Summer 8-2020

Model-Based Approach for Diffuse Glioma Classification, Grading, and Patient Survival Prediction

Zeina A. Shboul
Old Dominion University, zshbo001@odu.edu

MODEL-BASED APPROACH FOR DIFFUSE GLIOMA CLASSIFICATION, GRADING, AND PATIENT SURVIVAL PREDICTION

by

Zeina A. Shboul
B.S. December 2007, Yarmouk University, Jordan
M.E. May 2011, Jordan University of Science and Technology, Jordan

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL AND COMPUTER ENGINEERING

OLD DOMINION UNIVERSITY
August 2020

Approved by:

Khan M. Iftekharuddin (Director)

Chunsheng Xin (Member)

Jiang Li (Member)

Norou Diawara (Member)

ABSTRACT

MODEL-BASED APPROACH FOR DIFFUSE GLIOMA CLASSIFICATION, GRADING,

AND PATIENT SURVIVAL PREDICTION

Zeina A. Shboul
Old Dominion University, 2020
Director: Dr. Khan M. Iftekharuddin

The work in this dissertation proposes model-based approaches for molecular mutations classification of gliomas, grading based on radiomics features and genomics, and prediction of diffuse gliomas clinical outcome in overall patient survival. Diffuse gliomas are types of Central Nervous System (CNS) brain tumors that account for 25.5% of primary brain and CNS tumors and originate from the supportive glial cells. In the 2016 World Health Organization's (WHO) criteria for CNS brain tumor, a major reclassification of the diffuse gliomas is presented based on gliomas molecular mutations and the growth behavior. Currently, the status of molecular mutations is determined by obtaining viable regions of tumor tissue samples. However, an increasing need to non-invasively analyze the clinical outcome of tumors requires careful modeling and co-analysis of radiomics (i.e., imaging features) and genomics (molecular and proteomics features). The variances in diffuse Lower-grade gliomas (LGG), which are demonstrated by their heterogeneity, can be exemplified by radiographic imaging features (i.e., radiomics). Therefore, radiomics may be suggested as a crucial non-invasive marker in the tumor diagnosis and prognosis. Consequently, we examine radiomics extracted from the multi-resolution fractal representations of the tumor in classifying the molecular mutations of diffuse LGG non-invasively. The proposed radiomics in the decision-tree-based ensemble machine learning molecular prediction model confirm the efficacy of these fractal features in glioma prediction. Furthermore, this dissertation proposes a novel non-invasive statistical model to classify and predict LGG molecular mutations based on radiomics and count-based genomics data. The performance results of the proposed statistical model indicate that fusing radiomics to count-based genomics improves the performance of mutations prediction. Furthermore, the radiomics-based glioblastoma survival prediction framework is proposed in this work. The survival prediction framework includes two survival prediction pipelines that combine different feature selection and regression approaches. The framework is evaluated using two recent widely used benchmark datasets from Brain Tumor Segmentation (BraTS) challenges in 2017 and

2018. The first survival prediction pipeline offered the best overall performance in the 2017 Challenge, and the second survival prediction pipeline offered the best performance using the validation dataset. In summary, in this work, we develop non-invasive computational and statistical models based on radiomics and genomics to investigate overall survival, tumor progression, and the molecular classification in diffuse gliomas. The methods discussed in our study are important steps towards a non-invasive approach to diffuse brain tumor classification, grading, and patient survival prediction that may be recommended prior to invasive tissue sampling in a clinical setting.

ACKNOWLEDGMENTS

First and Foremost, I am immensely grateful to Almighty Allah (God) for granting me with the health, patience, and knowledge to complete my Ph.D. program. The work of this dissertation has been possible because of the support of many people. I would like to express my gratitude to my advisor Dr. Khan M. Iftekharuddin for his guidance, support, encouragement, and for giving me the opportunity to work in his research group. My special thanks to the members of my dissertation committee, Dr. Chunsheng Xin, Dr. Jiang Li, and Dr. Norou Diawara for their time, helpful suggestions, and willingness to advise and review the dissertation. My special thanks to all the colleagues of the Old Dominion University Vision Lab for supporting me throughout my Ph.D. program.

Members of the family have contributed in a very special and unique way. I am grateful for my beloved siblings, Omar, Fardous, Mohammad, and Nour Shboul for their endless love. I am deeply grateful to my parents Dr. Mahmoud Shboul and Shanaz Shboul for their constant inspiration, support, and encouragement. Last but not least, I would like to thank my husband Dr. Luay Wahsheh for his tremendous support and patience during my Ph.D. program.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Diffuse, or infiltrative, gliomas are types of Central Nervous System (CNS) brain tumors that account for 25.5% of primary brain and CNS tumors and 80% of malignant tumors [1]. Gliomas are considered the most malignant primary brain and CNS tumors. Gliomas originate from the supportive glial cells that consist of three major types: oligodendrocytes, astrocytes, and ependymal cells [2]. Glial cells and neurons are the two major types of cells in the neural tissues. However, only glial cells can undergo cell division, and if the process of cell division is not carefully controlled and happens too fast, glioma forms and grows [3]. The broad range of glial cells causes different forms of gliomas: astrocytomas, oligodendrogliomas, and mixed gliomas such as oligoastrocytomas [4]. The 2007 World Health Organization (WHO) classification of gliomas was based on microscopic similarities with different cells of origin and their levels of malignancy. For example, Gliomas' main types were classified based on their malignancy into grades II, III, and IV [5, 6]. The WHO grade II gliomas (low-grade gliomas) are generally infiltrating and have low mitotic activity (cell division and proliferation) and tend to progress to higher grades. The WHO grade III gliomas (high-grade gliomas) have a rapid mitotic activity and infiltrative capacity. The WHO grade IV gliomas (high-grade gliomas) have high mitotic activity, appear with necrosis, and usually are accompanied by a rapid progression and fatal outcome [5, 7].

Recently, studies have revealed the importance of the genetic elements in tumor formation and growth; such understanding may lead to greater diagnostic accuracy, as well as improved patient management and more accurate determination of prognosis and treatment response [8-11]. Major revision of diffuse gliomas of the 2007 CNS WHO classification have been introduced in the 2016 CNS WHO classification [12], which is established based on a combined phenotypic and genotypic classification, and the generation of "integrated" diagnoses. An updated classification of the diffuse lower-grade gliomas (LGG) is presented in the 2016 WHO Classification of Tumors of the CNS [12]. The new classification of the diffuse LGG (WHO Grade II and III gliomas) depends on the genetic driver mutations (*IDH* mutations, 1p/19q codeletion, and *ATRX* mutations), and their histological type (astrocytoma and oligodendroglioma) as illustrated in Figure 1. Note

that oligoastrocytomas are considered as not otherwise specified (NOS) categories unless there is no diagnostic molecular testing. This new classification correlates with patients' treatment and survival. The histological types have different clinical behavior; oligodendroglioma tumors are associated with longer survival when compared to astrocytoma [13].

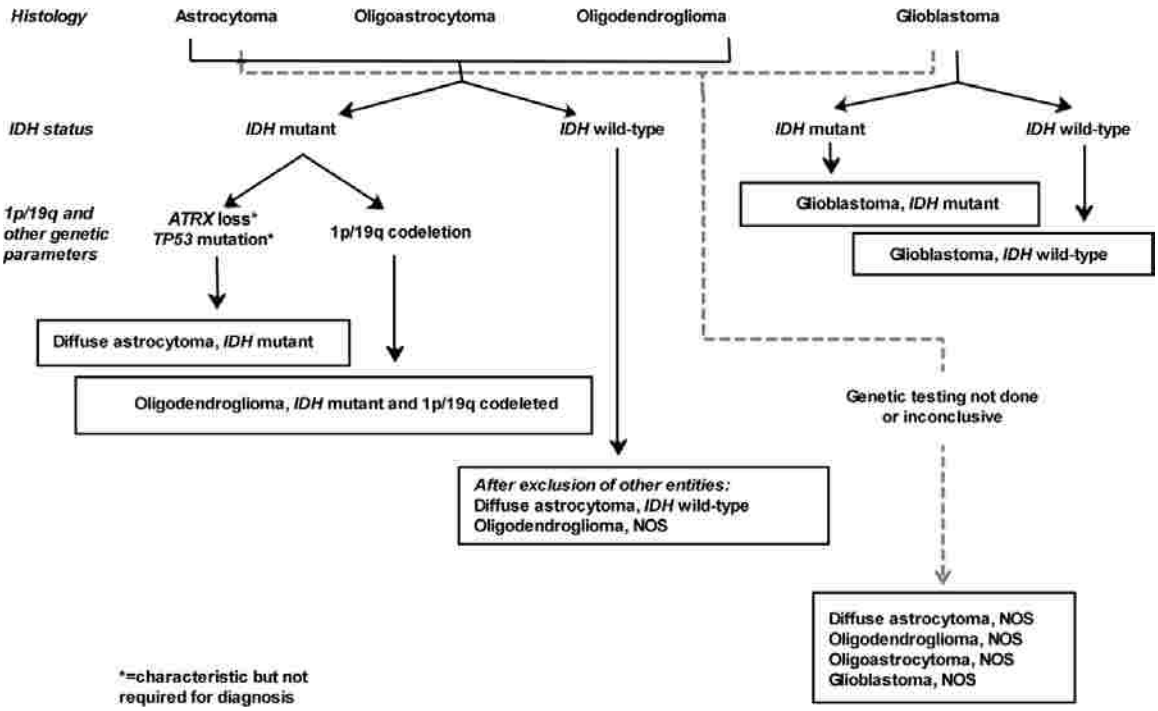


Figure 1: The 2016 WHO Classification of the diffuse gliomas based on their histology and genetic mutations [12].

Isocitrate Dehydrogenase mutations, *IDH1*, and *IDH2* have been found in gliomas [14-17]. Classifying gliomas based on their molecular profiling of IDH status (mutated vs. wildtype) creates clinically distinct groups. IDH wild-type gliomas behave aggressively when compared with the *IDH* mutant gliomas. As a result, patients with IDH mutant gliomas tend to have a better prognosis [18]. A 1p/19q codeletion is considered a molecular marker of oligodendroglioma and is associated with IDH mutation [19]. This genetic alteration happens when the short arm of chromosome 1 (1p), and the long arm of chromosome 19 (19q) are deleted. The existence of 1p/19q codeletion in diffuse LGG is associated with improved survival and with *IDH* mutations [16].

Using IDH mutation and 1p/19q codeletion, the diffuse LGG is classified into three groups [20]: 1p/19q codeletion with IDH mutant, 1p/19q non-codeletion with IDH mutant, and 1p/19q non-codeletion with IDH wild-type. *ATRX* is a somatic mutation in the Alpha-Thalassemia/mental Retardation syndrome, X-linked. This somatic mutation is reported in GBM [21]. This type of mutation frequently occurs in diffuse Astrocytoma and is associated with a significantly better prognosis [22, 23]. In addition, ATRX mutation has often occurred with IDH mutations and almost mutually exclusive with 1p/19q codeletion. Another molecular alteration that has a high prevalence of LGG is O6-methylguanine-DNA methyltransferase (MGMT) gene promoter methylation [24]. Patients with a methylated MGMT promoter are associated with better overall survival [25, 26]. It has a better impact on overall survival if MGMT methylation is combined with IDH mutation and 1p/19q codeletion [25]. Diffuse LGG is known for its heterogeneous characteristic that reveals variances in tumor biology. This heterogeneity can be seen through the histological types: astrocytoma, oligoastrocytoma, and oligodendroglioma [5, 12]. The heterogeneity can be characterized by magnetic resonance imaging (MRI) features [27-29], which suggests using MRI features as a non-invasive marker in tumor grading and classification [30-32].

Another diffuse infiltrative glioma is Glioblastoma multiforme (GBM). GBM accounts for 14.6% of primary brain and CNS tumors, 48.3% of primary malignant brain tumors, and 57.3% of gliomas [1]. GBM is staged as a WHO grade IV tumor [5] and arises from astrocytes glial cells. The extensive infiltrative growth pattern of Glioblastoma makes the curative treatment impossible and, hence, reduces the median survival rate to 15 - 16 months [33]. The naming “multiforme” in GBM emphasizes the heterogeneous nature of this type of glioma. The updated 2016 WHO classification of GBM is divided into GBM with IDH-wildtype, GBM with IDH-mutant, and NOS GBM, which is reserved for those tumors for which full IDH evaluation cannot be performed [34]. GBM IDH-wildtype accounts for ~90% of GBM, most frequently corresponds to primary (de novo) glioblastoma, and predominates in patients with a mean age of 61 years old. The median overall survival of GBM IDH-wildtype patients who are treated with surgery and radiotherapy is 9.9 months and 15 months when treated with surgery and radio/chemotherapy. GBM IDH-mutant accounts for 10% and corresponds to secondary GBM (i.e. with a history of prior lower grade diffuse gliomas) in patients with a mean age of 48 years. The median overall survival of GBM IDH-mutant patients is 24 months when treated with surgery and radiotherapy and 31 months when treated with surgery and radio/chemotherapy. Glioblastomas heterogeneity may show significant

intertumoral (tumor by tumor) and intra-tumoral (within a tumor) heterogeneity. The tumor heterogeneity impacts the clinical outcome and overall survival [35, 36]. Thus, analyzing this heterogeneity is important for associating tumor phenotype with prognostic and predictive purposes [37].

1.1 DISSERTATION GOALS AND PROBLEM STATEMENT

In current radiology practice, radiomics have a critical role in disease diagnosis, monitoring, treatment planning, and personalized medicine [38, 39]. Radiomics acquisition is extracted from various radiological imaging modalities of the disease's region of interest, which translates the radiographical images into useful data. Fusing crucial radiomics with clinical data in the proper machine/deep learning model may help to obtain a more comprehensive disease diagnosis and reliable early assessment of prognosis and treatment planning, and more personalized or precise medicine [40, 41]. Improving personalized treatment and precision medicine requires an accurate survival prediction, a fast and precise tumor detection, improved molecular grading and classification, and a detailed therapeutic response assessment of oncologic patients. Improvements in personalized medicine have succeeded as a consequence of advances in biologic and genomic technologies that rely on invasive methods of tissue sampling [42-44]. However, tissue sampling may also be associated with high cost, morbidity, and even mortality [45]. Also, in heterogeneous tumors such as glioblastomas, tissue sampling may not reflect the entire heterogeneity of the tumor. Consequently, developing alternative methods and non-invasively analyzing patients' survival-time prediction and classifying diffuse gliomas into subtypes using imaging features and machine learning techniques have emerged as vital and promising areas of research.

This study introduces a non-invasive automatic gliomas molecular classification and grading prediction based on radiomics extracted from conventional magnetic resonance imaging and clinical data. Also, we develop a glioma grading model based on genomics count-based data and radiomics utilizing statistical modeling. Additionally, we introduce a radiomics guided non-invasive gliomas clinical outcome of overall survival in patients with glioblastoma.

The first goal of this dissertation addresses a non-invasive analysis of low-grade glioma grading and classification using radiomics features based on the 2016 WHO Classification of

Tumors of the CNS [12]. The new classification of the diffuse LGG depends on the genetic driver mutations (MGMT methylation, *IDH* mutation, 1p/19q codeletion, *ATRX* mutation, *TERT* mutation, and *TP53* mutation). Because diffuse LGG is known for its heterogeneous characteristic, MRI features can be used to characterize this heterogeneity. The second goal of this dissertation is to propose a statistical radiogenomics model that utilizes both RNA sequencing (RNAseq) read counts data and radiomics for glioma molecular grading and prediction. The Negative Binomial (NB) distribution is fitted to RNAseq read counts to preserve the count-based nature of these data, and a log-linear regression modeling is fitted to the estimated mean of the NB distribution and linked with radiomics. Thus, a complete characterization radiogenomics prediction model is presented. The third goal of this dissertation is to obtain a radiomics-guided machine learning model for survival regression and prediction. In this work, we propose two representative survival prediction pipelines that combine different feature selection and regression approaches. The framework is evaluated using two widely used benchmark datasets from Brain Tumor Segmentation (BraTS) global challenges in 2017 and 2018.

In general, alternative non-invasive radiomics-based machine learning models are critical in the body of research to analyze and predict glioma clinical outcomes prior to invasive tissue sampling. Although invasive tissue sampling is effective, non-invasive radiomics-based models may be preferred if these models offer competitive performance. The efficacy of the radiomics-based models proposed in this work indicates the potential of correlating radiomics (computed imaging features) with glioma molecular mutations types and overall survival. These models also identify candidates that may be considered potential predictive biomarkers of glioma molecular classification and overall survival. Furthermore, the recent interest in associating genomic phenotypes and radiomics has shifted the research towards utilizing both types of data in a single computational modeling framework to analyze different aspects of glioma clinical outcomes.

1.2 CONTRIBUTIONS

The main contributions of this dissertation are discussed as follows:

1. *Effective non-invasive analysis of diffuse low-grade gliomas grading and classification using radiomics features based on the 2016 WHO Classification of Tumors of the CNS*

The 2016 WHO classification of diffuse LGGs heavily weighs molecular mutations classifying primary brain tumors with particular importance assigned to *IDH* mutation, 1p/19q codeletion, *ATRX* mutation, *TERT* mutations, and MGMT methylation. The second goal proposes and develops an imaging feature-based, computational machine learning model for diffuse LGGs classification and grading. Our model on diffuse LGG is largely able to predict the presence of the important molecular mutations as identified by WHO based on MR imaging features. Therefore, prediction of tumor aggressiveness (based on molecular mutations) may be achieved through non-invasive imaging features as an adjunct to traditional visual morphologic diagnosis and invasive tissue sampling. The proposed model addresses possible overfitting, model instability, and the efficacy of fractal and multifractal textures on the performance of the proposed molecular mutation prediction models. To the best of our knowledge, this is the first study that addresses the potency of fractal and multi-resolution fractal features in molecular mutations prediction. The experimental demonstrations of the proposed molecular mutation prediction show promise when compared to different methods and models in the literature. The proposed methods are important steps towards non-invasive imaging classification of diffuse LGG based on molecular mutations prior to invasive tissue sampling.

2. *A novel RNA Sequencing and Radiomics statistical fusion model (radiogenomics-Negative Binomial) for Glioma Grading and Prediction*

In current radiology practice, radiomics plays a key role in disease diagnosis, monitoring, treatment planning, and grading [38, 39]. Radiomics is extracted from various radiological images of a targeted area of the disease. Fusing the important radiomics and genomics information in the computational machine learning (ML) model may help in achieving a comprehensive disease diagnosis, prognosis, and treatment plan. Consequently, such fusion may yield a personalized or precision medicine model [40, 41]. The rise of radiogenomics introduces more comprehensive and better profiling of personalized medicine for patients [27, 46-48]. Thus, associating between the radiomics phenotypes and underlying disease molecular mutations enhances clinical oncology care [49-51]. Different studies evaluate the association between glioma molecular subtypes and radiomics (e.g., tumor shape and size) [52-54] or between different forms of genomics (e.g., RNA sequencing (RNAseq) gene expression, protein expression, copy number, molecular mutations, or DNA methylation) and glioma subtypes [55-57]. These associations help to pave the way for ML-

based classification of glioma molecular subtypes using radiomics [58-60], genomics [57, 61], or both [62-64].

Conventional ML models do not adequately model the count-based nature of the RNA-sequence data as these models are usually designed to work with data that has a normal distribution. In order to alleviate the lack of appropriate ML models, researchers propose to transform the RNAseq read-count data to approximate a normal distribution. The transformation to normal distribution allows the use of existing methods such as the nearest shrinkage method [65, 66] and Random Forest for classification. However, such transformation still removes the count-based nature of the RNAseq read counts data and, hence, lacks the ability to fully preserve the strong mean-variance relationship that is otherwise useful for glioma classification and prediction [67, 68]. In order to appropriately model the RNAseq read-count data, Negative Binomial (NB) and Poisson distributions are commonly used [69]. Poisson distribution is a single parameter distribution with its mean equals to its variance, which makes Poisson distribution rather restrictive. On the other hand, NB is similar to a Poisson distribution with an additional parameter called “dispersion” that allows the NB distribution to modify its variance without affecting the mean. On the other hand, existing works suggest the efficacy of using both genomics and radiomics data in the classification and prediction of different disease modality [70-73].

Consequently, this work proposes fusing RNAseq read counts data with radiomics in NB distribution to classify and predict glioma molecular mutations. Moreover, log-linear regression modeling is fitted to the estimated mean of the NB distribution and is linked with radiomics. We introduce this step to fuse the continuous radiomics data with the RNAseq count-based data without the need to transform RNAseq data into a normal distribution.

Finally, we investigate the effect of using different numbers of differentially expressed genes from the count-based data as well as the gender factor by developing a gender-specific radiogenomics-NB model. Moreover, we compare our radiogenomics-NB model performance with that of different genomics and radiogenomics state-of-the-art methods in the literature.

3. *An efficient non-invasive machine learning for radiomics-based Glioblastoma survival regression and classification*

Different studies [74-76] have discussed different methods of predicting the survivability of patients with brain tumors. Pope et al. [74] use different subtype tumor volumes, the extent of resection, location, size, and other imaging features in order to assess the potential of these features

in predicting survival probabilities using the Kaplan-Meier method. Gutman et al. [75] use a comprehensive visual features set known as Visually AcceSAble Rembrandt Images (VASARI) in their study with glioblastoma patients. The authors have assessed the relationship between VASARI features and patients' survival using multivariate Cox regression models and correlated these features with genetic alterations and molecular subtypes by using the Fisher exact test. Aerts et al. [76] predict survival using a multivariate Cox proportional hazards regression model by quantifying a large number of radiomics image features including shape and texture in computed tomography images of lung and head-and-neck cancer patients. Several of the survival prediction studies have been utilizing regression survival [77, 78] models such as the proportional hazard method. Lately, machine learning is replacing these proportional hazard methods to predict survival [79-81]. In this work, we utilize Random Forest and XGBoost in the survival regression and classification model. A feature-guided non-invasive machine learning approach is expected to benefit from known imaging features that are already proven effective to guide the efficacy of the proposed pipelines. Consequently, this work proposes a fully automated overall survival regression classification using imaging features extracted from the tumor volume of the raw structural MRI and three different texture characterizations of the tumor region. Finally, the framework is evaluated using two recent widely used benchmark datasets and global challenges in Brain Tumor Segmentation (BraTS) global challenges in 2017 and 2018, respectively. The overall contributions of this dissertation are summarized in Table 1.

Table 1. Summary of the research findings related to the proposed methods.

Chapter	Dissertation goals	Proposed Methods	Novel Contributions
3	Prediction and evaluation of tumor aggressiveness using molecular mutations through non-invasive fractal and multi-resolution fractal imaging features as adjunct marker.	A XGBoost classification and prediction model using fractal and non-fractal features as input and molecular mutation information as the target output.	Non-invasive molecular prediction model and efficacy validation of fractal features for glioma molecular prediction using traditional MRI data based on the latest 2016 WHO classification of LGG of the CNS.
4	Glioma grading and prediction using joint statistical modeling of RNA sequencing and radiomics.	A radiogenomics-Negative Binomial model that fuses radiomics (volumetric imaging features) with RNAseq (genes) for glioma grading and prediction. Negative Binomial distribution is proposed to fit RNAseq counts data.	A novel RNA Sequencing and Radiomics statistical fusion model (radiogenomics-Negative Binomial) for Glioma grading and prediction, where a log-linear regression model links between the estimated NB mean and radiomics.
5	Glioma survival regression and classification using novel radiomics-based machine learning (ML) methods	Two ML models for survival prediction are developed: a. A three-step feature selection along with RF regression classifier. b. A recursive feature selection along with XGBoost for classification step followed by a Cox regression along with three risk-based regression models.	A novel framework for fully automated radiomics-based Glioblastoma survival prediction that provides competitive results when compared with the state-of-the-art methods in BraTS 2017 and BraTS 2018 challenges and ranked first in BraTS 2017 Challenge.

1.3 ORGANIZATION OF THE DISSERTATION

The rest of the dissertation is organized as follows. Chapter 2 provides a background review on Gliomas' survival and molecular mutation analysis and prediction using different sources of

features. Additionally, this chapter provides a background review on utilizing RNAseq read counts data in molecular mutations in diffuse LGG. Chapter 3 studies different molecular prediction models utilizing fractal and multi-resolution fractal texture features and other MR imaging features. The molecular models include the *IDH*, 1p/19q codeletion, MGMT, *ATRX*, and *TERT* prediction. In Chapter 4 we propose an RNA-radiomics (henceforth referred to as radiogenomics) -NB model for glioma grading and prediction. Our radiogenomics-NB model is developed based on differentially expressed RNAseq and selected radiomics features. Chapter 5 proposes a fully automated glioblastoma survival prediction framework based on MR imaging features. The survival prediction step includes two representative survival prediction pipelines that combine different feature selection and regression approaches. Chapter 6 provides the dissertation concluding remarks and future research.

CHAPTER 2

BACKGROUND REVIEW

This chapter discusses the relevant previous work on glioma survival analysis and prediction, molecular mutations prediction in diffuse LGG based on MRI features, and molecular mutations classification based on RNAseq read counts data and MRI features.

2.1 GLIOBLASTOMA SURVIVAL PREDICTION

The World Health Organization (WHO) identifies Glioblastoma as a highly diffusive aggressive grade IV glioma, which is known for the presence of anaplastic glial cells, high mitotic activity, and dense cellularity, as well as an increase in microvascular proliferation [5, 82, 83]. The highly infiltrative growth pattern of Glioblastoma makes curative care difficult with a median survival time of less than 2-years [84]. Recently, research interests have been focused on replacing invasive methods of patients' diagnosis and prognosis outcome with non-invasive methods [31, 32, 49]. Glioblastoma heterogeneity and its implication on patients' clinical outcomes [29, 36, 85] can be explored through radiology images such as MRI [31, 86, 87]. Thus, accurate detection and segmentation of different abnormal tumor tissues are essential in planning treatment therapy, diagnosis, grading, and survival prediction.

Few works [74-76] have proposed different methods for assessing and predicting the survivability of patients with brain tumors using clinical data and imaging features. Pope et al. [74] evaluate and analyze the capability of different subtype tumor volumes, the extent of resection, location, size, and other imaging features in predicting the survivability of patients with grade III and grade IV GBM. The study reveals that patients' age at diagnosis and Karnofsky Performance Status (KPS) shows a significant prognostic for both GBM grades. Additionally, a univariate Cox analysis shows that the tumor imaging features of non-contrast enhancing tumor, edema, satellites, and multifocality are statistically significant prognostic indicators in patients with GBM. Gutman et al. [75] conduct a comprehensive analysis of the visual features set known as Visually AcceSAble Rembrandt Images (VASARI) to study their association with GBM genetic mutations, molecular subtypes, and patient survival rate. The authors utilized four cardinal VASARI MR

imaging features along with a single measure of lesion size. Their analysis reveals that the volume of the contrast-enhancing tumor and the longest axis length of the GBM are statically associated with poor survival. The analysis also reveals that VASARI imaging features are associated with genetic alterations and molecular subtypes. However, the authors only address the visual features set and focus on the semi-quantitative measurements of the different tumor compartments and have not addressed other texture features. Aerts et al. [76] study the prognostic value of a large number of radiomics image features in computed tomography images of lung and head-and-neck cancer patients. These features include tumor image intensity, shape, texture, and multiscale wavelet. The authors perform a Kaplan-Meier survival analysis to study such an association. Additionally, the authors select the most important radiomics feature from each feature set, then utilize these four features in a multivariate Cox proportional hazards regression model for prediction of survival. Nicolasjilwan et al. [88] utilize clinical data, VASARI features, and genomics (represented by copy number variations) of GBM patients and combine these features into a stepwise multivariate Cox model for prediction of overall survival time. Prasanna et al. [89] extract radiomics texture features that characterize three tumor regions: enhanced tumor, peritumoral brain zone, and necrosis from MR images. These features are assessed for overall survival prediction. Itakura et al. [49] utilize quantitative MRI features that describe tumor histogram statistics, texture, edge sharpness, compactness, and roughness. Then the authors cluster the features into three MRI phenotypic imaging subtypes: pre-multifocal, spherical, and rim-enhancing tumor. The three distinct subtypes are then correlated with overall survival and associated with molecular pathways. The prevailing survival analysis and prediction studies have employed statistical survival regression techniques such as Kaplan-Meier and proportional hazard methods [76-78, 88]. Statistical survival analysis methods such as Kaplan-Meier and Cox regression analysis have difficulty capturing the complex relationship between the covariates and the outcome and suffer from cumulative incidence overestimation [90]. Recently, research has shifted towards replacing these statistical methods with improved machine learning methods because of the latter's computational power, accuracy, and ability to employ a large number of different types of features.

In this study, we hypothesize that radiomics imaging features along with the proposed machine learning techniques may offer superior efficacy in selecting important radiomics imaging features and predicting overall survival. In this study, we propose two fully automated survival prediction frameworks for patients with glioblastoma: overall survival regression using Random

Forest model [91] and overall survival classification and regression using Extreme Gradient Boosting (XGBoost) [92]. Radiomics imaging features are obtained from 1) raw structural MRI data, 2) Texton [93] characterization of the raw MRI, 3) and fractal and multi-resolution fractal representations of the raw MRI. Fractal and multi-resolution fractal have shown their efficacy in brain tumor segmentation in prior studies [94-98]. The proposed framework is evaluated using two recent widely used benchmark datasets from Brain Tumor Segmentation (BraTS) global challenges in 2017 and 2018, respectively. Our results suggest that the proposed framework achieves better overall survival prediction performance compared to the state-of-the-art methods.

2.2 MOLECULAR MUTATIONS IN DIFFUSE LOW-GRADE GLIOMAS

The 2016 WHO classification of diffuse LGGs is based on molecular profiling to *IDH* mutation, 1p/19q codeletion, *ATRX* mutation, *TERT* mutations, and MGMT methylation. Such profiling requires invasive tumor tissue sampling However, non-invasive analysis of tumor aggressiveness using Imaging features is emerging to replace the invasive tissue sampling. An MGMT methylation prediction study by Kanas et al. [59] for patients with GBM reports the size of the tumor with respect to necrosis as one of the significant features. A GBM study conducted by Kanas et al.[59] proposes an MGMT prediction model using volumetric, morphological, and locational MR imaging features, respectively. The authors in [59] report the size of the tumor with respect to necrosis as one of the significant features. However, the entire process of the prediction model is not fully automated. Another study by Han et al. [99] where the authors use a bi-directional convolutional recurrent neural network to predict MGMT methylation status. One major drawback of their study is that their model mainly utilizes imaging features that cannot be correlated with tumor biology. In an *IDH* prediction model by Yu et al. [100] on 110 patients with Grade II glioma, the authors use 110 imaging features and SVM to classify *IDH* status with Grade II glioma patients. However, the performance would reflect higher reliability of the performance of the *IDH* status prediction model if the dataset they use was more diverse with patients from Grade II and Grade III gliomas. A different study by Ding et al. [101] on 76 LGG patients utilizes MR imaging features along with MR spectroscopic data to predict *IDH* mutations using a binary logistic regression model. The authors achieved the best performance when utilizing MR spectroscopic data. A study by Akkus et al.[102] with LGG patients proposes 1p/19q codeletion

prediction using a convolutional neural network (CNN). In their study, the authors represent each patient's MRI with only 3 MRI slices. The authors in Akkus et al. [102] do not consider the global information of the tumor since their dataset uses only 3 slices of the MRI sequence of each patient as input, not the whole volume of the tumor. A different recent study by van der Voort et al. [103] utilizes MR imaging features along with patients' age and sex using an SVM classifier to predict 1p/19q codeletion in LGG patients. The authors use 284 LGG patients for training and another 129 LGG patients for testing. Their analysis reveals that the cranial/caudal location of the tumor is one of the most important features in predicting 1p/19q codeletion. Recently, Wang et al. [104] explore survival prediction and *TERT* mutations in 39 LGG (30 WT, and 9 Mutant). The authors propose a *TERT* prediction model using 24 imaging features selected using Principle Component Analysis (PCA) and classified using the Partial Least Squares (PLS). One major concern about their study is that the number of selected features (24 imaging features) is relatively high when compared to the dataset size (39 patients), which would increase the possibility of overfitting problems. In this work, we propose a fully automated computation prediction method of LGG molecular mutations utilizing radiomics imaging features. In our study, we utilize a more diverse dataset using Grade II and III, where significant and imaging features 1) are extracted from conventional MR imaging volumes that are clinically available and 2) are associated with overall survival and tumor biology. Additionally, we utilize fractal and multi-resolution fractal modeling. In this study, we hypothesize that the fractal and multi-resolution fractal modeling may be related to the underlying structure of molecular mutations. To the best of our knowledge, this is the first study that addresses the potency of fractal and multi-resolution fractal features in molecular mutations prediction. Several studies have shown the efficacy of fractal and multi-resolution fractal feature analysis for characterization, segmentation, and classification of the complex abnormal brain tissues in MRI [94, 97, 105, 106]. The spatial intensity distribution which is a basic property that characterizes medical images (such as MRI) have a degree of noise and randomness that allows the use of fractal and multi-resolution fractal texture modeling efficiently. The multi-resolution fractal modeling captures the randomly varying complex structure of the tumor texture on a different scale. Few studies have shown an association between different types of imaging features such as the grey-level co-occurrence matrix (GLCM) for texture, volume and area related features, and intensity-based features to the tumor classification [107-110]. While GLCM features may capture the grey-level spatial variation in an image, these deterministic features may not be effective in the analysis of the random surface

structure variation of abnormal tumor tissues in MRI. Wavelet features, on the other hand, examine the intensity variation of the tumor tissues in different image resolutions [111, 112]. In comparison, the multi-resolution fractal modeling mathematically combines the capabilities of regular texture analysis (e.g., GLCM) and multi-resolution analysis (e.g., wavelets) and, hence, may capture the randomly varying complex structure of the tumor tissue texture at different scales. The spatial intensity distributions of abnormal brain tissues in MRI have a degree of randomness that is amenable to fractal and multi-resolution fractal texture modeling.

2.3 RNA SEQUENCING AND RADIOMICS

RNAseq is a new technology that employs next-generation sequencing technology (NGS). RNAseq has developed as a novel technique to replace microarrays to measure gene expression. In RNAseq, the gene expression level is quantified by counting the number of times the RNAseq reads map (i.e. align) to one gene [113]. One of the major drawbacks of microarrays is background noise. RNAseq starts with isolating RNA from the organism, then converts RNA fragments to complementary DNA (cDNA), then prepares the sequencing library, and finally sequences the cDNA using NGS platform. RNAseq is a very sensitive technique that provides high resolution and a thorough understanding of the transcriptome and has revealed many novel gene structures.

RNAseq distribution requires an appropriate model that adapts and preserves the nature of RNAseq read counts data, and such classification models that preserve the nature of RNAseq are lacking in the traditional ML literature. The NB distribution is an appropriate choice to model such discrete read counts data [69]. Even though traditional ML tools that are developed based on NB are lacking, the choice of using NB distribution in differential gene expression and RNAseq analysis has been adapted by different studies in literature such as in EdgeR [114-116], DESeq [117], and NBPSeq [118]. In EdgeR [114-116], the authors estimate the NB dispersion parameter using a quantile-adjusted conditional maximum likelihood estimator assuming a common dispersion between the different genes and replicates. However, a common dispersion between the different genes is generally not true. Therefore, reference [115] defines a weighted conditional log-likelihood as a combination of “individual” and weighted “common” likelihoods. The "common" likelihood assumes all genes have a common dispersion value. The "individual" likelihood is estimated by conditioning on the sum of counts of a gene for each class.

An example of a count-based classifier that fits NB distribution is the Negative Binomial Linear Discriminant Analysis (NBLDA). NBLDA is a well-known classifier that is developed by fitting NB to RNAseq and the mean and dispersion parameter are estimated from the RNAseq data [119]. In the NBLDA method, the dispersion parameter is estimated using the shrinkage estimator that is proposed by Yu et al. [120]. A different type of classifier, known as VoomNSC, is developed based on the transformed count data. VoomNSC is a combination of voom (an acronym for mean-variance modeling at the observational level) transformation [65] and the nearest shrunken centroids classifier (NSC) [121]. Voom is designed as follows: first RNAseq read counts are transformed as the log-counts per million (log CPM). Then, a relationship between the square-root-standard-deviations and the mean of the log-count. Then, a precision weight (i.e., the inverse of the variance) of each sample is estimated. The estimated weights and the log-counts are used to build the NSC classifier. In NSC, a soft-thresholding shrinkage method is used to shrink each weighted difference score between a particular gene centroid and the overall centroid toward zero. Then, the updated weighted difference is used to update the weighted centroid for each remaining gene (i.e. important genes). In the prediction step, the updated weighted centroids of the important genes are used to predict the gene class.

CHAPTER 3

PREDICTION OF MOLECULAR MUTATIONS IN DIFFUSE LOW-GRADE GLIOMAS USING RADIOMICS

3.1 CHAPTER OVERVIEW

The 2016 WHO classification of CNS presents a major restructuring of diffuse gliomas. The updated classification of diffuse gliomas group tumors based on the shared genetic driver mutations, in addition to their growth pattern and behavior. Currently, the status of molecular mutations such as *IDH* mutations, *ATRX*, mutation, *TERT* mutation, and *1p/19 codeletion* is determined by obtaining tissue samples that represent viable regions of the tumor with elevated proliferation and neovascularization [122].

A shift into developing alternative methods to non-invasively classify diffuse LGG into its different subtypes has inspired different research groups to employ the developing field of radiomics and machine learning techniques. In this chapter, we address diffuse LGG grading and classification prediction based on molecular mutations using imaging features that are extracted from multimodality raw MRI sequences (T1, T1Gd, T2 FLAIR, and T2) of the anatomically depicted tumor volume and texture representations of the tumor MRI sequences. The extracted features describe the multi-resolution fractal features, texture features, volumetric, and area-based characteristics. In this study, different molecular (*IDH*, *1p/19q codeletion*, *ATRX*, and *TERT*), and MGMT methylation prediction models are introduced. In addition, our study investigates the efficacy of our novel texture features the fractal and multi-resolution fractal modeling on the performance of the non-invasive prediction of molecular mutation in LGG.

Several studies have shown the efficacy of fractal and multi-resolution fractal feature analysis for characterization, segmentation, and classification of complex abnormal brain tissues in MRI [94, 97, 105, 106]. Consequently, in this study, we hypothesize that the fractal and multi-resolution fractal modeling may relate to the underlying structure of molecular mutations. To the best of our knowledge, this is the first study that addresses the potency of fractal and multi-resolution fractal features in molecular mutations prediction.

3.2 RADIOMIC-BASED LGG MOLECULAR MUTATION PREDICTION MODEL

3.2.1 DATASET

In this study, we use a total of 108 pre-operative LGG patients described in [123-125]. Four sequences of the MRI are provided with the dataset: pre-contrast T1-weighted (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR). These scans are skull-stripped, re-sampled to $1\ mm^3$ resolution, and co-registered to the T1 template. The dataset provides the segmented sub-regions of the LGG: Gadolinium enhanced tumor (ET), the peritumoral edema (ED), and the necrosis along with non-contrast enhancing tumor (NCR/NET). Molecular alterations (*IDH* mutation, 1p/19q codeletion, *ATRX*, and *TERT* mutation), grade (II and III), and clinical data are downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). Clinical data are de-identified by the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The distribution of the data is as follows: (i) *IDH* mutation: 85 Mutant (of which 27 cases are co-deleted) and 23 wildtype (WT), (ii) 1p/19q codeletion: 27 codeletion and 81 non-codeletion, (iii) *ATRX* status: 43 Mutant and 65 WT, (iv) *TERT* status: 46 Mutant and 62 WT, and (v) O6-methylguanine-DNA methyltransferase (MGMT) promoter methylation: 91 methylated and 14 un-methylated. The patient age range of the at the time of diagnosis is 20 – 75 years, and the median value is 46.5 years.

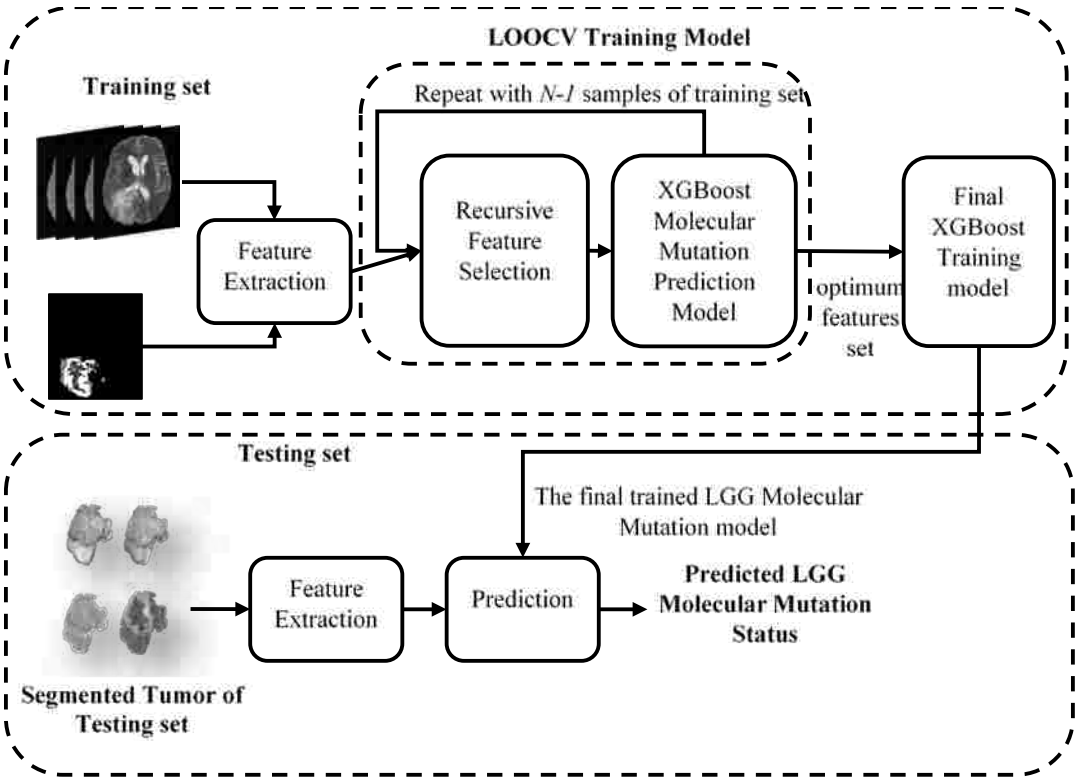


Figure 2. An overview of the proposed LGG Molecular Mutation prediction Model.

3.2.2 METHODOLOGY

In this study, we introduce different molecular prediction models based on fractal and multi-resolution fractal texture features and other MR imaging features. These molecular models include the *IDH*, 1p/19q codeletion, MGMT, *ATRX*, and *TERT* prediction. A classic way to avoid overfitting is to divide the dataset into training, validation, and testing datasets [126]. The dataset is randomly partitioned into training (75% of the entire dataset = 81 cases) and testing (25% of the entire dataset = 27 cases) subsets. A balanced distribution of the target molecular mutation is ensured in the training and testing sets in each molecular prediction model. The features are extracted from multimodality MRI sequences of the tumor volume in the training partition. Then, a recursive feature selection is performed to select the number of features and is validated with Leave-One-Out Cross-Validation (LOOCV). The selected features are then trained using an Extreme Gradient Boosting (XGBoost) method along with LOOCV. Then, a prediction performance is evaluated using the testing partition.

Furthermore, we study the efficacy of fractal and multi-resolution fractal texture features (e.g., piecewise-triangular prism surface area (PTPSA), multi-resolution fractional Brownian motion (mBm), and Holder Exponent (HE)) extracted from tumor volumes on the performance of the molecular mutation prediction models. Figure 2 shows the overall pipeline of the proposed LGG-XGBoost prediction model for different molecular mutations in LGG.

3.2.3 FEATURE EXTRACTION

In this study, around 680 features are extracted to represent texture, volume, and area of the tumor and its sub-regions (edema, enhanced tumor, and necrosis). These features include 41 texture features [127] extracted from the tumor volume in raw MRI (T1Gd, T2, and FLAIR) sequences and an additional three different texture characterizations of the tumor region. The three texture characterizations are as follows:

- fractal characterization using our PTPSA[96] modeling,
- multi-resolution mBm[97] modeling,
- and the characterization Holder Exponent (HE) [128, 129] modeling of the tumor region.

The computational algorithm of the PTPSA, mBm, and the HE is found in [96, 97, 130]. The 41 texture features are: global variance, skewness, kurtosis, energy, contrast, correlation, homogeneity, variance, sum average, entropy, short-run emphasis, long-run emphasis, gray-level non-uniformity, run-length non-uniformity, run percentage, low gray-level run emphasis, high gray-level run emphasis, short-run low gray-level emphasis, short-run high gray-level emphasis, long-run low gray-level emphasis, long-run high gray-level emphasis, gray-level variance, run-length variance, small zone emphasis, large zone emphasis, gray-level non-uniformity, zone-size non-uniformity, zone percentage, low gray-level zone emphasis, high gray-level zone emphasis, small zone low gray-level emphasis, small zone high gray-level emphasis, large zone low gray-level emphasis, large zone high gray-level emphasis, gray-level variance, zone-size variance, coarseness, contrast, busyness, complexity, and strength.

Furthermore, six histogram-based statistics (mean, variance, skewness, kurtosis, energy, and entropy) features are also extracted from the different tumor sub-regions (edema, enhanced tumor, and necrosis), respectively.

In addition, we extracted 12 volumetric features: the volume of the whole tumor, the volume of the whole tumor with respect to the brain, the volume of sub-regions (edema, enhanced tumor, and necrosis) divided by the whole tumor, the volume of sub-regions (edema, enhanced tumor, and necrosis) divided by the brain, the volumes of the enhanced tumor and necrosis divided by the edema, the summation of the volume of the edema and enhanced tumor, the volume of the edema divided by the summation of the volume of enhanced tumor and necrosis, and the volume of the necrosis divided by the summation of the volume of the edema and enhanced tumor. Finally, nine-volume properties (area, bounding box, centroid, perimeter, major axis length, minor axis length, eccentricity, orientation, solidity, and extent) are extracted from the tumor volume and from three viewpoints (*x*, *y*, and *z*-axes) of the tumor and its sub-regions (edema, enhanced tumor, and necrosis).

Texture features are analyzed using MATLAB-based software developed by Vallières et al.[127]. Fractal characterization, multi-resolution fractal characterization, HE characterization, and volumetric features are analyzed using MATLAB-based in-house software.

3.2.4 MOLECULAR MUTATION PREDICTION MODEL AND FEATURE SELECTION

The molecular mutation prediction model is performed on the training set using nested LOOCV to avoid an optimistic performance estimate. Recursive feature selection is performed in the inner loop to find the optimum features set. The LOO cross-validated performance of the molecular mutation prediction model is estimated in the outer loop. The molecular mutation prediction model is performed using the R statistical packages Caret and XGBoost (www.r-project.org).

Recursive feature selection (RFS) is implemented by first fitting a Random Forest (RF) model to all features. Each feature is ranked by its importance, and the least important features are removed from the current feature set. Then, this step is repeated recursively until the optimum features set that has the best performance is reached. In our implementation of recursive feature selection, the number of features in the features' sets are 2, 3, 5, 7, 9, 11, 13, and 15 features. In addition, the best performance is determined by maximizing the area under the receiver operating characteristic (ROC) metric. The feature set that provides the combination of features that maximize

the area under the ROC (AUC) is chosen for training in the prediction model. Using recursive feature selection, the maximum number of selected features is fifteen so that the training samples (eighty-one cases) is at least 5 times the number of features to reduce model overfitting.

In our study, XGBoost is utilized as a classification and prediction model using the optimum features set as input and molecular mutation information as the target output. XGBoost [131] is an advanced tree boosting supervised machine learning technique that is effective in handling imbalanced datasets. XGBoost is widely used in classification and regression tasks. For a given dataset D with *s* samples and *m* features $D = \{(x_i, y_i)\}(|D| = s,\ x_i \in \mathbb{R}^m, y_i \in \mathbb{R})$, a tree ensemble model uses *K* additive functions to predict the output as follows,

$$\widehat{y_i} = \sum_{k=1}^{K} f_k(x_i)\,, f_k \in F, \tag{1}$$

where x_i is the feature/input vector, y_i is the target/output variable, and $f(x)$ is a function in the functional space F, and F is a set of all possible classification and regression trees. One of the major advantages of using XGBoost is that XGBoost provides L1 and L2 regularization. L1 regularization handles sparsity, whereas L2 regularization reduces overfitting. In addition, we choose XGBoost because it is known for handling an imbalanced dataset. A detailed mathematical derivation of the XGBoost algorithm is found in Chen et al. [131]. The final molecular mutation prediction model (that is used for the testing set) is obtained by fitting the optimum features set that maximizes performance in the inner loop (over all the outer cross-validation loops). Note, if more than one feature sets maximize inner loop performance, the common features between the features' sets are used. The prediction performance of the final molecular mutation model is tested using the paired testing sets (partitions).

Finally, in order to study the efficacy of fractal and multi-resolution fractal texture features used in this study (e.g., PTPSA, mBm, and Holder Exponent) on the performance of the proposed prediction models as shown in Figure 2, we perform molecular prediction analyses with and without these texture features, respectively. The entire process in Figure 2 is repeated *n* times independently with *n* different training/testing set pairs. The *n* number of repetitions is a random number between 10 and 15 that is generated for each model.

3.2.5 MOLECULAR MODELS EVALUATION

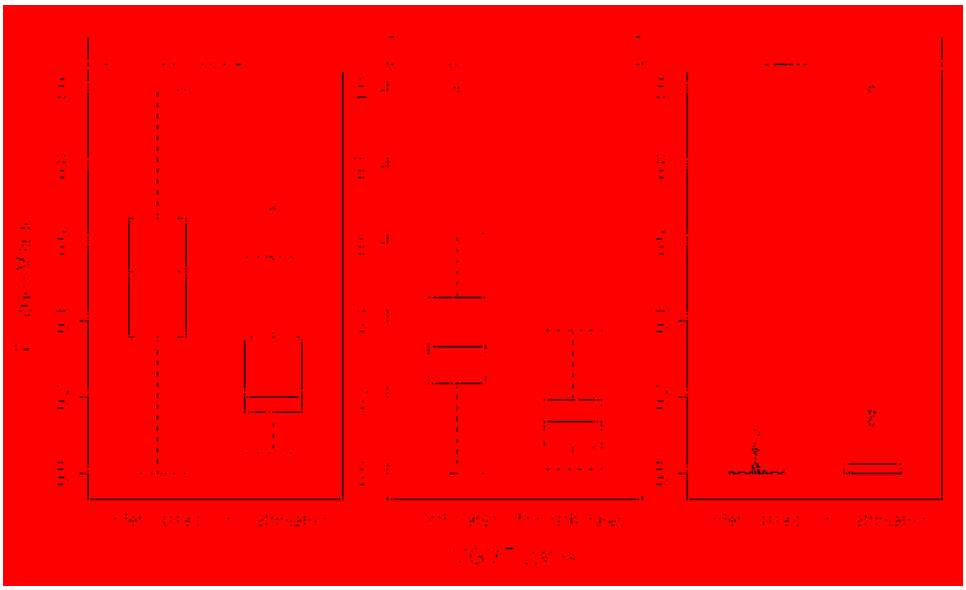
The molecular models are validated using separate testing sets, and the prediction performance (test performance) of the trained XGBoost model is estimated using AUC, sensitivity, and specificity. After *n* times of independent repetitions, the ANOVA test is used to compare the difference in the prediction performance between two models with and without the fractal and multi-resolution fractal texture in the prediction models. In addition, ANOVA is used to analyze the significant association between features and the different molecular mutations. The survival groups that are formed using the significant features are compared using Kaplan-Meier curves and the log-rank test. The hazard ratio of features is determined using the Cox proportional hazards model and assessed using the likelihood-ratio test. Finally, the evaluation step for Survival is conducted using R statistical packages.

3.3 EXPERIMENTAL RESULTS

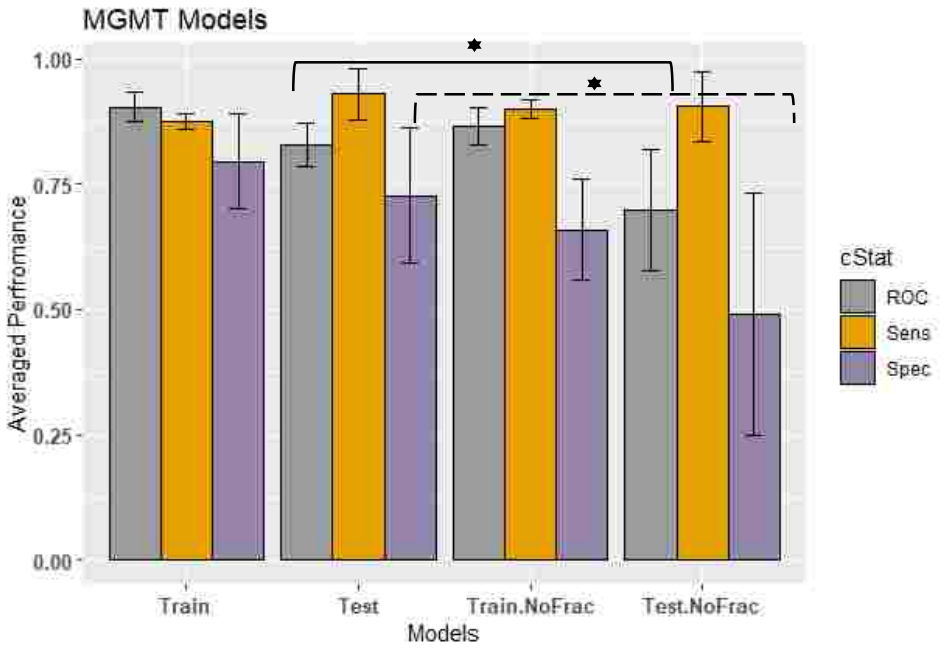
Around 680 imaging features are extracted from multimodality MRI sequences. Recursive feature selection is used to find the optimum number of significant features for each molecular mutation prediction model. Our analyses of the different prediction models are repeated independently *n* times with different training and testing pairs (partitions). Table 2 displays the number of repetitions *n*, LOOCV performance, and the test performance of the different prediction models when including/removing texture characterization of the fractal and multi-resolution fractal of PTPSA, mBm, and Holder Exponent characterization.

3.3.1 MGMT METHYLATION MODEL

The most frequent features that are selected are the necrosis width in the *Z* planar, the histogram entropy of the mBm characterization on the whole tumor, and size ratio between the enhanced tumor and the necrosis. The necrosis width in the *Z* planar, the histogram entropy of the mBm characterization on the whole tumor features are significantly (ANOVA test, p-value < 0.05) associated with methylated MGMT, whereas the size ratio between enhanced tumor and necrosis is associated significantly with un-methylated MGMT as shown in Figure 3a.



(a)



(b)

Figure 3. MGMT methylation models. a) Distribution of the most selected features in discriminating MGMT mutated and WT, and b) MGMT prediction model performance using the train and test partitions with and without fractal texture features. Error bars represent standard deviation. The asterisk * illustrates the significant difference between two measurements.

The LOOCV performance and the prediction performance on the testing set for predicting the MGMT methylation status using imaging features are illustrated in Table 2 and Figure 3b. The prediction performance using the test partitions achieves an AUC, a sensitivity, and a specificity of 0. 0.83±0.04, 0.93±0.05, and 0.73±0.13, respectively.

Removal of our fractal and multi-resolution fractal features from the MGMT methylation prediction model has an effect on the prediction performance on the testing set. AUC and specificity drop significantly (ANOVA test, *p*-value = 0.003, and 0.01, respectively) when the fractal features are removed (Figure 3b).

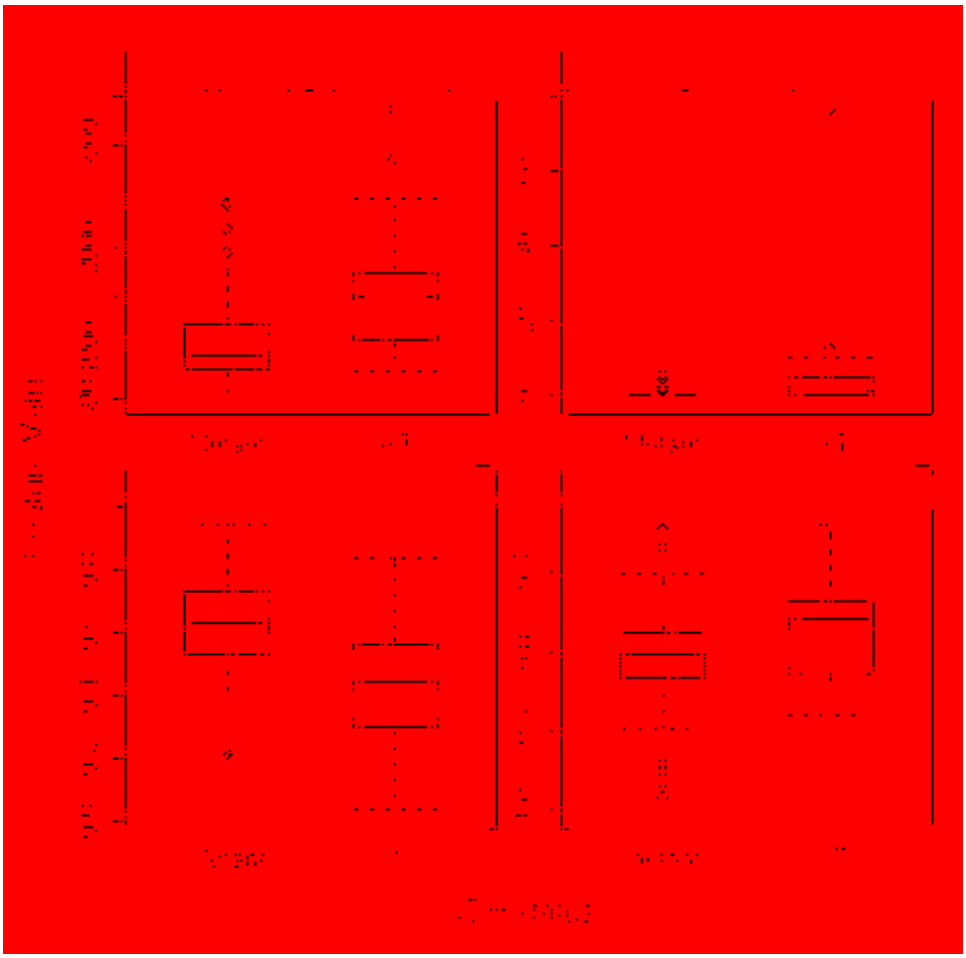
Table 2. LOO cross-validated performance of the outer-loop, and the predictive/test performance of the different LGG molecular prediction models.

Cross-Validated performance	***n* repetition**	**With fractal & multi-resolution fractal features**			**Without fractal & multi-resolution fractal features**		
		AUC.	Sens.	Spec.	AUC.	Sens.	Spec.
MGMT Methylation	11	0.86±0.03	0.88±0.02	0.80±0.09	0.87±0.04	0.90±0.02	0.66±0.10
***IDH* mutation**	13	0.85±0.04	0.90±0.03	0.75±0.05	0.79±0.07	0.89±0.08	0.75±0.07
1p/19q codeletion	15	0.83±0.03	0.78±0.08	0.83±0.03	0.80±0.05	0.63±0.08	0.87±0.02
***ATRX* mutation**	10	0.77±0.06	0.62±0.09	0.80±0.03	0.80±0.04	0.77±0.06	0.76±0.03
TERT	14	0.82±0.04	0.70±0.06	0.83±0.04	0.82±0.04	0.82±0.04	0.76±0.05
Prediction/test Performance							
MGMT Methylation	11	**0.83±0.04**	0.93±0.05	**0.73±0.13**	**0.70±0.12**	0.90±0.07	**0.50±0.24**
***IDH* mutation**	13	**0.84±0.03**	0.90±0.06	**0.79±0.09**	**0.75±0.07**	0.83±0.11	**0.66±0.18**
1p/19q codeletion	15	**0.80±0.04**	**0.75±0.08**	0.85±0.06	**0.75±0.07**	**0.67±0.12**	0.84±0.10
***ATRX* mutation**	10	0.70±0.09	0.69±0.06	**0.83±0.10**	0.66±0.10	0.65±0.16	**0.68±0.18**
TERT	14	0.82±0.04	0.77±0.12	0.86±0.09	0.78±0.07	0.77±0.11	0.79±0.11

3.3.2 IDH MUTATION MODEL

Our analysis reveals that the tumor correlation, the vertical orientation of edema major axis (the angle between the edema major axis and the vertical axis), the size ratio between the enhanced tumor and the necrosis, and the complexity of the holder exponent of the tumor are among the most frequently selected features to predict *IDH*-mutated status in LGG. Higher values of the complexity of holder exponent of the tumor, the size ratio between the enhanced tumor and the necrosis, and higher values of the vertical orientation of edema major axis associate significantly (ANOVA test, *p-value* < 0.005) with WT *IDH* status, whereas the tumor correlation associates significantly (ANOVA test, *p-value* < 0.005) with mutated *IDH* status as illustrated in Figure 4a. Figure 4b shows the clustering of IDH status using the most frequent features in Figure 4a. The clustering between the mutated *IDH* and WT *IDH* is demonstrated using t-Distributed Stochastic Neighbor Embedding (tSNE) [132].

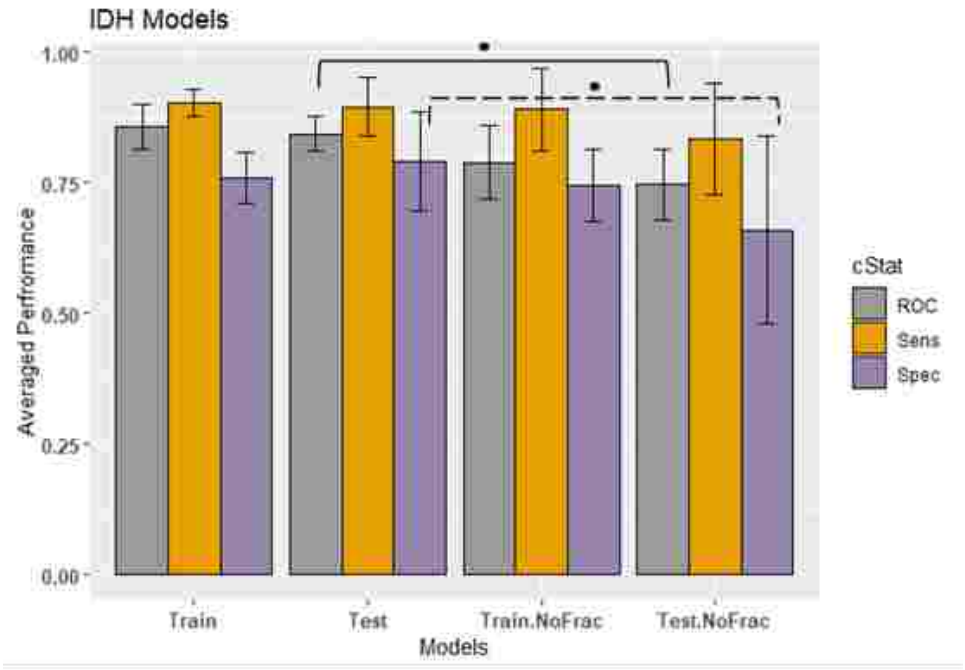
In addition, the tumor correlation, the vertical orientation of edema major axis, and the complexity of holder exponent of the tumor features carry a hazard ratio [HR] of 0.562 (95% CI, 0.381-0.828), 2.655 (95% CI, 1.617-4.36), and 1.553 (95% CI, 1.165-2.07) with likelihood ratio test *p*-value = 0.005, 0.0001, 0.008, respectively. Because these features are continuous features, the HRs interpolate as follows: the risk of death increases (or decreases if HR<1) by (HR-1) ×100% for every 1-standard deviation increase in that feature. The LOOCV and the test performance of the proposed IDH models are illustrated in Table 2. The prediction performance using the testing partitions achieves an AUC, sensitivity, and specificity of 0.84±0.03, 0.90±0.06, and 0.79±0.09, respectively. Note that the AUC and specificity of the IDH status prediction model drop significantly to 0.75±0.07 and 0.66±0.18 (ANOVA test, $p = 0.0001$ and $p = 0.028$, respectively) after removing features extracted from texture extracted from fractal and the multi-resolution modeling (Figure 4c).



(a)

Figure 4. IDH models. a) Distribution of the most selected features in discriminating among IDH mutated and WT cases in LGG, b) 2-D t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization using only the 4 features in A, and c) Performance comparison of IDH prediction model using the train and test partitions with and without fractal features. Error bars represent 2 standard deviations. The asterisk * illustrates the significant difference between two measurements.

(b)

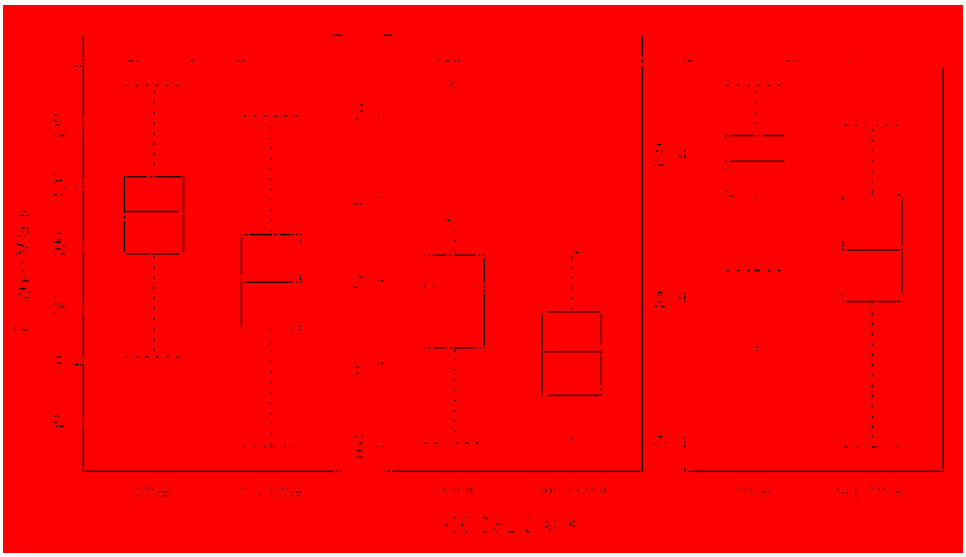


(c)

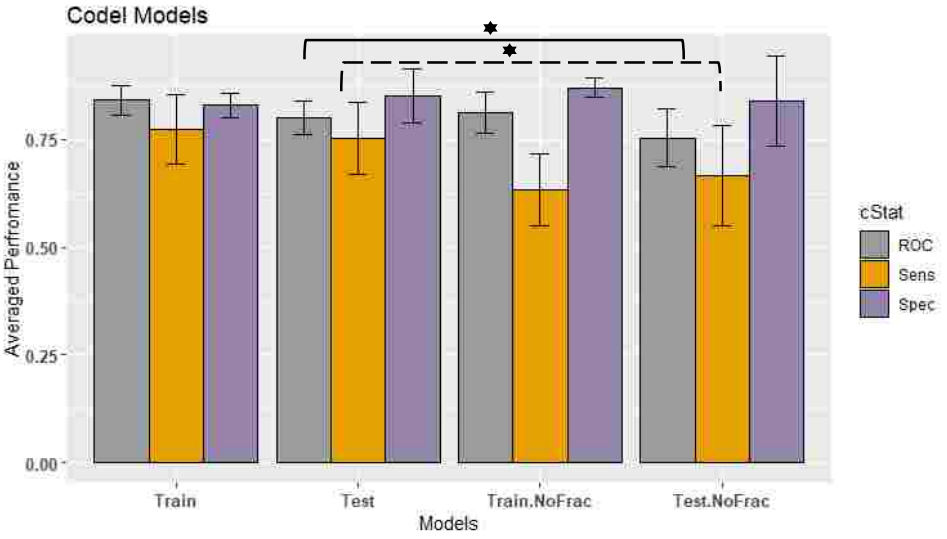
Figure 4. (continued)

3.3.3 1P/19Q CODELETION MODEL

The necrosis upper-left bounding box location, the histogram entropy of mBm characterization of the whole tumor, and the horizontal coordinate of necrosis centroid are the most frequent optimum features that are selected in our proposed 1p/19q codeletion models. These three features show significance (ANOVA test, p-value < 0.0001) associated with the existence of the 1p/19q codeletion. Our analysis shows frontal tumors are associated significantly (ANOVA test, p-value < 0.0001) with 1p/19q codeletion mutations (Figure 5.a). The performance of the proposed codeletion prediction LOOCV model is illustrated in Table 2. The 1p/19q codeletion performance using the test partitions achieves an AUC of 0.80±0.04, a sensitivity of 0.75±0.08, and a specificity of 0.85±0.06. In addition, the efficacy of our fractal and multi-resolution fractal texture features on the performance of the codeletion prediction model is significant as shown in Figure 5b and Table 2. The AUC and the sensitivity of the codeletion prediction model drop significantly (ANOVA test, p-value of 0.024 and 0.029, respectively) after removing features extracted from our fractal and multi-resolution fractal features in the 1p/19q codeletion prediction model to 0.75±0.07 and 0.67±0.12 (without fractal features). Figure 5c illustrates the location of the centroid and the upper-left bounding box of the necrosis. The histogram entropy of mBm of the tumor volume offers HR of 0.59 per standard deviation (95% CI, 0.35-0.97) with a likelihood ratio test p-value of 0.037.

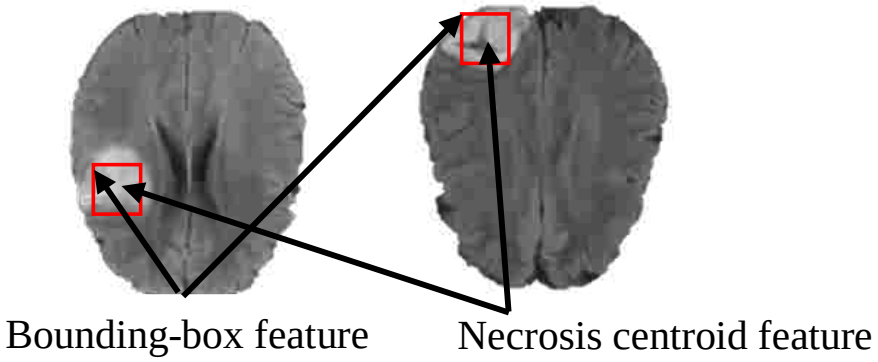


(a)



(b)

Figure 5. 1p/19q codeletion models. a) Distribution of the most selected features in discriminating 1p/19q codeletion and non-codeletion, b) Performance comparison of codeletion classifier models using the train and test partitions with and without fractal features. c) Example of FLAIR images illustrates the location of the necrosis centroid and the upper-left location of the necrosis bounding-box. Error bars represent 2 standard deviations. The asterisk * illustrates the significant difference between two measurements.

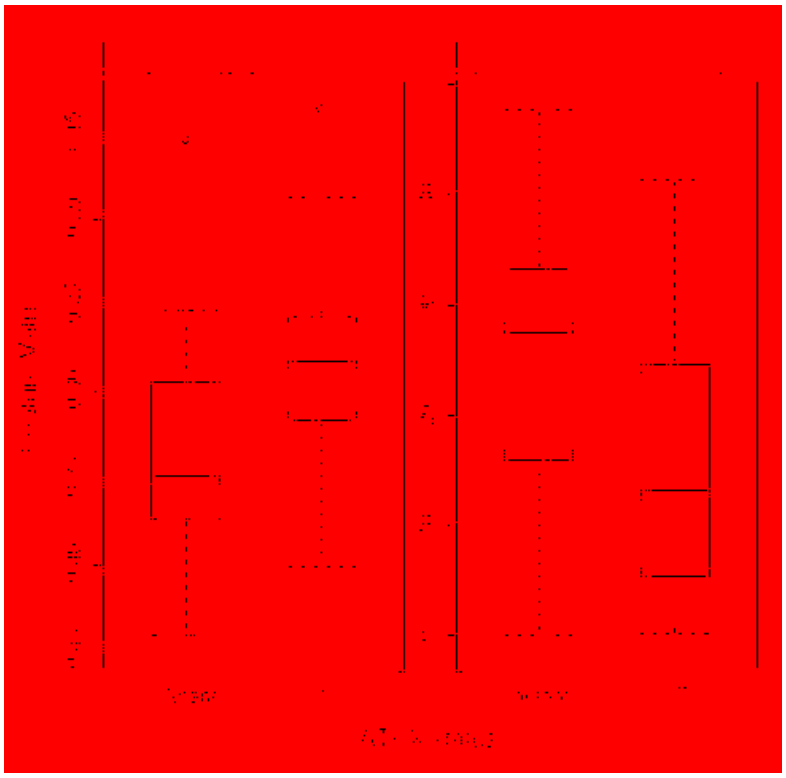


(c)

Figure 5. (continued)

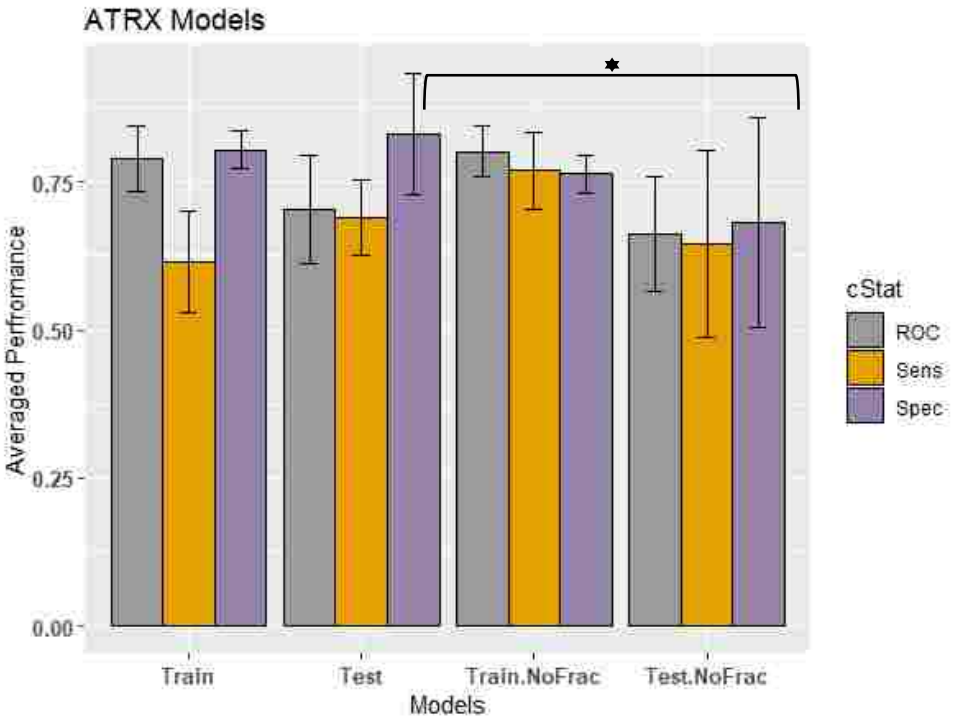
3.3.4 ATRX MUTATION MODEL

The information content of correlation and the histogram mean of the tumor volume of the most frequently selected features are employed in the XGBoost model to train and predict *ATRX* status. The distribution of the most frequent features is illustrated in Figure 6.a. Higher values of information of correlation are associated significantly (ANOVA test, p-value < 0.001) with *ATRX* wildtype, whereas higher values of histogram mean are associated significantly (ANOVA test, p-value < 0.001) with mutated *ATRX*. The *ATRX* prediction model achieves prediction performance of an AUC of 0.70±0.09, a sensitivity of 0.70±0.06, and a specificity of 0.83±0.10 using the test partitions. Removing features extracted from our fractal and multi-resolution fractal modeling from the *ATRX* prediction model has a significance specificity drop to 0.68±0.18 performance of the model with p-value = 0.03 (ANOVA test) as shown in Figure 6b.

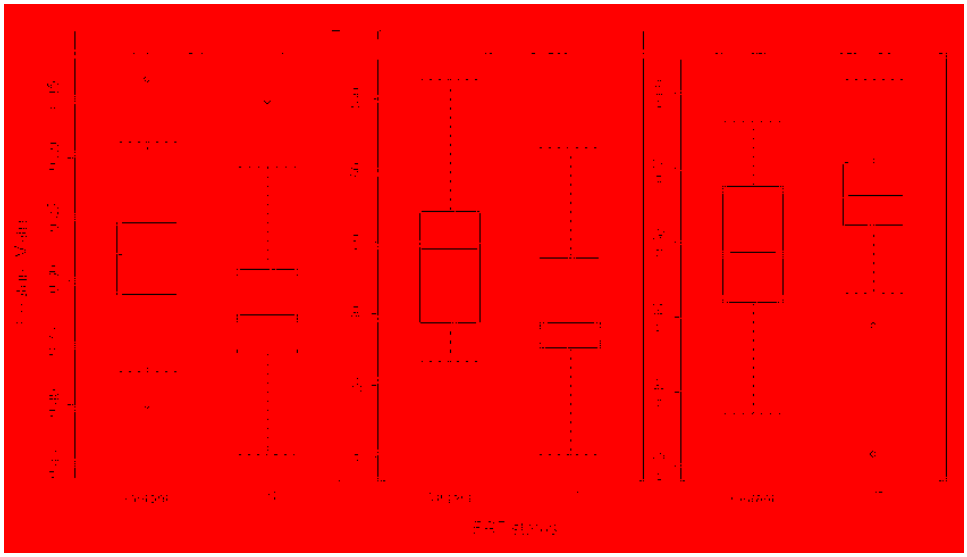


(a)

Figure 6. Distribution of the most selected feature in discriminating a) mutated ATRX and WT, b) Performance comparison of *ATRX* classifier models using the train and test partitions with and without fractal features. Error bars represent 2 standard deviations. The asterisk * illustrates the significant difference between two measurements, and c) mutated TERT and WT.



(b)



(c)

Figure 6. (continued)

3.3.5 TERT MUTATION MODEL

After reviewing the most frequently selected imaging features of the TERT mutation prediction model, we notice that the information content of correlation of the tumor volume, the edema upper-left bounding box location, and the inverse difference moment of HE characterization of tumor volume are the most frequently selected features (Figure 6c). The inverse difference moment of HE characterization of tumor volume offers HR of 0.612 per standard deviation (95% CI, 0.405-0.924) with a likelihood ratio test *p-value* = 0.027. The TERT prediction models' performances are illustrated in Table 2. The TERT prediction performance using the test partitions achieves an AUC of 0.82±0.04, a sensitivity of 0.77±0.12, and a specificity of 0.86±0.09, respectively. Removing our fractal and multi-resolution fractal texture modeling has no significant effect on the TERT performance of prediction models when using the test partitions as shown in Table 2.

3.4 DISCUSSIONS

The 2016 WHO classification of diffuse LGGs heavily weighs molecular mutations classifying primary brain tumors with particular importance assigned to *IDH* mutation, 1p/19q codeletion, *ATRX* mutation, *TERT* mutations, and MGMT methylation. Our study on diffuse LGG is largely able to predict the presence of these important molecular mutations based on MR imaging features. Therefore, prediction of tumor aggressiveness (based on molecular mutations) may be achieved through non-invasive imaging features as an adjunct to traditional visual morphologic diagnosis and invasive tissue sampling.

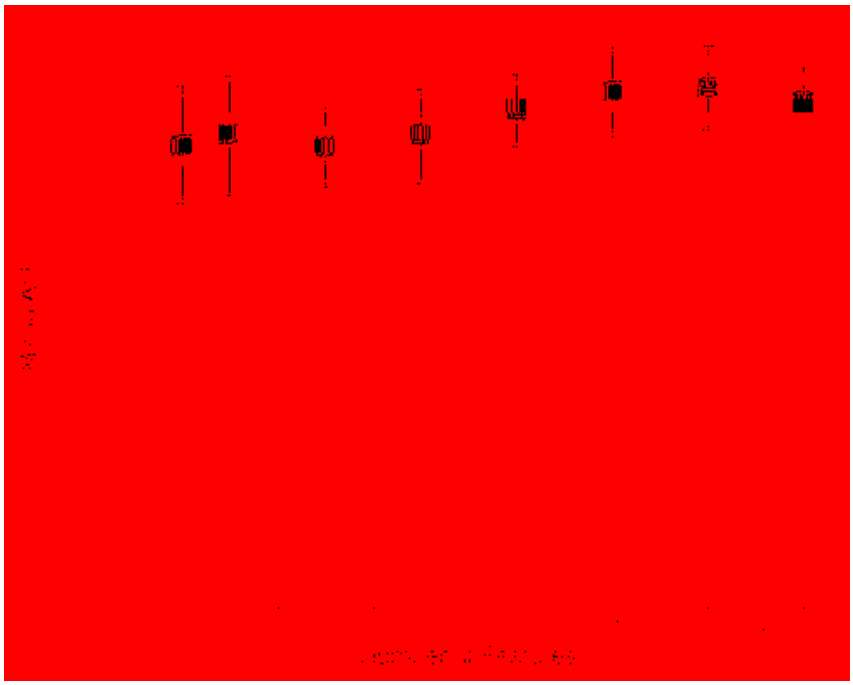
In this work, the number of originally extracted imaging features (six hundred eighty features) is higher than the number of samples (eighty-one cases) in the training dataset, which may cause overfitting. To address possible overfitting, we implement feature selection in the training model that offers a maximum of fifteen selected features. Figure 7 illustrates the effects of the number of features on the cross-validated performance of the different mutation prediction models. The average AUC performances and the standard error of the different prediction models improve when the number of features is greater than 9. Note that the standard error mostly plateaus when the number of features varies between 9-15 (standard error reflects instability).

For the fractal and multifractal texture model in Table 2, the AUC predictive performance of MGMT, 1p/19q codeletion, and *ATRX* models drop to 0.83±0.04, 0.80±0.04, and 0.70±0.09, respectively. This statistically non-significant drop in performance (ANOVA test, *p-value* = 0.076, *p-value* = 0.073, and *p-value* = 0.071 respectively) when compared to their AUC cross-validated performance, may suggest minimal overfitting for these models. The AUC predictive performances of *IDH* and *TERT* models are almost comparable to their cross-validated performance as shown in Table 2 that suggests there is no overfitting in these two models.

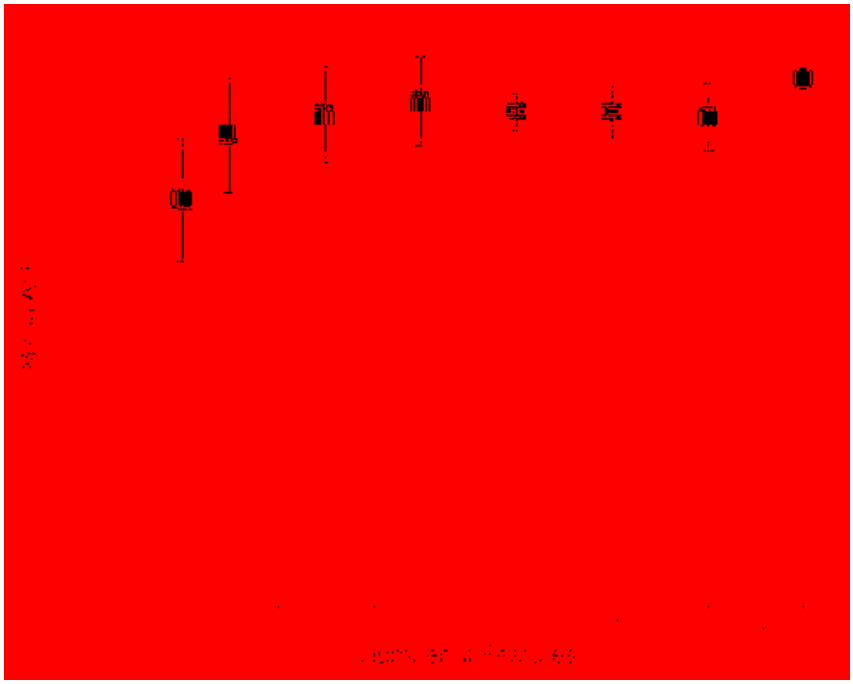
For the non-fractal models in Table 2, the AUC predictive performance of *IDH*, 1p/19q codeletion, and *TERT* models drops to 0.75±0.07, 0.75±0.07, and 0.78±0.07 respectively. This statistically non-significant drop (ANOVA test, *p-value* = 0.182, *p-value* = 0.056, and *p-value* = 0.062 respectively) when compared to their AUC cross-validated performance, may suggest minimal overfitting for these models. However, the poor predictive performances of the non-fractal MGMT and non-fractal ATRX models when compared with their non-fractal cross-validated performances are a sign of overfitting.

Note optimally chosen features in the non-fractal models are not selected by simply replacing the fractal features with alternate features (or by using the same number of predictors as in the fractal models). In each mutation model (fractal or non-fractal), RFS (and thus selecting optimally chosen features) is performed independently. In our implementation of RFS for fractal or non-fractal models, the possible number of features in the features' sets could be 2, 3, 5, 7, 9, 11, 13, and 15, respectively. The maximum number of selected features is set to 15 such that the training samples (81 cases) are at least 5 times the number of features to reduce the possibility of model overfitting.

Finally, we compare the performances of our prediction models with a list of state-of-the-art studies as illustrated in Table 3. However, a direct comparison between the performances of our proposed models and these studies may not be relevant because of the different datasets.

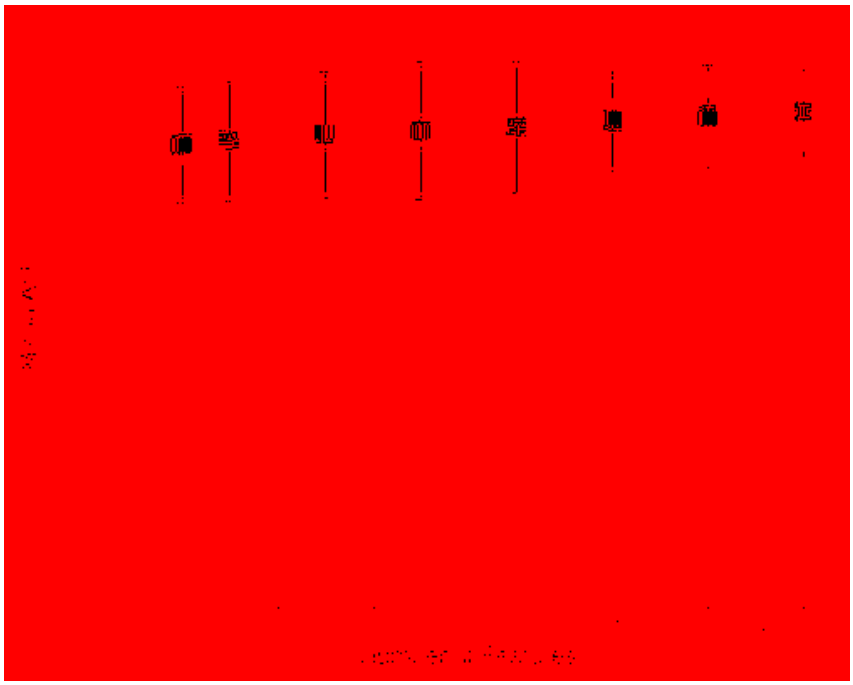


(a)

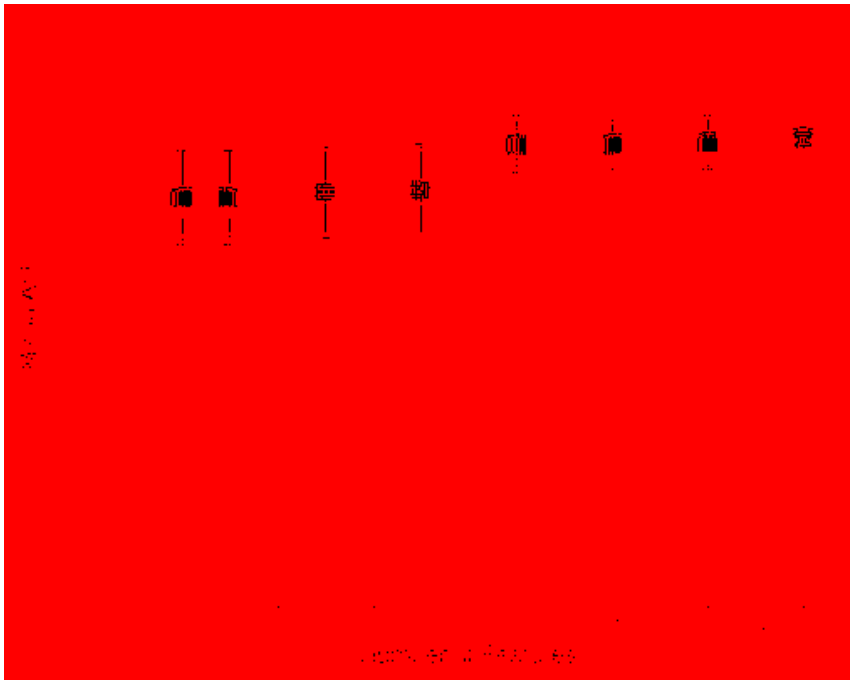


(b)

Figure 7. The effect of number of features on cross-validated performance of the different mutation models. a) MGMT prediction model, b) IDH prediction model, c) 1p/19q codeletion prediction model, d) ATRX prediction model, and e) TERT prediction model. The y axis represents mean AUC of every feature set which is computed from all independent n repetitions, and error bars represent 2 standard deviations. The x axis represents the number of selected features.



(c)

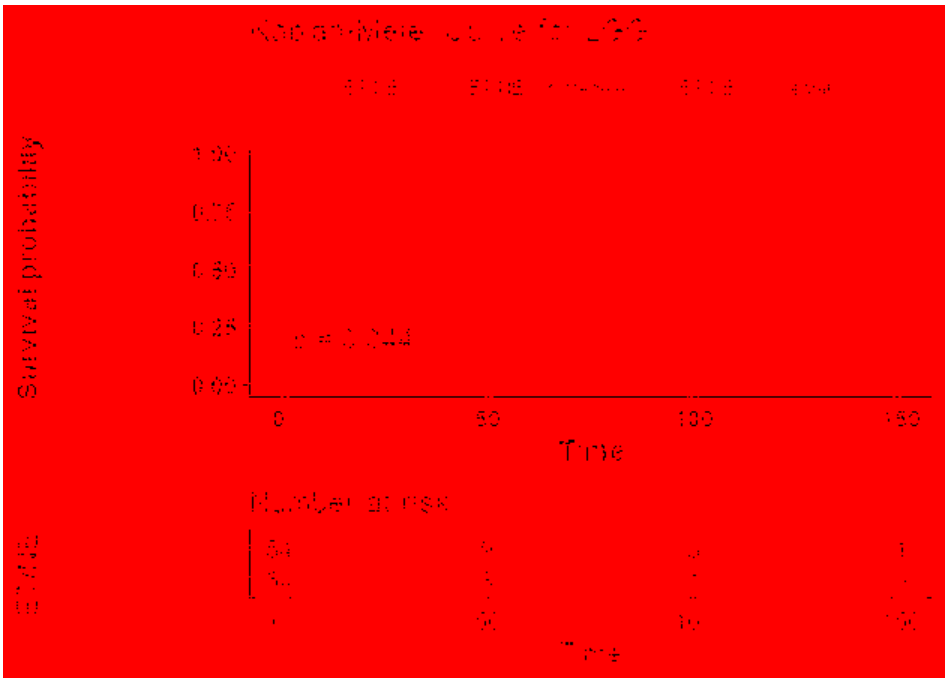


(d)

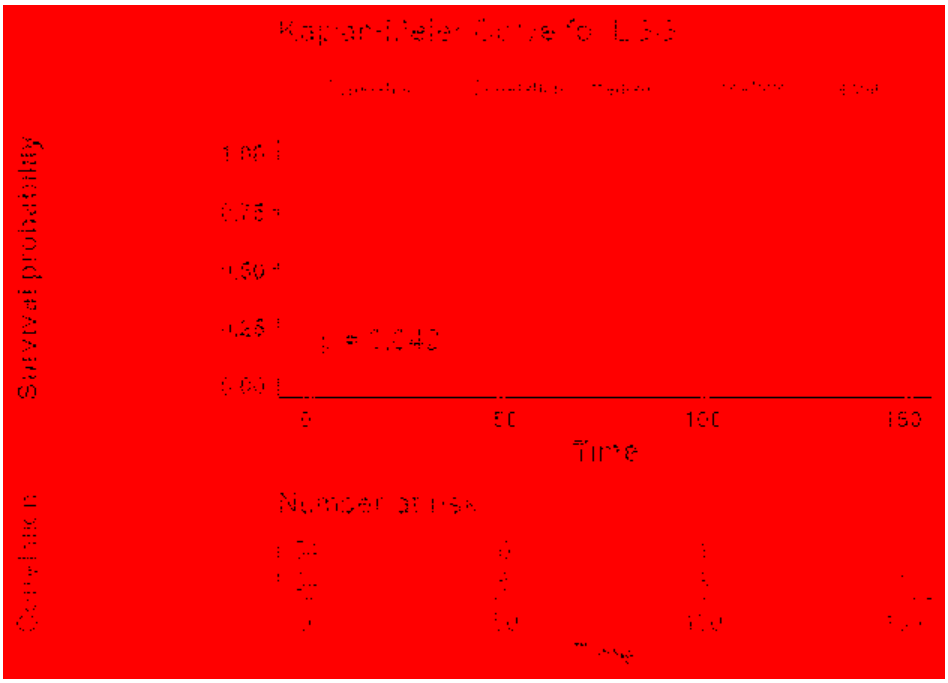
Figure 7. (continued)

(e)

Figure 7. (continued)

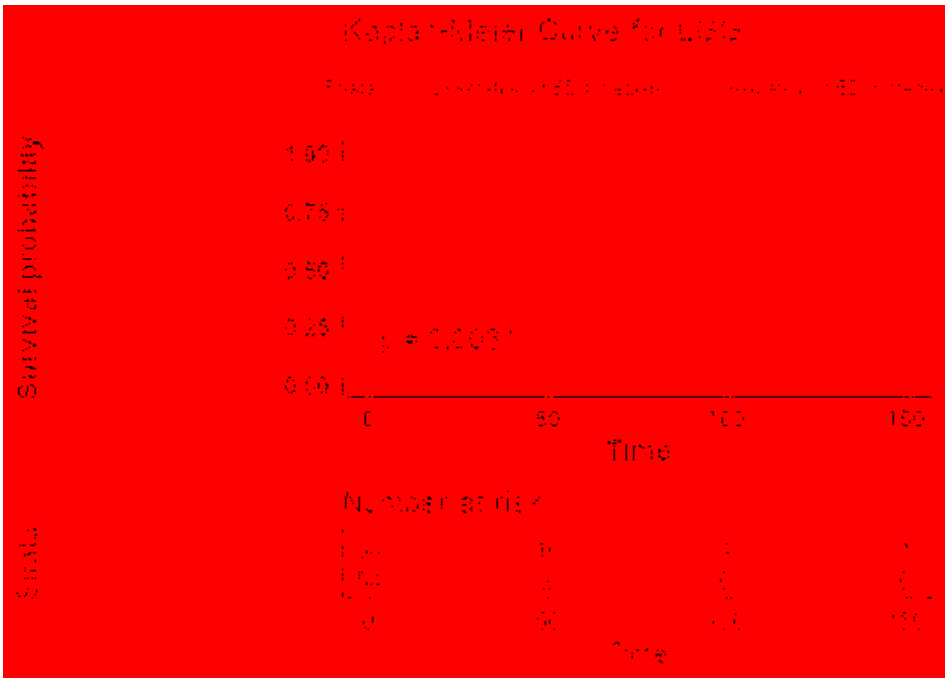


(a)



(b)

Figure 8. The effect of thresholding a) the size ratio between the enhanced tumor and necrosis, b) tumor correlation, and c) vertical orientation of edema major axis around the median. P-values are computed using the likelihood ratio test.



(c)

Figure 8. (continued)

Our study on MGMT methylation prediction shows that MGMT methylation correlates with high values of fractal texture features such as histogram entropy of mBm for tumor volume. Entropy measures randomness or uncertainty in the tumor. The analysis reveals that high histogram entropy of mBm associates with the less aggressive methylated MGMT status and carries HR of 0.579 per standard deviation (95%CI, 0.345, 0.969) with likelihood ratio test *p*-value = 0.035. The study further shows that the size ratio between enhanced tumor and necrosis correlates significantly with un-methylated MGMT, which indicates that the higher the aggressive MGMT un-methylated LGG, the higher the values of the size ratio.

The MGMT methylation prediction study by Kanas et al. [59] for patients with GBM reports the size of the tumor with respect to necrosis as one of the significant features. Our analysis shows that removing the texture features such as fractal and the multi-resolution fractal (of PTPSA, mBm, and Holder Exponent) characterization is significant on the AUC and specificity performance of the MGMT methylation model. The GBM study conducted by Kanas et al. [59] (Table 3) proposes an MGMT prediction model using volumetric, morphological, and locational MR imaging features, respectively. In our study, we use texture features and volumetric features.

Moreover, the entire process of the prediction model in our current study including feature extraction is automated, unlike the proposed work by Kanas et al. [59]. In another study by Han et al. [99] (Table 3), the authors use a bi-directional convolutional recurrent neural network to predict MGMT methylation status. A major difference between our MGMT prediction model and the method proposed by Han et al. [99] is that our model mainly utilizes quantitative imaging features that can be correlated with tumor biology.

In addition, our *IDH* mutation prediction model indicates that the tumor correlation associates significantly with mutated *IDH* and offers HR of 0.562 per standard deviation with a likelihood ratio test *p*-value =0.005. In addition, our analysis shows that the complexity of HE of enhanced tumor associates significantly with WT IDH status with HR of 1.553 per standard deviation with a likelihood ratio test *p*-value =0.008. Complexity is related to the visual information content and the shape of the object. Texture with higher information content and with a large number of edges is considered complex [133]. This outcome is in agreement with another glioma study by Wang et al. [134] which reports that the enhancement patterns predict the prognosis in *IDH1* mutations in Anaplastic gliomas. Our analysis also shows that the size ratio between enhanced tumor and necrosis is a significant predictive feature of the *IDH* status. This feature is also a significant predictor of MGMT status, which can be explained by the fact that MGMT methylation is associated with *IDH* status as reported by different studies [14, 135].

The *IDH* prediction model by Yu et al. [100] (Table 3) uses 110 imaging features and SVM for Grade II glioma. Even though the dataset we use in the *IDH* mutation prediction is not the same as the dataset used by Yu et al. [100], the dataset used in our study is more diverse with data from both Grade II and III, and this reflects higher reliability of the performance of the *IDH* status prediction model.

Furthermore, our study on the 1p/19q codeletion prediction model indicates that the location of the upper-left necrosis bounding box and horizontal coordinate of the necrosis centroid (illustrated in Figure 5c) are among the most predictive features. This outcome is in agreement with different studies [136-138] which report that gliomas with 1p/19q codeletion are associated with the tumor location. In addition, our analysis shows that higher values of histogram entropy of mBm texture of tumor volume associate significantly with the existence of 1p/19q codeletion. Moreover, our analysis reveals that removing the texture representation of fractal and multi-resolution fractal from the 1p/19q codeletion prediction model decreases the AUC and specificity

significantly. The test prediction performance of the 1p/19q codeletion prediction model drops (after removing the fractal features). A study by Akkus et al. [102] (Table 3) with LGG patients (where each patient has 3 MRI slices) proposes 1p/19q codeletion prediction using a convolutional neural network (CNN). Their method achieves better sensitivity; however, our method achieves slightly better specificity. In addition, Akkus et al. [102] do not consider the global information of the tumor, since their dataset uses only 3 slices of the MRI sequence of each patient as input, not the whole volume of the tumor.

Our analysis of *ATRX*-status prediction shows that tumor information-measure of correlation imaging feature and histogram mean tumor volume are the most frequently selected features. Higher values of information-measure of correlation associate significantly with WT *ATRX* status. This is in agreement with an *ATRX* mutation prediction study by Li et al. [58] (Table 3) in patients with a low-grade glioma, where the authors use MRI texture feature and the LASSO regression model. In their model, information-measure of correlation is one of the features that is used to predict *ATRX* mutation. In addition, our analysis shows that the tumor information-measure of correlation is one of the most frequent features in the *TERT* model as well. This can be explained by the fact that *ATRX* and *TERT* mutations are mutually exclusive.

The *TERT* prediction analysis shows that tumor information-measure of correlation and upper-left edema bounding box are the most frequently selected features. Higher values of these two features are significantly associated with mutated *TERT* status. The information-measure of correlation assesses the correlation/dependency between two gray-levels using mutual information content. Higher values of information-measure of correlation are associated with mutated *TERT*. In addition, our analysis suggests that the higher values of Inverse difference moment of HE associates significantly with WT *TERT* and offers HR = 0.612 per standard deviation with a likelihood ratio test, p-value = 0.03. Inverse difference moment measures local homogeneity. High values of inverse difference moment of an HE tumor predict the less aggressive WT *TERT*.

Overall, our analysis shows that the necrosis location and the necrosis volume-related features are very important (most frequently selected features) in MGMT, *IDH*, and 1p/19q codeletion prediction. Edema volume-related features are very important in *IDH* and *TERT* prediction models. Fractal features have a significant effect on MGMT, *IDH*, 1p/19q codeletion, and *ATRX* prediction models. Further analysis of the most frequent features in each prediction model indicates that the effect of thresholding the value of the standardized feature around the

median can stratify the 108 cases significantly (log-rank test, *p-value* < 0.05) into two survival groups (Figure 8a-c). The features and the median survival of each group are:

- the size ratio between the enhanced tumor and necrosis stratifies the 108 cases into two groups with a median survival of 87.4 months vs 30.7 months,
- the correlation of the tumor volume stratifies the 108 cases into two groups with a median survival of 114 months vs 46 months,
- and the vertical orientation of edema major axis stratifies the 108 cases into two groups with a median survival of 114 vs 44 months.

In summary, this study presents molecular prediction model designs from traditional MRI data based on the 2016 update of the WHO classification of LGG of the CNS. Our prediction model performance shows promise when compared to other methods and models in the literature. An association among computed MR imaging features and the molecular mutations LGG was established. The methods discussed in our study are important steps towards non-invasive imaging classification of diffuse LGG based on molecular mutations prior to invasive tissue sampling. In this work for the first time in the literature, we hypothesize that fractal and multi-resolution fractal features have an association with molecular prediction. The feature selection using RFS and the subsequent prediction results in Table 1 confirm our hypothesis by showing the efficacy of these fractal features in glioma prediction. Therefore, this work may be considered as a validation of previously hypothesized fractal biomarkers and, hence, may have the potential for generalizability for other types of tumors.

Table 3. Comparison between our proposed molecular mutations models and state-of-the art glioma grading models.

Our Proposed Prediction Models 108 LGG (75% training, 25 testing)					**Other Models**	
Models	**Test Performance (using testing sets)** AUC, Sens., Spec.			***n***	**Study (dataset)**	**Performance**
MGMT	0.83 ± 0.04	0.93 ± 0.05	0.73 ± 0.13	11	Kanas et al.[59] (86 GBM patients)	Acc, Sens., Spec. 0.736, 0.85, 0.66
					Han et al. [99] (159 LGG; 70% training, 15% validation, and 15% testing)	AUC, Accuracy, Prec., Recall Testing: 0.61, 0.62, 0.67, 0.67 Validation: 0.66, 0.67, .67,0.73
IDH	0.84 ± 0.03	0.90 ± 0.06	0.79 ± 0.09	13	Yu et al.[100] (110 training, 30 independent validation)	Accuracy, AUC, Sens., Spec. Training: 0.80, 0.86, 0.83, 0.74 Validation: 0.83, 0.79, 0.88, 0.67
1p/19q co-del.	0.80 ± 0.04	0.75 ± 0.08	0.85 ± 0.06	15	Akkus et al.[102] (159 LGG (252 slices), validation (68 slices), and testing (90 slices)).	Accuracy Sens., spec. 0. 877, 0.933, 0.822
ATRX	0.70 ± 0.09	0.69 ± 0.06	0.83 ± 0.10	10	Li et al.[58] (95 LGG, 63 training, 32 validation, and 91 external validation	Accuracy, AUC, Sens. Spec. Validation: 0.938, 0.925,0.83.3,1.00 External: 0.769, 0.725, 0.571, 0.857

CHAPTER 4

GLIOMA GRADING AND PREDICTION USING JOINT MODELING OF RNA SEQUENCING AND RADIOMICS IN A NEGATIVE BINOMIAL DISTRIBUTION

4.1 CHAPTER OVERVIEW

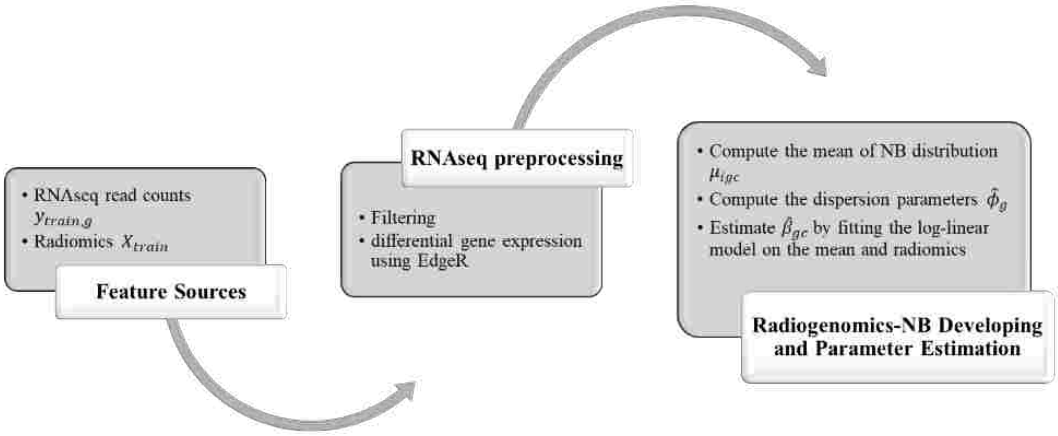
RNA sequencing (RNAseq) is a recent technology that profiles gene expression by measuring the relative frequency of the RNAseq reads. Contemporary literature lacks appropriate modeling for RNAseq distribution where the model is adaptive and also preserves the nature of RNAseq read counts data for glioma grading and prediction. Negative Binomial (NB) distribution may be useful to model RNAseq read counts data that addresses these shortcomings. In addition to RNAseq read counts data, radiomics has long played a pivotal role in current radiology practice such as disease diagnosis, monitoring, and treatment planning.

In this study, we propose an RNA-radiomics (henceforth, radiogenomics) -NB model for glioma grading and prediction. Our radiogenomics-NB model is developed based on differentially expressed RNAseq and selected radiomics/volumetric features which characterize the tumor volume and its sub-regions. The NB distribution is fitted to RNAseq counts data, and a log-linear regression model is assumed to link between the estimated NB mean and radiomics. Three radiogenomics-NB molecular mutation models (e.g., IDH mutation, 1p/19q codeletion, and ATRX mutation) are investigated. Additionally, we explore the gender-specific effect on the radiogenomics-NB models.

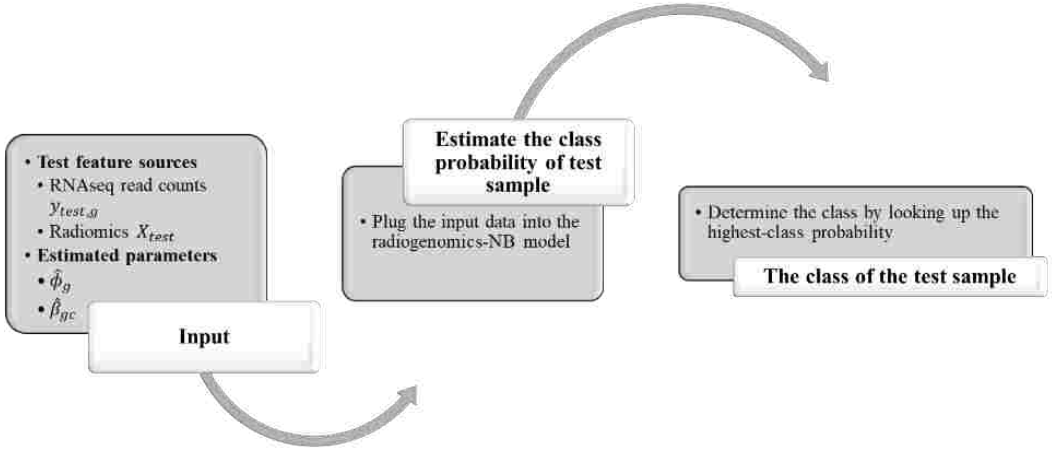
Finally, we compare the performance of the three mutation prediction radiogenomics-NB models with different well-known methods in the literature: Negative Binomial Linear Discriminant Analysis (NBLDA), differentially expressed RNAseq with Random Forest (RF-genomics), radiomics and differentially expressed RNAseq with Random Forest (RF-radiogenomics), and voom-based count transformation combined with the nearest shrinkage classifier (VoomNSC). Our analysis shows that the proposed radiogenomics-NB model significantly outperforms (ANOVA test, $p\text{-}value < 0.05$) for prediction of IDH and ATRX mutations and offers similar performance for prediction of 1p/19q codeletion, when compared to these competing models in the literature, respectively.

4.2 PREDICTION USING NEGATIVE BINOMIAL REGRESSION MODEL

In this study, we propose a radiogenomics-NB method for glioma molecular grading and prediction. Figure 9 illustrates an overall flow diagram of the proposed radiogenomics model. In Figure 9a, we fit NB distribution to RNAseq read counts data of the training dataset and estimate the model mean and dispersion parameter. Then, we use the estimated mean along with the predictor radiomics vector in a log-linear regression model to estimate the model regression coefficients. The dispersion parameter is estimated using the weighted likelihood empirical Bayes [115]. In **Figure 9**b, the estimated parameters of regression coefficients and the dispersion parameters along with the sample radiomics and its RNAseq read counts are utilized to predict the class label of a future test sample. A complete mathematical derivation of the radiogenomics-NB model is presented in the following subsection.



(a)



(b)

Figure 9. Overall Flow diagram of the proposed radiogenomics-NB prediction model.a) radiogenomics-NB model utilizing the training data. b) class prediction of a test sample using the developed radiogenomics-NB model.

4.2.1 Prediction using Negative Binomial regression model

To fuse radiomics with RNAseq read counts data in an NB model, the following parametrization is defined:

Let C be the total number of classes, and $I_c \in (1, \ldots, n_c)$ the indices of samples in class c for $c = 1, \ldots, C$. The examples of different classes include:

- *IDH* mutated vs wildtype *IDH* ($C = 2$),
- *1p/19q codeletion*: codeletion vs *non-codeletion* (in this case $C = 2$),

- Mutated *ATRX* vs wildtype (in this case $C = 2$).

let $Y_i = (y_{i1}, y_{i2}, \ldots, y_{iG})$ be the RNAseq read counts training sample in the class label c and G is the total number of RNAseq. The purpose of this study is to predict the class label c of a future observation Y_t using training samples associated with known class labels: $p(c|Y_t) \propto p(Y_t|c)p_c$, where p_c is the probability of class c. Using Bayes' rule, we have,

$$p(c|Y_i) \propto p(Y_i|c)p_c; \tag{2}$$

where $p(Y_i|c)$ is the pdf of the sample Y_i in class c, and p_c is the prior probability that one sample comes from class c. The pdf of class-specific c of RNAseq read counts of sample Y_i and gene g is

$$P(Y_{ig} = y_{ig}|Z = c) = \frac{\Gamma(\phi_g^{-1} + y_{ig})}{\Gamma(\phi_g^{-1}) y_{ig}!} \left(\frac{\phi_g \mu_{igc}}{1 + \phi_g \mu_{igc}}\right)^{y_{ig}} \left(\frac{1}{1 + \phi_g \mu_{igc}}\right)^{\phi_g^{-1}}. \tag{3}$$

In this parameterization, Y_{ig} represents a count response of RNAseq, where μ_{igc} represents the mean, ϕ_g represents the dispersion parameter, $E(Y_{ig}) = \mu_{igc}$, and $Var(Y_{ig}) = \mu_{igc} + \mu_{igc}\phi_g^2$. Note we assume that all RNAseq are independent of each other, so we have,

$$p(Y_i|Z = c) = \prod_{g=1}^{G} P(Y_{ig} = y_{ig}). \tag{4}$$

Evaluating equation 2 requires an estimation of $p(Y_i|c)$ and p_c. The model in equation 3 states that $Y_{ig} \sim NB(\mu_{igc}, \phi_g)$. We first estimate $\phi_1, \phi_2, \ldots, \phi_G$ and $\mu_{i1c}, \mu_{i2c}, \ldots, \mu_{iGc}$ of all the training samples n_c, and all RNAseq G. The mean is estimated as $\mu_{igc} = s_{ic}\lambda_{gc}$, where s_{ic} is the size factor [139, 140] which is used to scale RNAseq counts for the ith sample (in class c), λ_{gc} is the total number of reads of RNAseq g across all samples in class c. Regarding the prior p_c, we assume all classes are equally likely $p_c = 1/C$. Note that μ_{igc}, s_{ic}, and λ_{gc} are estimated for each class c.

Next, plugging these estimates into equation 3 and using the assumption of independent RNAseq, equation 2 yields,

$$log(p(c|Y_i)) = log(p(Y_i|c) + log(p_c). \tag{5}$$

The log-likelihood $log(p(Y_i|c))$ is written as,

$$log(p(Y_i|c)) = log(\prod_{g=1}^{G} P(Y_{ig} = y_{ig}|c)) = log(\prod_{g=1}^{G} \frac{\Gamma(\phi_g^{-1}+y_{ig})}{\Gamma(\phi_g^{-1})y_{ig!}} \times (\frac{\phi_g \mu_{igc}}{1+\phi_g \mu_{igc}})^{y_{ig}} \times (\frac{1}{1+\phi_g \mu_{igc}})^{\phi_g^{-1}}). \quad (6)$$

Equation 6 can be written as,

$$log(p(Y_i|c)) = \sum_{g=1}^{G} log(\frac{\phi_g \mu_{igc}}{1+\phi_g \mu_{igc}})^{y_{ig}} + \sum_{g=1}^{G} log(\frac{1}{1+\phi_g \mu_{igc}})^{\phi_g^{-1}} + \sum_{g=1}^{G} log(\frac{\Gamma(\phi_g^{-1}+y_{ig})}{\Gamma(\phi_g^{-1})y_{ig!}}). \quad (7)$$

Rewriting equation 7 yields,

$$log(p(Y_i|c)) = \sum_{g=1}^{G} y_{ig} log(\phi_g \mu_{igc}) - \sum_{g=1}^{G} y_{ig} log(1+\phi_g \mu_{igc}) - \sum_{g=1}^{G} \frac{1}{\phi_g} log(1+\phi_g \mu_{igc}) + \sum_{g=1}^{G} log(\frac{\Gamma(\phi_g^{-1}+y_{ig})}{\Gamma(\phi_g^{-1})y_{ig!}}). \quad (8)$$

The proposed NB model of genomics relates to the radiomics (imaging features) $\boldsymbol{X}$ through the mean parameters μ_{igc} (estimated mean of an ith sample and RNAseq g in class *c*). We assume a log-linear regression model for estimated the mean μ_{igc} in terms of the radiomics (imaging features) as follows:

$$log(\mu_{igc}) = X_i \beta_{gc}; \quad (9.a)$$

$$log(s_{ic}\lambda_{gc}) = X_i \beta_{gc}; \quad (9.b)$$

where X_i is a p-dimensional of radiomics, β_{gc} is a p-dimensional vector of unknown regression coefficients (translate the relationship between X and Y through μ_{igc}). The estimation of β_{gc} depends on class c and gene g of the ith sample. Hence, if there are two classes, we will need to estimate $\beta_{g1}$and β_{g2} (one from each class). Plugging equation 9.a into equation 8, yields,

$$log(p(Y_i|c)) = \sum_{g=1}^{G} y_{ig} log(\phi_g exp(X_i\beta_{gc} + \varepsilon)) - \sum_{g=1}^{G} y_{ig} log(1 + \phi_g exp(X_i\beta_{gc} + \varepsilon)) - \sum_{g=1}^{G} \frac{1}{\phi_g} log(1 + \phi_g exp(X_i\beta_{gc} + \varepsilon)) + \sum_{g=1}^{G} log(\frac{\Gamma(\phi_g^{-1} + y_{ig})}{\Gamma(\phi_g^{-1}) y_{ig}!}). \quad (10)$$

Using the estimated $\hat{\beta}_{gc}$, and $\hat{\phi}_g$ from the *training* data, if we classify a *test* observation Y_t as follows,

$$log(p(Z = c|Y_t)) = log(p(Y_t|C = c) + log(p_c) \quad (11)$$

and,

$$log(p(c|Y_t)) = \sum_{g=1}^{G} y_{tg} log(\hat{\phi}_g exp(X_t\hat{\beta}_{gc})) - \sum_{g=1}^{G} y_{tg} log(1 + \hat{\phi}_g exp(X_t\hat{\beta}_{gc})) - \sum_{g=1}^{G} \frac{1}{\phi_g} log(1 + \hat{\phi}_g exp(X_t\hat{\beta}_{gc})) + \sum_{g=1}^{G} log(\frac{\Gamma(\hat{\phi}_g^{-1} + y_{tg})}{\Gamma(\hat{\phi}_g^{-1}) y_{tg}!}) + log(p_c). \quad (12)$$

4.2.2 ESTIMATING DISPERSION PARAMETER ϕ_g USING WEIGHTED LIKELIHOOD EMPIRICAL BAYES.

Various methods for estimating the dispersion parameter ϕ_g are proposed in the literature. EdgeR applies a weighted conditional log-likelihood method to estimate dispersion parameter [115]. The weighted conditional log-likelihood (WL) for $\phi_g$is defined as a weighted combination of the individual (per-gene) likelihood $l_g(\phi_g)$ and common $l_C(\phi_g)$ likelihood:

$$WL(\hat{\phi}_g) = l_g(\phi_g) + \alpha l_C(\phi_g); \quad (13)$$

where α is the weight of $l_C(\phi_g)$.

In EdgeR, $\hat{\phi}_g$ is assumed to be normally distributed with means ϕ_g and known variance τ^2, and has the following hierarchical model:

$$\hat{\phi}_g|\phi_g \sim N(\phi_g, \tau^2), and\ \phi_g \sim N(\phi_0, \tau_0^2). \quad (14)$$

Under this hierarchical normal model, the maximum weighted conditional log-likelihood estimator is given as:

$$\hat{\phi}_g^{\,WL} = \frac{\hat{\phi}_g/\tau^2 + \alpha\sum_{i=1}^{G}\hat{\phi}_i/\tau_i^2}{1/\tau^2 + \alpha\sum_{i=1}^{G}1/\tau_i^2}; \tag{15}$$

where,

$$1/\alpha = \sum_{i=1}^{G}\tau_0^2/\tau_i^2, \tag{16}$$

and,

$$\phi_0 = \hat{\phi}_0 = \frac{\sum_{i=1}^{G}\hat{\phi}_i/\tau_i^2}{\sum_{i=1}^{G}1/\tau_i^2}. \tag{17}$$

4.2.3 COMPUTATION OF THE MEAN OF RNASEQ μ_{igc}

The size factor s_{ic} of sample i and class c is the total number of RNAseq read counts of that sample divided by the total number of all RNAseq read counts across all training samples (in class c). The size factor estimation is vital to account for the different sequencing depth (library size) that may be used to sequence different samples.

The mean μ_{igc} of sample i and RNAseq g in class c is then estimated as $\mu_{igc} = s_{ic}\lambda_{gc}$, where λ_{gc} is the total number of reads per RNAseq in class c. Using the estimated value of μ_{igc}, the values of β_{gc} are computed using equation 9.a. The algorithm in Figure 10 illustrates the steps of estimating the different parameters in the radiogenomics Negative Binomial classification model.

Input: RNA, imaging, and clinical data of Training data and Testing data
Using Training data, do

1. Prepare genetic data if data needs preprocessing such as filtering and differential gene expression
2. Compute the mean of RNAseq of training samples $\mu_{igc} = s_{ic}\lambda_{gc}$ of all classes C
3. Estimate the dispersion parameters of all RNAseq $\hat{\phi}_g$ using EdgeR (i.e., Equation 15)
4. Estimate $\hat{\beta}_{gc}$ by fitting the log-linear model as in Equation 9.a

Using Testing data, do

a. Estimate the class probability of a test sample by plugging the estimated parameters of $\hat{\beta}_{gc}$ and $\hat{\phi}_g$ in steps 3, and 4 into Equation 12
b. Determine the class by looking up the highest-class probability.

Output: the class of a test sample

Figure 10. Algorithm of prediction using radiogenomics Negative Binomial classification model.

4.3 EXPERIMENTAL RESULTS

4.3.1 DATASET

The dataset in this study consists of 108 pre-operative LGG patients that are described in [123, 124, 141]. Four sequences of the MRI are provided with the dataset: pre-contrast T1-weighted (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR). These scans are skull-stripped, re-sampled to 1 mm^3 resolution, and co-registered to the T1 template. The dataset provides the segmented sub-regions of the LGG: Gadolinium enhanced tumor (ET), the peritumoral edema (ED), and the necrosis along with non-contrast enhancing tumor (NCR/NET). RNAseq read counts data (with a total number of 56830 RNAseq), molecular alterations (*IDH* mutation, *1p/19q codeletion*, and *ATRX*), grade (II and III), and the clinical dataset is downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). The clinical dataset is de-identified by the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The distribution of the data is as follows: (i) *IDH* mutation: 85 Mutant and 23 wildtype (WT), (ii) *1p/19q codeletion*: 27 *codeletion* and 81 *non-*

codeletion, and (iii) *ATRX* status: 43 Mutant and 65 WT. The range of the patients' age at diagnosis is 20 – 75 years, and the median age is 46.5 years.

4.3.2 DATA PREPARATION

In this study of glioma molecular grading and prediction, for the dataset of 108 LGG cases, we obtain two types of features (i.e. radiomics and RNAseq read counts data) and 3 molecular mutation types. The three molecular types are (i) *IDH* mutation: 85 Mutant and 23 wildtype (WT), (ii) *1p/19q codeletion*: 27 codeletion and 81 *non-codeletion*, and (iii) *ATRX* mutation: 43 Mutant and 65 WT.

We first filter RNAseq read counts to remove RNAseq with a very low value of read counts before performing any statistical analysis. RNAseq with very low read counts holds very little information because an RNAseq of biological importance needs to be expressed at some minimal level. We utilize a quantile filter [142] with a quantile threshold of 0.25. This step returns each RNAseq that has a mean across all samples higher than the defined quantile threshold of 0.25. Then, we reduce the number of RNAseq that are used in the radiogenomics-NB models by utilizing EdgeR [116] to extract the differentially expressed RNAseq (DERs). DERs reflect the significance of a gene in a certain biological condition. In this study, we select the top 10, 20, 30, 50, 100, and 150 DERs.

Furthermore, we use eight volumetric radiomics features as illustrated in Table 4. ANOVA analysis for radiomics in Table 4 shows that feature numbers 1, 3, 4, 7, and 8 are significantly associated (ANOVA test, $p\text{-value} < 0.05$) with *IDH* mutations. Our analysis also indicates that feature number 2 is marginally associated (ANOVA test, $p\text{-value} = 0.07$) with *1p/19q codeletion*. Furthermore, our analysis indicates that feature numbers 3, 4, 6, 7, and 8 are significantly associated (ANOVA test, $p\text{-value} < 0.05$) with *ATRX* mutations. Few other studies suggest that these volumetric imaging features and their ratios are associated with and predictive of several mutations in Gliomas [143-146].

Table 4. Radiomics features description and their ANOVA *p-value* association with *IDH* mutations, *1p/19q codeletion*, and *ATRX* mutations.

Feature Number	**Feature Description**	***p-value* of *IDH* mutation**	***p-value* of *1p/19q codeletion***	***p-value* of *ATRX* mutation**
1	the size of the enhanced tumor to the necrosis size	0.0033	0. 393	0.178
2	the size of the enhanced tumor to the size of enhanced tumor and necrosis	0.8630	0.070	0.239
3	the size of the enhanced tumor to the edema size	<0.005	0.600	0.002
4	the size of the enhanced tumor to the whole tumor size,	<0.005	0.707	0.027
5	the size of the edema to the necrosis size,	0.188	0.996	0.114
6	the size of the edema to the size of enhanced tumor and necrosis	0.138	0.789	0.0237
7	the size of the edema to the whole tumor size	<0.005	0.131	<0.005
8	and the size of the necrosis to the whole tumor size	<0.005	0.221	<0.005

The 108 LGG cases are randomly split into 80% training and 20% testing sets, and a balanced distribution of the target molecular alteration is ensured in the training and testing sets in each molecular classifier. The trained model classifier is developed using the training set. Model performance prediction is estimated and reported using the testing sets in terms of accuracy, balanced accuracy, F1 score, sensitivity, specificity, negative predictive value, and positive predictive value. The training set is utilized to build our radiogenomics-NB classifier as shown in steps 1-4 in Figure 10. The testing set is used to estimate the performance of the classifier as shown in steps a and b in Figure 10. Authors in [119, 147-149] repeat training and testing analysis for a specific number of times to ensure the robustness of the model performance. Consequently, in this work, we repeat the whole procedure 100 times independently for the 3 molecular alterations and then report the mean and standard deviation of the classifiers' performance using the testing sets.

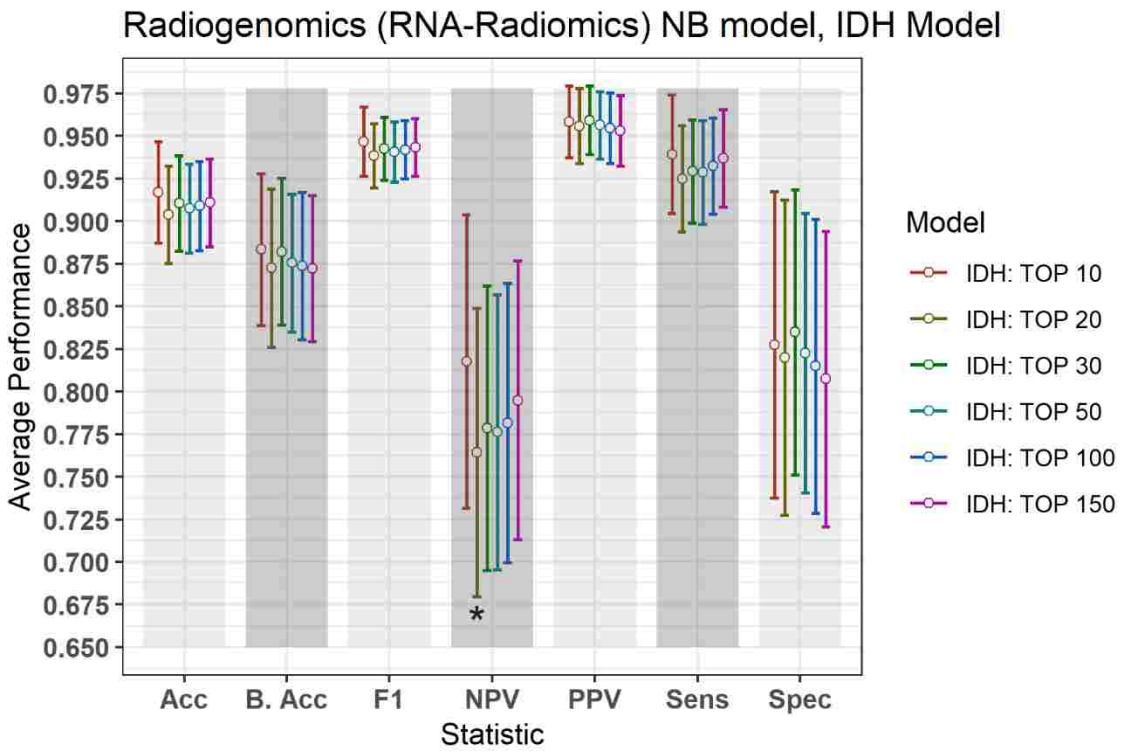
4.4 RADIOGENOMICS-NB MODELS USING DIFFERENT NUMBERS OF DIFFERENTIALLY EXPRESSED RNAS

In this section, we investigate the importance of using different numbers of DERs on the performance of the radiogenomics-NB model. LGG radiogenomics-NB mutation prediction models are developed based on the top 10, 20, 30, 50, 100, and 150 DERs.

The performance of the radiogenomics-NB *IDH* model using the top 10 DERs achieves slightly high performance. However, such improvement is not statistically significant (ANOVA test, *p-value* > 0.05) when compared to the performance of the *IDH* models with the other number of DERs (Figure 11.a) except for NPV performance when using the top 20 DERs. Using the top 20 DERs in the *IDH* model achieves significantly (ANOVA test, *p-value* < 0.05) worse NPV when compared to the NPV achieved using the top 10 DERs. Radiogenomics-NB *IDH* model with the top 10 DERs (red line in Figure 11.a) achieves an overall accuracy (Acc) of 0.92±0.06, sensitivity (Sens) of 0.94±0.07, specificity (Spec) of 0.83±0.18, positive predictive value (PPV) of 0.96±0.04, negative predictive value (NPV) of 0.82±0.17, F1 score of 0.95±0.04, and balanced accuracy (B. Acc) of 0.88±0.09, respectively.

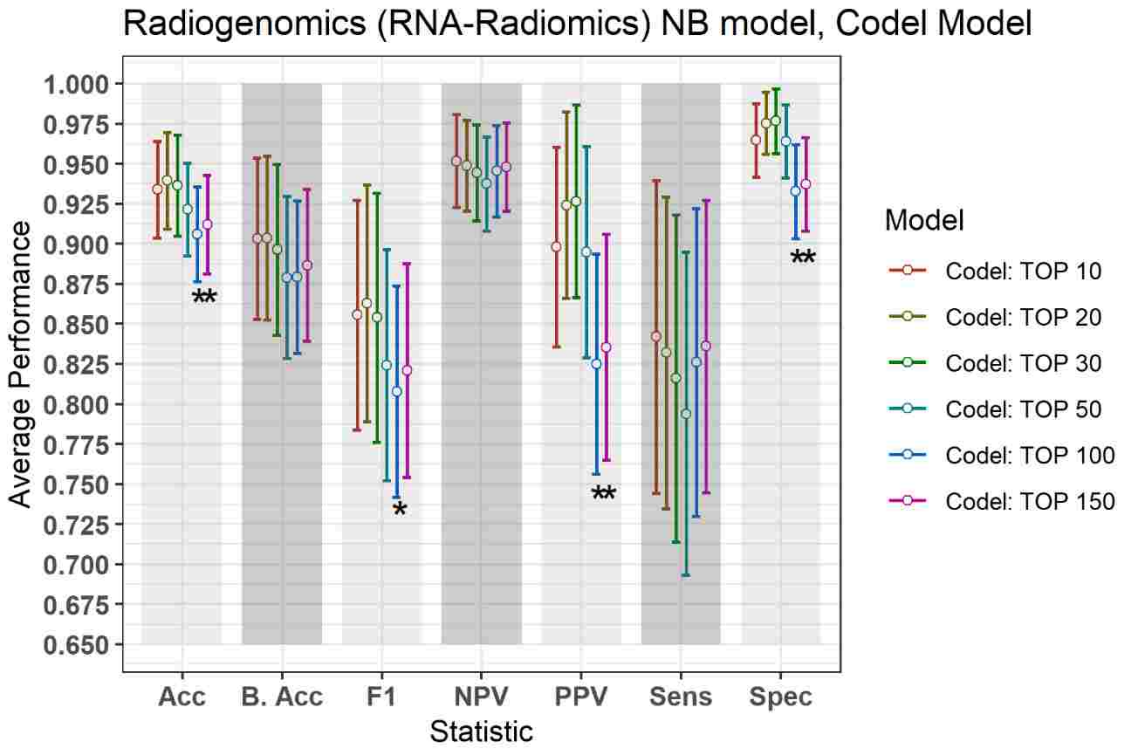
Radiogenomics-NB *codeletion* models significantly achieve similar performance (ANOVA test, *p-value* > 0.05) using the top 10, 20, 30, and 50 DERs as shown in Figure 11.b. Furthermore, using the top 100 and 150 DERs in the *codeletion* model achieves worse performance (ANOVA test, *p-value* > 0.05) when compared to the performance of using the top 10 DERs. Using the top 10 DERs, the radiogenomics-NB *codeletion* model achieves an accuracy of 0.93±0.06, a balanced accuracy of 0.90±0.10, F1 score of 0.86±0.14, a sensitivity of 0.84±0.19, a specificity of 0.96±0.04, an NPV of 0.95±0.06, and a PPV of 0.90±0.12, respectively.

The radiogenomics-NB *ATRX* model achieves similar performance (ANOVA test, *p-value* > 0.05) using the top 10, 20, and 30 DERs, even though the performance when using the top 10 DERs is slightly better as illustrated in Figure 11.c. Using the top 10 DERs, the *ATRX* model achieves an accuracy of 0.85±0.07, a balanced accuracy of 0.85±0.07, an F1 score of 0.82±0.08, a sensitivity of 0.86±0.13, a specificity of 0.85±0.09, an NPV of 0.91±0.08, and a PPV of 0.80±0.10, respectively. Table 5 illustrates the confusion matrix of the test performance of the radiogenomics-NB *IDH*, *Codeletion* and *ATRX* models using the top 10 DERs over the 100 repetitions.

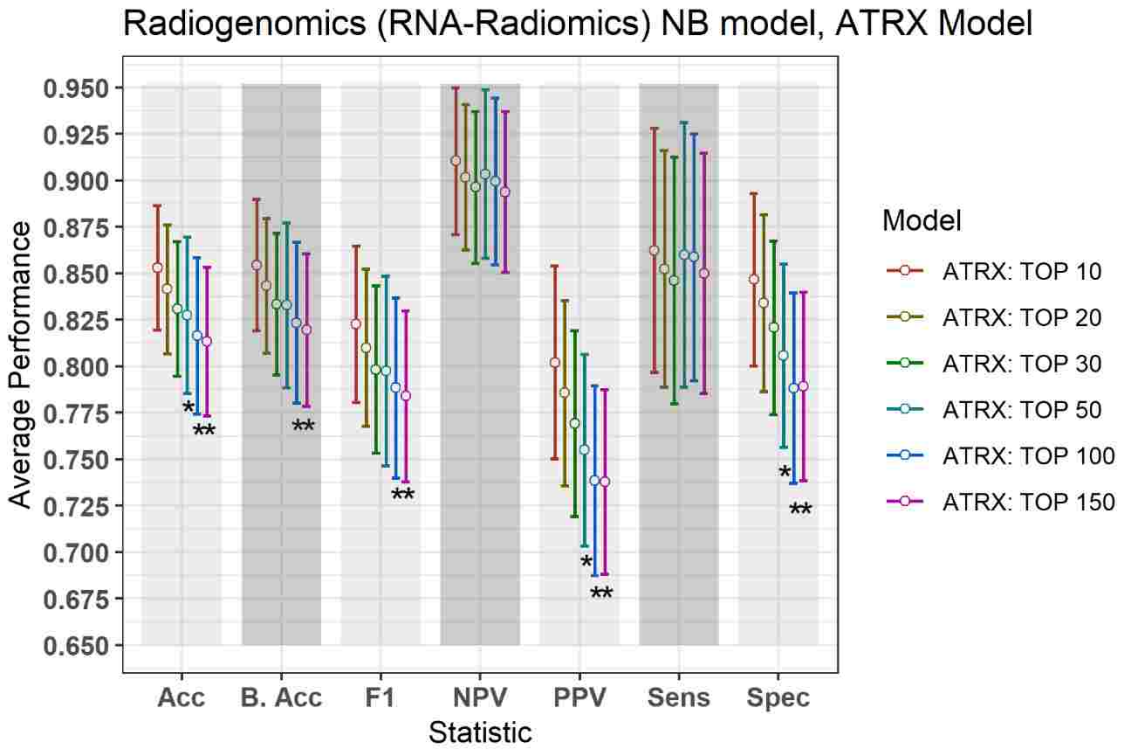


(a)

Figure 11. Performance of the proposed radiogenomics-NB model using a different number of DERs. a) Radiogenomics-NB *IDH*, b) Radiogenomics-NB *Codeletion*, and c) Radiogenomics-NB *ATRX* models. The average performance (of the Acc, B. Acc, F1, NPV, PPV, Sens, and Spec) is computed across 100 testing sets/splits. Y-axis represents the average performance of the different statistics on the X-axis. Different colors represent the radiogenomics-NB model with different numbers of DERs. The error bar represents one standard deviation. Asterisk "*" represents a statically significant difference between the performance achieved when using the top 10 DERs (in red) and using the number of DER where the star is located.



(b)



(c)

Figure 11 (continued)

Table 5. Confusion matrix of test performance over 100 repetitions of radiogenomics-NB *IDH*, *Codeletion*, and *ATRX* models using the top 10 DERs.

Radiogenomics-NB *IDH* model			**Radiogenomics-NB *Codeletion* model**			**Radiogenomics-NB *ATRX* model**		
	References			**References**			**References**	
Predictions	**Mutant**	**WT**	**Predictions**	**Codel.**	**Non-codel**	**Predictions**	**Mutant**	**WT**
Mutant	1503	69	**Codel.**	421	53	**Mutant**	690	184
WT	97	331	**Non-codel**	79	1447	**WT**	110	1016

4.5 COMPARATIVE ANALYSIS

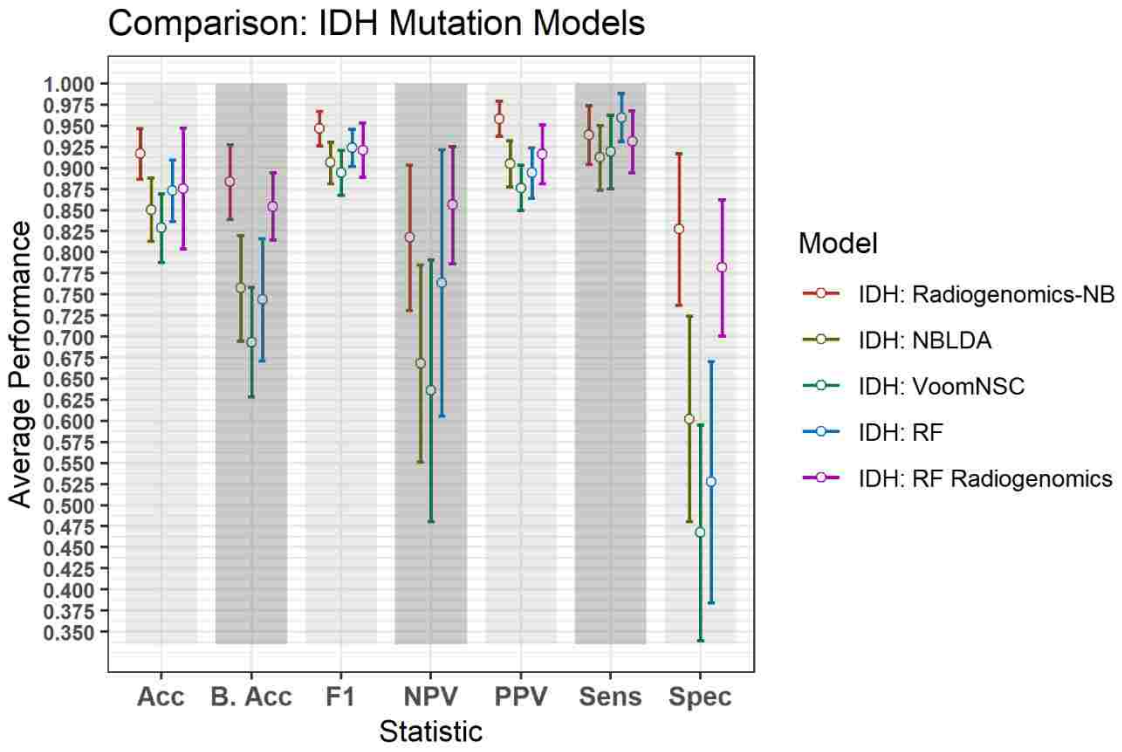
Figure 12 illustrates a graphical performance comparison between our radiogenomics-NB model with that of four different classifiers in the literature: NBLDA [119], VoomNSC [65, 66], RF-genomics where we first log-transformed [116] the RNAseq into a normal distribution, and RF-radiogenomics. Note that the number of DERs that we apply to develop these classifiers is 10 DERs. Moreover, when developing these classifiers, the 108 LGG cases are randomly split into 80% training and 20% testing sets, and balanced distribution is ensured when developing the different classifiers. The trained model classifier is developed using the training set, and 10-fold cross-validation is performed to identify the tuning parameters in the different classifiers. Model performance prediction is estimated and reported using the testing sets. Additionally, to ensure the robustness of the different classifiers' performance, we repeat the whole procedure 100 times independently and every training/testing set is utilized to develop and estimate the performance of each classifier.

The NBLDA [119] classifier is developed by fitting NB to the top 10 DERs; then the mean and dispersion parameter are estimated from these DERs. In RF-genomics, the top 10 DERs of the training sets are first log-transformed into normal distribution and then fed into RF to build the RF-genomics classifier. In RF-radiogenomics, radiomics (eight volumetric features described previously in Table 4) are utilized with the log-transformed DERs and then fed into RF to build the RF-radiogenomics classifier. VoomNSC [65, 121] is developed by first applying the voom-based transformation on the 10 DERs and then applying the NSC classifier as illustrated in [65, 121].

Comparing the performance of our radiogenomics-NB *IDH* model with that of NBLDA, RF-genomics, and VoomNSC, the radiogenomics-NB *IDH* significantly outperforms (ANOVA test, *p-value* < 0.05) these methods as shown in Figure 12.a and Table 6. Additionally, our radiogenomics-NB *IDH* model significantly outperforms (ANOVA test, *p-value* < 0.05) the F1 score, balanced accuracy, and PPV performance of the RF-radiogenomics method whereas it achieves a similar (ANOVA test, *p-value* > 0.05) accuracy, sensitivity, and specificity. Our radiogenomics-NB *IDH* model archives an accuracy of 0.92±0.06, a sensitivity of 0.94±0.07, a specificity of 0.93±0.18, an F1 score of 0.95±0.04, and a balanced accuracy of 0.88±0.09, respectively. The RF-radiogenomics-*IDH* model achieves an accuracy of 0.88±0.17, a sensitivity of 0.93±0.07, a specificity of 0.78±0.16, an F1 score of 0.92±0.06, and a balanced accuracy of 0.85±0.08, respectively.

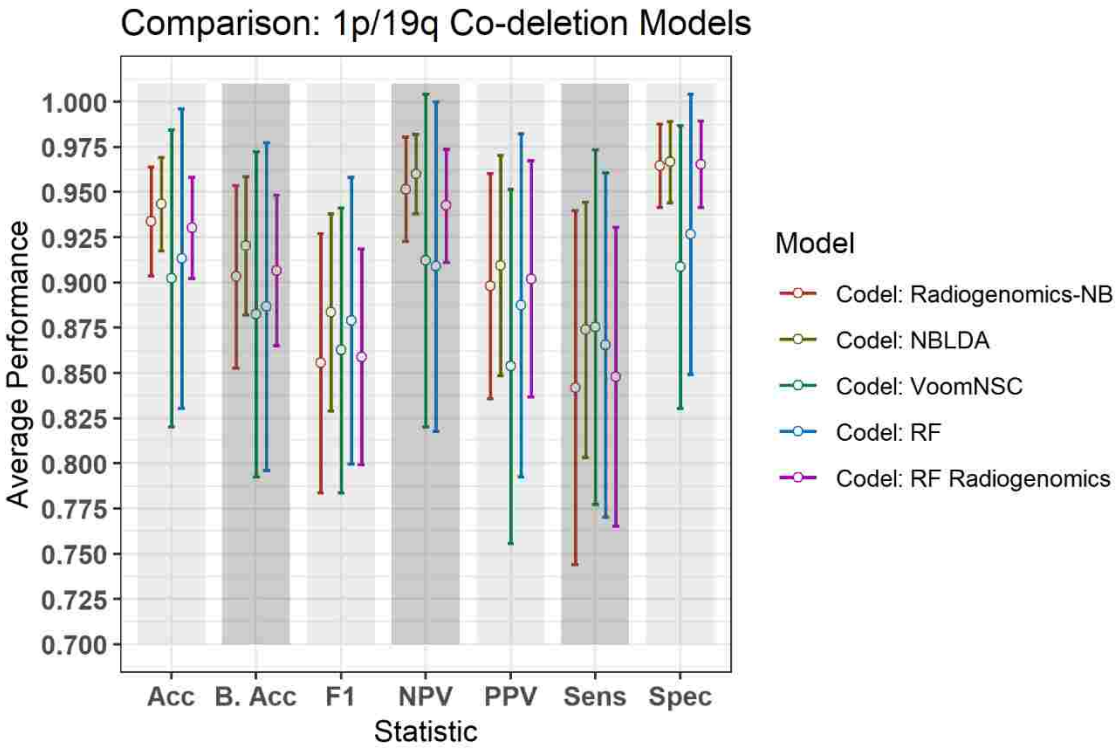
Our radiogenomics-NB *codeletion* model (Figure 12.b and Table 7) performance is similar to NBLDA, RF-genomics, VoomNSC, and RF-radiogenomics models, except for the specificity and NPV performance when using RF-genomics and VoomNSC. The specificity and NPV of our model are significantly higher than those achieved by RF-genomics and VoomNSC. Our radiogenomics-NB *codeletion* model achieves an accuracy of 0.93±0.06, a sensitivity of 0.84±0.20, a specificity of 0.96±0.5, an F1 score of 0.86±0.14, and a balanced accuracy of 0.90±0.10, respectively.

The performance of our radiogenomics-NB *ATRX* model as shown in Figure 12.c and Table 8 outperforms both NBLDA and VoomNSC significantly (ANOVA test, *p-value* < 0.05). However, comparing our *ATRX* model to RF-genomics, our model achieves significantly better balanced-accuracy, F1 score, NPV, and sensitivity. Additionally, comparing our *ATRX* model to RF-radiogenomics, our model achieves significantly (ANOVA test, *p-value* < 0.05) better sensitivity but achieves similar accuracy, balanced-accuracy, F1 score, and sensitivity. Our radiogenomics-NB *ATRX* model achieves an accuracy of 0.85±0.07, a sensitivity of 0.86±0.13, a specificity of 0.85±0.09, an F1 score of 0.82±0.08, and a balanced accuracy of 0.85±0.07, respectively. The RF-radiogenomics *ATRX* model achieves an accuracy of 0.84±0.08, a sensitivity of 0.80±0.14, a specificity of 0.86±0.10, an F1 score of 0.80±0.09, and a balanced accuracy of 0.83±0.08, respectively.

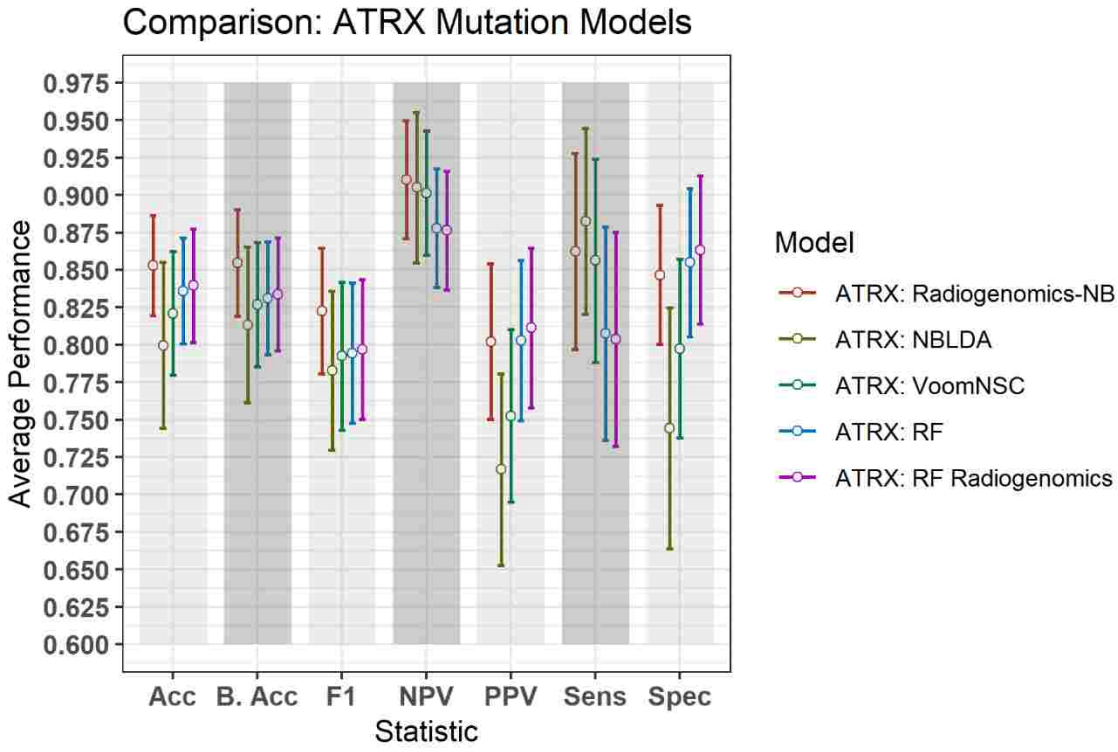


(a)

Figure 12. Comparison of performance between our radiogenomics-NB model and different classifiers. The comparison is performed using the a) *IDH* mutation, b) *1p/19q codeletion*, and c) *ATRX* mutations dataset. The average performance (of the Acc, B. Acc, F1, NPV, PPV, Sens, and Spec) is computed across 100 test sets. The error bar represents one standard deviation. RNAseq that are used in developing all classifiers represent the top 10 DERs in the training sets between mutated and WT *IDH* group, codeleted and non-codeleted groups, and mutated and WT *ATRX* mutation, respectively. Y-axis represents the average performance of the different statistics on the X-axis. Different colors represent different classifiers.



(b)



(c)

Figure 12 (continued)

Table 6. Probability of significant difference using ANOVA test between the differentially expressed radiogenomics-NB model and different classifiers using IDH dataset. A statistically significant difference exists if *p-value* < 0.05. Values in bold show a significant improvement of our radiogenomics-NB IDH over the compared one.

IDH	**Accuracy**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**F1**	**Balanced Accuracy**
radiogenomics-NB VS NBLDA	**0.000**	**0.010**	**0**	**0**	**0.000**	**0.000**	**0**
radiogenomics-NB VS VoomNSC	**0.000**	0.075	**0**	**0**	**0.000**	**0.000**	**0**
radiogenomics-NB VS RF	**0.001**	0.023	**0**	**0**	0138	**0.000**	**0**
radiogenomics-NB VS RF-radiogenomics	0.069	0.432	0.061	**0**	0.084	**0.001**	**0.01**

Table 7. Probability of significant difference using ANOVA test between the differentially expressed radiogenomics-NB model and different models using *1p/19q codeletion* dataset. A statistically significant difference exists if *p-value* < 0.05. Values in bold show a significant improvement of our radiogenomics-NB codeletion over the compared one.

CODEL	**Accuracy**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**F1**	**Balanced Accuracy**
radiogenomics-NB VS NBLDA	0.232	0.186	0.756	0.514	0.253	0.123	0.181
radiogenomics-NB VS VoomNSC	0.072	0.228	0.001	0.057	0.042	0.742	0.317
radiogenomics-NB VS RF	0.242	0.390	0.020	0.636	0.027	0.271	0.42
radiogenomics-NB VS RF-radiogenomics	0.671	0.815	0.893	0.825	0.282	0.855	0.792

Table 8. Probability of significant difference using ANOVA test between the differentially expressed radiogenomics-NB model and different models using ATRX dataset. A statistically significant difference exists if *p-value* < 0.05. Values in bold show a significant improvement of our radiogenomics-NB ATRX over the compared one.

ATRX	**Accuracy**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**F1**	**Balanced Accuracy**
radiogenomics-NB VS NBLDA	**0.000**	0.269	**0.000**	**0.000**	0.677	**0.004**	**0.001**
radiogenomics-NB VS VoomNSC	**0.003**	0.741	**0.001**	**0.002**	0.432	**0.021**	**0.012**
radiogenomics-NB VS RF	0.083	**0.005**	0.540	0.960	**0.004**	**0.026**	**0.025**
radiogenomics-NB VS RF-radiogenomics	0.183	**0.003**	0.215	0.561	**0.003**	0.052	0.053

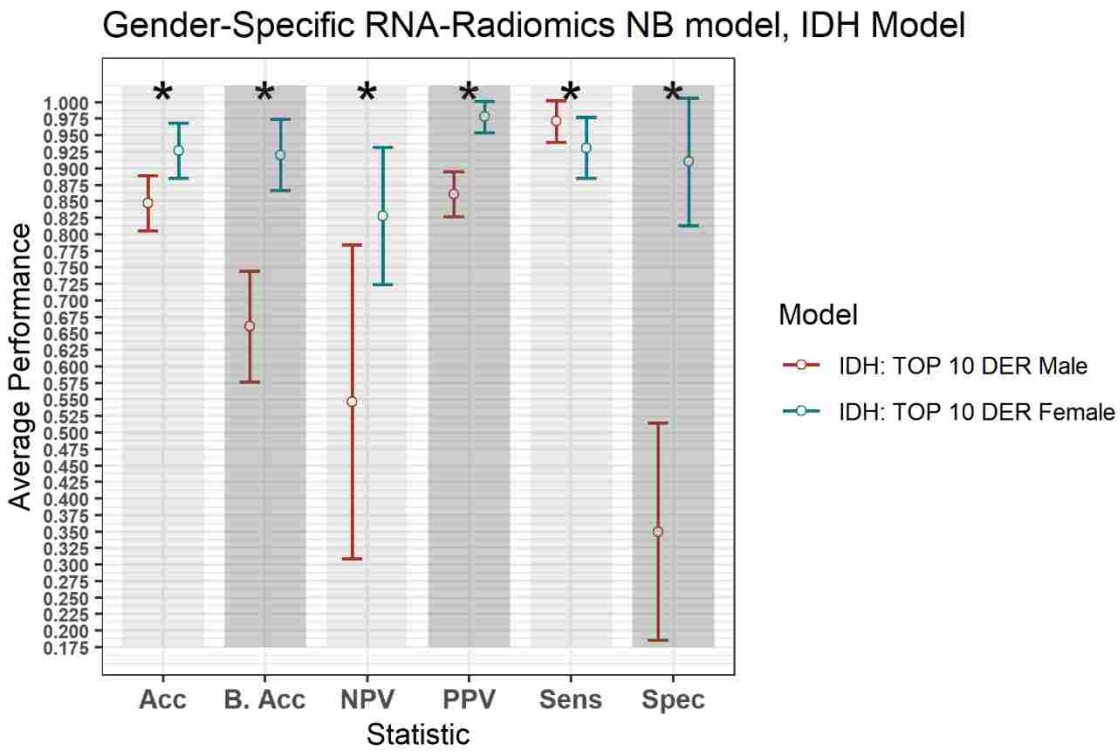
4.6 GENDER-SPECIFIC EFFECT ANALYSIS OF RADIOGENOMICS-NB

In our LGG dataset, *IDH* mutated patients, unlike *IDH* WT patients, have significantly longer survival (65.7 vs 19.9 months, log-rank test *p-value* = 0.004). The association between *IDH* status and overall survival remains significant after stratifying for gender (likelihood ratio test *p-value* = 0.015). However, the association between 1p/19q codeletion and *ATRX* status and overall survival is not significant. Additionally, the chi-square test shows no significant association (*p-value* > 0.05) between gender and *IDH* status, 1p/19q codeletion, and *ATRX* status. Table 9 shows patient *IDH* status, 1p/19q codeletion, and *ATRX* status distribution based on gender.

To explore the gender-specific effect in the performance of the radiogenomics-NB, we build two radiogenomics-NB models based on gender; male-specific radiogenomics-NB and female-specific radiogenomics-NB. Our analysis indicates that female-specific models outperform significantly (ANOVA test, *p-value* < 0.05) male-specific models as illustrated in Figure 13. Female-specific models achieve an accuracy of 0.93±0.08 and a balanced accuracy of 0.92±0.11 in the radiogenomics-NB *IDH*, an accuracy of 0.91±0.09 and a balanced accuracy of 0.84±0.17 in the radiogenomics-NB *codeletion*, and an accuracy of 0.80±0.11 and a balanced accuracy of 0.80±0.12 in the radiogenomics-NB ATRX, respectively.

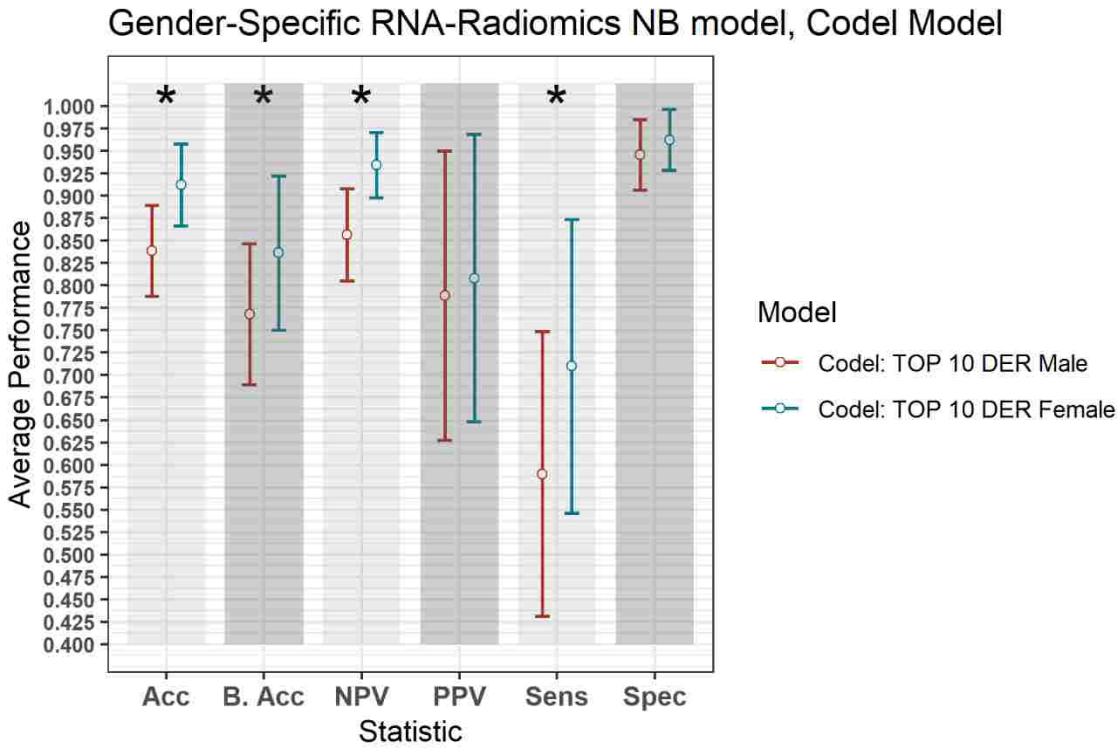
Table 9. Gender-based distribution of IDH status, 1p/19q codeletion, and ATRX status in the LGG dataset.

	IDH status		*1p/19q codeletion*		*ATRX* status	
	Mutant	**WT**	***Codeletion***	***Non-Codeletion***	**Mutant**	**WT**
Female	43	14	14	43	24	33
Male	42	9	13	38	19	32

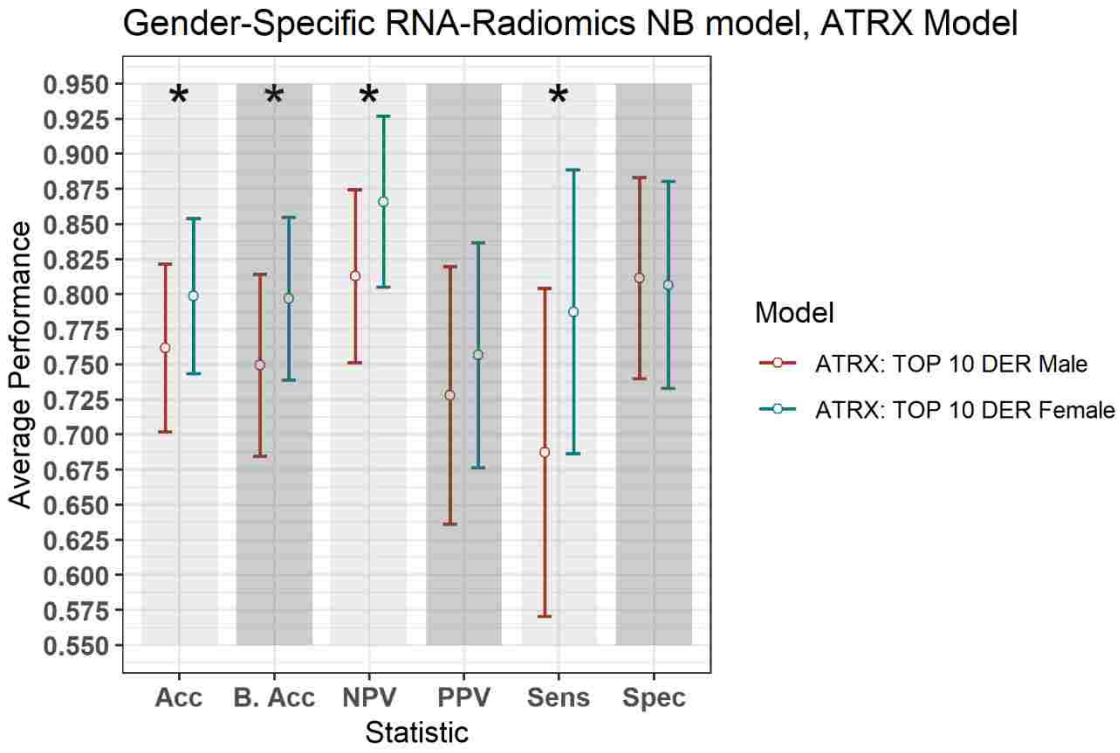


(a)

Figure 13. Gender-based radiogenomics-NB models performance of a) IDH mutations, b) 1p/19q codeletion, and c) ATRX mutations which is computed across 100 testing sets. The error bar represents one standard deviations. The asterisk * illustrates a significant difference between two measurements. Y-axis represents the average performance of the different statistics on the X-axis. Different colors represent the female- and male-specific radiogenomics-NB models.



(b)



(c)

Figure 13 (continued)

4.7 DISCUSSION

In this study, we propose a novel radiogenomics-NB model to fuse radiomics (imaging features) with RNAseq (genes) for glioma grading and prediction. NB distribution is appropriate for modeling RNAseq discrete read counts data and for preserving the count-based nature of this data. In the proposed radiogenomics-NB model, log-linear regression modeling is fitted to the estimated mean of the NB distribution and is linked with radiomics. We introduce this step to fuse the continuous radiomics data with the RNAseq count-based data without the need to transform the RNAseq data into a normal distribution.

NB, unlike Poisson distribution, has two parameters; the mean (e.g., the expected value of the RNAseq read counts data) and dispersion (e.g., a parameter that helps in capturing the variability of the RNAseq read counts). If the dispersion of NB is zero, the model reduces to Poisson distribution. In Poisson distribution, the mean is equal to the variance, which makes it rather restrictive. However, variation is usually observed in the real data of RNAseq counts data that the Poisson distribution cannot handle properly. On the other hand, NB has an additional parameter called the “dispersion” that allows the NB distribution of RNAseq counts data to modify its variance without affecting the mean. Thus, NB serves as a practical approximation to model RNAseq count data with variability different from its mean.

The mean of the proposed radiogenomics-NB model is estimated as the size factor multiplied by the total number of reads per RNAseq. Moreover, we utilize EdgeR to estimate the dispersion of the proposed radiogenomics-NB assuming RNAseq variability is assessed using the weighted conditional log-likelihood model. In the weighted conditional model, RNAseq counts data is assumed to have a distinct and individual dispersion for each RNAseq in addition to a common dispersion. Such an assumption can be more reliable when estimating the dispersion of real data of RNAseq counts data.

The performance evaluation of the proposed work indicates that linking radiomics (i.e. tumor volumetric features) to RNAseq improves the performance of *IDH* and *ATRX* mutations prediction. The radiomics features utilized in the proposed radiogenomics-NB model that are described in Table 1 mainly depend on volumetric features. Our analysis shows that these features are associated with glioma mutations. This outcome supports previous studies that show the association between volumetric features and glioma mutations [143-146].

The efficacy of the proposed radiogenomics-NB model is further investigated using the top 10, 20, 30, 50, 100, and 150 DERs, respectively as illustrated in Figure 12. Our analysis shows that the smaller the number of DERs (fewer than 30 DERs) utilized in radiogenomics-NB, the better the radiogenomics-NB model performance. Our analyses indicate that using fewer than 30 DERs in our analysis offers the best performance (statically significant) in the radiogenomics-NB *codeletion* and *ATRX* prediction model. This suggests that using large numbers of DERs (more than 30) in the proposed radiogenomics-NB would over parametrize the dataset and create model fitting problems and thus degrade the performance.

Comparing our radiogenomics-NB model to NBLDA, RF-genomics, FR-radiogenomics, and VoomNSC, our model significantly outperforms NBLDA, RF-genomics, and VoomNSC for prediction of *IDH* and *ATRX* mutations. Our radiogenomics-NB model offers similar performance as NBLDA, RF-genomics, RF-radiogenomics, and VoomNSC models for prediction of *1p/19q codeletion*. Specifically, for prediction of *IDH* mutations, while the proposed radiogenomics-NB model achieves significantly better balanced-accuracy, F1 score, and PPV than RF-radiogenomics, our model achieves similar accuracy, sensitivity, and specificity. Such results indicate the power of fusing radiomics and genomics data to develop radiogenomics models for classification and prediction models. The findings in this work indicate that the radiomics volumetric features may be vital for the prediction of *IDH* and *ATRX* mutations along with the genomics.

Different studies have revealed that gender is a significant factor in identifying cancer survival, prognosis, and treatment response [150-152]. Hence, improved glioma molecular mutation prediction may require the development of gender-specific models. In this study, we explore the gender-specific effect on the radiogenomics-NB models. Our analysis reveals that *IDH* mutated patients have significantly longer survival and remain significant after stratifying for gender, unlike *1p/19q codeletion* and *ATRX* status. Moreover, our analysis indicates that no association is found between gender and the three mutations (*IDH* mutations, *1p/19q codeletion*, and *ATRX* status) using the Chi-square test. This result is in agreement with the findings in [153-155]. However, a gender-specific radiogenomics-NB model shows that female-specific radiogenomics-NB models significantly outperform the male-specific radiogenomics-NB models for prediction of *IDH* status, *1p/19q codeletion*, and *ATRX* status, respectively.

There are several limitations to this study. First, the radiogenomics-NB model is proposed to employ the count-based nature of the RNAseq dataset. We utilize both the sensitivity (the

probability of test positive when the test is positive) and the specificity (the probability of test negative when the test is negative) to assess the predictivity of the model. However, using the measurements of sensitivity and specificity may be not adequate when the normality of the data cannot be obtained [156]. Second, we assume that all RNAseq counts dataset are independent of each other in the derivation of the radiogenomics-NB model. This assumption may not be always realistic, and a gene interaction may affect a certain glioma outcome.

In conclusion, we present a glioma mutations radiogenomics-NB prediction model that preserves the count nature of RNAseq counts data in the NB model and utilizes radiomics to develop a complete and a better characterization prediction model of patient data. Our analysis shows the superiority of utilizing both genomics and radiomics data when compared to only genomics models. Finally, this study shows the efficacy of volumetric radiomics features in the radiogenomics-NB model for glioma molecular grading and prediction.

CHAPTER 5

RADIOMICS-GUIDED PREDICTION OF OVERALL SURVIVAL IN GLIOBLASTOMA PATIENTS

5.1 CHAPTER OVERVIEW

Radiomics is defined as quantitative imaging features that are extracted from radiographic images (e.g., MRI). Radiomics have provided insights into personalized medicine and, thus, provided accurate survival prediction (i.e. risk stratification), tumor detection, subtype classification, and therapeutic response assessment of oncologic patients [157]. Additionally, different studies [30, 75, 88, 89, 158] study tumor heterogeneity using different types of imaging such as MRI. As a result, MRI is suggested as a potential non-invasive imaging biomarker for gliomas diagnostic, prognostic, and survival prediction. This chapter proposes a fully automated MRI-based glioblastoma survival prediction framework. The survival prediction framework includes two representative survival prediction pipelines that combine different feature selection and regression approaches. The framework is evaluated using two recent widely used benchmark datasets from Brain Tumor Segmentation (BraTS) global challenges in 2017 and 2018.

5.2 GLIOBLASTOMA SURVIVAL PERDITION USING RADIOMICS

In this section, two different survival prediction models are proposed for survival prediction. The first model is a tree-based method for overall-survival regression prediction using a three-step feature selection and Random forest (RF) regression model (Figure 14.a). The second survival prediction model (Figure 14.b) includes a prediction of survival risk classification (short, medium, and long-term survival). Subsequently, an overall survival regression is performed based on the survival risk class label. Both classification and regression models are trained on quantitative- radiomics features obtained from the segmented tumor. The recursive feature selection method is used to select the features that are used in the classification model. Finally, Cox regression is used as a feature selection method in the overall survival regression model. Three

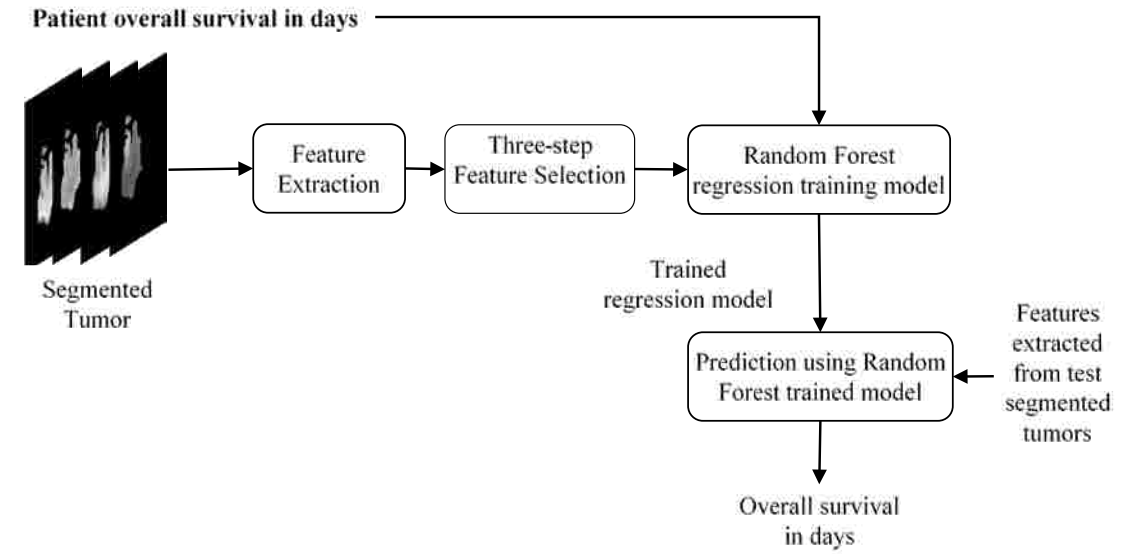
overall regression models are trained: long-regression model, mid-regression model, and short regression model.

5.3 METHODOLOGY

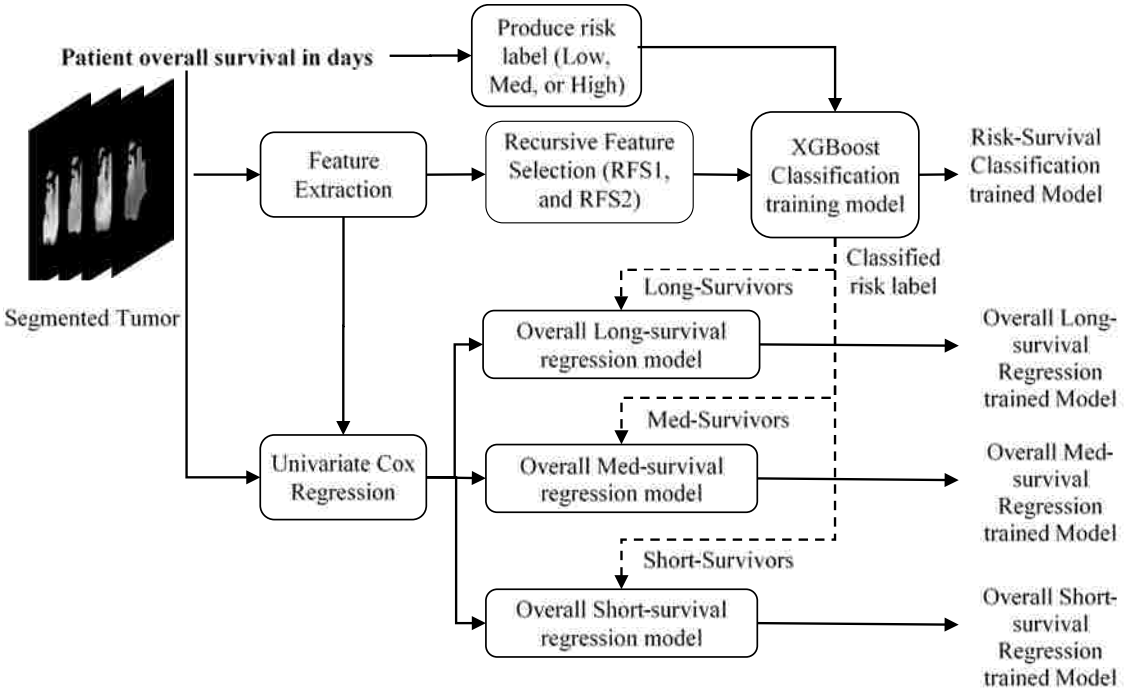
5.3.1 MATERIALS

This study uses BraTS18 training, validation, and testing dataset [123-125, 159], and BraTS17 training, validation, and testing datasets for patient survival prediction analysis. Both BraTS17 and BraTS18 datasets contain a total of 163 Glioblastoma (high-grade glioma (HGG)) cases for training, with overall survival, defined in days, and the age of the patient at diagnosis, defined in years. The training dataset provides 4 modalities (T1, post-contrast T1-weighted (T1Gd), T2-weighted (T2), T2 Fluid Attenuated Inversion Recovery (FLAIR)) along with the ground truth segmentation of multiple abnormal tissues (enhanced (EN), edema (ED), necrosis and non-enhanced (NCR/NET)) in the tumor. Overall survival risk is classified into three survival groups: long (greater than 15 months), medium (between 10 to 15 months), and short (less than 10 months).

In addition, for validation purposes, we use the validation datasets of BraTS17 and BraTS18. BraTS17 validation dataset consists of thirty-three cases while that for BraTS18 consists of twenty-eight cases for overall survival prediction purposes. BraTS17 testing dataset consists of ninety-five cases while that for BraTS18 offers seventy-seven cases for testing the overall survival prediction performance. A total of 135 pre-operative GBM BraTS18 training patients are provided by The Cancer Imaging Archive (http://www.cancerimagingarchive.net/). The overall survival (OS), age at diagnostic (median age, 59 years; range, 17 – 84 years), and other clinical data, molecular mutations, and genomics data are downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/).



(a)



(b)

Figure 14. Glioblastoma Survival Prediction Model Outline using Radiomics. a) The first survival prediction model (SP1) pipeline using RF regression classifier, and b) the second survival prediction model (SP2) pipeline using XGBoost.

5.3.2 FEATURE EXTRACTION

Feature extraction is the first step in the overall survival prediction task. Different quantitative imaging features (of around 31 thousand) are extracted from the different types of segmented abnormal tissues (edema, enhanced tumor, and tumor core). These features include texture, volumetric and area-related features, histogram-graph features, and Euler characteristics (vertices, edges, and faces). The heterogeneity in Glioblastoma may be quantified using texture and histogram-graph features while the shape of the tumor may be effectively captured using volumetric and Euler characteristic features [74, 76, 160].

TEXTURE FEATURES

A detailed breakdown of the extracted features is as follows: a total of 1107 texture features [127] are computed from the raw MRI sequences and different texture representation of the tumor volume. The texture representations of the raw MRI are 1) Texton filters [93] 2) texture-fractal characterization using piecewise-triangular prism surface area (PTPSA) [96] modeling, 3) fractal characterization of multi-fractional Brownian motion (mBm) [97] modeling, 4) and the Holder Exponent [129] modeling. The extracted texture imaging features are described by 1) the histogram, 2) the co-occurrence matrix, 3) the neighborhood gray-tone difference matrix, and 4) the Size Zone Matrix. Histogram features describe the first-order statistical properties of the image [161]. The features (that are described by the co-occurrence matrix) measure the texture of the image using Haralick statistical features [162]. The features (that are described by the neighborhood gray-tone difference matrix) measure a grayscale difference between pixels with certain grayscale and their neighboring pixels [133]. The features (that are described by Size Zone Matrix) are estimated using run-length techniques [163].

Furthermore, six histogram-based statistics (mean, variance, skewness, kurtosis, energy, and entropy) features are extracted from the edema, enhanced tumor and necrosis tissues.

VOLUMETRIC AND AREA-RELATED FEATURES

Thirteen volume-related features are considered: the volume of the whole tumor, the volume of the whole tumor with respect to the brain, the volume of sub-regions (edema, enhanced tumor, and necrosis) divided by the whole tumor, the volume of sub-regions (edema, enhanced

tumor, and necrosis) divided by the brain, the volumes of the enhanced tumor and necrosis divided by the edema, the summation of the volume of the edema and enhanced tumor, the volume of the edema divided by the summation of the volume of enhanced tumor and necrosis, and the volume of the necrosis divided by the summation of the volume of the edema and enhanced tumor. The tumor locations and the spread of the tumor in the brain are computed. Another nine area-related properties (area, centroid, perimeter, major axis length, minor axis length, eccentricity, orientation, solidity, and extent) are computed from three viewpoints (x, y, and z-axes) of the whole tumor.

HISTOGRAM-GRAPH FEATURES

A total of 832 features extracted from the histogram graph of the different modalities of the whole tumor, edema, enhancing, and necrosis regions. These features represent the frequency at different intensity bins (of 11,15, and 23) and the bins of the max frequency.

EULER CHARACTERISTICS

Finally, we compute the Euler characteristic [164] of the whole tumor, edema, enhancing, and necrosis, for each slice. The Euler characteristic features are computed on the tumor curve, at 100 points, and at 72 different angles. Then, the Euler characteristic features are integrated over all slices. As a result, each patient is represented by 4 (whole tumor, edema, enhancing, and necrosis) Euler characteristic feature vectors. Each vector has a size of 7200 (100 points × 72 angles).

5.3.3 SURVIVAL PREDICTION MODELS

Two different survival prediction models are proposed for survival prediction. The first model is a tree-based method for overall-survival regression prediction using the Random Forest (RF) regression model. We have employed RF due to its efficiency, robustness, and flexibility in utilization for both multi-class classification and regression tasks [165]. Additionally, RF does not require extensive hyper-parameter tuning and is resilient to overfitting. These traits make RF preferable over more common models such as artificial neural networks especially when the training data is limited. The complete pipeline for the survival regression using RF is illustrated in

Figure 14.a. This model uses significant, predictive, and important features selected from the above-mentioned texture, histogram-graph, and volumetric and area-related features. A three-step feature selection method is utilized as follows. Univariate Cox regression is fitted on every extracted feature, and features with *p-value* < 0.05 are considered significant. Second, univariate Cox regression is fitted on the quantitative copy of the significant features. The quantitative copy is obtained by thresholding the significant feature around its median value. The last step is performed to ensure that each significant feature is also able to split the data set into long vs. short survival. Then, recursive feature selection is applied to the 240 features. Finally, the RF regression model with tenfold cross-validation is used to evaluate the model at each iteration.

The model in Figure 14a is used as a baseline to obtain a second more comprehensive survival prediction pipeline as shown in Figure 14.b. We incorporate additional features such as Euler characteristics. The features for the updated model are then selected using the recursive feature selection method as follows. First, we perform recursive feature selection (RFS1) on the Euler features alone. Next, another recursive feature selection (RFS2) on the remaining features (texture, volumetric, histogram-graph based) is performed. In addition, the overall-survival regression model uses Cox regression to select significant features with a *p-value* < 0.05. Moreover, we introduce a state-of-the-art Extreme Gradient Boosting (XGBoost) [131] based regression technique for stepwise survival risk classification and overall-survival regression prediction using the selected features. The XGBoost based regression model is applied to each of the three groups (short, medium, and long) to obtain survival duration in the number of days, respectively. One of the major advantages of XGBoost is utilization of L1 and L2 regularization. L1 regularization handles sparsity, whereas L2 regularization reduces overfitting [131]. It is worth noting that we have not utilized any neural network models for the survival prediction because the sample size in this study is not large enough to ensure good training in a neural network setting.

5.4 OVERALL SURVIVAL PREDICTION FRAMEWORK EVALUATION

The proposed framework consists of two distinct radiomics based automated survival prediction pipelines. Accordingly, we obtain extensive performance evaluation using two pipelines: the first one comprises three-step feature selection and the RF-based survival prediction method (henceforth referred to as SP1), while the second consists of XGBoost based survival

prediction algorithm (henceforth referred to as SP2). We first participated in the BraTS 2017 challenge, and the specific combination of machine learning methods with RF survival prediction model (known as SP1) offered the best overall performance in this challenge. We subsequently participated in the BraTS 2018 challenge, and the augmented model (known as SP2) offered the best performance using the validation dataset.

Table 10. Performance of SP1, SP2, and modified-SP2 methods with BratS17 and BraTS18 datasets. The evaluation of validation is performed using the online evaluation platform of CBICA IPP (https://ipp.cbica.upenn.edu).

	Survival Prediction Performance	
Model/Dataset	**Accuracy**	**Mean Square Error**
SP1/ BraTS17 training	0.67	78,929
SP1/ BraTS17 validation	0.667	209,908
SP1/ BraTS17 test	0.579	245,780
SP2/ BraTS18 training	0.73	91,585
SP2/ BraTS18 validation	0.679	153,466
SP2/ BraTS18 test	0.519	367,240
RF-SP1/ BraTS18 validation	0.464	170,737
XGBoost-SP2/ BraTS17 validation	0.636	218,097
Modified-SP2/ BraTS18 training	0.718	99,358
Modified-SP2/ BraTS18 validation	0.679	127,697

For SP1 the survival prediction features are the age and 40 texture and volumetric features. The distribution of the 40 features is as follows: 12 features extracted from Texton of the tumor, 9 features extracted from the Holder exponent representations of the tumor, 6 features represent the histogram of the abnormal tissues, 5 from the raw MR modality of the tumor and sub-regions, 4 describe the volume of the tumor and the sub-regions, and 4 features are extracted from the tumor area and major axis length.

In comparison, as discussed above and shown in Figure 14.b for SP2, all relevant features are extracted from the ground truth cases available with the BraTS18 training dataset. The subsequent recursive feature selection for Euler features (28 thousand) alone generates 39 features.

The distribution of the 39 Euler features includes 16 features computed around the contour of ET, 16 features computed around that of WT, and 7 features computed around that of edema, respectively. The application of recursive feature selection on the remaining features produces an additional 23 texture features, 4 histogram graph features, and 8 area features of the edema, ET, and WT, respectively. The XGBoost with leave-one-out cross-validation (LOOCV) is employed on the selected 74 features and the age to predict three corresponding survival classes (short, medium, long). This yields a classification accuracy of 0.73 (95% confidence intervals [CI]: 0.655-0.797) for the BraTS18 training dataset.

Table 11. Confusion matrix of SP1, SP2, and modified-SP2, and some statistics derived from the confusion matrix based on each survival label in the training model.

	SP1 2017			**SP2 2018**			**Modified-SP2 2018**		
	Reference			**Reference**			**Reference**		
Predictions	**Long**	**Med**	**Low**	**Long**	**Med**	**Low**	**Long**	**Med**	**Low**
Long	32	7	10	43	13	4	44	11	4
Med	24	34	12	5	18	3	7	18	6
Low	0	1	43	8	11	58	5	13	55
Total number of cases	56	42	65	56	42	65	56	42	65
Statistics	**SP1 2017**			**SP2 2018**			**Modified-SP2 2018**		
Sensitivity	0.571	0.810	0.662	0.768	0.429	0.892	0.786	0.429	0.846
Specificity	0.841	0.702	0.990	0.841	0.934	0.806	0.860	0.886	0.816
Balanced Accuracy (Sen. + Spec.)/2	0.706	0.756	0.826	0.804	0.681	0.849	0.823	0.657	0.831
Positive Prediction Value (PPV)	0.653	0.486	0.977	0.717	0.692	0.753	0.745	0.581	0.753
Negative Prediction Value (NPV)	0.789	0.914	0.815	0.874	0.825	0.919	0.885	0.817	0.889

First, we establish the performance of both SP1 and SP2 methods using the BraTS17 and BraTS18 training and validation datasets. The training dataset performance is obtained through leave-one-out cross-validation analysis. The performance evaluation of methods using BraTS validation datasets is restricted to the online evaluation platform of the organizer of the BraTS challenge (CBICA IPP at https://ipp.cbica.upenn.edu) and must be performed during a specific

time period during the challenge. Note that the second pipeline (SP2) is developed after the BraTS 2017 challenge is concluded; hence, the 2017 validation portal is no longer available for evaluation. However, a fair comparison between the pipelines can still be obtained through the training data evaluations and the validation evaluations of respective challenge years. The results are summarized in Table 10 and Table 11.

Table 12. Performance of LOOCV of the three regression models in SP2 and modified-SP2 in the XGBoost overall survival model.

Regression model	SP2			Modified-SP2		
	Root Mean Square Error	**Mean Square Error**	**Mean Absolute Error**	**Root Mean Square Error**	**Mean Square Error**	**Mean Absolute Error**
Long-regression model	294.177	86,540	217.714	302.069	91,246	209.253
Medium-regression model	35.629	1,269	28.190	40.702	1,657	34.971
Short-regression model	61.449	3,776	50.402	80.340	6,455	65.094

The results in Table 10 for training and validation illustrate that the SP2 model offers better performance in accuracy over that of the SP1 model. The SP2 model also obtains improvement over SP1 in validation MSE. Note, the SP1 model has been ranked the first in the BraTS 2017 challenge for survival prediction category among seventeen teams globally. The overall high MSE for survival prediction is particularly due to the wide range within the long-term survival category resulting in large prediction errors. Further note that the MSE of SP2 for BraTS18 training is the sum of the three MSE (Table 12) values obtained for the short-, medium-, and long-regression models shown in Table 12. Finally, the test results for both SP1 for BraTS17 and SP2 for BraTS18 in Table 10 show that SP1 performed better in patient-survival prediction than that for SP2. This performance difference for SP1 and SP2 models is further analyzed below.

5.4.1 COMPARATIVE EVALUATION OF SURVIVAL PREDICTION PERFORMANCE WITH SP1 AND SP2

Table 11 shows the confusion matrix of both SP1 and SP2 and relevant statistics for each class in the classification training model for survival risk prediction. The sensitivity and balanced accuracy of the medium survival group in SP2 is the lowest when compared to the other two survival groups. The top four important features as ranked by XGBoost are tumor extent in the *z-axis*, the width of the enhanced tumor computed from *x-axis* point of view, and contour around the edema contour and enhanced tumor. The mean value of each of the four features is able to significantly (*p-value* < 0.05) stratify the 163 cases into two risk groups (low-risk and high-risk) as illustrated in Figure 15.

Our analysis of Figure 15 reveals that thresholding the value of features around the mean can stratify the 163 cases significantly (log-rank test, *p-value* < 0.05) into two survival groups. The features and the median survival of each group are:

- the tumor extent from *z-axis* viewpoint stratifies the 163 cases into two groups with a median survival of 394 days vs 318 days,
- the enhanced tumor width from *x-axis* viewpoint stratifies the 163 cases into two groups with a median survival of 437 days vs 268 days,
- the contour around the edema stratifies the 163 cases into two groups with a median survival of 387 days vs 330 days,
- and contour around the enhanced tumor stratifies the 163 cases into two groups with a median survival of 421 days vs 278 days.

The second step in the survival prediction is to obtain individual regression training models corresponding to the short, medium, and long survival classes. These short-, medium-, and long-regression models use features selected distinctly for each survival class using Cox regression (with *p-value* < 0.05). The number of significant features selected for the short-, medium-, and long-regression models is 83, 51, and 148, respectively. Table 12 illustrates the performance of LOOCV with XGBoost for the selected features using specified survival risk cases in BraTS18 training cases.

Note that the wide range of the overall survival of the long-survival group (greater than 15 months) may cause the RMSE of the long-regression model to have the highest RMSE (Table 12).

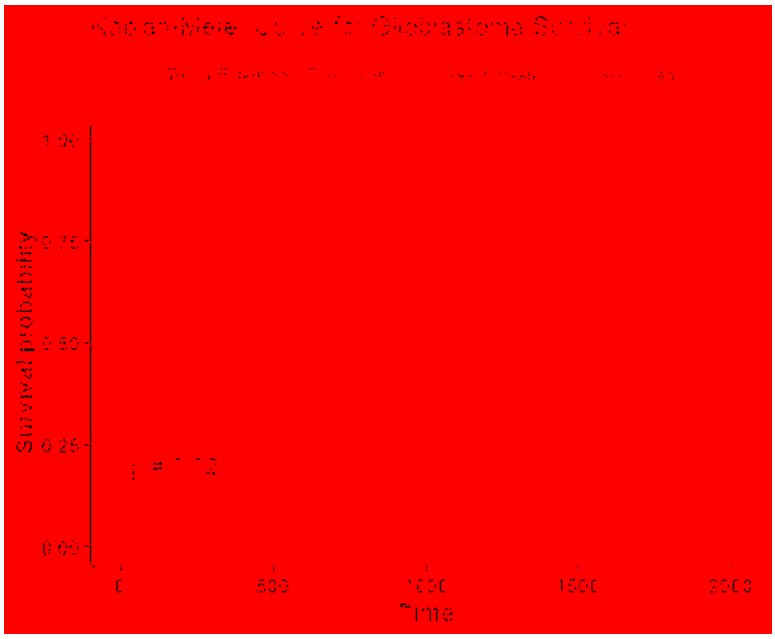
This also may cause a high mean square error when using the validation dataset (Table 10). The range of the overall survival of the short-survival group is 10 months, whereas the medium-survival group is 5 months.

5.4.2 MODIFIED-SP2

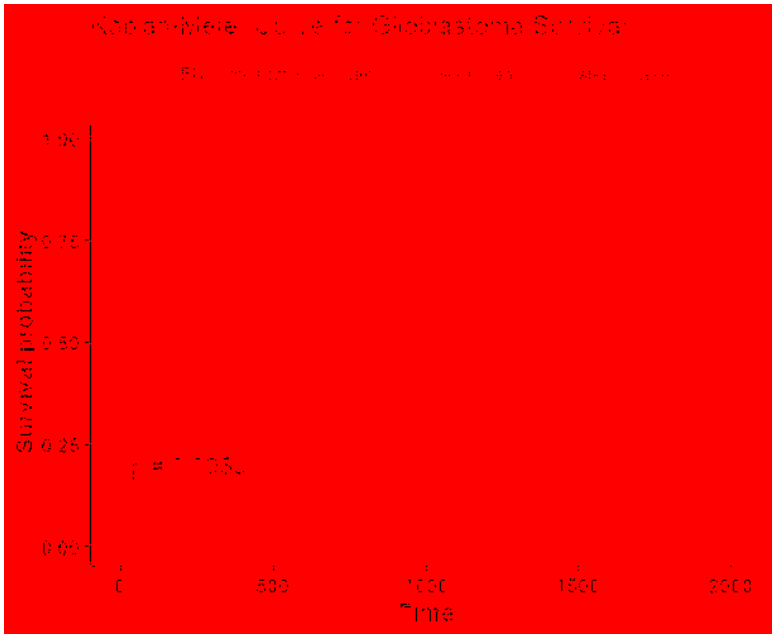
In order to reduce the high dimensionality of the features in SP2 classification and regression steps, we modify SP2 in Figure 14.b as follows: 1) calculate and rank the feature importance for each classification and regression model; 2) select features that have relative scaled importance greater than 50%; 3) train the modified selected features in a new classification and regression training models utilizing XGBoost. The resulting 30 significant features are applied in the classification step of the modified-SP2. The distribution of these features is as follows: 13 features represent Euler characteristics, 7 features represent volumetric and area-related properties, 4 histogram-graph based features, 5 texture features, and one feature with age information.

The number of significant features used in the short-, medium-, and long-regression models of the modified-SP2 is 11, 9, and 11, respectively. The distribution of the features in the modified short-regression model is as follows: 2 volumetric and area-related features, 1 histogram-graph based features, 7 texture features, and one feature with age information. The features employed in the modified med-regression model are 5 volumetric and area-related features, 3 texture features, and Age. The features of the modified long-regression model are 2 volumetric and area-related features, 8 texture features, and one feature with age information.

The modified-SP2 achieves cross-validated accuracy of 0.718 as illustrated in Table 10. Table 11 illustrates the statistics of its confusion matrix in the classification training model. Table 12 illustrates the performance of the modified regression training models. Additionally, the modified-SP2 is validated using the BraTS18 validation set, and its performance is illustrated in Table 10. Note that the different performances of SP2 and modified-SP2 are almost similar when using the BraTS18 training and validation dataset statistics of each class in SP2 and the modified-SP2 are almost similar. This can be explained by the fact that XGBoost provides L1 and L2 regularization. Additionally, the modified-SP2 is validated using the BraTS18 validation set, and its performance is illustrated in Table 10.

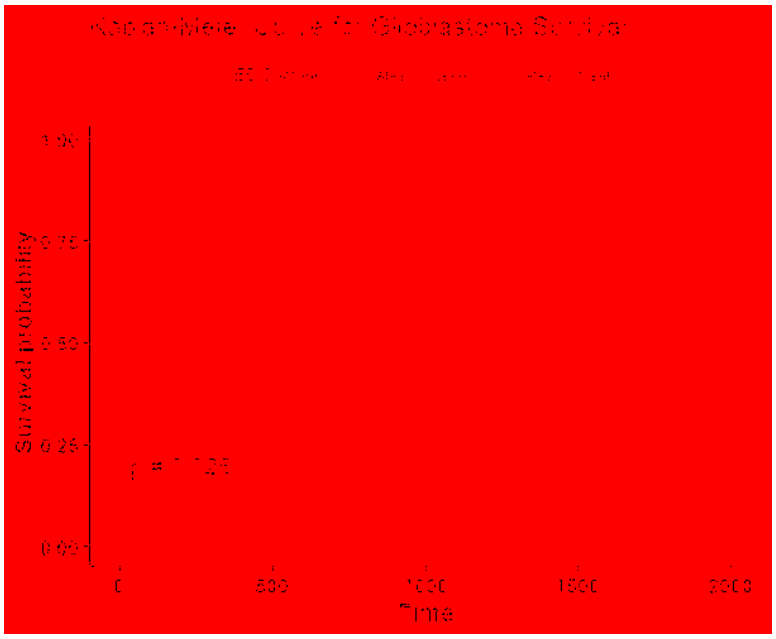


(a)

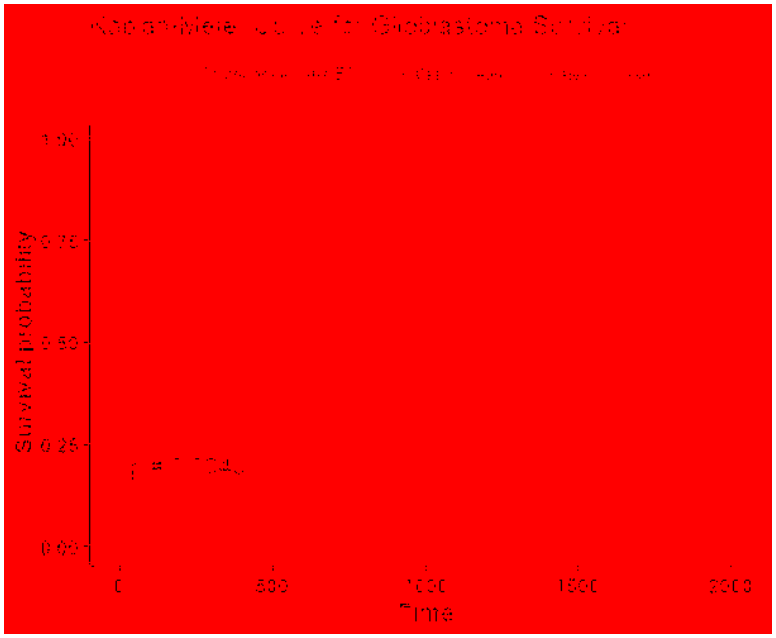


(b)

Figure 15. Kaplan Meier of the top four important features used in SP2. The features are divided around its mean value to stratify the 163 subjects into two groups: high risk group (red line), and low risk group (blue line). The features are (a) tumor extent; (b) the enhanced tumor width; (c) contour around the edema; and (d) contour around the enhanced tumor. The shaded area indicates the 95% confidence interval. The time is measured in days.



(c)



(d)

Figure 15. (continued)

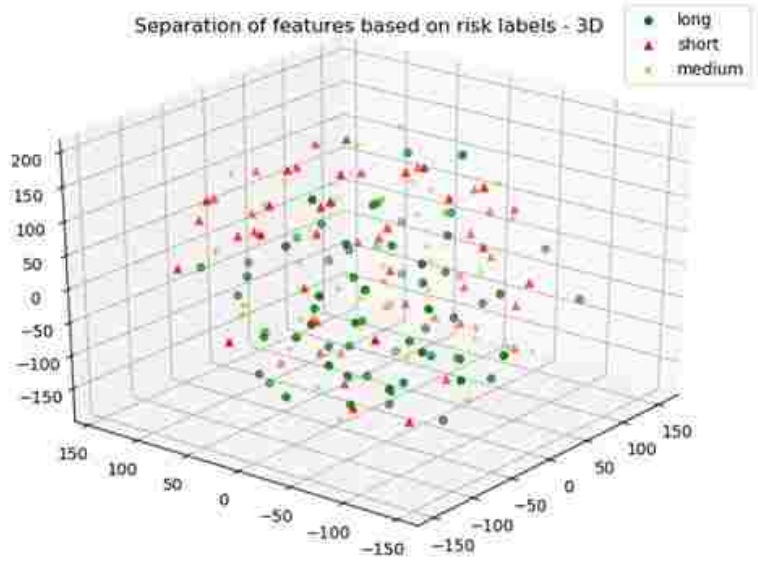
5.5 CRITICAL ANALYSIS OF FEATURES UTILIZED IN THE SURVIVAL PREDICTION PIPELINES

This section provides a critical analysis of the features and their effects on survival prediction performance. As mentioned in the previous sections, the features that are derived from different abnormal tissue types in the tumor of the segmented tumor region significantly contribute to the survival prediction performance. Accordingly, we visualize the features extracted from different abnormal tissue types of the segmented tumor. The visualization is performed using one of the most widely-used high-dimensional data visualization techniques known as t-Distributed Stochastic Neighbor Embedding [132] (t-SNE). First, t-SNE is used to explore the features obtained from different abnormal tissue types from the segmented tumor region and analyze the effects of these features on the performance of the survival prediction task using the BraTS 2017 and BraTS 2018 dataset.

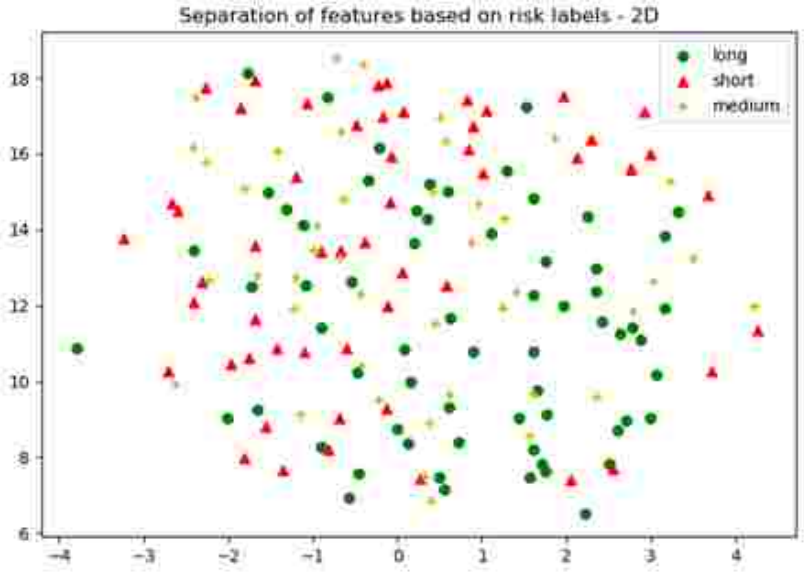
For the SP1 pipeline, we extract a total of 40 features from the sub-tissue types of the segmented tumor region. The features extracted in SP1 are as follows: 36 features for the whole tumor, 2 features for the enhanced tumor, and 2 features for edema. We visualize the feature clusters for patient survival categories: long, medium, and short term. In this case, we consider all 40 features obtained from the 163 BraTS17 training data as mentioned above and explore the grouping against the tumor risk labels using the t-SNE technique. Figure 16 shows the visualization of the corresponding features for long, medium, and short risk labels. Note that all the visualization outcomes shown are obtained after extensive hyper-parameter tuning of t-SNE to produce the best possible results. Figure 16 demonstrates that though there is some separation of corresponding features between the long and short categories, the medium category is mixed with both long and short categories. This suggests that it is still difficult to visualize a clear separation of extracted features for the survival prediction task with the available patient dataset for this study. The corresponding survival prediction performance of the SP1 pipeline using the testing dataset is as shown in Table 10 and Table 12. As mentioned above, though the SP1 pipeline was ranked first place in the BraTS 2017 challenge, the feature distribution in Figure 16 suggests an inherent challenge in extracting representative features for the survival prediction task.

Next, we explore the features and their effects on the performance of our SP2 pipeline using the BraST18 dataset. We extract a total of 74 features and the age for the SP2 pipeline. The

features extracted in SP2 are as follows: 43 features for the whole tumor, 22 features for the enhanced tumor, 8 features for edema, and 1 feature for necrosis. Figure 17 shows the visualization of the 74 features in terms of long, medium, and short risk labels using the 163 sample BraTS18 training data. Our analysis suggests that the tSNE technique again fails to group the features in long, medium, and short categories. Though there is some separation between the corresponding features for long and short categories, the features for the medium category mix with both short and long categories for multiple subjects, quite similar to the visualization of SP1. This poor separation may still be due to the lack of sufficient representative strength of the features for categorizing different risk labels. Consequently, Table 10 shows that our proposed SP2 pipeline achieves 0.73, 0.679, and 0.519 accuracies on the BraTS18 training, validation, and testing data.

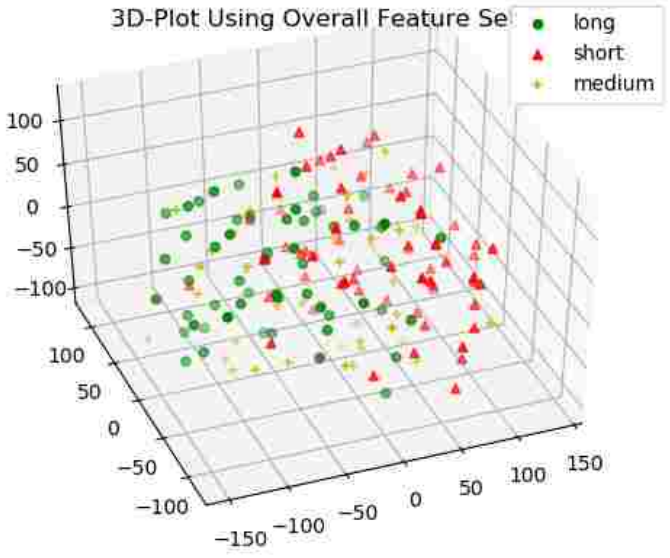


(a)

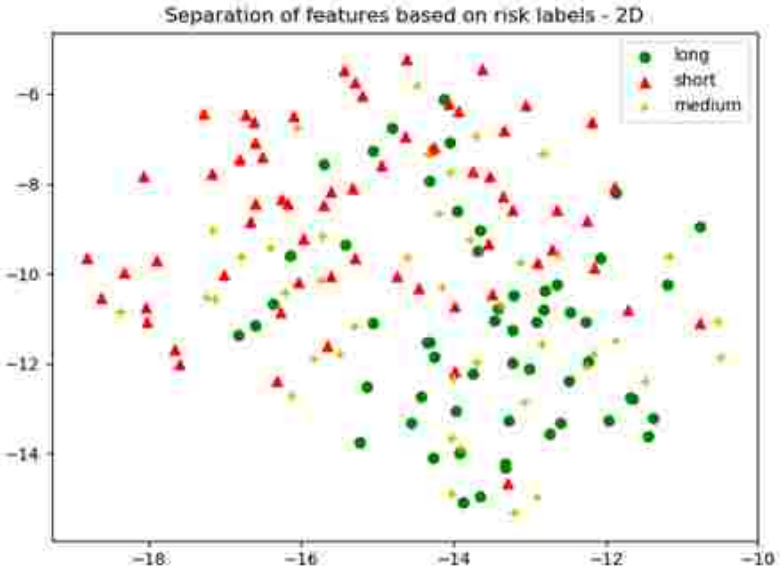


(b)

Figure 16. SP1 feature visualization. (a) The 3D; and (b) the 2D plot of t-Distributed Stochastic Neighbor Embedding (t-SNE) of the selected features of SP1 clustered based on the long, medium, and short risk labels using BraTS17 training dataset.



(a)



(b)

Figure 17. SP2 feature visualization. **(a)** The 3D plot of the t-Distributed Stochastic Neighbor Embedding (t-SNE) of the selected features of SP2 clustered based on the long, medium, and short risk labels using BraTS18 training dataset. **(b)** The 2D plot of the same training dataset.

Additionally, we validate our RF survival prediction in SP1 (RF-SP1) using the BraTS18 validation set. We also validate XGBoost survival prediction in SP2 (XGBoost-SP2) using the BraTS17 validation dataset. The results are summarized in Table 1. Using the BraTS17 validation

dataset, the RF-SP1 model achieves 67.7% accuracy, whereas the XGBoost-SP2 model achieves 63.6%. Using the BraTS18 validation dataset, the RF-SP1 model achieves 46.4% accuracy, whereas the XGBoost-SP2 model achieves 67.9% accuracy. These results indicate that the XGBoost-SP2 combination performs considerably better than that of the RF-SP1 with the BraTS18 dataset and reasonably well with the BraTS17 dataset, respectively. Note that the ground truth of the BraTS17 and BraTS18 validation dataset is not provided. As a result, we have segmented the BraTS17 and BraTS18 validation dataset using the semantic label fusion model of CNN and RF [95] and the semantic label fusion of U-Net and FCN [92], respectively.

5.6 COMPARISON OF SURVIVAL PREDICTION WITH STATE-OF-THE-ART WORKS

Comparison of the proposed survival prediction pipelines SP1 and SP2 with some of the state-of-the-art methods in the literature is discussed next. Table 13 summarizes the performances of these state-of-the-art models and presents a comparison with our proposed framework (SP2). Reference [166] proposes using histogram features extracted from denoised MR images (by using 2 levels of the Daubechies wavelet transform) in a support vector machine to predict overall survival. Their method achieves a 10-fold cross-validation accuracy of 0.667 using the BraTS17 training dataset. Reference [167] extracts volumetric, spatial, morphological, and tractography features from MR images. Feature normalization and selection is performed, and the selected features are trained in a support vector machine model. Their proposed model achieves an accuracy of 0.7 using the BraTS18 training dataset and an accuracy of 0.5 using the BraTS18 validation dataset. Reference [168] utilizes volumetric features along with Random Forest to predict overall survival. Their method achieves five-fold cross-validation accuracy of 0.638 using the BraTS17 training dataset. The results demonstrate that our proposed framework achieves higher accuracy in overall survival prediction compared to the current-state-of-the-art models applied to the same datasets. Note, unlike our proposed SP1 and SP2 pipelines, the reported performance for all these other methods in Table 13 are obtained by the authors themselves.

Table 13. Comparison of our proposed survival prediction pipeline with state-of-the-art methods in literature.

Reference	Algorithm	Validation method	Performance	Dataset
Chato et al.[166]	histogram features along with SVM	10-fold cross validation	accuracy of 0.667	BraTS17 training dataset
Kao et al.[167]	volumetric, spatial, morphological, and tractography features along with SVM	5-fold cross-validation	Accuracy of 0.7	BraTS18 training dataset
Soltaninejad et al. [168]	volumetric features along with Random Forest	5-fold cross validation	Accuracy of 0.638	BraTS17 training dataset
XGBOOST overall survival prediction model (SP2)	Texture, Volumetric, histogram-graph, and Euler features Along with XGBoost	LOOCV	Accuracy of 0.73 and MSE of 91585.51	BraTS18 training dataset
		Validation dataset	Accuracy of 0.679 and MSE of 153466.3	BraTS18 validation dataset

5.7 DISCUSSION

This work proposes a novel framework for fully automated radiomics-based Glioblastoma survival prediction. The overall framework is designed as a two-step process where feature selection is carried out in the first step, and then the survival prediction is carried out in the second step. The framework includes two survival prediction algorithms SP1 and SP2, represented using feature types, feature selection, regression, and classification methods.

The primary survival pipeline (SP1) combines three-step feature selection and the RF-based survival prediction algorithm to obtain the final output. The second pipeline (SP2) consists of XGBoost based survival prediction algorithm. As shown in Table 10 and Table 13 the pipelines used in both SP2 and SP1 offer a comparative survival prediction performance. The functionality of SP2 is further enhanced by using additional features extracted from the sub-tissues (edema,

enhanced tumor, and necrosis) and a two-step classification and regression method. Different studies [158, 169-171] correlate between survival prediction in glioblastoma and different sub-tissues. SP2 shows improvements over our primary survival prediction model (SP1) [91] with LOOCV accuracy increase to 0.73 from 0.67 for training datasets, whereas the modified-SP2 achieves cross-validation accuracy of 0.718 using the training dataset.

CHAPTER 6

SUMMARY AND FUTURE WORK

Computational modeling for non-invasive methods that analyze clinical outcomes for patients has emerged as a vital and promising body of research. For example, these non-invasive methods are important in tumor progression analysis and classification of molecular mutations in diffuse gliomas prior to invasive tissue sampling. Radiomics may be used to quantify tumor heterogeneity that varies between patients and even within a patient. Consequently, radiomics has an essential role in current radiology practice such as disease diagnosis, monitoring, and treatment planning. This dissertation proposes computational models for non-invasive survival prediction, molecular classification, and grading prediction based on radiomics and clinical data for diffuse glioma patients. Furthermore, statistical models are proposed for molecular mutation classification and grading based on genomics count-based data and radiomics utilizing the negative binomial regression. To evaluate the efficacy of the proposed methods discussed in this dissertation, we cross-validate, test, and compare the performance with state-of-the-art methods.

In Chapter 3, we investigate the efficacy of our novel fractal and multi-resolution fractal radiomics features on the performance of the non-invasive prediction of molecular mutation in diffuse LGG. We address diffuse LGG molecular grading and classification using radiomics features extracted from multimodality MRI of the segmented tumor volume. The extracted radiomics features describe the multi-resolution fractal modeling, texture features, volumetric, and area-based characteristics. In the LGG molecular grading and classification model, different molecular (*IDH*, *1p/19q codeletion*, *ATRX*, and *TERT*), and MGMT methylation prediction models are introduced. Our analysis of the MGMT classification reveals that the less aggressive methylated MGMT status associates significantly with higher values of the histogram entropy (i.e., uncertainty) of tumor mBm and carries HR of 0.579 per standard deviation (likelihood ratio test p-value = 0.035). The study further shows that the highly aggressive MGMT un-methylated LGG associates with the higher values of the size ratio between enhanced tumor and necrosis. The histogram entropy of tumor mBm and the size ratio between enhanced tumor and necrosis are among the most predictive features of MGMT methylation status. Our *IDH* mutation classification

indicates that the tumor correlation associates significantly with mutated *IDH* and the complexity of HE of the enhanced tumor associates significantly with WT IDH status. Furthermore, our analysis reveals that the size ratio between the enhanced tumor and necrosis is a significant predictor feature of the *IDH* status which is also a significant predictor of MGMT status. Furthermore, our analysis of the 1p/19q codeletion classification model indicates that the location of the upper-left necrosis bounding box and horizontal coordinate of the necrosis centroid are among the most predictive features and that high values of histogram entropy of tumor mBm associate significantly with the existence of 1p/19q codeletion. Our analysis of the *ATRX* status classification model reveals that the tumor information-measure of correlation and the mean of the tumor histogram are the most frequently selected features and that the higher values of information-measure of correlation are associated significantly with WT *ATRX*. Moreover, the *TERT* classification model analysis shows that tumor information-measure of correlation and the upper-left coordinate in the edema bounding box are the most frequently selected features. Higher values of these two features are significantly associated with mutated *TERT* status. In addition, our analysis suggests that the higher values of the inverse difference moment of HE associates significantly with WT *TERT* and offers HR = 0.612 per standard deviation (likelihood ratio test, *p*-value = 0.03). High values of the inverse difference moment of an HE tumor predict the less aggressive WT *TERT*. Finally, our molecular mutations classification models reveal that fractal features significantly affect MGMT, *IDH*, 1p/19q co-deletion, and *ATRX* prediction models. Features analysis and the performance of the molecular mutations classification models in Table 2 confirm our hypothesis by showing the importance of the fractal features in diffuse LGG molecular mutations prediction.

A thorough study utilizing the genomics count-based data along with radiomics in a negative binomial regression model (i.e., radiogenomics-NB) is discussed in Chapter 4. We employ both RNAseq counts data and radiomics features to perform molecular mutations classification and prediction in diffuse LGG. The NB distribution is proposed to model RNAseq counts data, and the log-linear regression model is utilized to model the relationship between the estimated model mean and radiomics. The proposed radiogenomics-NB model preserves the count-based nature of RNAseq data, fuse radiomics and RNAseq counts data, and develop a complete radiogenomics-NB model with better characterization of patient data. This study investigates three radiogenomics-NB molecular mutation models; IDH mutation, 1p/19q

codeletion, and ATRX mutation. We investigate the efficacy of using the top different number of DERs. Our analysis reveals the best performance of the radiogenomics-NB models is achieved using 10 DERs. Moreover, we investigate the gender-specific effect on the radiogenomics-NB models. Our analysis reveals that female-specific radiogenomics-NB models significantly outperform the male-specific radiogenomics-NB models for prediction of *IDH* status, *1p/19q codeletion*, and *ATRX* status, respectively. Additionally, we compare our radiogenomics-NB model to NBLDA, RF-genomics, FR-radiogenomics, and VoomNSC. Our model significantly outperforms NBLDA, RF-genomics, and VoomNSC for prediction of *IDH* and *ATRX* mutations, which indicates the power of fusing radiomics into genomics data to develop classification and prediction models.

Furthermore, in Chapter 5, a fully automated radiomics-based overall survival prediction framework for glioblastoma patients is discussed. The overall survival prediction framework comprises two different pipelines that combine different feature selection and regression approaches. The survival prediction framework is assessed using two recent widely used benchmark datasets: BraTS17 and BraTS18. The framework predicts the overall survival in days and classifies the overall survival risk into three survival groups: long (greater than 15 months), medium (between 10 to 15 months), and short (less than 10 months). Our analysis using the t-SNE tool reveals that selected features are able to separate between the long- and short-term survival categories, while the medium-term survival category is mixed with both long and short categories. This poor separation of the medium category may still be due to the lack of the sufficient representative strength of the radiomics for categorizing different risk labels The performances of the overall survival pipelines (SP1 and SP2) demonstrate that our proposed framework achieves higher accuracy in overall survival prediction compared to the current-state-of-the-art models applied to the same datasets. The proposed pipeline SP1 achieves the best performance in the BraTS 2017 in the survival prediction task.

6.1 FUTURE WORK

The analysis of glioma survival prediction and molecular mutations classification and predictions that are discussed in this dissertation show competitive performance when compared with the state-of-the-art methods. However, there are a few limitations to this work. Future studies

are necessary to address these limitations and enhance the performance of the glioma outcome analysis and prediction models proposed in this work. Future studies are necessary for the following areas. In Chapter 3, the performances of the proposed LGG molecular mutation classification and prediction model show promise when compared to different methods and models in the literature. In the study, we target radiomics features in LGG patients. These molecular mutations are not carried by LGG patients only. Therefore, in the future, a full-scale study of a more diverse dataset of glioma patients (not only LGG patients) may reflect higher reliability of the performance of the proposed molecular mutations prediction model. Additionally, utilizing important clinical data (e.g., age at diagnosis, gender, ethnicity, tumor morphology, and Karnofsky Status) along with the existing radiomics may offer more powerful non-invasive molecular mutation classifying models.

In Chapter 4, we propose volumetric radiomics in the radiogenomics-NB model. However, different mutations may be better predicted in the radiogenomics-NB model using different clinical information and/or radiomics. In the future, a superior framework/mechanism is needed to investigate a specific radiomics and associate them with the different mutations and their RNAseq read counts data through the NB model. Moreover, in the derivation of the radiogenomics-NB model, we assume that all RNAseq are independent of each other. This assumption may not always be realistic, and gene interaction may affect a certain glioma outcome. Therefore, it is necessary to consider the gene interaction effect to extend the radiogenomics-NB model.

Even though in Chapter 5, the SP1 ranked first in the BraTS 2017 challenge and SP2 achieved a competitive performance, there are a few limitations of the proposed models. First, the total number of cases for the survival training dataset is 163, and both BraTS 2017 and BraTS 2018 required that the data must be divided into three separate survival groups. Consequently, the number of training cases are divided among three groups as follows: 65 cases for short, 42 cases for medium and 56 cases for the long survival group, respectively. Therefore, future work includes a larger dataset for training each regression model to improve model performance. Second, this study may benefit from additional clinical data such as gender, race, and Karnofsky Status to strengthen the reliability of the different survival regression and classification models. Finally, the overall survival risk classification performance of the state-of-the-art methods in the literature, including the pipelines proposed in this work, may be improved further. The visualization of survival features suggests the difficulty in separating the high dimensional data into the three

distinctive risk classes. In addition, there is a need for further research in fusing clinical data with radiomics and novel feature engineering for survival prediction.

Finally, the work in this dissertation mainly presents multiple models for glioma outcome survival analysis and prediction, and molecular mutations classification. However, the therapeutic outcome prediction requires a careful investigation of the imaging changes before and after the therapeutic session in a proper computational machine learning model. Furthermore, to develop a full and robust clinical outcome framework, the therapeutic outcome should be assessed and predicted. Such a study requires an understanding of the therapeutic modality and the relevant clinical data (e.g. age, gender, the severity of the tumor), radiomics (e.g., tumor volume, tumor location, the proportion of necrosis), and genomics (e.g., existed mutations), and other omics data predictors of the treatment outcome.

BIBLIOGRAPHY

[1] Q. T. Ostrom *et al.*, "CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012–2016," *Neuro-oncology,* vol. 21, no. Supplement_5, pp. v1-v100, 2019.

[2] A. Volterra and C. Steinhäuser, "Glial modulation of synaptic transmission in the hippocampus," *Glia,* vol. 47, no. 3, pp. 249-257, 2004.

[3] G. M. Cooper, R. E. Hausman, and R. E. Hausman, *The cell: a molecular approach*. ASM press Washington, DC, 2000.

[4] C. Walker, A. Baborie, D. Crooks, S. Wilkins, and M. Jenkinson, "Biology, genetics and imaging of glial cell tumours," *The British journal of radiology,* vol. 84, no. special_issue_2, pp. S90-S106, 2011.

[5] D. N. Louis *et al.*, "The 2007 WHO classification of tumours of the central nervous system," *Acta neuropathologica,* vol. 114, no. 2, pp. 97-109, 2007.

[6] P. Kleihues, P. C. Burger, and B. W. Scheithauer, "The new WHO classification of brain tumours," *Brain pathology,* vol. 3, no. 3, pp. 255-268, 1993.

[7] A. Claes, A. J. Idema, and P. Wesseling, "Diffuse glioma growth: a guerilla war," *Acta neuropathologica,* vol. 114, no. 5, pp. 443-458, 2007.

[8] D. N. Louis, "The next step in brain tumor classification:“Let us now praise famous men”… or molecules?," ed: Springer, 2012.

[9] A. Korshunov *et al.*, "Integrated analysis of pediatric glioblastoma reveals a subset of biologically favorable tumors with associated molecular prognostic markers," *Acta neuropathologica,* vol. 129, no. 5, pp. 669-678, 2015.

[10] N. Nonoguchi, T. Ohta, J.-E. Oh, Y.-H. Kim, P. Kleihues, and H. Ohgaki, "TERT promoter mutations in primary and secondary glioblastomas," *Acta neuropathologica,* vol. 126, no. 6, pp. 931-937, 2013.

[11] D. N. Louis *et al.*, "International Society of Neuropathology-Haarlem consensus guidelines for nervous system tumor classification and grading," *Brain pathology,* vol. 24, no. 5, pp. 429-435, 2014.

[12] D. N. Louis *et al.*, "The 2016 World Health Organization classification of tumors of the central nervous system: a summary," *Acta neuropathologica,* vol. 131, no. 6, pp. 803-820, 2016.

[13] H. Ohgaki and P. Kleihues, "Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas," *Journal of Neuropathology & Experimental Neurology,* vol. 64, no. 6, pp. 479-489, 2005.

[14] A. Mukasa *et al.*, "Significance of IDH mutations varies with tumor histology, grade, and genetics in Japanese glioma patients," *Cancer science,* vol. 103, no. 3, pp. 587-592, 2012.

[15] H. Yan *et al.*, "IDH1 and IDH2 mutations in gliomas," *New England Journal of Medicine*, vol. 360, no. 8, pp. 765-773, 2009.

[16] A. L. Cohen, S. L. Holmen, and H. Colman, "IDH1 and IDH2 mutations in gliomas," *Current neurology and neuroscience reports,* vol. 13, no. 5, p. 345, 2013.

[17] J.-R. Chen, Y. Yao, H.-Z. Xu, and Z.-Y. Qin, "Isocitrate dehydrogenase (IDH) 1/2 mutations as prognostic markers in patients with glioblastomas," *Medicine,* vol. 95, no. 9, 2016.

[18] C. Hartmann *et al.*, "Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: implications for classification of gliomas," *Acta neuropathologica,* vol. 120, no. 6, pp. 707-718, 2010.

[19] N. Hu, R. Richards, and R. Jensen, "Role of chromosomal 1p/19q co-deletion on the prognosis of oligodendrogliomas: A systematic review and meta-analysis," *Interdisciplinary Neurosurgery,* vol. 5, pp. 58-63, 2016.

[20] 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, 2015.

[21] J. Schwartzentruber *et al.*, "Driver mutations in histone H3. 3 and chromatin remodelling genes in paediatric glioblastoma," *Nature,* vol. 482, no. 7384, pp. 226-231, 2012.

[22] Y. Jiao *et al.*, "Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas," *Oncotarget,* vol. 3, no. 7, p. 709, 2012.

[23] H. E. Leeper, A. A. Caron, P. A. Decker, R. B. Jenkins, D. H. Lachance, and C. Giannini, "IDH mutation, 1p19q codeletion and ATRX loss in WHO grade II gliomas," *Oncotarget,* vol. 6, no. 30, p. 30295, 2015.

[24] J. F. Costello, B. W. Futscher, K. Tano, D. M. Graunke, and R. O. Pieper, "Graded methylation in the promoter and body of the O6-methylguanine DNA methyltransferase

(MGMT) gene correlates with MGMT expression in human glioma cells," *Journal of Biological Chemistry,* vol. 269, no. 25, pp. 17228-17237, 1994.

[25] S. Leu *et al.*, "IDH/MGMT-driven molecular classification of low-grade glioma is a strong predictor for long-term survival," *Neuro-oncology,* vol. 15, no. 4, pp. 469-479, 2013.

[26] K. Zhang, X.-q. Wang, B. Zhou, and L. Zhang, "The prognostic value of MGMT promoter methylation in Glioblastoma multiforme: a meta-analysis," *Familial cancer,* vol. 12, no. 3, pp. 449-458, 2013.

[27] A. M. Rutman and M. D. Kuo, "Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging," *European journal of radiology,* vol. 70, no. 2, pp. 232-241, 2009.

[28] N. Just, "Improving tumour heterogeneity MRI assessment with histograms," *British journal of cancer,* vol. 111, no. 12, pp. 2205-2213, 2014.

[29] L. S. Hu *et al.*, "Multi-parametric MRI and texture analysis to visualize spatial histologic heterogeneity and tumor extent in glioblastoma," *PloS one,* vol. 10, no. 11, p. e0141506, 2015.

[30] H. Itakura *et al.*, "Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities," *Science translational medicine,* vol. 7, no. 303, pp. 303ra138-303ra138, 2015.

[31] 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," *Medical physics,* vol. 42, no. 11, pp. 6725-6735, 2015.

[32] R. Brown *et al.*, "The use of magnetic resonance imaging to noninvasively detect genetic signatures in oligodendroglioma," *Clinical Cancer Research,* vol. 14, no. 8, pp. 2357-2362, 2008.

[33] W. L. Bi and R. Beroukhim, "Beating the odds: extreme long-term survival with glioblastoma," *Neuro-Oncology,* vol. 16, no. 9, p. 1159, 2014.

[34] H. Ohgaki and P. Kleihues, "The definition of primary and secondary glioblastoma," *Clinical cancer research,* vol. 19, no. 4, pp. 764-772, 2013.

[35] J. P. O'Connor, C. J. Rose, J. C. Waterton, R. A. Carano, G. J. Parker, and A. Jackson, "Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome," *Clinical Cancer Research,* vol. 21, no. 2, pp. 249-257, 2015.

[36] Y. Liu, X. Xu, L. Yin, X. Zhang, L. Li, and H. Lu, "Relationship between glioblastoma heterogeneity and survival time: An MR imaging texture analysis," *American Journal of Neuroradiology,* vol. 38, no. 9, pp. 1695-1701, 2017.

[37] D. Molina *et al.*, "Tumour heterogeneity in glioblastoma assessed by MRI texture analysis: a potential marker of survival," *The British journal of radiology,* vol. 89, no. 1064, p. 20160242, 2016.

[38] J. P. O'Connor *et al.*, "Imaging biomarker roadmap for cancer studies," *Nature reviews Clinical oncology,* vol. 14, no. 3, p. 169, 2017.

[39] R. Weissleder, M. C. Schwaiger, S. S. Gambhir, and H. Hricak, "Imaging approaches to optimize molecular therapies," *Science translational medicine,* vol. 8, no. 355, pp. 355ps16-355ps16, 2016.

[40] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology,* vol. 278, no. 2, pp. 563-577, 2015.

[41] A. Giardino *et al.*, "Role of imaging in the era of precision medicine," *Academic radiology,* vol. 24, no. 5, pp. 639-649, 2017.

[42] G. S. Ginsburg and H. F. Willard, "Genomic and personalized medicine: foundations and applications," *Translational research,* vol. 154, no. 6, pp. 277-287, 2009.

[43] F. Meric-Bernstam, C. Farhangfar, J. Mendelsohn, and G. B. Mills, "Building a personalized medicine infrastructure at a major cancer center," *Journal of Clinical Oncology,* vol. 31, no. 15, p. 1849, 2013.

[44] W. K. Redekop and D. Mladsi, "The faces of personalized medicine: a framework for understanding its meaning and scope," *Value in Health,* vol. 16, no. 6, pp. S4-S9, 2013.

[45] T. W. Hakkarainen, N. M. Kopari, T. N. Pham, and H. L. Evans, "Necrotizing soft tissue infections: review and current concepts in treatment, systems of care, and outcomes," *Current problems in surgery,* vol. 51, no. 8, pp. 344-362, 2014.

[46] J. Wu, K. K. Tha, L. Xing, and R. Li, "Radiomics and radiogenomics for precision radiotherapy," *Journal of radiation research,* vol. 59, no. suppl_1, pp. i25-i31, 2018.

[47] E. Sala *et al.*, "Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging," *Clinical radiology*, vol. 72, no. 1, pp. 3-10, 2017.

[48] M. D. Kuo and N. Jamshidi, "Behind the numbers: decoding molecular phenotypes with radiogenomics—guiding principles and technical considerations," *Radiology,* vol. 270, no. 2, pp. 320-325, 2014.

[49] H. Itakura *et al.*, "Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities," *Science translational medicine,* vol. 7, no. 303, 2015.

[50] J. Wu *et al.*, "Unsupervised clustering of quantitative image phenotypes reveals breast cancer subtypes with distinct prognoses and molecular pathways," *Clinical Cancer Research,* vol. 23, no. 13, pp. 3334-3342, 2017.

[51] D. Fehr *et al.*, "Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images," *Proceedings of the National Academy of Sciences,* vol. 112, no. 46, pp. E6265-E6273, 2015.

[52] P. Kickingereder *et al.*, "Radiogenomics of glioblastoma: machine learning–based classification of molecular characteristics by using multiparametric and multiregional MR imaging features," *Radiology,* vol. 281, no. 3, pp. 907-918, 2016.

[53] M. A. Mazurowski, K. Clark, N. M. Czarnek, P. Shamsesfandabadi, K. B. Peters, and A. Saha, "Radiogenomics of lower-grade glioma: algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with The Cancer Genome Atlas data," *Journal of neuro-oncology,* vol. 133, no. 1, pp. 27-35, 2017.

[54] S. Rathore *et al.*, "Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1," *Scientific Reports,* vol. 8, no. 1, p. 5087, 2018/03/23 2018.

[55] X. Zhang *et al.*, "Long non-coding RNA expression profiles predict clinical phenotypes in glioma," *Neurobiology of disease,* vol. 48, no. 1, pp. 1-8, 2012.

[56] W.-J. Zeng *et al.*, "Integrative analysis of DNA methylation and gene expression identify a three-gene signature for predicting prognosis in lower-grade gliomas," *Cellular Physiology and Biochemistry,* vol. 47, no. 1, pp. 428-439, 2018.

[57] Y. Iwadate *et al.*, "Molecular Classification and Survival Prediction in Human Gliomas Based on Proteome Analysis," *Cancer Research*, vol. 64, no. 7, pp. 2496-2501, 2004.

[58] Y. Li *et al.*, "Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature," *European Radiology*, pp. 1-9, 2018.

[59] V. G. Kanas, E. I. Zacharaki, G. A. Thomas, P. O. Zinn, V. Megalooikonomou, and R. R. Colen, "Learning MRI-based classification models for MGMT methylation status prediction in glioblastoma," *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 249-257, 2017.

[60] H. Zhou *et al.*, "MRI features predict survival and molecular markers in diffuse lower-grade gliomas," *Neuro-oncology*, vol. 19, no. 6, pp. 862-870, 2017.

[61] B. Niu *et al.*, "Glioma stages prediction based on machine learning algorithm combined with protein-protein interaction networks," *Genomics*, 2019.

[62] S. M. Reza, M. D. Samad, Z. A. Shboul, K. A. Jones, and K. M. Iftekharuddin, "Glioma grading using structural magnetic resonance imaging and molecular data," *Journal of Medical Imaging*, vol. 6, no. 2, p. 024501, 2019.

[63] A. Chaddad, P. Daniel, S. Sabri, C. Desrosiers, and B. Abdulkarim, "Integration of Radiomic and Multi-omic Analyses Predicts Survival of Newly Diagnosed IDH1 Wild-Type Glioblastoma," *Cancers*, vol. 11, no. 8, p. 1148, 2019.

[64] N. Beig *et al.*, "Radiogenomic analysis of hypoxia pathway is predictive of overall survival in Glioblastoma," *Scientific reports*, vol. 8, no. 1, p. 7, 2018.

[65] C. W. Law, Y. Chen, W. Shi, and G. K. Smyth, "voom: Precision weights unlock linear model analysis tools for RNA-seq read counts," *Genome biology*, vol. 15, no. 2, p. R29, 2014.

[66] G. Zararsiz *et al.*, "voomDDA: discovery of diagnostic biomarkers and classification of RNA-seq data," *PeerJ*, vol. 5, p. e3890, 2017.

[67] D. J. McCarthy, Y. Chen, and G. K. Smyth, "Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation," *Nucleic acids research*, vol. 40, no. 10, pp. 4288-4297, 2012.

[68] A. Oshlack, M. D. Robinson, and M. D. Young, "From RNA-seq reads to differential expression results," *Genome biology*, vol. 11, no. 12, pp. 1-10, 2010.

[69] W. Gardner, E. P. Mulvey, and E. C. Shaw, "Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models," *Psychological bulletin,* vol. 118, no. 3, p. 392, 1995.

[70] M. Zanfardino *et al.*, "Bringing radiomics into a multi-omics framework for a comprehensive genotype–phenotype characterization of oncological diseases," *Journal of translational medicine,* vol. 17, no. 1, p. 337, 2019.

[71] Z. Shboul and K. Iftekharuddin, *Efficacy of radiomics and genomics in predicting TP53 mutations in diffuse lower grade glioma* (SPIE Medical Imaging). SPIE, 2020.

[72] X. Zhu *et al.*, "Imaging-genetic data mapping for clinical outcome prediction via supervised conditional gaussian graphical model," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016, pp. 455-459: IEEE.

[73] V. Subramanian, M. N. Do, and T. Syeda-Mahmood, "Multimodal fusion of imaging and genomics for lung cancer recurrence prediction," *arXiv preprint arXiv:2002.01982,* 2020.

[74] W. B. Pope, J. Sayre, A. Perlina, J. P. Villablanca, P. S. Mischel, and T. F. Cloughesy, "MR imaging correlates of survival in patients with high-grade gliomas," *American Journal of Neuroradiology,* vol. 26, no. 10, pp. 2466-2474, 2005.

[75] 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, 2013.

[76] H. J. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications,* vol. 5, p. 4006, 2014.

[77] J. Guinney *et al.*, "Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: development of a prognostic model through a crowdsourced challenge with open clinical trial data," *The Lancet Oncology,* vol. 18, no. 1, pp. 132-142, 2017.

[78] F. Passamonti *et al.*, "A clinical-molecular prognostic model to predict survival in patients with post polycythemia vera and post essential thrombocythemia myelofibrosis," *Leukemia,* vol. 31, no. 12, p. 2726, 2017.

[79] R. Shouval *et al.*, "An Integrative Scoring System for survival prediction following umbilical cord blood transplantation in acute leukemia," *Clinical Cancer Research,* 2017.

[80] M. Kirienko *et al.*, "EP-1362: Random forest analysis to predict Disease-Free Survival using FDG-PET and CT in Lung Cancer," *Radiotherapy and Oncology,* vol. 127, pp. S743-S744, 2018.

[81] L. Macyszyn *et al.*, "Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques," *Neuro-oncology,* vol. 18, no. 3, pp. 417-425, 2015.

[82] H. Ohgaki, "Genetic pathways to glioblastomas," *Neuropathology,* vol. 25, no. 1, pp. 1-7, 2005.

[83] F. E. Bleeker, R. J. Molenaar, and S. Leenstra, "Recent advances in the molecular understanding of glioblastoma," *Journal of neuro-oncology,* vol. 108, no. 1, pp. 11-27, 2012.

[84] D. R. Johnson, H. E. Leeper, and J. H. Uhm, "Glioblastoma survival in the United States improved after Food and Drug Administration approval of bevacizumab: A population-based analysis," *Cancer,* vol. 119, no. 19, pp. 3489-3495, 2013.

[85] A. Vartanian *et al.*, "GBM's multifaceted landscape: highlighting regional and microenvironmental heterogeneity," *Neuro-oncology,* vol. 16, no. 9, pp. 1167-1175, 2014.

[86] D. Yang *et al.*, "Cerebral gliomas: prospective comparison of multivoxel 2D chemical-shift imaging proton MR spectroscopy, echoplanar perfusion and diffusion-weighted MRI," *Neuroradiology,* vol. 44, no. 8, pp. 656-666, 2002.

[87] K. E. Emblem *et al.*, "Glioma grading by using histogram analysis of blood volume heterogeneity from MR-derived cerebral blood volume maps," *Radiology,* vol. 247, no. 3, pp. 808-817, 2008.

[88] M. Nicolasjilwan *et al.*, "Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients," *Journal of Neuroradiology,* vol. 42, no. 4, pp. 212-221, 2015.

[89] P. Prasanna, J. Patel, S. Partovi, A. Madabhushi, and P. Tiwari, "Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings," *European radiology,* vol. 27, no. 10, pp. 4188-4197, 2017.

[90] S. Lacny *et al.*, "Kaplan–Meier survival analysis overestimates cumulative incidence of health-related events in competing risk settings: a meta-analysis," *Journal of clinical epidemiology*, vol. 93, pp. 25-35, 2018.

[91] Z. A. Shboul, L. Vidyaratne, M. Alam, and K. M. Iftekharuddin, "Glioblastoma and Survival Prediction," in *International MICCAI Brainlesion Workshop*, 2017, pp. International MICCAI Brainlesion Workshop 358-368: Springer.

[92] Z. A. Shboul, M. Alam, L. Vidyaratne, L. Pei, M. I. Elbakary, and K. M. Iftekharuddin, "Feature-Guided Deep Radiomics for Glioblastoma Patient Survival Prediction," *Frontiers in Neuroscience*, vol. 13, 2019.

[93] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International journal of computer vision*, vol. 43, no. 1, pp. 29-44, 2001.

[94] S. Reza and K. Iftekharuddin, "Multi-fractal texture features for brain tumor and edema segmentation," in *Medical Imaging 2014: Computer-Aided Diagnosis*, 2014, vol. Proc. SPIE 9035, p. 903503: International Society for Optics and Photonics.

[95] L. Vidyaratne, M. Alam, Z. Shboul, and K. Iftekharuddin, "Deep learning and texture-based semantic label fusion for brain tumor segmentation," in *Medical Imaging 2018: Computer-Aided Diagnosis*, 2018, vol. 10575, p. 105750D: International Society for Optics and Photonics.

[96] K. M. Iftekharuddin, W. Jia, and R. Marsh, "Fractal analysis of tumor in brain MR images," *Machine Vision and Applications*, vol. 13, no. 5-6, pp. 352-362, 2003.

[97] A. Islam, K. M. Iftekharuddin, R. J. Ogg, F. H. Laningham, and B. Sivakumar, "Multifractal modeling, segmentation, prediction, and statistical validation of posterior fossa tumors," in *Medical Imaging 2008: Computer-Aided Diagnosis*, 2008, vol. Proc. SPIE 6915, p. 69153C.

[98] S. Ahmed, K. Iftekharuddin, R. Ogg, and F. Laningham, "Efficacy of texture, shape, and intensity features for robust posterior-fossa tumor segmentation in MRI," in *Medical Imaging 2009: Computer-Aided Diagnosis*, 2009, vol. 7260, p. 726020: International Society for Optics and Photonics.

[99] L. Han and M. R. Kamdar, "MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks," in *Pacific Symposium on*

Biocomputing. Pacific Symposium on Biocomputing, 2018, vol. 23, pp. 331-342: World Scientific.

[100] J. Yu *et al.*, "Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma," *European radiology*, vol. 27, no. 8, pp. 3509-3522, 2017.

[101] H. Ding *et al.*, "Prediction of IDH Status Through MRI Features and Enlightened Reflection on the Delineation of Target Volume in Low-Grade Gliomas," *Technology in cancer research & treatment*, vol. 18, 2019.

[102] Z. Akkus *et al.*, "Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images Using Machine Intelligence," *Journal of Digital Imaging*, pp. 1-8, 2017.

[103] S. R. van der Voort *et al.*, "Predicting the 1p/19q co-deletion status of presumed low grade glioma with an externally validated machine learning algorithm," *Clinical Cancer Research*, p. 1127, 2019.

[104] J. Wang *et al.*, "Pretreatment MRI Radiomics Analysis Allows for Reliable Noninvasive Prediction of Survival and TERT Promoter Mutation in Lower-Grade Gliomas," *Available at SSRN 3487723*, 2019.

[105] A. Islam, S. M. Reza, and K. M. Iftekharuddin, "Multifractal texture estimation for detection and segmentation of brain tumors," *IEEE transactions on biomedical engineering*, vol. 60, no. 11, pp. 3204-3215, 2013.

[106] S. M. Reza, R. Mays, and K. M. Iftekharuddin, "Multi-fractal detrended texture feature for brain tumor classification," in *Medical Imaging 2015: Computer-Aided Diagnosis*, 2015, vol. Proc. SPIE 9414, p. 941410: International Society for Optics and Photonics.

[107] R. Meier, S. Bauer, J. Slotboom, R. Wiest, and M. Reyes, "A hybrid model for multimodal brain tumor segmentation," *Multimodal Brain Tumor Segmentation*, vol. 31, pp. 31-37, 2013.

[108] A. Pitiot, A. W. Toga, N. Ayache, and P. Thompson, "Texture based MRI segmentation with a two-stage hybrid neural classifier," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, 2002, vol. 3, pp. 2053-2058: IEEE.

[109] S. Bauer, T. Fejes, J. Slotboom, R. Wiest, L.-P. Nolte, and M. Reyes, "Segmentation of brain tumor images based on integrated hierarchical classification and regularization," in *MICCAI BraTS Workshop. Nice: Miccai Society*, 2012, p. 11.

[110] X. Feng, N. Tustison, and C. Meyer, "Brain Tumor Segmentation Using an Ensemble of 3D U-Nets and Overall Survival Prediction Using Radiomic Features," Cham, 2019, pp. 279-288: Springer International Publishing.

[111] P. John, "Brain tumor classification using wavelet and texture based neural network," *International Journal of Scientific & Engineering Research,* vol. 3, no. 10, pp. 1-7, 2012.

[112] N. Nabizadeh and M. Kubat, "Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features," *Computers & Electrical Engineering,* vol. 45, pp. 286-301, 2015.

[113] K. R. Kukurba and S. B. Montgomery, "RNA sequencing and analysis," *Cold Spring Harbor Protocols,* vol. 2015, no. 11, p. pdb. top084970, 2015.

[114] M. D. Robinson and G. K. Smyth, "Small-sample estimation of negative binomial dispersion, with applications to SAGE data," *Biostatistics,* vol. 9, no. 2, pp. 321-332, 2007.

[115] M. D. Robinson and G. K. Smyth, "Moderated statistical tests for assessing differences in tag abundance," *Bioinformatics,* vol. 23, no. 21, pp. 2881-2887, 2007.

[116] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics,* vol. 26, no. 1, pp. 139-140, 2010.

[117] S. Anders and W. Huber, "Differential expression analysis for sequence count data," *Genome biology,* vol. 11, no. 10, p. R106, 2010.

[118] Y. Di, D. W. Schafer, J. S. Cumbie, and J. H. Chang, "The NBP negative binomial model for assessing differential gene expression from RNA-Seq," *Statistical Applications in Genetics and Molecular Biology,* vol. 10, no. 1, 2011.

[119] K. Dong, H. Zhao, T. Tong, and X. Wan, "NBLDA: negative binomial linear discriminant analysis for RNA-Seq data," *BMC bioinformatics,* vol. 17, no. 1, p. 369, 2016.

[120] D. Yu, W. Huber, and O. Vitek, "Shrinkage estimation of dispersion in Negative Binomial models for RNA-seq experiments with small sample size," *Bioinformatics,* vol. 29, no. 10, pp. 1275-1282, 2013.

[121] R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu, "Class prediction by nearest shrunken centroids, with applications to DNA microarrays," *Statistical Science,* vol. 18, no. 1, pp. 104-117, 2003.

[122] L. E. Jalbert *et al.*, "Metabolic profiling of IDH mutation and malignant progression in infiltrating glioma," *Scientific reports*, vol. 7, p. 44792, 2017.

[123] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993-2024, 2015.

[124] S. Bakas *et al.*, "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection," *The Cancer Imaging Archive*, 2017.

[125] S. Bakas *et al.*, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, p. 170117, 2017.

[126] J. Reunanen, "Overfitting in making comparisons between variable selection methods," *Journal of Machine Learning Research*, vol. 3, no. Mar, pp. 1371-1382, 2003.

[127] M. Vallières, C. R. Freeman, S. R. Skamene, and I. El Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in medicine and biology*, vol. 60, no. 14, p. 5471, 2015.

[128] A. Ayache and J. L. Véhel, "Generalized multifractional Brownian motion: definition and preliminary results," ed: Springer, 1999.

[129] A. Ayache and J. L. Véhel, "On the identification of the pointwise Hölder exponent of the generalized multifractional Brownian motion," *Stochastic Processes and their Applications*, vol. 111, no. 1, pp. 119-156, 2004.

[130] Z. A. Shboul, S. M. Reza, and K. M. Iftekharuddin, "Quantitative MR Image Analysis for Brain Tumor," in *European Congress on Computational Methods in Applied Sciences and Engineering*, 2017, pp. 10-18: Springer.

[131] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794: ACM.

[132] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579-2605, 2008.

[133] M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Transactions on systems, man, and Cybernetics*, vol. 19, no. 5, pp. 1264-1274, 1989.

[134] Y. Wang *et al.*, "Patterns of tumor contrast enhancement predict the prognosis of anaplastic gliomas with IDH1 mutation," *American Journal of Neuroradiology*, 2015.

[135] M. Zhang, W. Ma, W. Wu, and A. Maher, "The MGMT promoter methylation is correlated with gene mutation of IDH1 and 1p19q in glioma," ed: AACR, 2016.

[136] M. C. Zlatescu *et al.*, "Tumor location and growth pattern correlate with genetic signature in oligodendroglial neoplasms," *Cancer research*, vol. 61, no. 18, pp. 6713-6715, 2001.

[137] X. Ren *et al.*, "Co-deletion of chromosome 1p/19q and IDH1/2 mutation in glioma subsets of brain tumors in Chinese patients," *PloS one*, vol. 7, no. 3, p. e32764, 2012.

[138] K. Wang *et al.*, "Regional specificity of 1p/19q co-deletion combined with radiological features for predicting the survival outcomes of anaplastic oligodendroglial tumor patients," *Journal of neuro-oncology*, pp. 1-9, 2017.

[139] J. C. Marioni, C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad, "RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays," *Genome research*, vol. 18, no. 9, pp. 1509-1517, 2008.

[140] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, "Mapping and quantifying mammalian transcriptomes by RNA-Seq," *Nature methods*, vol. 5, no. 7, p. 621, 2008.

[141] A. H. Bakas S, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C., "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Nature Scientific Data [in press]*, 2017.

[142] A. Colaprico *et al.*, "TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data," *Nucleic acids research*, vol. 44, no. 8, pp. e71-e71, 2015.

[143] D. A. Gutman *et al.*, "Somatic mutations associated with MRI-derived volumetric features in glioblastoma," *Neuroradiology*, vol. 57, no. 12, pp. 1227-1237, 2015.

[144] P. Metellus *et al.*, "Absence of IDH mutation identifies a novel radiologic and molecular subtype of WHO grade II gliomas with dismal prognosis," *Acta neuropathologica*, vol. 120, no. 6, pp. 719-729, 2010.

[145] Y. 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, 2018.

[146] S. Thust *et al.*, "Apparent diffusion coefficient for molecular subtyping of non-gadolinium-enhancing WHO grade II/III glioma: volumetric segmentation versus two-dimensional region of interest analysis," *European radiology*, pp. 1-10, 2018.

[147] A. Maufroy, E. Chassot, R. Joo, and D. M. Kaplan, "Large-scale examination of spatio-temporal patterns of drifting fish aggregating devices (dFADs) from tropical tuna fisheries of the Indian and Atlantic Oceans," *PloS one,* vol. 10, no. 5, 2015.

[148] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PloS one,* vol. 14, no. 11, 2019.

[149] L. Pan *et al.,* "Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia," *Scientific reports,* vol. 7, no. 1, pp. 1-9, 2017.

[150] W. Yang *et al.,* "Sex differences in GBM revealed by analysis of patient imaging, transcriptome, and survival data," *Science translational medicine,* vol. 11, no. 473, p. eaao5253, 2019.

[151] Y. Yuan *et al.,* "Comprehensive characterization of molecular differences in cancer between male and female patients," *Cancer cell,* vol. 29, no. 5, pp. 711-722, 2016.

[152] J. E. Ippolito, A. K.-Y. Yim, J. Luo, P. Chinnaiyan, and J. B. Rubin, "Sexual dimorphism in glioma glycolysis underlies sex differences in survival," *JCI insight,* vol. 2, no. 15, 2017.

[153] M.-Y. Li *et al.,* "Isocitrate dehydrogenase 1 gene mutation is associated with prognosis in clinical low-grade gliomas," *PloS one,* vol. 10, no. 6, 2015.

[154] D. J. Brat *et al.,* "Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas," *New England Journal of Medicine,* vol. 372, no. 26, pp. 2481-2498, 2015.

[155] A. Ebrahimi *et al.,* "ATRX immunostaining predicts IDH and H3F3A status in gliomas," *Acta neuropathologica communications,* vol. 4, no. 1, p. 60, 2016.

[156] T. Coolen-Maturi and F. Coolen, "Non-parametric predictive inference for the validation of credit rating systems," *Journal of the Royal Statistical Society: Series A (Statistics in Society),* vol. 182, no. 4, pp. 1189-1204, 2019.

[157] P. Lambin *et al.,* "Radiomics: extracting more information from medical images using advanced feature analysis," *European journal of cancer,* vol. 48, no. 4, pp. 441-446, 2012.

[158] R. Jain *et al.,* "Outcome prediction in patients with glioblastoma by using imaging, clinical, and genomic biomarkers: focus on the nonenhancing component of the tumor," *Radiology,* vol. 272, no. 2, pp. 484-493, 2014.

[159] Bakas S *et al.,* "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection," *The Cancer Imaging Archive,* 2017.

[160] S. Rathore, H. Akbari, M. Rozycki, S. Bakas, and C. Davatzikos, "NIMG-20. IMAGING PATTERN ANALYSIS REVEALS THREE DISTINCT PHENOTYPIC SUBTYPES OF GBM WITH DIFFERENT SURVIVAL RATES," ed: Oxford University Press US, 2016.

[161] A. Materka and M. Strzelecki, "Texture analysis methods–a review," *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, pp. 9-11, 1998.

[162] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE,* vol. 67, no. 5, pp. 786-804, 1979.

[163] X. Tang, "Texture information in run-length matrices," *IEEE transactions on image processing,* vol. 7, no. 11, pp. 1602-1609, 1998.

[164] K. Turner, S. Mukherjee, and D. M. Boyer, "Persistent homology transform for modeling shapes and surfaces," *Information and Inference: A Journal of the IMA,* vol. 3, no. 4, pp. 310-344, 2014.

[165] L. Breiman, *Random forests* (Machine learning, no. 1). 2001, pp. 5-32.

[166] L. Chato, E. Chow, and S. Latifi, "Wavelet Transform to Improve Accuracy of a Prediction Model for Overall Survival Time of Brain Tumor Patients Based On MRI Images."

[167] P.-Y. Kao, T. Ngo, A. Zhang, J. Chen, and B. Manjunath, "Brain Tumor Segmentation and Tractographic Feature Extraction from Structural MR Images for Overall Survival Prediction," *arXiv preprint arXiv:1807.07716,* 2018.

[168] M. Soltaninejad, L. Zhang, T. Lambrou, G. Yang, N. Allinson, and X. Ye, "MRI Brain Tumor Segmentation and Patient Survival Prediction Using Random Forests and Fully Convolutional Networks," in *International MICCAI Brainlesion Workshop,* 2017, pp. 204-215: Springer.

[169] A. Pierallini *et al.,* "Radiological assessment of necrosis in glioblastoma: variability and prognostic value," *Neuroradiology,* vol. 40, no. 3, pp. 150-153, 1998.

[170] M. Lacroix *et al.,* "A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival," *Journal of neurosurgery,* vol. 95, no. 2, pp. 190-198, 2001.

[171] M. V. Maldaun *et al.,* "Cystic glioblastoma multiforme: survival outcomes in 22 cases," *Journal of neurosurgery,* vol. 100, no. 1, pp. 61-67, 2004.

VITA

Zeina A. Shboul
Department of Electrical and Computer Engineering
Old Dominion University
Norfolk, VA 23529

Zeina A. Shboul joined the Vision Lab at Old Dominion University in Fall 2013. She received a Master's degree in Electrical Engineering in 2011 from Jordan University of Science and Technology in Irbid, Jordan and a Bachelor's degree in Electronics Engineering in 2007 from Yarmouk University in Irbid, Jordan. The motivation of her current research is to develop efficient non-invasive methods for Glioma grading and classification and predicting overall survival.

Her research interests include biomedical image processing, gene expression analysis, molecular subtyping, machine learning, and computer vision.