol. 49, no. 4, pp. 939–954, Apr. 2019, doi: 10.1002/jmri.26534.

[113] A. Bohr and K. Memarzadeh, “The rise of artificial intelligence in healthcare applications,” in *Artificial Intelligence in Healthcare*, Elsevier, 2020, pp. 25–60. doi: 10.1016/B978-0-12-818438-7.00002-2.

[114] A. M. Rahmani *et al.*, “Machine Learning (ML) in Medicine: Review, Applications, and Challenges,” *Mathematics*, vol. 9, no. 22, p. 2970, Nov. 2021, doi: 10.3390/math9222970.

[115] Q. An, S. Rahman, J. Zhou, and J. J. Kang, “A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges,” *Sensors*, vol. 23, no. 9, p. 4178, Apr. 2023, doi: 10.3390/s23094178.

[116] M. Girolami, *A First Course in Machine Learning*. Chapman and Hall/CRC, 2011. doi: 10.1201/9781466506299.

[117] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004. doi: 10.1017/CBO9780511804441.

[118] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009. doi: 10.1007/978-0-387-84858-7.

[119] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," 2020.

[120] A. Antoniou, A. Storkey, and H. Edwards, "DATA AUGMENTATION GENERATIVE ADVERSARIAL NETWORKS".

[121] A. Anand, G. Pugalenthi, G. B. Fogel, and • P N Suganthan, "An approach for classification of highly imbalanced data using weighting and undersampling", doi: 10.1007/s00726-010-0595-2.

[122] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[123] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A Comprehensive Survey of Loss Functions in Machine Learning," vol. 9, pp. 187–212, 2022, doi: 10.1007/s40745-020-00253-5.

[124] J. Brabec and L. Machlica, "Bad practices in evaluation methodology relevant to class-imbalanced problems".

[125] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," 2015, doi: 10.1371/journal.pone.0118432.

[126] N. Mubin Zahari *et al.*, "Joint Learning with both Classification and Regression Models for Age Prediction Prediction of Future Ozone Concentration for Next Three Days Using Linear Regression and Nonlinear Regression Models Cycling performance prediction based on cadence analysis by using multiple regression", doi: 10.1088/1742-6596/1168/3/032016.

[127] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation”, doi: 10.7717/peerj-cs.623.

[128] S. He, R. Bao, P. E. Grant, and Y. Ou, “U-Netmer: U-Net meets Transformer for medical image segmentation”.

[129] H. Cao *et al.*, “Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation”, Accessed: Oct. 19, 2023. [Online]. Available: https://github.com/HuCaoFighting/Swin-Unet.

[130] J. Chen *et al.*, “TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation”, Accessed: Oct. 19, 2023. [Online]. Available: https://github.com/Beckschen/

[131] X. Jia, J. Bartlett, T. Zhang, W. Lu, Z. Qiu, and J. Duan, “U-Net vs Transformer: Is U-Net Outdated in Medical Image Registration?”, Accessed: Oct. 19, 2023. [Online]. Available: https://github.com/xi-jia/LKU-Net.

[132] J. Ma, “Segmentation Loss Odyssey,” 2020, Accessed: Oct. 19, 2023. [Online]. Available: https://github.com/JunMa11/SegLoss.

[133] S. Jadon, “A survey of loss functions for semantic segmentation”, Accessed: Oct. 19, 2023. [Online]. Available: https://github.com/shruti-jadon/

[134] M. Montazerolghaem, Y. Sun, G. Sasso, and A. Haworth, “U-Net Architecture for Prostate Segmentation: The Impact of Loss Function on System Performance,” *Bioengineering*, vol. 10, no. 4, p. 412, Mar. 2023, doi: 10.3390/bioengineering10040412.

[135] J. Heaton, “An Empirical Analysis of Feature Engineering for Predictive Modeling”.

[136] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” vol. 2, p. 160, 2021, doi: 10.1007/s42979-021-00592-x.

[137] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, “Improvements to Platt’s SMO Algorithm for SVM Classifier Design”, Accessed: Oct. 22, 2023. [Online]. Available: http://direct.mit.edu/neco/article-pdf/13/3/637/814768/089976601300014493.pdf?casa_token=NWBXz9xLad8AAAAA:33jRdEzuiCMG24B0X-gywCZPe9sWigomOZHdyEjSxok7Wg-Rxw4u5dw1uVx8Cm6GKFzBUXo

[138] S. Le Cessie and J. C. Van Houwelingen, “Ridge Estimators in Logistic Regression,” *Appl Stat*, vol. 41, no. 1, p. 191, 1992, doi: 10.2307/2347628.

[139] J. R. Quinlan, “Induction of Decision Trees,” *Mach Learn*, vol. 1, pp. 81–106, 1986.

[140] L. Breiman, “Random Forests,” vol. 45, pp. 5–32, 2001.

[141] D. Shen, G. Wu, and H.-I. Suk, “Deep Learning in Medical Image Analysis”, doi: 10.1146/annurev-bioeng-071516-044442.

[142] B. Sahiner *et al.*, “Deep learning in medical imaging and radiation therapy,” *Med Phys*, vol. 46, no. 1, pp. e1–e36, Jan. 2019, doi: 10.1002/MP.13264.

[143] S. Niyas, S. J. Pawan, A. Kumar, and J. Rajan, “Medical Image Segmentation with 3D Convolutional Neural Networks: A Survey”.

[144] K. Simonyan and A. Zisserman, “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION,” 2015, Accessed: Oct. 21, 2023. [Online]. Available: http://www.robots.ox.ac.uk/

[145] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning,” *Proceedings of the AAAI*

Conference on Artificial Intelligence, vol. 31, no. 1, pp. 4278–4284, Feb. 2017, doi: 10.1609/AAAI.V31I1.11231.

[146] C. Szegedy *et al.*, “Going deeper with convolutions”.

[147] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition”, Accessed: Oct. 21, 2023. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[148] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation”.

[149] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, Accessed: Oct. 22, 2023. [Online]. Available: http://lmb.informatik.uni-freiburg.de/

[150] R. Yousef *et al.*, “U-Net-Based Models towards Optimal MR Brain Image Segmentation,” 2023, doi: 10.3390/diagnostics13091624.

[151] S. Ali, A. Ismael, A. Mohammed, and H. Hefny, “An enhanced deep learning approach for brain cancer MRI images classification using residual networks,” 2019, doi: 10.1016/j.artmed.2019.101779.

[152] O. Oktay *et al.*, “Attention U-Net: Learning Where to Look for the Pancreas”.

[153] Z. Zhou, R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation”, doi: 10.1007/978-3-030-00889-5_1.

[154] F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, “nnU-Net for Brain Tumor Segmentation”, Accessed: Oct. 22, 2023. [Online]. Available: https://zenodo.org/record/3718904

[155] Y. Yu, X. Si, C. Hu, and J. Zhang, “A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures,” *Neural Comput*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019, doi: 10.1162/NECO_A_01199.

[156] J. Donahue *et al.*, “Long-term Recurrent Convolutional Networks for Visual Recognition and Description”, Accessed: Oct. 22, 2023. [Online]. Available: http://jeffdonahue.com/lrcn/.

[157] Z. Ye, K. Levick, A. Gilman, P. Cosman, Q. Peng, and L. Milstein, “Comparison of Neural Network Architectures for Spectrum Sensing,” 2019.

[158] L. Xu, Y. H. Yan, X. X. Yu, W. Q. Zhang, J. Chen, and L. Y. Duan, “LSTM neural network for solar radio spectrum classification,” *Res Astron Astrophys*, vol. 19, no. 9, p. 135, Sep. 2019, doi: 10.1088/1674-4527/19/9/135.

[159] Y. Sun, S. Brockhauser, and P. Hegedűs, “Comparing End-to-End Machine Learning Methods for Spectra Classification,” *Applied Sciences*, vol. 11, no. 23, p. 11520, Dec. 2021, doi: 10.3390/app112311520.

[160] L. E. Jalbert *et al.*, “Magnetic resonance analysis of malignant transformation in recurrent glioma,” *Neuro Oncol*, vol. 18, no. 8, pp. 1169–1179, Aug. 2016, doi: 10.1093/neuonc/now008.

[161] A. Verma, I. Kumar, N. Verma, P. Aggarwal, and R. Ojha, “Magnetic resonance spectroscopy — Revisiting the biochemical and molecular milieu of brain tumors,” *BBACLI*, vol. 5, pp. 170–178, 2016, doi: 10.1016/j.bbacli.2016.04.002.

[162] X. Li, Y. Xu, F. Xiang, Q. Liu, W. Huang, and B. Xie, “KINET: A NON-INVASIVE METHOD FOR PREDICTING KI67 INDEX OF GLIOMA,” in *Proceedings* -

International Conference on Image Processing, ICIP, IEEE Computer Society, 2021, pp. 150–154. doi: 10.1109/ICIP42928.2021.9506741.

[163] E. D. H. Gates *et al.*, "Guiding the first biopsy in glioma patients using estimated Ki-67 maps derived from MRI: Conventional versus advanced imaging," *Neuro Oncol*, vol. 21, no. 4, pp. 527–536, Mar. 2019, doi: 10.1093/neuonc/noz004.

[164] S. S. Gurbani, S. Sheriff, A. A. Maudsley, H. Shim, and L. A. D. Cooper, "Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting," *Magn Reson Med*, vol. 81, no. 5, pp. 3346–3357, May 2019, doi: 10.1002/mrm.27641.

[165] S. P. Kyathanahally, A. Döring, and R. Kreis, "Deep learning approaches for detection and removal of ghosting artifacts in MR spectroscopy," *Magn Reson Med*, vol. 80, no. 3, pp. 851–863, Sep. 2018, doi: 10.1002/MRM.27096.

[166] H. H. Lee and H. Kim, "Deep learning-based target metabolite isolation and big data-driven measurement uncertainty estimation in proton magnetic resonance spectroscopy of the brain," *Magn Reson Med*, vol. 84, no. 4, pp. 1689–1706, Oct. 2020, doi: 10.1002/mrm.28234.

[167] R. Rizzo, M. Dziadosz, S. P. Kyathanahally, M. Reyes, and R. Kreis, "Reliability of Quantification Estimates in MR Spectroscopy: CNNs vs Traditional Model Fitting," 2022, pp. 715–724. doi: 10.1007/978-3-031-16452-1_68.

[168] N. Schmid *et al.*, "Deconvolution of 1D NMR spectra: A deep learning-based approach," 2022, doi: 10.1016/j.jmr.2022.107357.

[169] N. Zarinabad, M. Wilson, S. K. Gill, K. A. Manias, N. P. Davies, and A. C. Peet, "Multiclass imbalance learning: Improving classification of pediatric brain tumors from

magnetic resonance spectroscopy," *Magn Reson Med*, vol. 77, no. 6, pp. 2114–2124, Jun. 2017, doi: 10.1002/MRM.26318.

[170] D. Zhao *et al.*, "Metabolite selection for machine learning in childhood brain tumour classification," *NMR Biomed*, vol. 35, no. 6, Jun. 2022, doi: 10.1002/NBM.4673.

[171] J. C. Crane, M. P. Olson, and S. J. Nelson, "SIVIC: Open-source, standards-based software for DICOM MR spectroscopy workflows," *International Journal of Biomedical Imaging*, vol. 2013. 2013. doi: 10.1155/2013/169526.

[172] A. Fedorov *et al.*, "3D Slicer as an image computing platform for the Quantitative Imaging Network.," *Magn Reson Imaging*, vol. 30, no. 9, pp. 1323–41, Nov. 2012, doi: 10.1016/j.mri.2012.05.001.

[173] P. Y. Wen and S. Kesari, "Medical Progress Malignant Gliomas in Adults," *N Engl J Med*, vol. 359, pp. 492–507, 2008, Accessed: Nov. 02, 2023. [Online]. Available: www.nejm.org

[174] A. Giese, R. Bjerkvig, M. E. Berens, and M. Westphal, "Cost of Migration: Invasion of Malignant Gliomas and Implications for Treatment," *Journal of Clinical Oncology*, vol. 21, no. 8, pp. 1624–1636, Apr. 2003, doi: 10.1200/JCO.2003.05.063.

[175] L. E. Gaspar *et al.*, "SUPRATENTORIAL MALIGNANT GLIOMA: PATTERNS OF RECURRENCE AND IMPLICATIONS FOR EXTERNAL BEAM LOCAL TREATMENT".

[176] M. C. Dobelbower *et al.*, "Patterns of failure for glioblastoma multiforme following concurrent radiation and temozolomide," 2011, doi: 10.1111/j.1754-9485.2010.02232.x.

[177] Y. R. Lawrence *et al.*, "Early toxicity predicts long-term survival in high-grade glioma," *Br J Cancer*, vol. 104, pp. 1365–1371, 2011, doi: 10.1038/bjc.2011.123.

[178] W. Stummer *et al.*, "Extent of resection and survival in glioblastoma multiforme: Identification of and adjustment for bias," *Neurosurgery*, vol. 62, no. 3, 2008, doi: 10.1227/01.neu.0000317304.31579.17.

[179] S. J. Nelson and S. Cha, "Imaging Glioblastoma Multiforme," *The Cancer Journal*, vol. 9, no. 2, pp. 134–145, Mar. 2003, doi: 10.1097/00130404-200303000-00009.

[180] D. B. Einstein *et al.*, "Phase II trial of radiosurgery to magnetic resonance spectroscopy-defined high-risk tumor volumes in patients with glioblastoma multiforme," *Int J Radiat Oncol Biol Phys*, vol. 84, no. 3, 2012, doi: 10.1016/j.ijrobp.2012.01.020.

[181] N. A. Parra *et al.*, "Volumetric spectroscopic imaging of glioblastoma multiforme radiation treatment volumes," *Int J Radiat Oncol Biol Phys*, vol. 90, no. 2, 2014, doi: 10.1016/j.ijrobp.2014.03.049.

[182] D. Heo *et al.*, "Deep learning based on dynamic susceptibility contrast MR imaging for prediction of local progression in adult-type diffuse glioma (grade 4)," *Sci Rep*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-41171-9.

[183] H. Mehrabian, K. L. Desmond, H. Soliman, A. Sahgal, and G. J. Stanisz, "Differentiation between Radiation Necrosis and Tumor Progression Using Chemical Exchange Saturation Transfer," *Clinical Cancer Research*, vol. 23, no. 14, pp. 3667–3675, Jul. 2017, doi: 10.1158/1078-0432.CCR-16-2265.

[184] A. Deviers *et al.*, "Evaluation of the lactate-to-N-acetyl-aspartate ratio defined with magnetic resonance spectroscopic imaging before radiation therapy as a new predictive marker of the site of relapse in patients with glioblastoma multiforme," *Int J Radiat Oncol Biol Phys*, vol. 90, no. 2, 2014, doi: 10.1016/j.ijrobp.2014.06.009.

[185] S. Angeli, K. E. Emblem, P. Due-Tonnessen, and T. Stylianopoulos, “Towards patient-specific modeling of brain tumor growth and formation of secondary nodes guided by DTI-MRI,” *Neuroimage Clin*, vol. 20, 2018, doi: 10.1016/j.nicl.2018.08.032.

[186] J. C. L. Alfonso *et al.*, “The biology and mathematical modelling of glioma invasion: a review,” 2017, doi: 10.1098/rsif.2017.0490.

[187] V. Anand Cuddapah, S. Robel, S. Watkins, and H. Sontheimer, “A neurocentric perspective on glioma invasion,” 2014, doi: 10.1038/nrn3765.

[188] J. Berberat, J. McNamara, L. Remonda, S. Bodis, and S. Rogers, “Diffusion tensor imaging for target volume definition in glioblastoma multiforme,” *Strahlentherapie und Onkologie*, vol. 190, no. 10, 2014, doi: 10.1007/s00066-014-0676-3.

[189] T. C. W. Mok and A. C. S. Chung, “Robust Image Registration with Absent Correspondences in Pre-operative and Follow-Up Brain MRI Scans of Diffuse Glioma Patients,” 2023. doi: 10.1007/978-3-031-33842-7_20.

[190] P. Y. Wen *et al.*, “Updated response assessment criteria for high-grade gliomas: Response assessment in neuro-oncology working group,” *Journal of Clinical Oncology*, vol. 28, no. 11. 2010. doi: 10.1200/JCO.2009.26.3541.

[191] J. M. Duarte-Carvajalino, G. Sapiro, N. Harel, and C. Lenglet, “A framework for linear and non-linear registration of diffusion-weighted MRIs using angular interpolation,” *Front Neurosci*, no. 7 APR, 2013, doi: 10.3389/fnins.2013.00041.

[192] M. Jenkinson and S. Smith, “A global optimisation method for robust affine registration of brain images,” *Med Image Anal*, vol. 5, no. 2, 2001, doi: 10.1016/S1361-8415(01)00036-6.

[193] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, “Improved optimization for the robust and accurate linear registration and motion correction of brain images,” *Neuroimage*, vol. 17, no. 2, 2002, doi: 10.1016/S1053-8119(02)91132-8.

[194] F. Isensee *et al.*, “Automated brain extraction of multisequence MRI using artificial neural networks,” *Hum Brain Mapp*, vol. 40, no. 17, 2019, doi: 10.1002/hbm.24750.

[195] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” Jun. 2016.

[196] T. Henry *et al.*, “Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: a BraTS 2020 challenge solution”, Accessed: Oct. 31, 2023. [Online]. Available: https://github.com/lescientifik/open

[197] Y. Wu and K. He, “Group Normalization”, Accessed: Oct. 31, 2023. [Online]. Available: https://github.com/facebookresearch/Detectron/

[198] L. Wright and N. Demeure, “RANGER21: A SYNERGISTIC DEEP LEARNING OPTIMIZER A PREPRINT,” 2021.

[199] L. Liu *et al.*, “ON THE VARIANCE OF THE ADAPTIVE LEARNING RATE AND BEYOND”, Accessed: Oct. 31, 2023. [Online]. Available: https://github.com/LiyuanLucasLiu/RAdam.

[200] M. R. Zhang, J. Lucas, G. Hinton, and J. Ba, “Lookahead Optimizer: k steps forward, 1 step back”, Accessed: Oct. 31, 2023. [Online]. Available: https://github.com/michaelrzhang/lookahead.

[201] E. Marton *et al.*, “Over ten years overall survival in glioblastoma: A different disease?,” 2019, doi: 10.1016/j.jns.2019.116518.

[202] M. Weller *et al.*, “Molecular Predictors of Progression-Free and Overall Survival in Patients With Newly Diagnosed Glioblastoma: A Prospective Translational Study of the German Glioma Network Written on behalf of the German,” *J Clin Oncol*, vol. 27, pp. 5743–5750, 2009, doi: 10.1200/JCO.2009.23.0805.

[203] N. F. Brown *et al.*, “Survival Outcomes and Prognostic Factors in Glioblastoma,” *Cancers (Basel)*, vol. 14, no. 13, p. 3161, Jul. 2022, doi: 10.3390/CANCERS14133161/S1.

[204] A. Armina Abedi *et al.*, “A Prognostic Model for Glioblastoma Patients Treated With Standard Therapy Based on a Prospective Cohort of Consecutive Non-Selected Patients From a Single Institution,” *Article*, vol. 11, p. 1, 2021, doi: 10.3389/fonc.2021.597587.

[205] E. Lee, R. L. Yong, P. Paddison, and J. Zhu, “Comparison of glioblastoma (GBM) molecular classification methods,” *Semin Cancer Biol*, vol. 53, pp. 201–211, Dec. 2018, doi: 10.1016/J.SEMCANCER.2018.07.006.

[206] E. Essock-Burns *et al.*, “Assessment of perfusion MRI-derived patients with newly diagnosed glioblastoma,” *Neuro Oncol*, vol. 13, no. 1, pp. 119–131, 2011.

[207] J. M. Lupo *et al.*, “Using susceptibility-weighted imaging to determine response to combined antiangiogenic, cytotoxic, and radiation therapy in patients with glioblastoma multiforme,” *Neuro Oncol*, vol. 15, no. 4, pp. 480–489, 2013, doi: 10.1093/neuonc/nos325.

[208] C. Larsson *et al.*, “Prediction of survival and progression in glioblastoma patients using temporal perfusion changes during radiochemotherapy,” 2020, doi: 10.1016/j.mri.2020.01.012.

[209] J. Lupo, S. Cha, S. M. Chang, and S. J. Nelson, “Characterization of Spatial Heterogeneity Imaging of High-Grade Gliomas: Dynamic Susceptibility-Weighted Perfusion Dynamic

Susceptibility-Weighted Perfusion Imaging of High-Grade Gliomas: Characterization of Spatial Heterogeneity," *AJNR Am J Neuroradiol*, vol. 26, no. 6, pp. 1446–1454, 1446, Accessed: Nov. 07, 2023. [Online]. Available: http://www.ajnr.org/content/26/6/

[210] J. M. Lupo, Q. Wen, J. J. Phillips, S. M. Chang, and S. J. Nelson, "Weighted-average model curve preprocessing strategy for quantification of DSC perfusion imaging metrics from image-guided tissue samples in patients with brain tumors".

[211] P. Y. Wen *et al.*, "RANO 2.0: Update to the Response Assessment in Neuro-Oncology Criteria for High- and Low-Grade Gliomas in Adults," *Journal of Clinical Oncology*, Sep. 2023, doi: 10.1200/JCO.23.01059.

[212] Q. Wen *et al.*, "Association of Diffusion and Anatomic Imaging Parameters with Survival for Patients with Newly Diagnosed Glioblastoma Participating in Two Different Clinical Trials," *Transl Oncol*, vol. 8, no. 6, pp. 446–455, Dec. 2015, doi: 10.1016/j.tranon.2015.10.001.

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:

D1881704F7314C3

Author Signature

12/1/2023

Date