Robust Brain Tumor Segmentation and Overall Survival Prediction

Thesis

Submitted in partial fulfillment of the requirement of

Doctor of Philosophy

by

Snehal Rajput

Enrollment No. 19RCP002

Under the guidance of

Mohendra Roy (ICT Department)

Mehul S. Raval (Ahmedabad University)

Department of Computer Science and Engineering

School of Technology

Pandit Deendayal Energy University

Gandhinagar – 382426. Gujarat - India

June, 2024

Approval Sheet

This thesis entitled **"Robust Brain Tumor Segmentation and Overall Survival Prediction"** by **Snehal Rajput** is recommended for the degree of Doctor of Philosophy.

Examiner

Dr. Jignesh N. Sarvaiya
Professor & Head of the Department
Department of Electronics Engineering
Sardar Vallabhbhai National Institute of Technology

Supervisor

Dr. Mohendra Roy
Assistant Prof., Department of ICT
School of Technology
Pandit Deendayal Energy University

Co-Supervisor

Dr. Mehul S. Raval
Prof. & Associate Dean, Experiential Learning
School of Engineering and Applied Science
Ahmedabad University

Chairman

Dr. Dhaval Pujara
Prof. & Director, School of Technology
Pandit Deendayal Energy University

Date:_____

Place:_____

Student Declaration

I, **Snehal Rajput**, affirm that the following written submission reflects solely my thoughts and expressions. Any inclusion of others' ideas or language is properly acknowledged through citations and references to the original sources. Furthermore, I attest to upholding the principles of academic honesty and integrity, ensuring that no misrepresentation, fabrication, or falsification of any idea, data, fact, or source has occurred in this submission. I am aware that failure to adhere to these standards may result in disciplinary measures from Pandit Deendayal Energy University and potential legal repercussions from improperly cited sources or those from whom consent was not obtained when necessary.

_____

Snehal Rajput

Enrollment No.: 19RCP002

Date: _____

Acknowledgment

I am extremely grateful for the opportunity to pursue my doctoral studies in the Computer Science and Engineering (CSE) department at the School of Technology (SoT) of Pandit Deendayal Energy University (PDEU), Gandhinagar. This doctoral journey has enlightened me on the process and impact of conducting good research in respective domains.

This journey has instilled in me a deeper understanding of the significance and potential impact of my research domains, as well as how I can contribute to society through my research efforts, both as an individual and through collaborative research endeavors. Additionally, I have acquired skills in clarity and articulation of thought, particularly in research writing knowing its profound impact. Viewed through the lens of personal growth, this doctoral journey has imparted invaluable lessons including discovering my strength by pushing my limit every time I have been knocked down, and taught me the profound magnitude of perseverance, hard work and manifestation. I was and am in love with research because as a researcher I was occupied with strategic plans to overcome challenges or chase personal or compelling journal deadlines.

This doctoral journey has been one of the most challenging yet enlightening phases of my life. Along this path, I have had the privilege of meeting numerous individuals—from fellow researchers to esteemed figures in my field and beyond—who have served as pillars of inspiration and motivation for my work. I have also learned the power of humility and kindness from them. At this moment, I want to express my heartfelt gratitude to all those who have supported and encouraged me throughout the process of working on this thesis.

First, I express my deep gratitude towards my mentors **Dr. Mehul S. Raval** and **Dr. Mohendra Roy**. My journey started with Dr. Raval, and one year later, Dr. Roy joined as a mentor. Their guidance, expertise, and unwavering support have been instrumental throughout my doctoral journey. I am profoundly grateful for their mentorship, encouragement, and invaluable insights, which have significantly contributed to the completion of this doctoral journey. I truly look up to them as a source of energy, continually inspiring me to persevere, and their wisdom has been a guiding light in my research domains. Moreover, I have learned from them the importance of numerous virtues, which contribute to their exceptional personalities, and I am striving to absorb and embody them. A special thanks to my external charismatic supervisors **Prof. Suman Mitra**, Professor of Computer Science at Dhirubhai Ambani Institute of Information and Communication Technology, Ganghinagar, India and **Prof. Mukesh A. Zaveri**, Professor of Computer Science at S. V. National Institute of Technology, Surat, India, for their unwavering support, constant suggestions, and guiding direction, which have greatly contributed to the improvement of my research. Further, I would like to extend my gratitude to another mentor, **Dr. Rupal Kapdi**, who has provided unwavering support and invaluable advice throughout my journey.

I extend my sincere gratitude to **Prof. S. Sundar Manoharan** (DG, PDEU), **Prof. Dhaval Pujara** (Director, SoT), and **Prof. Surendra Singh Kachhwaha** (Dean, SoT) for their invaluable support and guidance. I am also thankful to **Dr. Shakti Mishra** (HoD, CSE), **Dr. Ganga Prasad Pandey** (HoD, ICT) for their continuous encouragement. The insightful suggestions and unwavering support from the respected Doctoral committee panel members have been instrumental in shaping my journey. Your confidence in me has been a driving force, inspiring me

to strive for excellence in my endeavors.

At this moment, I cannot overlook the opportunity to extend my heartfelt gratitude to my family members and my dear friends, who are truly the gems of my life. Dear **Ram Achal Singh (Dad)**, you have been the hero of my life, a constant source of guidance, inspiration, and encouragement. Your unwavering support has always urged me to follow my heart and strive for excellence. **Baijanti Singh (Mother)**, thank you for your unconditional love, support and infinite sacrifices throughout my life. Gratitude to **Pavan Arya (Husband)** for steadfastly standing by my side through thick and thin, serving not only as my partner but also as my best friend, advisor, and pillar of support. Thanks to my elder brothers cums friends **Satyendra Rajput** (my role model) and **Shailendra Rajput** from whom I drew my inspiration to strive and excel.

Lastly, I would like to express my gratitude to my dear friends who bring laughter and joy into my life: Irfan Ayubi, Megha Suman, (To be Dr.) Ravi Kumar Ranjan, (To be Dr.) Kumkum Bhattacharjee, (To be Dr.) Varsha Kheradiya. Also thanks to Nilesh Bhai, Chirag Bhai, and Shailesh bhai, lab staff members (PDEU), for their respect and acknowledgment by addressing me as Dr. Snehal Rajput even before the completion of my doctoral degree, empowering me to navigate this journey. Lastly, I would like to express my appreciation to all the supporters, who have stood by me and aided me in various ways.

Abstract

Glioblastoma, also known as Gliomas, is a brain tumor and is graded as a type-IV tumor by the World Health Organization. This tumor is well known for its low survival rate and poor prognosis. It originates from glial cells and is the most prevalent form of neoplasm in the central nervous system. Annually, six cases of gliomas are reported per 100,000 individuals in the United States. Despite significant technological advancements, a definitive cure is not yet available. The current treatment approach involves the most extensive safe removal of the tumor, followed by simultaneous administration of radiation and chemotherapy. Currently, experts delineate tumors in each image slice for performing biopsy or resection across all the planar views (axial, coronal, and sagittal), which is laborious, lethargic, and subjective. In contrast, for predicting survival days (SD), the current practice includes multiple pathological and histological examinations of specimens obtained during surgery.

Implementing automated brain tumor segmentation (BTS) can significantly assist oncologists in optimizing treatment strategies, improving patient outcomes, and ensuring better overall care. Furthermore, it allows for efficient utilization of resources and reduces the potential for human error. Additionally, the prediction of SD is vital for tailoring treatment plans to each patient's unique needs and condition. This predictive data guides decisions regarding the timing and intensity of surgical interventions, selecting suitable therapeutic options, and managing post-operative care. Automation in the prediction of SD can aid clinicians in estimating a patient's prognosis and potential survival time, allowing them to make

well-informed selections about suitable treatment strategies and patient care. Developing an efficient, easily accessible, and robust method for BTS and SD prediction is essential and challenging in the current landscape. Therefore, this thesis focuses on formulating a robust methodology for BTS and SD prediction that is well-equipped to function effectively within a resource-constrained environment (with limited computing and memory resources).

We employed a triplanar (2.5D) network, which balances 2D and 3D segmentation networks regarding performance and computational requirements. We integrated lightweight attention mechanisms to enhance segmentation performance and mitigate the increase in trainable parameters. This integration improved the Dice score of the concurrent channel and spatial attention (CCSAv1) network for enhancing tumor (ET), increasing it from 0.618 to 0.679. Notably, the number of trainable parameters increased by $0.256, M$, reaching a total of $10.589 M$. Further, we post-processed the segmentation to remove false positive lesions. The Dice similarity coefficient for the Brain tumor segmentation (BraTS2020) challenge training set is 0.736 for enhancing tumor (ET) lesions, 0.896 for whole tumor (WT) lesions, and 0.841 for tumor core (TC) lesions. On the BraTS2020 validation set, these scores stand at 0.713 (ET), 0.873 (WT), and 0.778 (TC), which are comparable to those achieved by leading 3D models like the method proposed by Soltanineja et al., which achieved scores of 0.660 (ET), 0.870 (WT), and 0.800 (TC).

Again, for SD prediction, we extracted a comprehensive set of radiomics-based and image-based features. Subsequently, we conducted feature selection to identify a set of 29 dominant features. We trained ensemble-based learning models, e.g., random forest (RF) regressor and extra-random forest regressor (ERFR)

models, to predict the SD of patients. We validated the robust nature of the feature set by obtaining a correlation map and Spearman ranking coefficient (SRC). The best scores on the BraTS2020 training set are 0.504 (accuracy), 59927.38 (mean squared error (MSE)), 20101.86 (median squared error (medianSE)), 107835.95 (standard deviation of the squared error(stdSE)), and 0.725 SRC. On the validation set, the scores are 0.586 (accuracy), 76529.43 (MSE), 41402.78 (medianSE), 130978.47 (stdSE) and 0.52 (SRC). The performance of the proposed SD predictions on the validation set is better than cutting-edge techniques on multiple metrics, such as McKinley et al., who achieved scores of 0.414 (accuracy), 098704.66 (MSE), 36100.00 (median SE), 152176.00 (stderr), 0.253 (SRC).

Despite the remarkable achievements of machine learning (ML) models, their integration into the medical field is constrained, primarily because of the profound consequences on human life and the inherent black-box nature of these models. Therefore, to understand the decisions made by blackbox ML models, we used various post-hoc interpretation methods. The impact of the features is investigated on two levels: the overall sample or generic behavior and the individual sample or local behavior. Techniques such as Shapley additive explanations (SHAP), partial dependency plots (PDP), and accumulated local-effect (ALE) are utilized for observing generic behavior. Conversely, methods including SHAP, SHAP-force plot, SHAP waterfall plot, and local-interpretable model-agnostic explanations (LIME) are employed to explore the local impact.

Explanations obtained from these analyses were then utilized to extract biological significance from the features. The insights drawn from the visual graphs were consistent with the existing understanding in the medical domain. The feature set under investigation primarily comprises radiomic features renowned for

capturing heterogeneity and indicating tumor malignancy through pixel intensity. However, these features have limitations in directly correlating with tumor malignancy, which is more closely tied to the molecular landscape. To address this, we conducted an analysis linking radiomic features with gene expression in glioma patients, allowing for direct biological inference. This area is known as radiogenomic. Our radiogenomic study examined gene expressions with both strong correlations and those exhibiting uncorrelated characteristics. Additionally, we explored gene expressions exhibiting behaviors akin to our feature set. When gene expressions correlate with radiomic features, it validates the relevance of those features in capturing tumor characteristics or biological processes. Moreover, identifying gene expressions that mimic radiomic features may lead to discovering biomarkers for predicting tumor behavior, prognosis, or treatment response.

Contents

List of Figures

List of Tables

Abbreviations

Adagrad ..... adaptive gradient algorithm

AI ..... Artificial intelligence

ALE ..... Accumulated local effect

AP ..... Axial planar

AT ..... Active tumor

AUC ..... Area under the curve

BTS ..... Brain tumor segmentation

BraTS Challenge ..... Brain tumor segmentation challenge

CA ..... Channel attention

CBTRUS ..... The central brain tumor registry of the United States

CCSA ..... Concurrent channel and Spatial Attention

CCSAv1 ..... Concurrent channel and Spatial Attention version 1

CP ..... Coronal planar

CNN ..... Convolution neural network

CT ..... Computerized axial tomography

DA ..... Discriminant analysis

DNNs ..... Deep neural networks

DSC ..... Dice similarity coefficient

DTI ..... Diffusion tensor imaging

DWI ..... Diffusion-weighted imaging

EEG ..... Electroencephalogram

ELU ..... Exponential linear unit

ET ..... Enhancing tumor

ExtraRF . . . . . Extra-random forest regressor

ERP . . . . . Evoked related potentials

FC . . . . . Fully connected layer

FCM . . . . . Fuzzy C-means clustering

FLAIR . . . . . Fluid attenuated inversion recovery

FL . . . . . Focal loss

FO . . . . . First-order statistical

FOK . . . . . wavelet-HHH_firstorder_Kurtosis

fMRI . . . . . functional MRI

GAP . . . . . Global average pooling

GBR . . . . . Gradient boosting regressor

GDL . . . . . Generalized Dice loss

GLCM . . . . . Gray- level co-occurrence matrix

Glcm_Corr . . . . . log-sigma-1-0-mm-3D_glcm_Correlation

GLDM . . . . . Gray- level dependence matrix

GLSZ . . . . . Gray-level size-zone

GLRLM . . . . . Gray-level run-length matrix

GPU . . . . . Graphics processing unit

GTR . . . . . Gross total resection

HD . . . . . Hausdorff distance

HGG . . . . . High-grade gliomas

KM . . . . . Kaplan Meier

KPS . . . . . Karnofsky performance score

LFK . . . . . log-sigma-2-0-mm3D_firstorder_Kurtosis

LGG . . . . . Low-grade gliomas

LIME ..... Local-interpretable model-agnostic Explanations

LG ..... Logistic regression

LoG ..... Laplacian of gaussian

lncRNA ..... Long non-coding RNA

LReLU ..... Leaky ReLU

MSE ..... Mean squared error

MedianSE ..... Median squared error

ML ..... Machine learning

MLP ..... Multi-layered perceptron

MRF ..... Markov random field

MRI ..... Magnetic resonance imaging

MRSI ..... Magnetic resonance spectroscopic imaging

NCR ..... Necrosis

NET ..... Non-enhancing tumor

NGTDM ..... Neighboring gray-tone difference matrix

NLP ..... Natural language processing

PC ..... Pearson correlation

PET ..... Positron emission tomography

PI ..... Permutation importance

PRelu ..... Parametric ReLU

PD ..... Partial dependency

PDP ..... Partial dependency plot

PPM ..... Personalized, patient-tailored medicine

ReLU ..... Rectified Linear Unit

RMSProp ..... Root means square propagation

TA ..... Texture analysis

TCGA ..... The cancer genome atlas

T1-ce ..... Post- contrast T1-weighted

ROC-AUC ..... Receiver operating characteristic - area under the curve

RF ..... Random forest

RFR ..... Random forest regressor

ResNET ..... Residual network

SA ..... Spatial attention

SE ..... Squeeze and excitation

SP ..... Sagittal planar

SCSA ..... Sequential channel and spatial attention

SD ..... Survival days

SHAP ..... Shapley additive explanations

SRC ..... Spearman ranking coefficient

StdSE ..... Standard deviation squared error

TC ..... Tumor core

WCE ..... Weighted cross-entropy

WIR ..... Wavelet-LLL_firstorder_InterquartileRange

WHO ..... World health organization

WT ..... Whole tumor

Xtreme gradient boosting ..... XGB

Chapter 1

Introduction

This section provides essential information on brain tumors, including their characteristics and diagnostic imaging methods. It then outlines the problem statement, research objectives, and the dataset employed.

1.1 Brain tumors with their traits

A brain tumor is an aggregation of tissue formed by the uncontrolled growth of cells in the brain. More than 150 different forms of brain tumors have been documented, falling into two main categories known as primary and metastatic /secondary brain tumors. [AANS (2023)].

Primary brain tumors develop from the brain or their adjacent local tissues. These tumors are categorized as glial (comprising glial cells) or nonglial (arising within or on brain structures such as nerve fibers, blood channels, and glands) and can be classified as benign or malignant. Conversely, metastatic brain tumors stem in other regions of the body, such as the breast or lungs, before spreading

to the brain, frequently through the blood channels. These tumors are identified as cancerous and are classified as malignant. The nature of the tumor can be exhibited in Fig. 1.1.

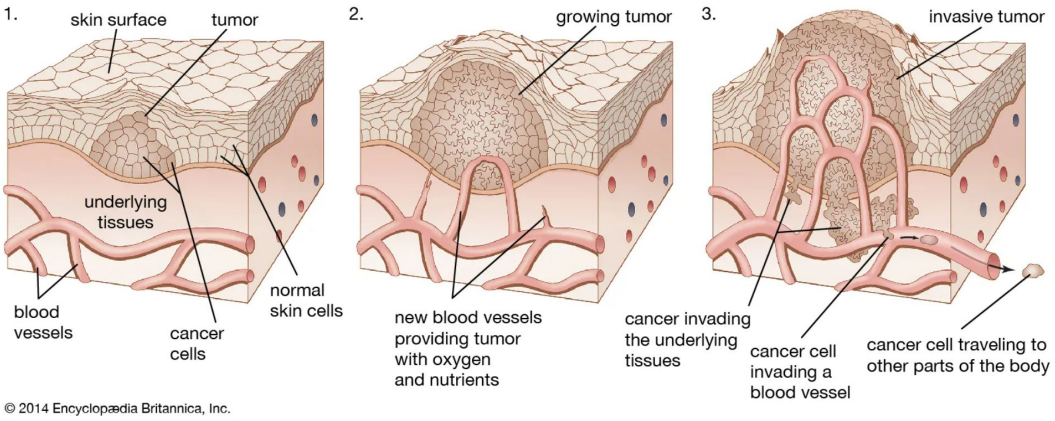


Figure 1.1. Tumor development and its evolution into complex stages [of Encyclopaedia Britannica (2024)].

The World health organization (WHO) devised a grading scheme to determine the malignancy or benignity nature of a tumor based on histological characteristics, including high malignancy, extensive invasiveness, rapid growth, susceptibility to aggressive necrosis, and rapid recurrence, which can be seen in Table1.1. Further, as per the Central Brain Tumor Registry of the United States (CBTRUS) statistical report, approximately 1 million Americans are believed to be living with a primary brain tumor, and about 28% of all brain tumors are malignant [Ostrom et al. (2022)].

Glioblastoma is a Grade-4 tumor and represents the most frequently occurring primary malignant brain tumor. It comprises over 60% of all brain tumors diagnosed in adults and accounts for approximately 80% of all malignant tumors [Hanif et al. (2017)]. The five-year survival rate is only 6.9%, with a me-

dian survival period of just eight months [*National Brain Tumor Society - Brain Tumor Facts* (n.d.)]. In the case of a new diagnosis, the typical treatment approach involves surgery (extent of resection) followed by concurrent radiation and chemotherapy. Resection is a surgical procedure aimed at eliminating a tumor, while "residual" pertains to the portion of the tumor that remains within the body post-surgery. A more favorable patient prognosis is typically associated with complete surgical removal of the entire tumor. Whenever the surgical team and the patient assess the risks as acceptable, it is advisable to consider surgery to remove all or most of the tumor. These surgical outcomes are categorized into four classes:

- Gross total resection (GTR): This denotes the removal of the entire visible tumor, but microscopic cells may remain.
- Near total or subtotal resection: In this case, a substantial portion of the tumor is excised.
- Partial: Only a part of the tumor is surgically removed.
- Biopsy only: This category involves removing a small tumor section exclusively for diagnostic testing.

Glioblastoma comprises multiple tumor lesions, including Enhancing/Active tumor (ET), Tumor core (TC), and Whole tumor (WT). Delineating gliomas into these distinct lesions is inherently difficult due to the extreme intrinsic heterogeneity in tumor boundaries, appearance, histology, shape, and location. Brain tumors, including gliomas, pose significant diagnostic challenges, are difficult to treat, and often exhibit resistance to conventional therapies. A significant difficulty arises

from the complexities of delivering drugs to the brain and the inherently high heterogeneity of these tumors across radiographic, morphologic, and molecular landscapes. Regrettably, there is no established treatment for recurrence. Despite this, depending on the patient's condition, operative intervention, radiation therapy, immunotherapy, chemotherapy, or systemic treatment with angiogenesis inhibitors are all potential avenues for treatment. [Kim & Lee (2022)].

Accurate segmentation, however, remains pivotal in strategically planning and monitoring treatment. Manual segmentation demands extensive medical knowledge, consumes significant time, and is susceptible to human error. Furthermore, the manual procedure lacks uniformity and the ability to be replicated, adversely impacting outcomes and potentially resulting in inaccurate prognoses and treatments. Typically, a radiologist takes more than 7 hours to label one such sample [G. Wu et al. (2021)]. Therefore, it is unsurprising that there is a strong demand for efficient and effective automatic segmentation techniques. Numerous machine learning (ML) approaches have been suggested for tumor segmentation and overall survival prediction. Nevertheless, these techniques often operate as black-box models, offering no insight into their decision-making processes. Strengthening the transparency and interpretability of these deep learning methods is vital for their widespread adoption in clinical applications.

TABLE 1.1
WHO Brain tumor types and grading [Louis et al. (2021)].

	Tumor Grade	Types	Traits
Low grade	Grade I	• Craniopharyngioma • Chordomas • Pilocytic astrocytoma • Gangliocytoma • Ganglioglioma	• Potentially treatable through surgical intervention alone • Long survival period • Least malignant/(benign) • Non-invasive
	Grade II	• Pineocytoma • "Diffuse" astrocytoma • Pure oligodendroglioma	• Mildly invasive • moderate tumor progression • May reoccur at a more advanced grade
High Grade	Grade III	• Anaplastic ependymoma • Anaplastic astrocytoma • Anaplastic oligodendroglioma	• Malignant • Invasive • Tend to recur as higher grade
	Grade IV	• Glioblastoma multiforme • Medulloblastoma • Ependymoblastoma • Pineoblastoma	• Most lethal • Rapid progression • Extensively invasive • Relapse of tumor • Likelihood for necrosis

1.2 Diagnostic brain scans/ Imaging modalities

Brain imaging, also known as neuroimaging, encompasses diverse techniques to directly or indirectly envision the brain's structure, function, or pharmacology [BrainLine (2023)]. By employing these methods, neuroscientists gain insight

into the inner workings of the living brain, enabling a deeper understanding of functions associated with specific brain regions. Brain MRI scans from a magnetic field intensity of 11.7 tesla were released in April 2024. The images exhibit remarkable resolution considering the brief acquisition time with 0.2 mm in-plane resolution and 1 mm slice thickness, corresponding to a volume equivalent to several thousand neurons. To achieve similar image quality, conventional MRI scanners available in medical facilities (1.5 or 3 tesla) would typically require hours of scanning [CEA (French Alternative Energies and Atomic Energy Commission) (April 2024)].

Brain imaging can be divided into two primary categories:

- Structural imaging - focuses on the brain's structure and aids in diagnosing significant intracranial diseases, such as tumors and injuries. Examples include Magnetic resonance imaging (MRI), Diffusion-based MRI (such as Diffusion tensor imaging (DTI)), and Computerized axial tomography (CT).

- Functional imaging - assesses various aspects of brain function to elucidate the connections between brain activity in specific regions and particular mental processes. This type of imaging primarily serves as a research tool in the fields of cognitive neuroscience and neuropsychology. Examples of this type of imaging are functional MRI (fMRI), Electroencephalogram (EEG), Positron emission tomography (PET), and Evoked-related potentials (ERP).

In recent decades, MRI has emerged as a widely employed technique for assessing brain tissues in clinical settings. The notable advantages of MRI encompass its non-invasiveness, non-exposure to ionized radiation, and capability to pro-

vide highly detailed images of the brain. This is facilitated by superior spatial resolution resulting from strong magnetic gradients. This heightened resolution is possible due to the distinct responses of tissues within grey and white matter and other brain parts to these magnetic gradients. This differentiation identifies various tissue types and structures [Hirsch et al. (2015)]. Unlike CT scans, the signal characteristics in MRI vary due to diverse acquisition protocols. Tumor cells exhibit varying intensity distributions when captured by different scanners, each having differences in field strength, voxel resolution, and field of view. MRI in three distinct clinical diagnosis plans, including the axial plane, sagittal plane, and coronal, is depicted in Figure 1.2.

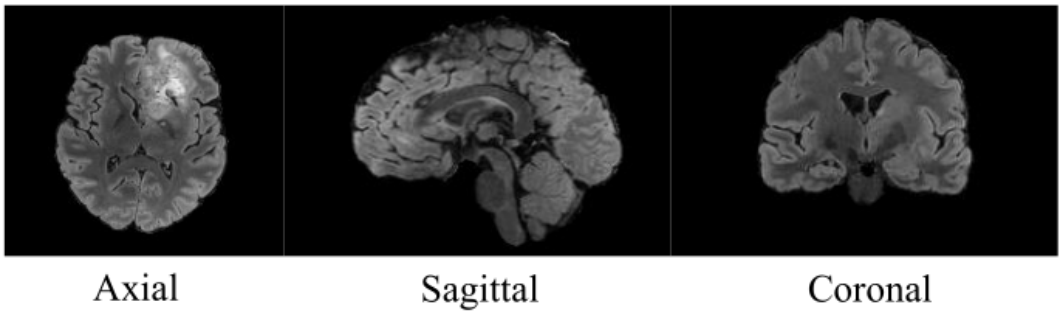


Figure 1.2. MRI in three diagnostic planar views: axial, sagittal, and coronal. The image is sourced from the BraTS dataset, acquired utilizing the ITK-SNAP software application [Yushkevich et al. (2006)].

The advantages of multimodal MRI images (including T1 (native-T1), T2-weighted, post-contrast (T1-ce), and fluid attenuation inversion recovery (FLAIR)) lie in their ability to provide a comprehensive view of the examined area by combining information from different MRI techniques or sequences. This approach enhances diagnostic accuracy and provides a more detailed and nuanced understanding of tissue characteristics. Multimodal MRI can offer improved sensitiv-

ity, specificity, and overall diagnostic capability compared to individual imaging modalities, allowing for a more thorough medical diagnosis and research assessment.

In the T1-ce modality, the ET lesion exhibits hyper-intensity, while the TC, including the ET and necrotic and non-enhancing parts, is also visible. The necrotic (NCR) and non-enhancing (NET) tumor core can be observed in the T1-ce modality, and the whole tumor (WT), encompassing the TC and edema (ED), is visible in FLAIR, with the ED appearing as a hyper-intense signal.

1.3 Motivation

Since the median SD are meager, accurate brain tumor segmentation and SD predictions are essential for various medical applications, including surgery and healthcare planning. Manual tracing by radiologists is inherently subjective and becomes impractical with a growing number of patients. This highlights the need for automated segmentation and SD prediction, especially in developing countries like India, which is heavily dense. Here, healthcare is strained, and the doctor-to-patient ratio is significantly lower for severe diseases like cancer, which require costly treatments. Similarly, predicting SD is a task that is both laborious and susceptible to error, yet it is crucial. In such cases, technology intervention becomes imperative in aiding experts with decision-making and surgical planning, facilitating more comprehensive and affordable treatments.

Therefore, we developed a methodology for BTS and SD that balances performance and computational resources. This technological support not only assists experts in their battle against cancers and in enhancing survival rates but also

serves densely populated countries by making quality healthcare more accessible. However, incorporating technology into the medical domain is challenging, considering its impact on human life. Understanding the rationale behind the model's decision is crucial before integrating it into practice. Therefore, we investigate the interpretability of decision-making of these ML models for SD prediction.

1.4 Problem specifications

The problem statement is "Achieving robust automatic segmentation of brain tumors and predicting survival days using volumetric multimodal MRI images." This statement breaks down into two tasks:

1. Robust automated segmentation of brain tumor-lesions in volumetric pre-operative multiparametric MRI scans [Menze et al. (2014)]. This involves segmenting tumors into ET, WT, and TC regions. The TC comprises NET/NCR and ET, while the WT encompasses the TC and edema, as illustrated in Figure 1.3.

2. SD prediction of patients is based on segmentation of tumors. This is further classified into three categories [Menze et al. (2014)]:

   (a) Short-term survivor, if the SD is < 300 days.

   (b) Mid-term survivor, if the SD is between 300 and 450 days.

   (c) Long-term survivor, if the SD is > 450 days.

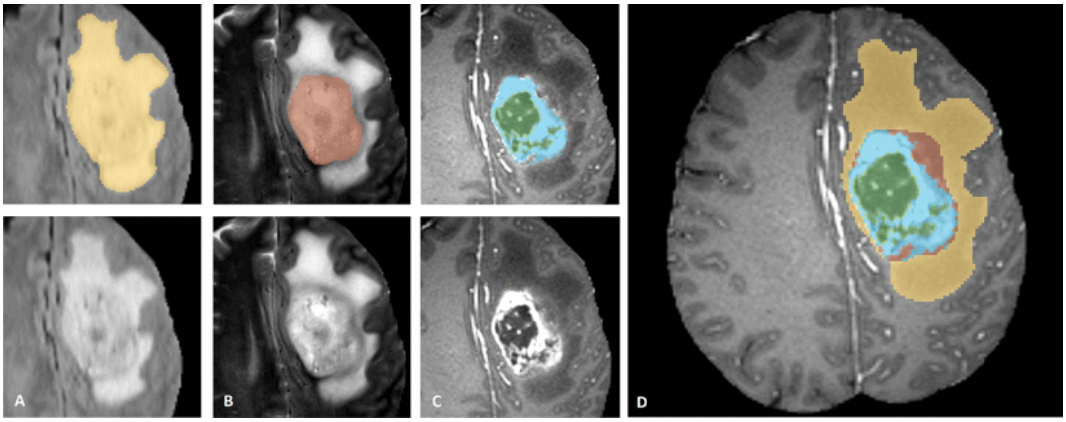


Figure 1.3. Clinical experts conducted manual annotation in the dataset. Progressing from left to right in the image patches: the entire tumor appears in FLAIR-MRI (A), the tumor core is evident in T2-MRI (B), and the enhancing tumor lesions can be observed in T1-ce (depicted in blue) encircling the necrotic components of the core (depicted in green) (C). These segmented tumor lesions are combined, shown in (D): edema (depicted in yellow), non-enhancing (solid tissue) (depicted in red), necrotic (fluid tissue) (depicted in green), and enhancing core (depicted in blue). [Menze et al. (2014)].

1.5 Research intents

This research aims to enhance network architecture, striving for reduced computational and memory requirements without compromising robust segmentation outcomes. The segmentation outcomes are refined using a post-processing technique. These refined tumor segmentation results will then be utilized to predict patients' SD precisely and to extract visual insights into the behavior of features and their significance in SD prediction. This process facilitates a deeper understanding of decision-making mechanisms employed by ML models in predicting SD. The specific objectives of this research are outlined below:

- Develop a robust network framework for segmenting brain tumors in resource-constrained environments.

- Develop an algorithm for accurately predicting the SD of patients diagnosed with Glioblastoma brain tumors.
- Deriving visual explanations and extracting insights from the decision-making process of the ML models for SD prediction.
- Deriving biological significance from the feature set used for SD prediction and connecting with genomic data (radio-genomics).

1.6 Data sources

The development of robust BTS and accurate prediction of SD heavily rely on the quality and availability of datasets. In the medical domain, obtaining standard and ample datasets is highly challenging. There are currently few 3D large volume datasets such as the BraTS (since 2012) [Bakas et al. (2017, 2018); Menze et al. (2014)], synthetic whole-head MRI BTS dataset [Dorjsembe & Xiao (2023)]. BraTS segmentation on sub-Sahara Africa population for adult glioma (since 2023) [Adewole et al. (2023)], and segmentation on pediatric tumors (since 2023) [Kazerooni et al. (2023)].

The BraTS, introduced in 2012, addresses this challenge by providing a substantial, volumetric, multimodal, and high-quality MRI-based labeled dataset. This initiative accelerates research and is a pivotal platform for developing robust segmentation models, offering significant assistance to the medical community. Over the years, BraTS has evolved consistently, becoming one of the widely used standard and large datasets for glioblastoma BTS. In the BraTS challenge since 2021, there are currently 1821 samples of MRI images. However, the BraTS2020

dataset was explicitly chosen for this study to align with the research objectives and accommodate system constraints. Moreover, the discontinuation of SD prediction in the BraTS challenge after 2020 reinforced the relevance of utilizing the BraTS2020 dataset for our investigation.

The BraTS2020 dataset encompasses 369 samples from 19 organizations, each with distinct image-acquiring standards and scanning devices. Experts annotated ground-truth segmentation labels for these samples and hand-crafted them. The validation set includes 125 samples, and the test set includes 166 samples. Each sample includes four 3D-MRI modalities, including native-T1, T1-ce, T2-weighted, and FLAIR, which can be seen in Figure 1.4. Details of modalities include [Menze et al. (2014)]:

- Native-T1: Originally acquired in sagittal or axial planes, with a slice thickness between 1 and 6 millimeters.
- T1ce: Native-T1 with contrast-enhanced agent gadolinium image, acquired in 3D form, and each voxel in dimensions of 1 millimeter in all directions, ensuring uniform spatial resolution for most patients.
- T2-weighted: Acquired in axial planes, with a slice thickness between 2 and 6 millimeters.
- FLAIR: Acquired in axial, coronal, or sagittal orientations with a thickness varying between 2 and 6 millimeters.

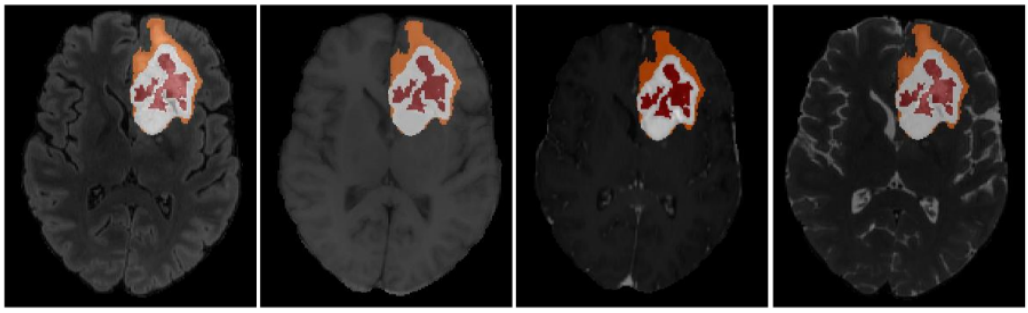


Figure 1.4. Multiparametric MRI through an axial plane mapped with ground truth image [Rajput et al. (2022)].

Achieving homogenization of these data involved rigid co-registering image volumes of all subjects to the T1-ce MRI, which typically possessed the highest spatial fidelity. Subsequently, all images were resampled to a standardized axial orientation with an isotropic resolution of 1 millimeter. Skulls were also removed from the images. The ground-truth MRI scans are annotated with unique values to delineate specific tumor lesions: background pixels are annotated as 0, NET/NCR lesions are annotated as 1, ED lesions as 2, and the ET/AT lesions as 4. The Ground truth for the validation and test sets is held back by the challenge organizers, who will use it to evaluate the model of the participants and their ranking. However, the developed models can be assessed using the online assessment system of BraTS2020 [1]. Each image's depth, width, and height are 155, 240, and 240, respectively.

In contrast, regarding SD prediction, the BraTS dataset comprises meta-data such as age, SD, and resection information. The training dataset includes 236 patients' samples, while the validation and test sets contain 29 and 107 samples,

[1] https://ipp.cbica.upenn.edu/

respectively. In contrast, we employed TCGA for radio-genomic analysis, obtaining clinical information and gene expression details corresponding to 35 samples that align with the BraTS dataset. The clinical information includes gender, race, and treatment information.

A brief history of BraTS's evolution is outlined in Table 1.2 and can be visualized in Figure 1.5.

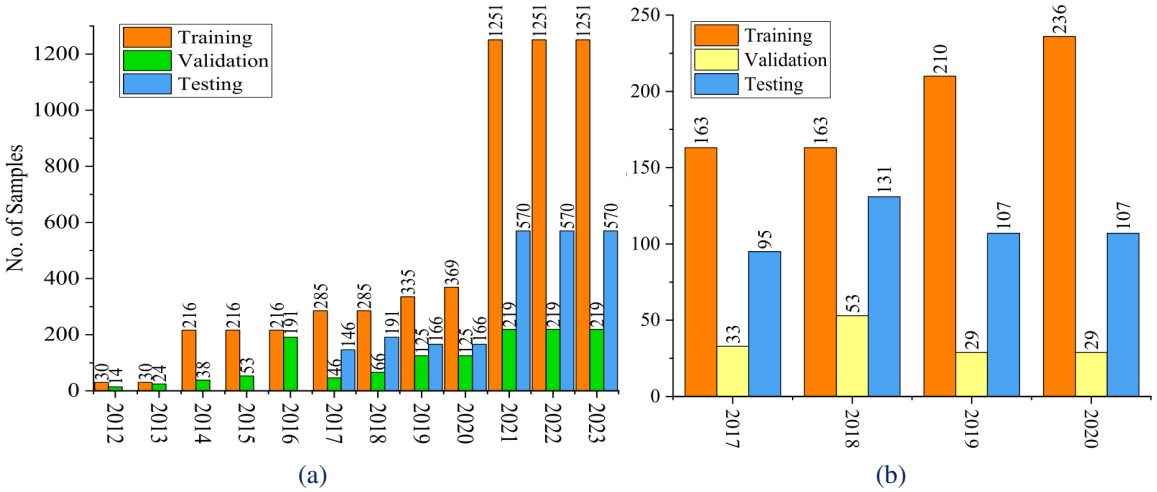


Figure 1.5. Timeline of dataset evolution for (a) BTS (b) SD prediction.

TABLE 1.2
A brief history of BraTS's evolution [Menze et al. (2014); Bakas et al. (2017)]. The dataset included synthetic images in the initial years of the BraTS challenge (2012 and 2013). However, starting in 2014, synthetic images were removed.

Year	Training (T), Validation (V), and Test (Tt) samples for BTS	T, V, Tt for SD prediction	Tumor lesions
2012	• T: 30 samples • V: 14 samples	N.A	Two tumor labels: "edema" and "core"

2013	• T: 30 samples • V: 24 samples	N.A	Same as BraTS 2012
2014	• T: 216 Samples • V: 38 Samples	NA	tumor labels: “necrosis”, “edema” “non-enhance core”, and “enhance core”
2015	• T: 274 Samples • V: 53 samples	NA	Same as BraTS 2014
2016	• T: 274 Samples • V: 191 samples	NA	Same as BraTS 2014
2017	• T: 285 samples • V: 46 samples • Tt: 146 samples	• T: 163 samples • V: 33 samples • Tt: 95 samples	Same as BraTS 2014
2018	• T: 285 samples • V: 66 samples • Tt: 191 samples	• T: 163 samples • V: 53 samples • Tt: 131 samples	Same as BraTS 2014
2019	• T: 335 samples • V: 125 samples • Tt: 166 samples	• T: 210 samples • V: 29 samples • Tt: 107 samples	Same as BraTS 2014
2020	• T: 369 samples • V: 125 samples • Tt: 166 samples	• T: 236 samples • V: 29 samples • Tt: 107 samples	Same as BraTS 2014
2021-2023	• T: 1251 samples • V: 219 samples • Tt: 570 samples	Discontinued	Same as BraTS 2014

1.7 Evaluation measures

The Dice-similarity coefficient (DSC), sensitivity, and specificity metrics assess the agreement between the segmented regions and the ground truth annotations at the voxel level. In contrast, another set of scores assesses the distance between segmentation boundaries, known as surface distance measures. A notable illustration is the Hausdorff distance (HD). DSC is particularly useful for evaluating the accuracy of the segmentation algorithms because it quantifies the overlap between the predicted and ground truth segmentations. A DSC of 1 signifies complete overlapping between actual and predicted pixel labels, whereas a score of 0 denotes no overlap. HD measures the alignment between the predicted and ground truth boundaries. A smaller HD value indicates better alignment between the predicted and ground truth boundaries, while a larger distance suggests poorer alignment. Both metrics offer unique perspectives on assessing segmentation performance. While the DSC evaluates the overall similarity between segmentations, the HD offers insights into the spatial accuracy of segmentation boundaries. By leveraging both metrics, one can gain a holistic understanding of the segmentation network's capabilities, effectively identifying their strengths and weaknesses.

In contrast, accuracy assesses the performance of SD classification models by measuring how often the model's predictions align with the actual class labels in the dataset. On the other hand, mean squared error (MSE) evaluates the average squared difference between the actual values and the predicted values generated by the SD model. Spearman ranking coefficient (SRC) is a statistical metric used to evaluate the strength and direction of relationships between two ranked features or variables. Accuracy is straightforward to interpret as it indicates the model's

ability to classify instances correctly. MSE provides a quantitative assessment of the degree to which the model's predictions match the true values.

Conversely, SRC can aid in variable selection and feature importance analysis in regression models. Variables with strong SRC values with the target variable may be considered significant predictors during the model training process. The evaluation schemes for BTS and SD prediction include the following measures.

1. (DSC) (or F1 measure): It measures the voxel-wise overlapping between two objects by dividing their intersection by the total size of both objects. True positive (TP) denotes the case where the model correctly predicts the positive label. While False positive (FP) occurs when the model incorrectly predicts the negative class as positive. False negative (FN) denotes instances where the model incorrectly predicts the positive label as negative. It is computed as depicted in Equation 1.1 below:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{1.1}$$

2. Jaccard Similarity Coefficient (Jaccard): It is referred to as the intersection over the union of two different sets as defined in Equation 1.2 below:

$$Jaccard = \frac{TP}{TP + FP + FN} \tag{1.2}$$

3. Sensitivity (True positive rate (TPR)): It is a metric that accurately determines tumorous voxels. This is defined in Equation 1.3 below:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1.3}$$

4. Specificity (True negative rate (TNR)): It is a metric that accurately locates non-tumorous voxels. This is defined in Equation 1.4 below:

$$Sensitivity = \frac{TN}{TN + FP} \tag{1.4}$$

5. HD: It computes the maximum distance between any point in one set and its nearest point in the other. If x and y be two non-empty subsets of a metric space (M,d), then their Hausdorff distance $d_H(x,y)$ can be explained in Equation 1.5 below:

$$d_H(x,y) = max\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\} \tag{1.5}$$

Where *sup* is the supremum and *inf* is the infimum.

6. Accuracy: It determines the ratio of correct predictions, encompassing both true positives and true negatives, among the total number of predictions. True negative (TN) is the outcome where the model correctly predicts the negative class as illustrated in Equation 1.6 below:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{1.6}$$

7. SRC: It is a statistical measure that assesses the strength and direction of the monotonic association between two features or variables. It is calculated by comparing the ranks of corresponding observations. The formula for the

SRC is specified in Equation 1.7 below:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{1.7}$$

Furthermore, interpretability tools like Shapley additive explanations (SHAP) and partial dependency plots (PDP) were employed to gain a comprehensive insight into the overall behavior of features. At the same time, SHAP, Accumulated local effect (ALE), and LIME were utilized to localize the understanding of feature behavior.

Chapter 2

Literature Survey

The delineation of brain tumors in medical images has become a notable focus of research in medical image analysis. The advent of automated and semi-automated brain tumor segmentation methods has introduced a significant paradigm shift in medical image analysis, offering the potential to reduce errors, enhance efficiency, and increase reproducibility. However, it is a challenging problem due to its inherent structure, location, and histology variations. The evolution of BTS methods can be broadly categorized into conventional and more recent approaches utilizing Deep neural networks (DNNs).

2.1 Conventional methods for BTS

Conventional methods primarily consist of non-deep learning-based approaches. Mohammed et al. [Mohammed et al. (2023)] classified them based on their evolutions or characteristics, such as

1. Primitive methods

2. Supervised-learning methods

3. Unsupervised-learning methods

2.1.1 Primitive methods

Among these traditional approaches are two prevalent image processing techniques: threshold-based methods and region-based methods.

Thresholding based technique: This method rapidly and efficiently segments images by comparing pixel intensities with predefined thresholds, including global and local thresholding approaches [Gordillo et al. (2013); J. Liu et al. (2014)]. Global thresholding, suitable for MR images with consistent intensity among objects, involves selecting a single threshold value for the whole image. While effective for homogeneous intensity objects [Wong (2005)], it may fail when tissue structures overlap in intensity. In contrast, local thresholding is effective when determining a threshold value from the entire image histogram is challenging. This is especially true when gradient effects are minimal or improved segmentation results require multiple thresholds [Bhargavi & Jyothi (2014)].

Region-growing techniques: This technique analyzes individual pixels in an image, grouping adjacent pixels with similar properties based on a predefined similarity measure to delineate distinct regions. The region-based approach encompasses methods such as “region growing” and “watershed”.

Region growing, a widely utilized technique in medical image analysis, particularly MRI segmentation, starts with a seed point in the area of interest. Neighboring pixels are then added based on predefined similarity criteria until no more pixels can be included. Despite its advantages, such as accurately distinguish-

ing regions with similar intensity, it faces challenges like the partial volume effect [Lakare & Kaufman (2000)]. To address this, a modified version incorporates gradient information for a more precise boundary definition. Overall, region growing is a valuable and less complex method, computation efficient, and remarkably effective for tissues, tumors, and homogeneous regions [Salman et al. (2006); Kaus et al. (2001)].

The watershed algorithm partitions an image into segments using topographic information, treating each pixel as a different height on a topographic surface. It identifies local minima as markers and floods the image with marker intensities, forming catchment basins akin to valleys on a map. While effective for object recognition and analysis, it may lead to over-segmentation in noisy medical images. Techniques to address this issue include merging regions based on specific criteria, as discussed in [Gies & Bernard (2004); Fisher & Aldridge (1999); Bhattacharya & Das (2008); Ratan et al. (2009)] for both 2D and 3D brain MRIs.

In summary, conventional brain tumor segmentation techniques struggle with the intricate task of delineating complex tumor structures, hampered by inherent limitations. Therefore, a growing focus is on advancing and promoting automated techniques to address these complexities effectively.

2.1.2 Supervised-learning methods

The algorithm receives input data (images) and associated labels in supervised methods, signifying the intended output or segmentation. The objective is for the algorithm to understand and establish a link from the input data to achieve the desired segmentation, drawing insights from the provided labels. The key advan-

tages of these methods are that they offer notable advantages in image analysis, particularly in achieving precise and customizable segmentation results. In existing literature, various researchers have categorized and discussed supervised segmentation techniques as outlined below:

Techniques based on pixel-level classification: These methods, employed in image processing and computer vision, involve categorizing individual pixels within an image into predefined categories or classes. The objective is to assign a label/class to every pixel according to its features, often facilitated by ML or deep learning algorithms. In the context of brain tumors, some of the commonly employed supervised classification methods found in the literature include the artificial neural networks (ANNs), the K-nearest neighbor (KNN), the random forest (RF), and the support vector machine (SVM).

Mishra et al. [Mishra (2010)] and Wang et al. [S. Wang et al. (2015)] utilized ANNs for brain tumor diagnosis, trained with statistical features. Ramteke et al. [Ramteke & Monali (2012)] used KNN for classification and anomaly detection with statistical features. [Havaei et al. (2014)] introduced a semi-automatic tumor segmentation method using KNN for voxel label classification. Moreover, RF proves to be particularly effective in addressing challenges related to high-dimensional feature vectors, multi-class classification, and high-grade gliomas (HGG), as demonstrated by the utilization of RF for tumor categorization in the study by [Koley et al. (2016)]. Phophalia et al. [Phophalia & Maji (2018)] proposed cascaded RF ensembles for voxel classification, while Sandabad et al. [Sandabad et al. (2016)] utilized SVMs for tumor detection and classification.

2.2 Unsupervised-learning methods

These techniques do not require labeled training data, and images are divided into homogeneous regions based on attributes such as pixel intensities or texture features. These methods accurately segment complex structures like heterogeneous tumors [Gordillo et al. (2013)]. However, in the segmentation of brain tumors, the absence of prior information regarding the structure of tumors or pixel intensity value makes unsupervised segmentation challenging, leading to limited research focused on unsupervised techniques. The unsupervised segmentation methods are categorized as follows:

Technique based on clustering: Clustering-based segmentation, an unsupervised method, involves dividing unlabeled image data into clusters of pixels or voxels. This process groups pixels with similar intensities, sharing common features within the same cluster while separating dissimilar pixels into different clusters. In MR brain images, k-means clustering has been employed by [M.-N. Wu et al. (2007)] to differentiate tumorous tissues from normal tissues. The authors achieved this by transforming a grayscale MRI image into an RGB image and enhancing its attributes through pseudo-chromatic transformation. The Fuzzy C-means clustering (FCM) algorithm, extensively utilized in MRI [Balafar (2014)], categorizes pixels into multiple classes by assigning memberships based on the distance between pixel units and class centers [X. Yang & Fei (2011)]. Notably, FCM allows pixels to belong to two or more classes, and its adaptability to data with multiple cluster solutions is attributed to the use of fuzzy membership functions [Ganesh et al. (2017)]. While the Markov random field (MRF) model is a powerful tool for image segmentation, its applicability for online processing is

limited due to the high computational cost. [Ahmadvand et al. (2017)] introduced a hybrid approach, combining the MRF model with the watershed algorithm to mitigate the drawbacks associated with MRF. Similarly, [Sun & Wang (2008); Yousefi et al. (2011)] proposed an optimized MRF-based model that can effectively segment tumors with less computational time.

Techniques based on model: Previous discussions have predominantly focused on 2D image segmentation, while segmenting 3D image data presents a notable challenge. This challenge is commonly addressed through model-based segmentation techniques like parametric deformable models, geometric deformable models, or level sets [Gordillo et al. (2013)]. These methods utilize prior knowledge about object characteristics, location, or regions of interest within an image. Parametric deformable models, also called active contour models, excel in segmenting, matching, and tracking images of anatomical structures, leveraging constraints from image data and prior knowledge about structure location, size, and shape. However, they face challenges in effectively handling topological changes, such as contour merging and splitting, in 3D images. This problem has been mitigated by introducing geometric deformable models or level sets [Gordillo et al. (2013)]. Various methods based on parametric deformable models include [C. Xu & Prince (1997, 1998)], while those based on geometric models include [J. Liu et al. (2014); H.-H. Chang & Valentino (2008)].

2.3 Deep learning methods

DNN is a form of "deep learning", which refers to the utilization of neural network models comprising multiple stacked functional layers, typically with more than

five layers [J. Liu et al. (2014)]. The versatility of DNNs spans a range of areas, such as pattern recognition, machine vision, healthcare, speech processing, and Natural language processing (NLP), rendering them widely applicable in various fields because:

1. DNNs automatically learn meaningful representations of input data, subsequently improving feature extraction and performance.

2. They excel in learning hierarchical features, enabling the capture of intricate patterns across multiple layers of abstraction.

3. DNNs excel in handling high-dimensional data, which enables their effectiveness in tasks such as image/pattern recognition and NLP.

4. They can approximate/estimate continuous mathematical functions, enabling the modeling of complex input-output mappings across various domains.

For this tremendous success across multiple domains, DNN demands challenging requirements such as:

1. DNNs demand substantial computational resources and can be computationally intensive, leading to longer training times and higher hardware requirements.

2. They heavily depend on extensive labeled data for training, which may not always be available for all possible scenarios.

3. Tuning hyperparameters for optimal performance can be time-consuming and require extensive experimentation.

4. Overfitting is a common problem with DNNs, especially when trained on small datasets or with overly complex architectures.

5. DNNs often function as black-box models, making interpreting their decisions or diagnosing issues challenging.

The typical structure of a DNN utilized for BTS is shown in Figure 2.1, where each component plays a crucial part in accurately segmenting tumor tissues. Input images can be single or multi-parametric. Additionally, preprocessing techniques are applied to remove inhomogeneity, and data augmentation is performed to amplify the diversity of the input images, which aids in enhancing the generalization and robustness of DNN models. Moreover, in the scenario of a 3D model, a 3D patch is selected to match the Graphics processing unit (GPU) requirements for training the DNN model. In contrast, for a 2D model, image slices are chosen. The loss function may be singular, combined, or weighted to enhance the training process. Conversely, images undergo preprocessing during testing, patches or slices are extracted, and labels are predicted using the trained DNN model. Finally, the segmented output can be further refined to adjust false positive labels.

2.3.1 Evolution of DNN

The evolution of DNN in the BTS field is rendered in Figure 2.2. Since the late 1990s, neural networks have been used for tumor segmentation. Still, their application was limited until the early 2000s due to the challenge of training them, stemming from a scarcity of high-end computational resources [J. Liu et al. (2014)]. Early research on BTS aimed to devise specialized deep Convolutional neural networks (CNN) for accurate tumor segmentation. Recent breakthroughs,

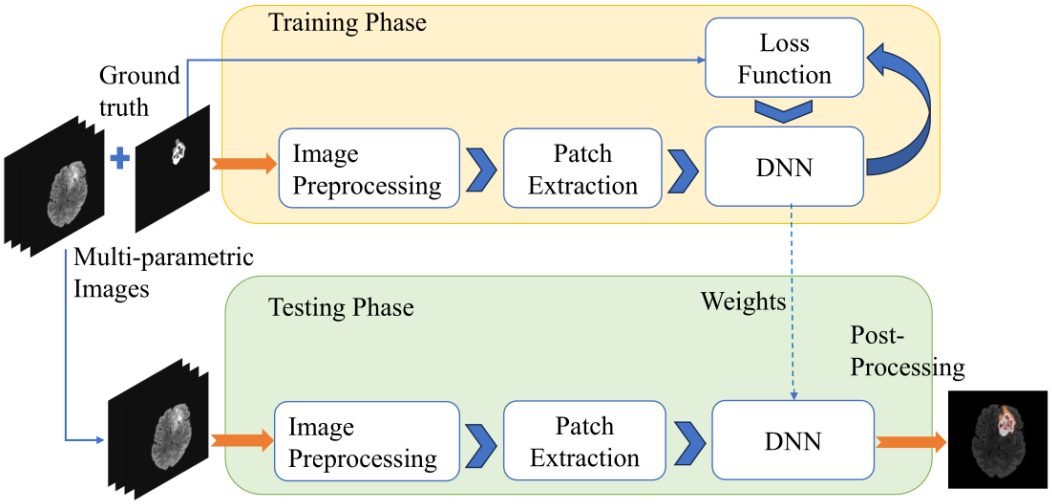


Figure 2.1. The typical structure of a DNN utilized for BTS.

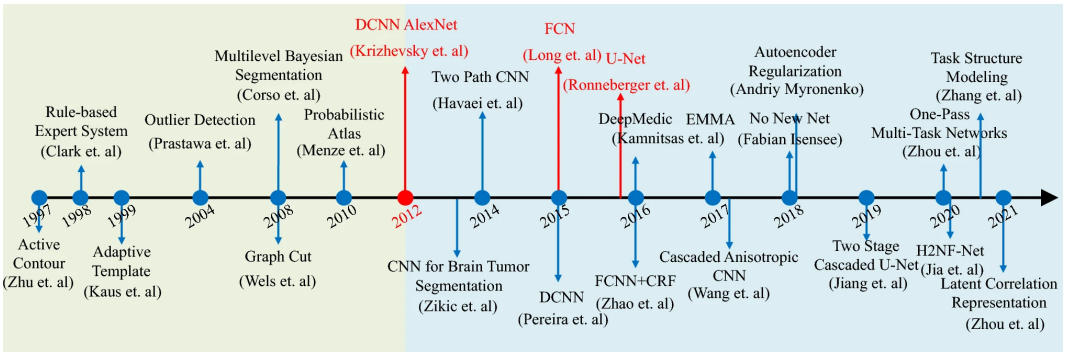


Figure 2.2. The evolution of DNN in the BTS field [J. Liu et al. (2014)].

including the Fully convolutional network (FCN) and U-Net and recent developments [J. Long et al. (2015); Ronneberger et al. (2015); Isensee et al. (2017); Strudel et al. (2021)], have spurred innovations in building encoder-decoder architectures, eliminating connected layers to enable end-to-end tumor segmentation.

CNN, a subtype of DNN, has emerged as a leading algorithm, consistently demonstrating its dominance in computer vision tasks. Its effectiveness extends across diverse fields, including biomedical imaging, where CNNs have showcased expert-level performance. CNN is designed to process data using a grid pattern,

such as images. Inspired by the organization of the animal visual cortex [Hubel & Wiesel (1968)], it is crafted to autonomously and dynamically learn layers of abstraction of features, progressing from low- to high-level patterns. The overview of CNN architecture is illustrated in Figure 2.3.

2.3.2 CNN design elements in BTS networks

In the BTS network, the development of CNN generally includes the following: 1. Basic Layers such as Convolution layer, Pooling/subsampling layer, Activation layer, Flatten layer, Normalization layer, and Dense layer 2) Techniques for regularization, 3) Optimization methods, 4) Loss functions, and 5) Layer weight initializations and regularizers. Brief discussions about each element are as follows:

- Convolution layer: This layer extracts essential features from the input data. It employs a weight-sharing procedure, utilizes local regions of input data, and offers some level of insensitivity to shifts.

- Activation layer: It introduces nonlinear behavior to the network, allowing it to capture and learn detailed/complex mappings from input to output. Some of the activation functions include Rectified linear unit (ReLU), Leaky ReLU (LReLU), Parametric ReLU (PReLU), Exponential linear unit (ELU), and its variants.

- Pooling/subsampling layer: This layer identifies significant features from non-overlapping neighborhoods. Its objectives encompass: 1. Decreasing parameter counts, 2. Mitigating overfitting, and 3. Attaining translation

invariance. Several prevalent pooling techniques comprise average pooling, max pooling, and global average pooling (GAP) functions.

- Normalization layer: It ensures consistent input scaling for each layer, promoting improved convergence and stability during training. Some of the popular layers are Batch, Group, and layer normalization.

- Flatten layer: This layer transforms the multi-dimensional output from previous layers into a one-dimensional (1D) feature vector, thereby setting it up for input into the fully connected layers.

- Fully connected/Dense layer: It converts the 1D feature vector into predictions or classifications. They map feature maps to the corresponding output classes or labels.

- Loss functions: They enhance the learning procedure by refining similarity within classes and separability between classes. Some commonly utilized loss functions are dice loss, focal loss, and cross-entropy loss. Combinations of these loss functions and their variants are also common.

- Techniques for regularization: These address the overfitting problem. Common techniques include dropout, L1/L2 regularization, reduced learning rate, learning-rate scheduler, and early stopping.

- Optimization: It fine-tune a model's parameters during training to minimize the loss function. Optimization algorithms iteratively adjust weights and biases during backpropagation, facilitating learning. Some optimization techniques are Root means square propagation (RMSProp), Adaptive gradient algorithm (Adagrad), and Nesterov accelerated gradient descent.

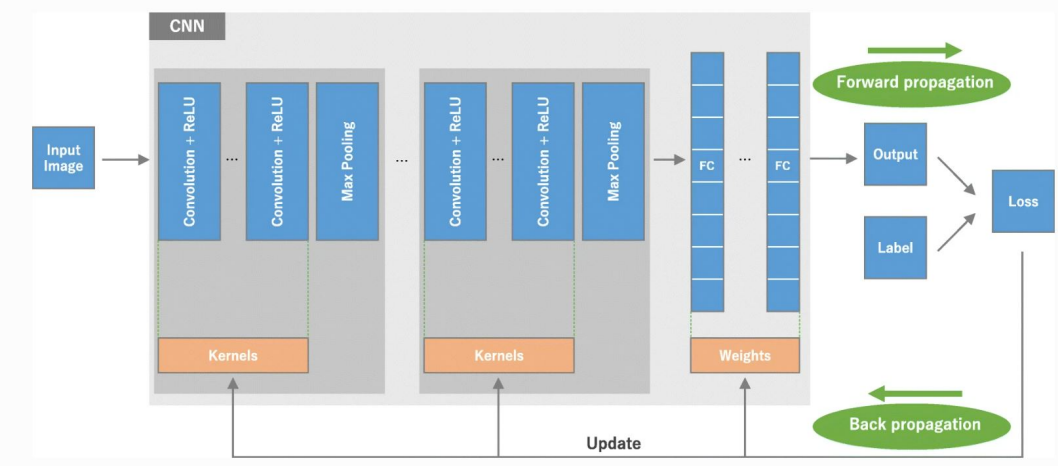


Figure 2.3. An overview of CNN architecture and the training process [Yamashita et al. (2018)].

- Initialization and normalization of layer weights: It accelerates the learning phase by ensuring proper initial values for weight updates.

2.3.3 Architectural variants of CNNs for BTS

Most of the architectural innovation in semantic segmentation comes from the success of DNN in image classification and transfer learning. A breakthrough came from a Fully convolution network (FCN) [J. Long et al. (2015)], where fully connected layers were replaced with convolution layers. This allowed 1. Accepting input of many sizes and generating output of different sizes, unlike dense layers, which require a fixed-size input. 2. Convolutional layers in neural networks produce feature maps as outputs and preserve the spatial arrangement of features in the input. Unlike dense layers, this compresses all spatial information into a single vector. 3. Skip connections facilitate the straightforward passage of information from earlier layers to later layers, minimize the vanishing gradient

problem, and strengthen the training process. The primary limitations of FCN include inefficient localization of labels, especially for objects with intricate shapes or fine details, the incapability to analyze global contextual information, and the lack of a mechanism for processing information across various scales.

UNet [Ronneberger et al. (2015)] addresses these challenges through its encoder-decoder architecture, wherein the encoder (contracting path) initiates with convolutional layers succeeded by max-pooling functions. Here, the dimensions of the spatial feature map gradually decrease while the feature channel counts increase, enabling the capture of contextual information.

In this encoder segment, the number of feature channels typically expands further into the network, enabling the extraction of progressively abstract features. while the decoder (expansive path) gradually restores object details and spatial dimensions using upsampling or deconvolution operation while reducing the count of feature channels. Each layer in the decoder is connected to the encoder's respective layer, called a skip connection. These skip connections enable the network to pertain to fine-grained details during upsampling by combining elementary/low-level features from the contracting path with abstract/high-level features from the expansive path. Finally, the final output layer typically includes a 1×1 convolution, succeeded by a sigmoid or softmax activation function for pixel-wise classification in semantic segmentation tasks. While softmax is commonly used for multi-class segmentation, sigmoid is often employed for binary segmentation tasks.

Other architectures like SegNet [Badrinarayanan et al. (2017)] and DeepLab [L.-C. Chen et al. (2017)] are alternative FCN architectures utilized for the BTS task. Recently, residual connections [K. He et al. (2016)] and attention mechanism-

based networks [Bahdanau et al. (2014)], including visual transformers [Vaswani et al. (2017)], have garnered attention. Residual connections entail incorporating the outcomes of a preceding layer into the input of a subsequent layer, thereby mitigating the vanishing gradient issue and aiding the training of deeper networks. On the other hand, attention mechanisms, such as those utilized in transformers, allow networks to focus on relevant input information while suppressing irrelevant parts of the input data. Unlike traditional attention mechanisms, transformers employ self-attention mechanisms that enable capturing relationships between all input tokens simultaneously, facilitating more effective learning of long-range dependencies in sequential data.

2.3.4 Strategies for addressing class imbalance

In medical imaging, class imbalance often occurs because abnormalities (such as tumors or lesions) are relatively rare compared to normal tissue. Addressing class imbalance is crucial to ensure precise and robust segmentation outcomes. Considering the BTS context, the class imbalance is very significant, as can be seen in Table 2.1.

TABLE 2.1
The distribution of sub-categories at the voxel level within the training data of the BraTS2020 dataset. [J. Zhang et al. (2020)].

Pixel-(labels)	**Percentage (%)**
Background-(0)	98.46%
Edema-(2)	1.02%
Enhancing tumor-(4)	0.29%
Necrotic and Non-enhancing tumor-(1)	0.23%

This issue is addressed using the following methods:

1. Image preprocessing: Removing non-brain pixels from images (such as skull and scalp regions).

2. Data augmentation: It plays a crucial role in addressing class imbalance by generating synthetic samples for minority classes, thereby balancing the dataset. This process helps the model avoid bias towards the majority class and improves its ability to generalize to all classes more effectively.

3. Patch extraction: By extracting patches from regions of interest containing both minority and majority classes, the approach ensures a more balanced representation of data, enhancing the model's accuracy in classifying all classes effectively.

4. Loss functions: They work by adjusting the share of each class (weighing) to the overall loss calculation during training. Among the frequently employed loss functions are the Generalized dice loss (GDL) [Sudre et al. (2017)] function, Focal loss (FL) function, and Weighted cross-entropy (WCE) function. Combining or weighting these functions is also prevalent, especially in addressing intricate challenges such as brain tumor segmentation. In the GDL, the weighting scheme is based on the inverse of the volume of each class that is introduced. In WCE, weight is computed based on the distribution of foreground class probabilities across all samples. In FL, weight is calculated based on predicted probability and the gamma parameter (positive value focusing on class).

2.3.5 Recent advancements in CNN architectures and trainable parameters for BTS

In the context of BTS, a concise summary of the proposed networks, encompassing both segmentation strategies and trends in network complexity, is as follows: [Urban et al. (2014)] in BraTS2014 utilized an ensemble of two 3D CNN networks having a four-layered architecture. The authors employed 15 filters in the initial layer and 25 for the subsequent two layers. In the final layer, six filters were utilized, each corresponding to one of the six different classes. Subsequently, a second network was trained, mirroring the first, adding an extra layer incorporating 40 filters of size $5 \times 5 \times 5$ preceding the last layer. The mean DSC on the validation set was 87 for WT, 77 for TC, and 73 for ET. Similarly, [Zikic et al. (2014)] proposed a five-layered 2D CNN. The first layer incorporated a convolutional layer with 64 filters sized 5×5, succeeded by a max-pooling function in the subsequent layer. The third layer included a convolutional layer with 64 filters of size 3×3, the fourth layer was composed of a fully connected layer having 512 nodes, and the fifth layer was a softmax layer for classification. The average DSC on the training set was 83.7±9.4 (WT), 73.6±25.6 (TC), and 69.0±24.9 (ET).

In the BraTS2015 challenge, [Pereira et al. (2016)] introduced two similar 3D-based architectures for HGG and Low-grade glioma (LGG) images. The HGG architecture comprised 10 layers, while the LGG architecture comprised an 8-layered CNN. The average DSC on the training set was 0.87, 0.73, and 0.68 for WT, TC, and ET, respectively. In 2016, [P. D. Chang et al. (2016)] proposed a 5-layered 2D CNN network with a larger input image size of 240×240, allowing the network to capture more detailed information from the input images. The network

includes a total of 130,400 trainable parameters. Only HGG images from the training set were utilized to train and validate the result. the DSC on the validation set was 0.87 (WT), 0.81 (TC), and 0.72 (ET). In 2016, [Kamnitsas et al. (2017)] introduced an efficient dual-pathway, multi-scale 11-layer deep network based on 3D CNN layers. The network structure comprised two parallel convolutional streams that handled the input data at various scales, aiming to achieve a broad receptive field for the ultimate classification task while maintaining computational efficiency. The DSC on cross-validation was 89.6 (WT), 76.3 (TC) 72.4 (ET). Notably, this model demanded a 12GB GPU for training, with input images of size $25 \times 25 \times 25$.

In 2017, [Alex et al. (2017)] proposed 23 layered FCNNs similar to UNet architecture. This FCNN model was trained on axial slices having dimensions 240×240 extracted from MRI modalities. The DSC on the validation set were 0.83 (WT), 0.69 (TC), and 0.69 (ET). In 2018, [Isensee et al. (2019)] proposed an efficient 3D-Unet with residual connections having an input image size of $128 \times 128 \times 128$. The model was trained on the additional institutional dataset. The authors ensembled 5 such networks for the label prediction, where the trainable parameters for each network were $14.4M$, and the total parameters were $72M$. The DSC on the validation set (66 cases) was 91.26 (WT), 86.34 (TC), and 80.87 (ET). Similarly, Myronenko et al. [Myronenko (2019)] employed an ensemble of 10, 3D Unet-based variational encoder-decoder architecture with large input patches sized $160 \times 192 \times 128$ and reconstructed raw data using a variational encoder branch. Each network had extensive GPU memory consumption, specifically on an NVIDIA V100 32 GB GPU with $18.8M$ trainable parameters. The total number of trainable parameters was $188M$. The DSC on the validation set

was 0.91(WT), 0.87 (TC), and 0.82(ET).

Recently, in 2020, a transformer-based 3D UNet [Hatamizadeh et al. (2021)] utilized a DGX-1 cluster with 8 NVIDIA V100 GPUs ($32GB \times 8 = 256GB$) for training, where they ensembled 8 such models for segmenting brain tumors. Each model in this setup had 61.98 M parameters, resulting in a total of 495.84 M trainable parameters. The DSC on the validation set was 0.926 (WT), 0.885 (TC), and 0.858 (ET).

Analyzing the trends in the proposed networks for brain tumor segmentation, performance has seen improvement due to the accessibility of extensive datasets and computational resources. The utilization of 3D models and ensembles is prevalent in brain tumor segmentation. However, it's crucial to acknowledge that the complexity of network architectures, while contributing to improved results, has introduced additional challenges. Besides longer training times, managing computational resources, and addressing potential overfitting, the current clinical settings underscore the imperative for efficient networks. The need is for enhanced accuracy and models that can operate efficiently without compromising precision, ensuring practical utility in clinical applications. Additionally, there is a need to bridge the gap between the machine and clinical experts by facilitating an understanding of the decisions made by these ML models. Furthermore, deciphering these highly complex models will pose additional challenges.

2.4 Importance of SD prediction for glioblastoma patients

Accurate prediction of overall SD for glioblastoma patients is imperative for effective treatment planning and monitoring. However, this task poses challenges due to its dependency on BTS, radio imaging parameters, the non-uniformity in radiomics, limited clinical information, and a relatively small dataset. The trend of the dataset evolution is shown in Section 1.6 in Table 1.2 and Figure 1.5.

In 2015, Chaddad et al. [Chaddad & Tanougast (2016)] utilized a texture-based Gray-level co-occurrence matrix (GLCM) to highlight the connection between glioma phenotype and overall survival prediction. The author used classifiers such as Discriminant analysis (DA), Naïve bayes (NB), Decision trees (DT), and an SVM. The study was conducted on 40 samples from The Cancer Imaging Archive (TCIA). The study underscored the importance of GLCM in the realm of SD prediction. Discriminating GBM phenotypes using GLCM features achieved the highest accuracy, sensitivity, and specificity, with values of 79.31%, 91.67%, and 98.75%, respectively. Similarly, [L. Liu et al. (2016)] utilized fMRI and DTI imaging to analyze 147 HGG patients, categorizing them as either "good" (if SD$>$ 650 days) and "bad" (if SD $<$ 650 days). Structural and functional brain networks were constructed for each sample, and numerous graph-theory-based features were extracted. The dataset also encompassed diverse clinical information, including the patient's age, tumor size, gender, tumor grading, histopathological type, spatial location, presence of epilepsy, precise location in the brain lobes, and hemispheres of tumor lesions. Using the SVM classifier, the model

achieved a prediction accuracy of 75%.

In 2019, Agravat et al. [Agravat & Raval (2019)] employed an RF regressor model trained on statistical, shape features, and age information to classify the SD of patients into long, medium, and short survival classes. Meanwhile, they utilized a 2D Dense Unet network for segmentation using the BraTS dataset. The model attained an accuracy 51.7% for survival prediction on the validation set. In 2020, McKinley et al. [McKinley et al. (2020)] proposed an ensemble of linear and RF regressor models, utilizing patient age, the number of tumor cores, and whole tumors for SD prediction derived from the segmented results. Their study also proposed an ensemble of 3D-to-2D networks for BTS segmentation using the BraTS dataset.

Similarly, Asenjo and Larraz Solıs [Marti Asenjo & Martinez-Larraz Solís (2021)] introduced an ensemble of classification and regressor models, incorporating geometric, textural, statistical, and tumor location features. The accuracy on the validation set is 41.4%. They utilized an ensemble of 2D and 3D Unet networks for BTS using the BraTS dataset. In 2022-2023, Rajput et al. [Rajput et al. (2022, 2023a)] employed gradient-boosting and RF regressors, incorporating statistical, shape, location, and texture features, and utilized a 3D Unet model for BTS using BraTS dataset. The accuracy on the validation set is 62.1% and 55.2%, respectively. Cepeda et al. [Cepeda et al. (2021)] utilized diverse classifiers to classify patients into short-term survivors. The classifiers employed in their analysis included a logistic regressor (LR), KNN, SVM, RF, and a Multilayer perceptron (MLP). The authors conducted feature extraction, yielding a total of 15,750 features. These features encompass first-order statistics, histogram-based features, and morphological, volumetric, and textural features. The NB classifier

demonstrated superior performance on the test dataset, achieving an area under the curve (AUC) of 0.769 and an accuracy rate of 80%.

In summary, existing literature suggests the importance of morphological, spatial location, and radiomics-based features. Meanwhile, ensemble-based models such as RF regressor were widely used. The RF regressor algorithm comprises an ensemble of decision trees, each trained on a random subset of the training data. This randomness helps in reducing overfitting and improving generalization and makes it more robust [Rajput et al. (2021)]. Radiomics is an evolving discipline of study that emphasizes extracting quantitative features from radiological scans. It captures characteristics of tissues and lesions, including their shape and heterogeneity. Furthermore, it can track changes over time, such as during treatment or monitoring, through serial imaging [Mayerhoefer et al. (2020)]. Therefore, it can provide valuable insights into precision and personalized medicine (PPM).

Further, from the existing literature survey we identified a research gap for BTS and SD prediction enumerated as follows:

- Due to high inhomogeneity between tumor lesions and within tumor lesions, conventional MRI images were not able to capture physiological, metabolic, and biological details of tumor lesions.
- Although various deep learning architectures hold promise for BTS, challenges persist due to tumor morphology complexity, imaging artifacts, and the strength of MRI scanners.
- Existing segmentation techniques are very computationally and memory intensive especially for real-time applications on edge devices.
- For SD prediction availability of ample dataset along with clinical and other omics/multimodality information is essential.

Chapter 3

A Proposed Holistic Approach to Efficient Segmentation of Brain Tumor and Predicting Survival Days

Considering the clinical significance and the reliance of both the BTS and SD prediction tasks on factors such as diagnosis, treatment planning, and post-surgery monitoring, our emphasis is on formulating a comprehensive approach for predicting BTS and SD within a constrained environment. In pursuit of our goal, i.e., to optimize a BTS architecture for resource constraint setup, we began with a 3D U-Net based on the top-performing segmentation model [Isensee et al. (2017)] and reduced the number of filters. This model comprised 8.3M trainable parameters, necessitating a system setup with 256 GB RAM and 16 GB GPU. The DSCs for BTS on the BraTS2020 validation set were 0.880 (WT), 0.858 (TC), and 0.759 (ET). Meanwhile, for SD prediction, the accuracy was 0.552 for the validation set.

Probing further optimization for both tasks and studying the trade-off between

the parameter and accuracy, we explored the possibility of 2D-based and 2.5D models. Since MRI is not inherently a 3D modality and is derived by fusing 2D slices of individual imaging planes (i.e., axial, sagittal, and coronal), some factors may contribute to the possibility of 2D BTS (trained on the individual plane) [McHugh et al. (2021)]. Considering the challenge, 3D MRI images were reconstructed from 2D MRI images, where thicker 2D slices were preferred in specific clinical situations to enhance the signal-to-noise ratio [McHugh et al. (2021); Hausmann et al. (2018)]. This leads to resolution anisotropy, requiring interpolation between slices for the 3D reconstruction of non-volumetric imaging. In non-volumetric imaging, the resolution in the coronal and sagittal planes is significantly lower than in the axial plane [McHugh et al. (2021)]. Also, in the 2.5D model, segmentation results from all the planes are fused to obtain final segmentation, including inter-slice information missing in 2D models [McHugh et al. (2021)]. This aids 2D models in including some depth information like 3D models. Therefore, we investigated 2D-based U-net models for BTS and SD prediction.

The proposed holistic approach for BTS and survival prediction can be seen in Figure 3.1. It includes a segmentation module where segmentation networks were utilized to segment tumors, as elaborated in the subsequent section 3.1. Further, from the segmentation results, features are extracted and selected strategically to find robust feature sets to predict SD. Lastly, an analysis is conducted on the trained model to understand the decisions of the complex models and find the biological relevance of the features.

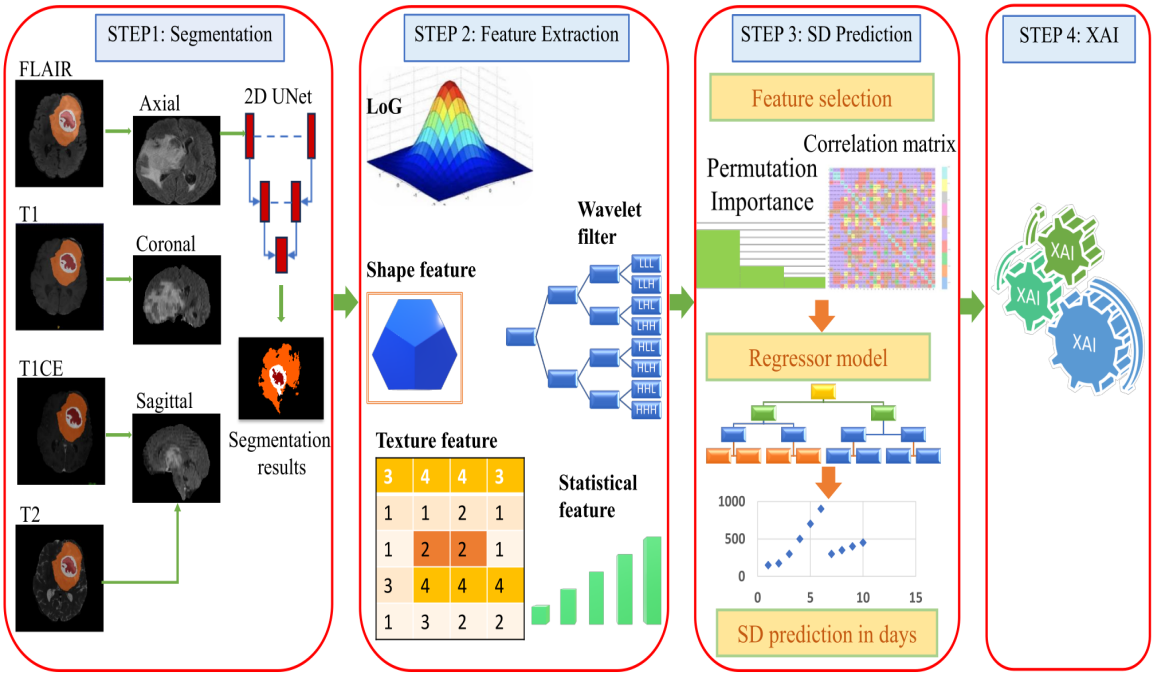


Figure 3.1. The complete pipeline to determine SD and its interpretability, utilizing BTS outcomes [Rajput, Kapdi, Raval, et al. (2024)].

3.1 Segmentation networks

The comprehensive architecture of the proposed segmentation network includes data preprocessing, segmentation models, fusing segmentation probability maps from multiple segmentation models, and post-processing, illustrated in Figure 3.2. As discussed previously, our initial approach involved utilizing a 3D U-Net network structure for segmentation. Subsequently, we proposed a novel 2.5D segmentation network to address specific challenges and optimize performance. The same preprocessing techniques were applied to both networks. This preprocessing involved bias correction, a crucial step due to the inherent uneven intensity distributions commonly found in MRI scans caused by bias fields. Therefore, our

preprocessing pipeline began with bias correction utilizing the N4ITK package [Tustison et al. (2010)] to address this.

Furthermore, non-brain regions were excluded from all multi-parametric MRI scans. Then, the highest and lowest 1% of intensity values were eliminated. Following that, the intensity levels of each image slice were standardized using mean and standard deviation values. For 3D segmentation, we have utilized the network proposed by [Isensee et al. (2017)]. Random flipping and distortion within [0 - 0.25] factor were used as augmentation techniques.

The 2.5D segmentation network, which primarily consists of an ensemble of 2D models trained on axial, sagittal, and coronal planes, is presented in Figure 3.2. The basic architecture of the 2D model is the same for all the utilized models. The input image slices had the following dimensions: $192\,(W) \times 152\,(H)$ for the axial model, $192\,(W) \times 144\,(H)$ for the sagittal model, and $152\,(W) \times 144\,(H)$ for the coronal model. These variations in size arose from the orientation of the imaging plane and the method used for MRI processing. During training, we implemented random vertical and horizontal flips to reduce memory consumption and enhance training efficiency on the fly. The proposed segmentation network is based on an attention mechanism, enabling it to prioritize important aspects of the input data while suppressing irrelevant information, enhancing its ability to extract meaningful features and make accurate predictions.

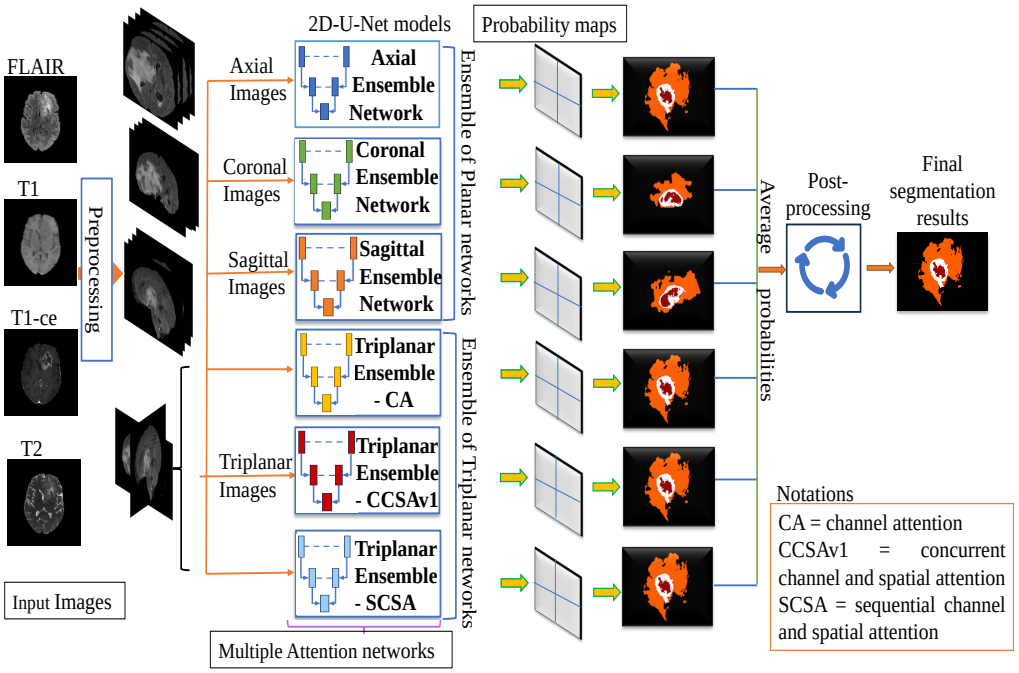


Figure 3.2. The holistic structure of the proposed 2.5D based segmentation architecture [Rajput et al. (2023b); Rajput, Kapdi, Roy, & Raval (2024)].

In this research, we investigate and analyze the effect of attention methodologies: Concurrent channel and spatial attention (CCSA) mechanism, Channel-attention (CA), and Sequential-channel and spatial attention (SCSA) on variants of planar and triplanar models (listed below). CA allows the network to modulate and weight each channel based on its relevance to the task at hand. CCSA allows the network to capture spatial and channel correlations simultaneously, while SCSA derives attention maps along channel and space and merges them. The probability outcomes from these networks are averaged to derive the final segmentation results.

1. Variants of planar are axial planar (AP), coronal planar (CP), and sagittal planar (SP).

2. Variants of triplanar based on CA, CCSA, and SCSA.

AP represents the ensemble of three 2D-based networks integrated with CA, CCSA, and SCSA and trained on axial plane images. Likewise, CP and SP denote the ensembles of networks, incorporating CA, CCSA, and SCSA, trained on the coronal and sagittal image planes, respectively. Lastly, we called it the *Super-ensemble* model, which is the ensemble of all the variants of planar and triplanar models.

The details of each of these variants of segmentation models are discussed as follows:

3.1.1 Architecture of planar (2D) and 2.5D models

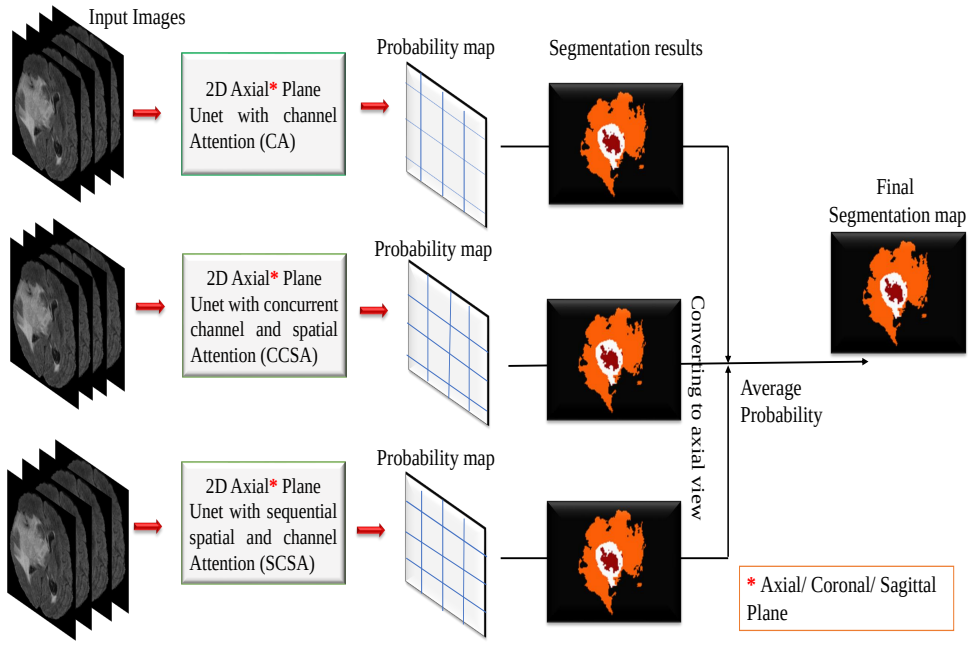


Figure 3.3. The axial ensemble network (2D), where segmented outcomes CA, CCSA, and SCSA networks trained on axial images are combined to generate an outcome. Similarly, coronal-ensemble and sagittal-ensemble planar networks can be constructed and trained on coronal and sagittal images, respectively [Rajput et al. (2023b)].

The base network (shown in Figure 3.5) is trained with input images from the axial plane. We created three models for the AP based on CA, CCSA, and SCSA. The attention modules are implemented on the decoder side (with circular A symbols in Figure 3.5). Similarly, three models are created for the CP and three for the SP. Thus, nine models (three planes $\times$ three attention per plane) are trained with the same hyperparameters and network structure. Different ensembles are created using a suitable combination of the outputs from these nine networks, and they are shown in Table 3.1. The planar ensemble combines outputs for the same plane but with three different attention mechanisms. The outputs of three models from the AP are combined to create an Axial - Ensemble. Similarly, outputs from three CP and three SP models are combined to produce a Coronal - Ensemble and a Sagittal - Ensemble, respectively. The axial planar ensemble is shown in Figure 3.3. The output of each model generates probability maps for each label for a sample. Subsequently, averaging the probability maps for each label will generate the final segmentation labels.

In the triplanar ensemble, we combine outputs from three planes - AP, CP, and SP- and each has the same attention mechanism. For example, the model's outputs for AP, SP, and CP with CA attention are combined. Similar ensembles are created using three orthogonal planes with CCSA and SCSA attention methodologies. The triplanar ensemble is shown in Figure 3.4. We also created a "Super-ensemble" combining planar and triplanar ensembles.

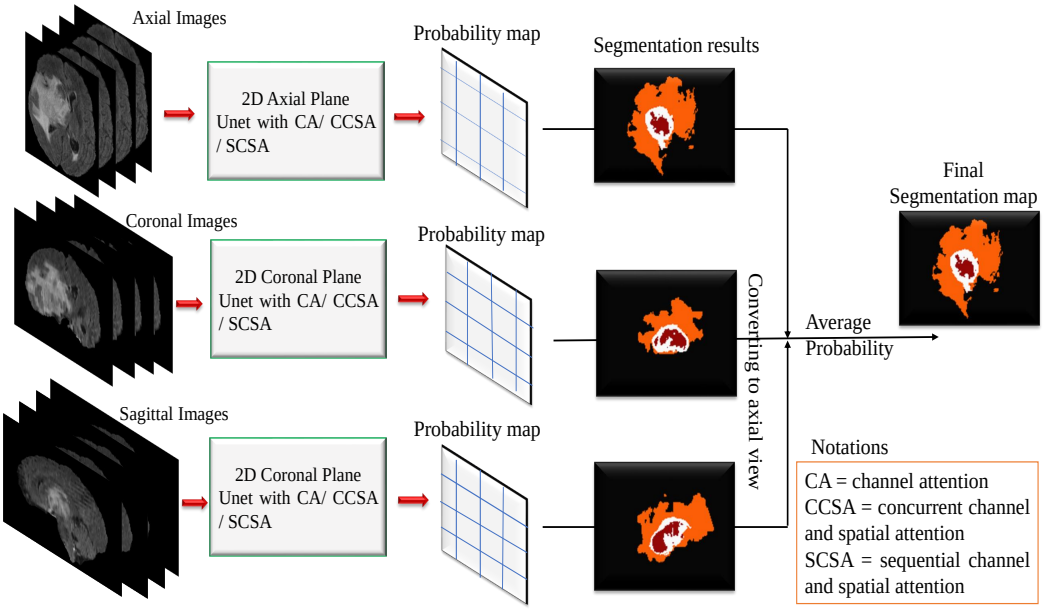


Figure 3.4. An overview of the 2.5D network, where segmented outcomes generated from individual attention networks (e.g., CA network) trained on axial, coronal, and sagittal images are combined to generate an outcome. Similar segmented outcomes are generated from CCSA and SCSA attention networks trained on three planes [Rajput et al. (2023b)].

The primary architecture for all the models is the same as shown in the subsequent section below:

3.1.2 Basic architecture and training details: 2D U-Net

Figure 3.5 illustrates the fundamental architecture employed in the proposed segmentation methods. This architecture incorporates three attention mechanisms designed for computational efficiency. They are depicted as the attention blocks in the figure and are explained in section 3.1.3. The network structure of the suggested 2D UNet can be observed in Figure 3.5. The architecture has 4 layered

encoder-decoder paths; each layer consists of a ResNet [K. He et al. (2016)] like convolution block.

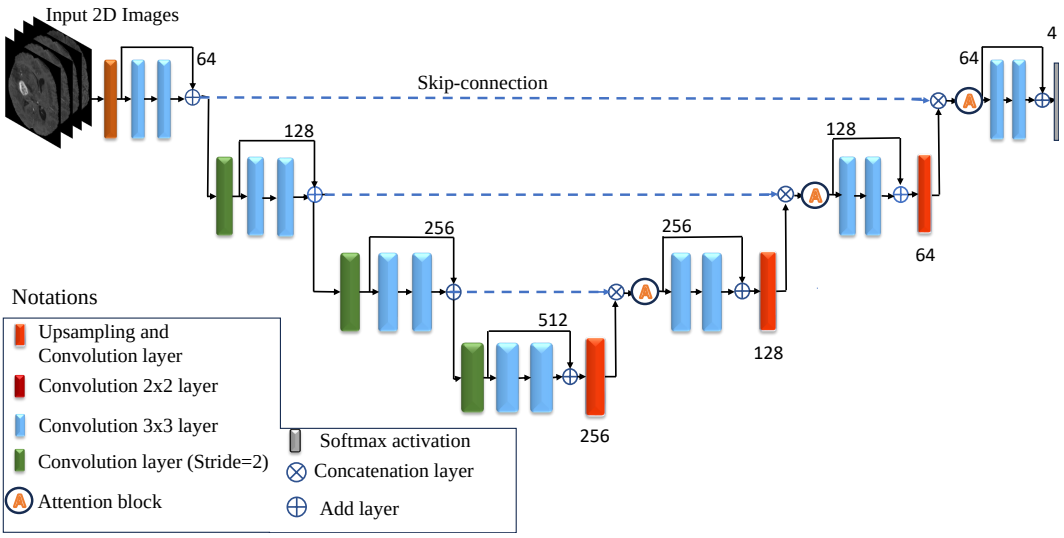


Figure 3.5. The basic structure of the proposed 2D-UNet [Rajput et al. (2023b); Rajput, Kapdi, Roy, & Raval (2024)].

The 2D image slices have been randomly chosen to serve as the inputs for the encoder path of every planar model. The input image size for each planar model is specified in the preceding paragraph. In each layer of the encoding path, strided convolution is utilized to halve spatial resolution while simultaneously doubling the count of channels. The initial count of channels is 64; whereas similar to the approach in Noori et al. [Noori et al. (2019)], each ResNet block consists of two convolution blocks with a kernel size of 3×3, along with batch normalization and a PRelu unit [K. He et al. (2015)]. On the decoder side, each layer decreases the count of feature maps by half and doubles their size, using an upsampling layer and a 2×2 convolution layer. Further, each layer feature map on the decoder side

is concatenated to the encoder layer feature map. Finally, it is passed to the attention block, which recalibrates each feature map (channel-wise and/or spatially) and forwards it to the subsequent layers. The network uses softmax activation to output 4 channels, each for the NCR/NET, ET, ED, and background. Further, these channels are combined to dissect the tumors into ET, TC, and WT regions.

This BTS network, built utilizing TensorFlow and Keras, runs on a Quadro RTX5000 with a 16GB GPU and 128GB RAM. During the training phase, random slices of dimensions $192(W) \times 152(H)$ are produced for the axial model, $192(W) \times 144(H)$ for the sagittal model, and $152(W) \times 144(H)$ for the coronal model, extracted from the input scans, where W is the width, and H is the height of an image slice. To accommodate memory constraints, a batch size of 16 is set. The network undergoes training utilizing a stochastic gradient descent optimizer with a learning rate of 0.008. A combination of cross-entropy (CE_{loss}) and generalized dice-loss functions (GD_{loss}) [Sudre et al. (2017)] is used as a loss function to minimize class imbalance, as explained in the equation 3.1. Further, if the validation loss fails to decrease for 30 epochs, the learning rate is reduced by 0.5. The Keras early-stopping mechanism also halts training if the validation loss does not decrease for 50 epochs.

$$
\begin{aligned}
Loss &= GD_{loss}(G_t, P_s) + CE_{loss}(G_t, P_s) \\
GD_{loss}(G_t, P_s) &= 1 - 2\frac{\sum_{l=1}^{C}(W_l \times \sum_{i=1}^{N} g_{li} \times p_{li})}{\sum_{l=1}^{C}(W_l \times \sum_{i=1}^{N} g_{li} \times p_{li})} \\
CE_{loss}(G_t, P_s) &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{C}(g_{li} \times \log p_{li})
\end{aligned}
\tag{3.1}
$$

where $W_l = \frac{1}{(\sum_{i=1}^{N} g_{li})^2}$ represents the adaptive weight for the l^{th} channel, p_{li}

denotes the estimated likelihood of class l for voxel i, g_{li} denotes the actual label for class l for voxel i, $g \varepsilon G_t$ where G_t denotes the actual label and $p \varepsilon P_s$ where P_s denotes the predicted label. Lastly, post-processing using the connected component algorithm is performed on the segmented outcomes obtained from the *super-ensemble* BTS model.

3.1.3 Attention mechanism: CA, CCSA, SCSA, CCSAv1

The attention mechanism is employed to boost the accuracy of the encoder-decoder network. It explores the interdependencies between the channels or spatial locations. The primary purpose of this attention mechanism is to enable the network to use the most pertinent portions of the input feature sequence in an adaptable way, depending on the context it carries. Those input features can be within the feature map (spatial attention) and across the feature map (called channel attention), which can further focus on enhancing performance. The most pertinent input feature vectors receive the highest weights; less informative vectors receive lower weights [Vaswani et al. (2017)]. The attention techniques of this work are simple in structure and marginally impact the model complexity [Hu et al. (2018); Roy et al. (2018); Woo et al. (2018)].

Channel attention (CA)

Each channel is weighted equally in standard convolution when producing the output feature maps. It produces activation maps that collectively capture the spatial and channel information by learning filters that obtain local spatial patterns across every input channel. The channel attention mechanism weights each channel sep-

arately, allowing it to recalibrate semantic attributes per salient features it carries [Hu et al. (2018); L. Chen et al. (2017)]. While significant efforts are invested in refining the combined representation of spatial and channel feature information, there remains a substantial gap in the exploration of encoding spatial-wise and channel-wise information separately. Contemporary studies have tried to resolve this problem by formally demonstrating the interdependencies between feature map channels. Recently, an architecture framework called "squeeze and excitation (SE)" network was proposed, which has a simple structure and is easy to integrate into the existing network. Therefore, we have also used SE techniques to implement channel attention.

SE works on two principles: First, it utilizes GAP to compress the feature maps into a single numeric value (number of channels) to gain global statistics of the channels. Second, the excitation operator captures nonlinear and non-mutual exclusive relations between the channels and a gating mechanism that utilizes a sigmoid activation to assign weights to each channel based on the information it holds. As depicted in Figure 3.6, the initial fully connected (FC) layer reduces the size of the channel (C) to $C/ratio$. While the native method [Hu et al. (2018)] recommended using $ratio = 16$. Based on empirical findings, we determine that $ratio = 8$ produces the most favorable outcomes for our method. Opting for $ratio = 16$ would notably decrease the channel counts, constraining the model's capacity to capture the context effectively. In summary, it squeezes across the spatial dimensions and weights across the channel dimensions to improve feature representations by exploring interdependencies between feature channels.

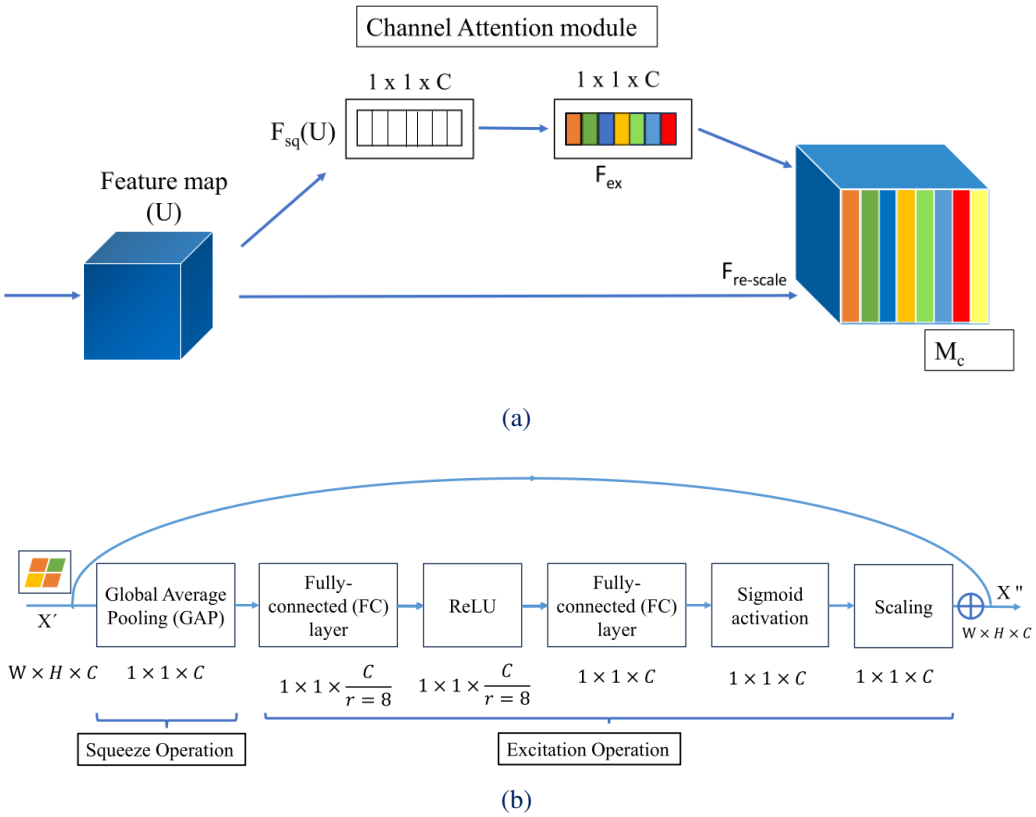


Figure 3.6. (a) A channel attention (CA) module [Hu et al. (2018)]. (b) The modified *CA* mechanism with *ratio* value is 8 [Rajput et al. (2023b); Rajput, Kapdi, Roy, & Raval (2024)].

Mathematically, a squeeze operation on the feature map U can be defined as:

$$z = F_{sq}(U) = \sum_{c=1}^{C} \left[\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \right] = GAP \tag{3.2}$$

where, $(U) = \{u_1, u_2, ...u_c\}$ is a feature-map which consists of multiple filters $\{1, 2, ...c^{th}\}$. The excitation can be defined as:

$$F_e x = F(z, W) = \sigma \left[\overbrace{W_2(\underbrace{\delta(W_1(z))}_{1^{st}\text{FC layer}})}^{2^{st}\text{FC layer}} \right] \tag{3.3}$$

where σ is a sigmoid function and W_1 and W_2 are the weights of 1^{st} and 2^{st} FC layers, respectively, and z is the feature map after the squeeze operation. The CA module is presented in Figure 3.6 where Figure 3.6 (a) defines operations of the CA mechanism, and detailed working is presented in Figure 3.6 (b).

Concurrent channel and spatial attention (CCSA)

Similarly, taking clues from the study mentioned in [Roy et al. (2018)] suggests calibrating the feature maps parallelly along spatial and channel dimensions. Combining the results can enhance the spatial and channel information. Therefore, we experiment with infusing concurrent channel and spatial squeeze-and-excitation blocks into the segmentation network. The CCSA module is illustrated in Figure 3.7.

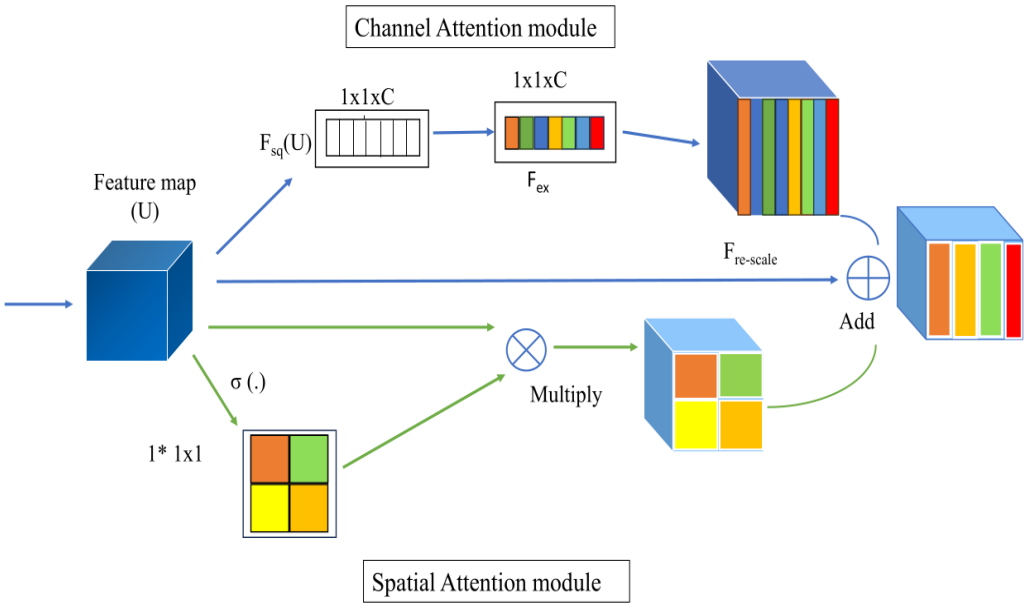


Figure 3.7. Concurrent channel and spatial attention (CCSA) [Roy et al. (2018); Rajput et al. (2023b); Rajput, Kapdi, Roy, & Raval (2024)].

Sequential channel and spatial attention (SCSA)

This module sequentially creates attention maps spanning the channel and spatial dimensions from an intermediary feature map. It subsequently aggregates these attention maps with the input feature map, allowing for an adaptable recalibration of features [Woo et al. (2018)]. In the SCSA architecture, the channel attention mechanism is employed through the *CA* module. In this scenario, rather than using GAP, two specific components (max pooling function and average pooling function) with identical architectures are constructed and combined to estimate the weight of the feature maps. The first FC layer's output channel (C) is also reduced to $C/ratio$. Here, we also empirically found that $ratio = 8$ generates the best results. It incrementally derives a 1D channel attention map (A_c), such that A_c ε $\{C \times 1 \times 1\}$ and a 2D spatial-attention map (A_s), such that A_s ε $\{1 \times H \times W\}$. For the channel and spatial module, channel and spatial information aggregation uses max-pool and average-pool operations that capture effective features across the channel and spatial dimensions. Mathematically, the whole attention procedure can be described as:

$$F' = A_c(F) \oplus F \tag{3.4}$$

$$F'' = A_s(F') \oplus F' \tag{3.5}$$

where, $\oplus$ indicates elementwise multiplication, F is a feature map, $A_c(F)$ is the channel attention map, $A_s(F')$ is the spatial-attention map and F'' is the ultimate output.

Whereas $A_c(F)$ and $A_s(F')$ can be defined as:

$$A_c(F) = \sigma\left[mlp\{AP(F)\} + mlp\{MP(F)\}\right] \tag{3.6}$$

$$A_s(F') = \sigma\left[f_c^{7\times7}\{mlp\{AP(F')\} \cdot mlp\{MP(F')\}\}\right] \tag{3.7}$$

where σ is the sigmoid function, mlp is the multi-layered perceptron, AP, MP are the average and max pooling operations on the feature map, respectively, $\cdot$ is the concatenation operation, and $f_c^{7\times7}$ indicates a convolution operation with a 7×7 kernel size. The SCSA module is illustrated in Figure 3.8.

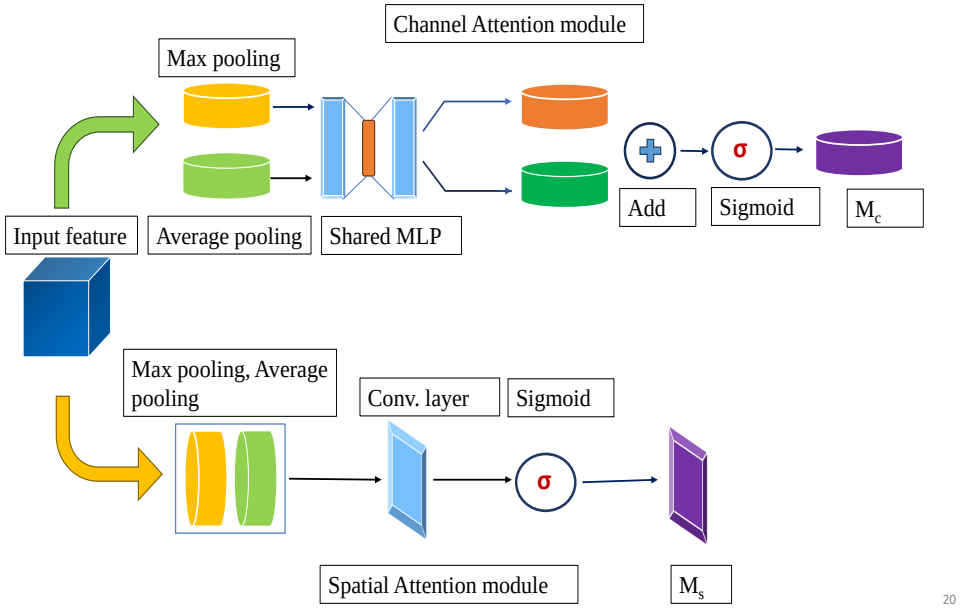


Figure 3.8. Sequential channel and spatial attention (SCSA) [Woo et al. (2018); Rajput et al. (2023b); Rajput, Kapdi, Roy, & Raval (2024)].

Improved Attention mechanism: *CCSAv1*

The framework of the improved *CCSAv1* mechanism is presented in Figure 3.9, showcasing the concurrent integration of channel and spatial attention mechanisms. We implement channel attention using the CA module. Whereas to implement spatial attention, we employ a convolution layer with a 1×1 kernel size to perform the squeeze operation, followed by an excitation operation where a sigmoid function is applied to generate attention weights used for modulating the original feature maps [Rajput, Kapdi, Roy, & Raval (2024)]. We observe that (*CCSA*) underperforms in the experimental trials of the CA and SCSA attention techniques. Hence, we adapt the Spatial and Channel Squeeze and Excitation Block component from Roy et al. [Roy et al. (2018)] into the *CCSA*, which is referred to as *CCSAv1*.

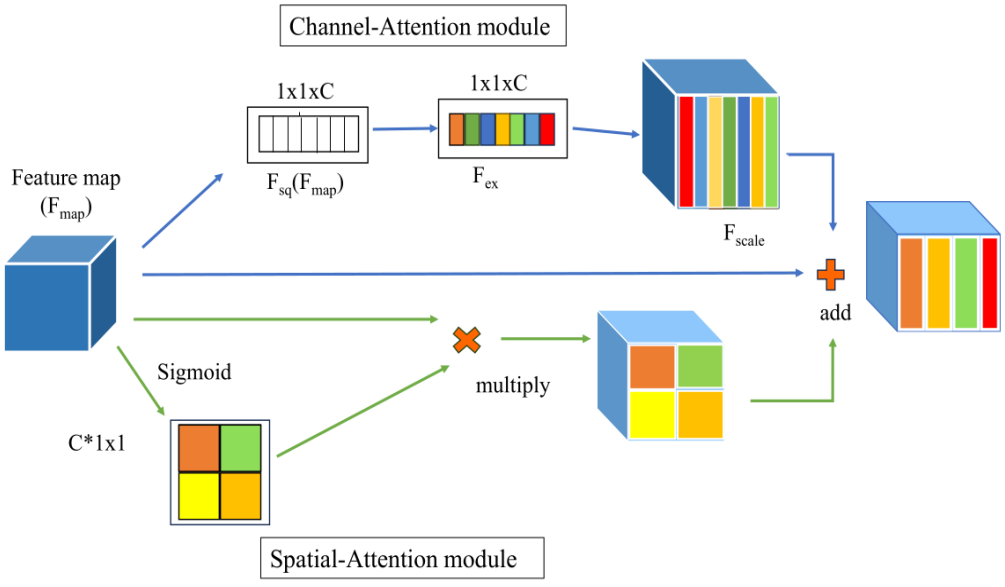


Figure 3.9. The improved *CCSAv1* mechanism [Roy et al. (2018); Rajput et al. (2023b); Rajput, Kapdi, Roy, & Raval (2024)].

In the native approach by Roy et al. [Roy et al. (2018)], spatial attention compression includes utilizing $C \times K$, where C is the count of filters, and K is kernel size with 1×1. However, unlike the native approach where C remains constant at 1 for each level of U-Net, our enhanced *CCSAv1* mechanism dynamically adjusts C, increasing with the corresponding depth of the UNet. It starts at a base value of 64 and increases gradually to 512. This improved *CCSAv1* method conserves more data by compressing it into a slightly larger form, thus retaining vital details without overly enlarging the model's dimensions. By dynamically modifying the count of filters according to the levels of the U-Net, the model becomes more adaptable and proficient at capturing features suitable for different complexities, which may lead to enhanced understanding and segmentation capabilities. Moreover, this dynamic adaptation of filter numbers, corresponding to the U-Net's depth, enables the model to handle diverse scales of features and objects present in images, thereby enhancing the overall performance of segmentation.

In summary, for the triplanar BTS network, three variants of networks are studied and optimized:

- *Super-ensemble* using CA, CCSA, SCSA attention mechanism.
- *Super-ensemble* using CA, CCSAv1, SCSA attention mechanism.
- *Super-ensemble* using CA, CCSAv1, SCSA attention mechanism and post-processing.

3.1.4 Results and discussions

As discussed in Section 3.1.1, we train nine models comprising three distinct attention-based architectures across three anatomical planes in 2D slices. Further, ensembles are created by combining their output probabilities and averaging them to generate final segmentation results. Table 3.1 presents a quantitative evaluation of each variant on validation datasets. In the recent BraTS2023 challenge, the benchmark scores achieved were DSC of 0.900 (WT), 0.867 (TC), and 0.850 (ET), along with HD scores of 14.940 (WT), 14.467 (TC), and 17.699 (ET) on the validation set [Ferreira et al. (2024)]. These scores were achieved through rigorous data augmentation and ensemble segmentation models to produce segmentation results. In contrast, on the BraTS2020 validation dataset, the benchmark DSC scores were 0.912 (WT), 0.857 (TC), 0.798 (ET), and HD scores were 3.730 (WT), 5.640 (TC), 26.410 (ET) [Isensee et al. (2021)]. In their segmentation method, the authors utilized rigorous data augmentation, region-based training, an increased batch size, and an ensemble of models.

In our proposed segmentation method, ensemble models give the best segmentation results from all the proposed variants. Table 3.1 showcases the top five performing models across various variants, distinctly marked in red based on their performance on the validation datasets, and their respective training outcomes can be observed in Table3.2. Considering the DSC improvement of *Super-Ensemble (with/without post-processing)* among all variants shown in Table 3.1, we use it for comparison with the leading BraTS2020 methods in subsequent analysis.

TABLE 3.1
The performance assessment of each variant of models was conducted on the validation set from BraTS2020. The most favorable outcomes are emphasized in bold font. Training and validation outcomes from identical models are differentiated utilizing consistent colors, with their corresponding training outcomes detailed in Table 3.2. The top two performing models from the BraTS2020 challenge are listed in the initial two rows. Matching colors are employed for the *CCSA* and *CCSAv1* models of the same category (exempting the initial two rows). All the results presented herein are acquired from the BraTS-challenge evaluation portal[+]

Dataset	Model Type	Model Name	Parameter Numbers in millions (M)	DSC			HD		
				ET	WT	TC	ET	WT	TC
Validation	Ensemble of five 3D U-Nets	Isensee et al. (2021)	$30.2 \times 5 =$ 156 M	0.798	0.912	0.857	26.410	3.730	5.640
Validation	Ensemble of eleven 3D U-Nets	Yuan (2021)	$16.5 \times 11 =$ 181.5 M	0.793	0.911	0.853	18.196	4.097	5.888
Validation	Axial	CA	10.246 M	0.610	0.873	0.754	56.627	9.532	14.038
Validation	Axial	CCSA	10.247 M	0.618	0.854	0.738	52.666	13.759	17.896
Validation	Axial*	CCSAv1*	10.589 M	0.679	0.867	0.788	42.230	10.885	11.496
Validation	Axial	SCSA	10.333 M	0.670	0.866	0.760	46.227	9.328	15.22
Validation	Coronal	CA	10.246 M	0.633	0.836	0.747	54.469	19.712	22.547
Validation	Coronal	CCSA	10.247 M	0.629	0.827	0.743	58.344	23.447	22.538
Validation	Coronal*	CCSAv1*	10.589 M	0.662	0.840	0.740	42.864	16.491	25.476
Validation	Coronal	SCSA	10.333 M	0.642	0.825	0.738	52.114	25.978	19.806
Validation	Sagittal	CA	10.246 M	0.663	0.850	0.737	48.509	10.414	15.969
Validation	Sagittal	CCSA	10.247 M	0.652	0.835	0.759	49.567	14.381	15.010
Validation	Sagittal*	CCSAv1*	10.589 M	0.663	0.833	0.747	48.829	10.955	18.954
Validation	Sagittal	SCSA	10.333 M	0.680	0.853	0.756	46.402	17.448	18.758
Validation	Axial Ensemble	CA-CCSA-SCSA	30.825 M	0.661	0.879	0.767	41.232	6.754	12.034
Validation	Axial Ensemble*	CA-CCSAv1-SCSA*	31.167 M	0.683	0.883	0.784	42.620	6.285	11.863
Validation	Coronal Ensemble	CA-CCSA-SCSA	30.825 M	0.648	0.849	0.762	53.748	16.502	15.713
Validation	Coronal Ensemble*	CA-CCSAv1-SCSA*	31.167 M	0.664	0.850	0.759	47.828	14.301	18.964
Validation	Sagittal Ensemble	CA-CCSA-SCSA	30.825 M	0.682	0.863	0.762	43.714	11.391	16.799
Validation	Sagittal Ensemble*	CA-CCSAv1-SCSA*	31.167 M	0.681	0.857	0.758	48.279	10.795	14.991
Validation	Triplanar Ensemble	CA	30.737 M	0.669	0.871	0.771	44.512	6.929	12.558
Validation	Triplanar Ensemble	CCSA	30.739 M	0.667	0.860	0.773	45.395	9.764	13.708
Validation	Triplanar Ensemble*	CCSAv1*	31.767 M	0.699	0.863	0.777	35.713	7.173	14.529
Validation	Triplanar Ensemble	SCSA	30.999 M	0.693	0.868	0.776	42.274	10.120	13.989
Validation	Super-Ensemble	CA-CCSA-SCSA	92.47 M	0.699	0.875	0.782	36.752	8.037	14.846
Validation	Super-Ensemble (without postprocessing)*	CA-CCSAv1-SCSA*	93.50 M	0.704	0.875	**0.785**	35.440	6.962	14.560
Validation	Super-Ensemble-pp (with postprocessing)*	CA-CCSAv1-SCSA*	93.50 M	**0.713**	**0.873**	0.778	**31.563**	**6.347**	13.274
Validation	Native-UNet *	without attention mechanism*	7.698 M	0.669	0.862	0.786	44.046	11.750	12.841

[+] https://ipp.cbica.upenn.edu/
[*] is an improved model/

TABLE 3.2
The results on the training set are exclusively displayed for the best-performing models on the validation set (highlighted in red). The most favorable outcomes are indicated in bold text. Results from models are represented using consistent colors for both training and validation sets. (As depicted in Table 3.1).

Dataset	Model Type	Model Name	Parameter Numbers in millions (M)	DSC			HD		
				ET	WT	TC	ET	WT	TC
Training	Triplanar Ensemble	SCSA	30.999 M	0.727	0.885	0.826	39.745	9.099	9.734
Training	Triplanar Ensemble*	CCSAv1*	31.767 M	0.727	0.892	0.829	34.538	6.828	6.128
Training	Super-Ensemble	CA-CCSA-SCSA	92.470 M	0.712	0.897	0.837	40.310	6.378	7.114
Training	Super-Ensemble*	CA-CCSAv1-SCSA (without postprocessing)*	93.500 M	0.724	0.898	0.843	39.489	5.884	5.710
Training	Super-Ensemble-pp*	CA-CCSAv1-SCSA (with postprocessing)*	93.500 M	**0.736**	**0.896**	**0.841**	**35.842**	6.005	7.041

* is a modified models/

Assessment of planar networks performance and parameter counts

Comparing the planar models, we observe that each model excels in segmenting different regions. Training models using different planar views allows models to capture distinct properties of lesions. For instance, among AP model types, the CA model can dissect WT (DSC: 0.873) and TC (DSC: 0.754) regions well, whereas the SCSA model dissects all the regions well, such as for ET (DSC: 0.670), WT (DSC: 0.866), TC (DSC:0.760). Similarly, among CP model types, we observe that all the models (CA, CCSA, and SCSA) have moderate DSCs for all the lesions. However, across all model types, the DSC for ET is better than that of AP model types.

Whereas, among SP model types, the CA model dissects ET (DSC: 0.663) and WT (DSC: 0.850) regions well, the CCSA dissects TC (DSC: 0.759) well, and SCSA dissects all the lesions well (DSC: 0.680(ET), 0.853 (WT), 0.756(TC)). The parameter count for all the model types, on average, is $10.27M$.

Further, a comparison is also conducted among planar ensembles. We observe that the *Axial ensemble* and the *Sagittal ensemble* can dissect all the lesions well. The DSC for *Axial ensemble* is 0.661 for ET, 0.879 for WT, and 0.767 for TC. Whereas for *Sagittal ensemble*, they are 0.682 (ET), 0.863 (WT), and 0.762 (TC). Furthermore, *Coronal ensemble* models trained on coronal view images perform less than the other two planar images. The reason can be that some modalities are acquired in 2D axial or sagittal view, and each slice has a specific thickness. So, reconstructing it to coronal views requires an interpolation technique to fill that thickness area. It causes anisotropy in resolution [Z. Wu et al. (2022)], meaning discrepancy in resolution along different planes, which occurs due to differences in the voxel size and acquisition parameters. It can be explicitly seen in Figure 3.10, where Figure 3.10(a) is the axial view of the Flair image, and the resolution is intact compared to coronal and sagittal slice views shown in Figure 3.10(b) and Figure 3.10(c). The coronal slice has the least detailed structure. In their work on BTS, McHugh et al. [McHugh et al. (2021)] also highlighted the issue of data loss caused by the coronal plane view. On average, the parameter count of our planar ensemble models is $30.8M$.

Assessment of triplanar networks performance and parameter counts

Triplanar is an ensemble of multi-view (axial, coronal, and sagittal) model types. The triplanar ensemble of all the models has been able to dissect all the regions well. The triplanar ensemble with CA has DSCs of 0.669 (ET), 0.871 (WT), and 0.771 (TC), respectively, whereas, with CCSA and SCSA, it is 0.667 (ET), 0.860 (WT), 0.773 (TC) and 0.693 (ET) 0.868 (WT), 0.776 (TC) respectively. Likewise, comparing triplanar models, both CA and SCSA exhibit superior per-

formance compared to the CCSA model. On average, the parameter counts for all the triplanar ensemble models is $30.8M$.

Performance impact of attention mechanisms

Table 3.1 showcases the results of the Native U-Net segmentation network in the last row, which does not incorporate an attention mechanism and is exclusively trained on axial plan images. This model comprises $7.698M$ trainable parameters. In contrast, the suggested *CCSAv1* network trained on axial images has a marginally higher parameter count, with an increment of $2.89M$ compared to the Native U-Net. It displays a notable performance enhancement compared to the Native U-Net model, with an increase of nearly 0.01 in the DSC for the ET area, recognized as the most challenging region for precise segmentation. Observing the HD metric, the proposed network exhibits a 0.865 *mm* enhancement compared to the Native-UNet model.

Impact of improved *CCSAv*1 mechanism

An analysis of Table 3.1 highlights that the results of the *CCSA* model appear relatively lower compared to the *CA* and *SCSA* models, even with the integration of spatial and channel attention mechanisms. For example, in Table 3.1, the information regarding the *CCSA* model trained on the axial images (4^{th} row) exhibits Dice-scores of 0.618 (ET), 0.854 (WT), and 0.738 (TC). In contrast, the *CA* model (3^{rd} row) demonstrates 0.610 (ET), 0.873 (WT), and 0.754 (TC), and the *SCSA* model (6^{th} row) showcases 0.670 (ET), 0.866 (WT), and 0.760 (TC) respectively.

Similar observations can be obtained on sagittal (11^{th}, 12^{th}, and 14^{th} rows) and coronal planes (7^{th}, 8^{th}, and 10^{th} rows). The potential causes for the subpar

performance of *CCSA* could stem from a combination of factors. This includes integrating the attention mechanism, which might not be effectively harmonizing with the U-Net framework. It's possible that specific modulation to hyperparameters, such as size and depth of filters, are necessary. Additionally, it could be less responsive to the complexities of the BraTS2020 dataset.

We improved the spatial attention mechanism of *CCSA* by increasing the number of feature maps at each corresponding layer, resulting in *CCSAv1*. This adjustment has amplified the feature representation and strengthened the network's discriminating abilities with minimal changes to the network trainable parameters [Roy et al. (2018)]. In Table 3.1, a contrast between *CCSA* and *CCSAv1* across various model types are presented with consistent coloring (excluding the first two rows). The *CCSAv1* model showcases notable improvements over other attention models (every permutation of orthogonal planar and triplanar) while incurring only a slight rise in the parameter count. It only has $0.342; M$ additional parameters compared to the previous iteration, *CCSA*. This further validates the robustness of the *CCSAv1* model.

The *Triplanar Ensemble* of *CCSAv1*, comprising axial, sagittal, and coronal models with the *CCSAv1* attention mechanism, has achieved notable performance. The performance improvement observed in the *CCSAv1* model can be determined by its triplanar nature, which grants access to some depth insights. The model types and names highlighted in red indicate the top-performing models in our approach. Further, the *super ensemble model* with *(CA, SCSA, CCSAv1)* attention blocks has Dice-score $0.704,\ 0.875,\ 0.785$ for ET, WT, and TC with total parameters $93.50\,M$ which is $1.7\times$ and $2\times$ lesser than top ranking models shown in first two rows of the Table 3.1.

Furthermore, applying postprocessing techniques using the connected components algorithm effectively eliminates false positive labels, improving 0.009 in the DSC for ET lesion segmentation (ET: 0.713). This improvement is supported by the HD, which calculates the dissimilarity between predicted and actual segmented boundaries. Notably, the *Super ensemble-pp** model exhibits the lowest HD following post-processing (pp), emphasized in bold font. While *CCSAv1* as a standalone model has made substantial contributions across all combinations, the *Super-ensemble model** (incorporating *CCSAv1*) [1] exhibits only marginal improvement when compared to the *Super-ensemble model* (utilizing *CCSA*). For instance, the enhancement for ET is minimal, and the DSC remains unchanged for WT and TC. This phenomenon could be attributed to the averaging of probabilities obtained from multiple models, which tends to moderate significant changes from any particular model.

Segmentation observations on training data

The training outcomes of the highest-performing models, encompassing triplanar and *Super-ensemble* configurations, on a set of 75 randomly selected samples are presented in Table 3.2. The triplanar ensemble models, with approximately 31 million parameters, can accurately predict segmentation labels, achieving DSCs of 0.727 for ET, 0.892 for WT, and 0.829 for TC. In contrast, the *Super-ensemble* model with *CCSAv1* demonstrates improvements in lesion segmentation across all categories. Furthermore, post-processing of the segmentation outcomes derived from this *Super-ensemble* model results in improved DSCs and HD metrics across all lesion types, as highlighted in bold font.

[1] modified model is shown by *

Figure 3.10 visually represents the performance of the *Super ensemble-pp* (performing post-processing), illustrating challenges encountered in distinguishing extremely complex tumor legions, as illustrated in Figure 3.10 (g)-(i). Nonetheless, the model demonstrates precise segmentation of other tumor lesions. This visual evaluation offers a valuable understanding of the merits and constraints of the model in accurately segmenting tumor lesions.

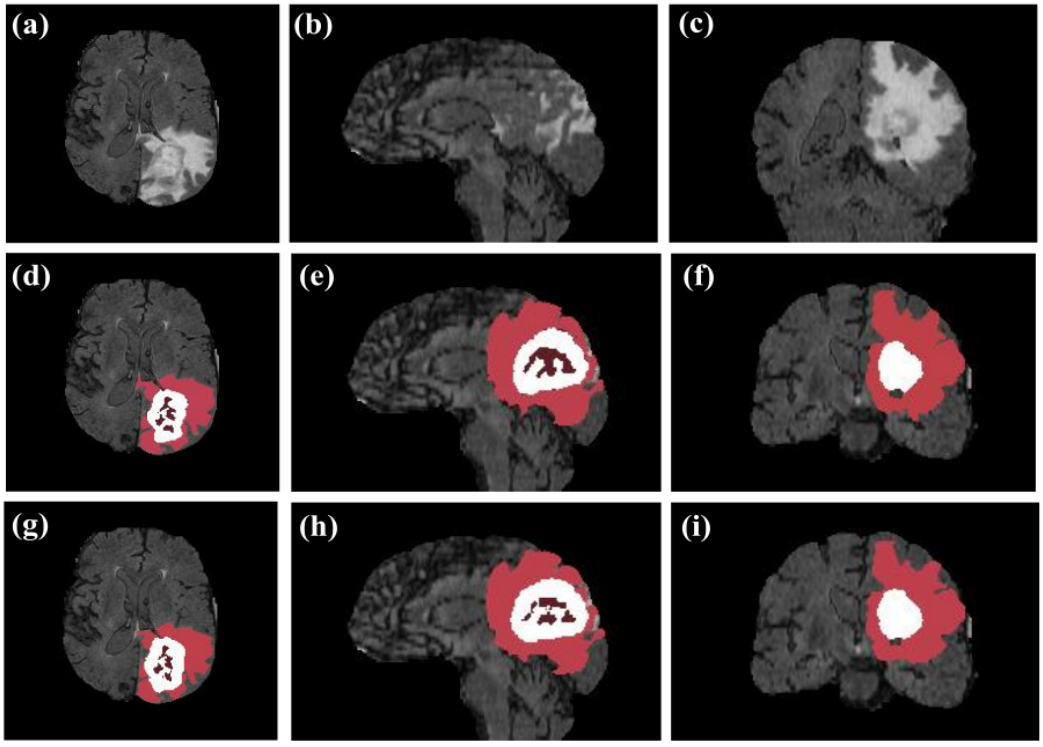


Figure 3.10. (a), (b), and (c) depict the orthogonal views, namely axial, sagittal, and coronal, of the FLAIR image, respectively. Correspondingly, (d), (e), and (f) represent the corresponding actual segmentation labels. Similarly, (g), (h), and (i) show the corresponding predicted labels. In this representation, the WT is indicated by light-brown, white, and dark-brown colors, the ET lesions are represented by white, and the TC is depicted by a combination of dark brown and white. Upon observation of slices presented in (a), (b), and (c), it is evident that the axial slice contains the most detailed anatomical information. In contrast, the coronal slice provides the least [Rajput et al. (2023b)].

The comparative analysis of computation and memory requirements between proposed models and Native U-Net are shown in Table 3.3. We can observe marginal differences among the proposed methods and between the proposed method and the Native U-Net. Despite a slight increase in memory and computation requirements of the proposed triplanar model, segmentation performance has improved significantly.

TABLE 3.3
Comparative analysis of computation and memory requirements between proposed models and Native U-Net.

	CA	*CCSAv1*	*SCSA*	*Native U-Net*
Trainable parameters	10.24 M	10.589 M	10.333 M	7.69 M
Training time per epoch	1290 ± 5 seconds	1302 ± 5 seconds	1302 ± 5 seconds	1260 ± 5 seconds
Memory size	15680 MB	15684 MB	15691 MB	15676 MB
Inference time per sample	1 second	1 second	1 second	1 second

Finally, we evaluate the proposed network against the leading 2D or 3D segmentation methods of the BraTS2020 challenge in Table 3.4. The proposed network performs better in accurately segmenting respective lesions indicated by the highlighted regions. The methods presented here outperform several 2D and 3D models, as shown in Table 3.4. This emphasizes the network's effectiveness in addressing complexities associated with lesions and delivering more reliable and accurate results.

TABLE 3.4
Performance evaluation on the validation (BraTS2020) dataset, the proposed network is assessed against the leading BTS networks from the challenge. The first row presents the details of the proposed network, with subsequent comparison results highlighting its superior performance. Specifically, the proposed network surpasses other networks, indicated by cells highlighted in yellow, with the best outcomes emphasized in bold.

		DSC			HD			Mean sensitivity	Mean specificity
Model name	Model Type	ET	WT	TC	ET	WT	TC	ET, WT, TC	ET, WT, TC
Proposed method	2D Ensemble U-Net	0.713	0.873	0.778	31.563	6.347	13.274	**0.796**	0.999
[Messaoudi et al. (2021)]	3D-U-Net	0.654	0.841	0.680	NA	NA	NA	NA	NA
[Ma et al. (2021)]	2D-U-Net	0.704	0.879	0.773	NA	NA	NA	NA	NA
[M. J. Ali et al. (2021a)]	2D, 3D Ensemble U-Net	0.748	0.871	0.748	**3.929**	**9.428**	10.090	0.780	0.996
[Agravat & Raval (2021)]	3D-U-Net	**0.763**	0.873	0.753	27.704	07.038	10.873	0.809	0.999
[J. H. Xu et al. (2021)]	2D-U-Net	0.673	0.861	0.704	40.608	7.942	15.750	NA	NA
[Soltaninejad et al. (2021)]	3D-U-Net	0.660	0.870	**0.800**	47.330	6.910	7.800	0.773	**1**
[Colman et al. (2021)]	2D-U-Net	0.676	0.886	0.672	47.620	12.110	15.740	NA	NA
[Ballestar & Vilaplana (2021)]	3D-U-Net	0.720	0.840	0.790	37.970	10.930	12.240	NA	NA
[Tarasiewicz et al. (2021)]	2D Ensemble U-Net	0.703	**0.888**	0.749	40.132	4.552	10.678	0.791	0.999
[Wacker et al. (2019)]	2D U-Net	0.685	0.878	0.753	40.385	7.936	17.675	NA	NA
[McHugh et al. (2021)]	2D U-Net	0.712	0.881	0.789	40.600	6.720	10.200	NA	NA

Ablation study

We carried out an ablation to study the impact of convolution layer kernel size k at the convolution block in yellow and upsample block in orange in Figure 3.5. At the preliminary level, we experiment with 3×3 and 2×2 kernel size *(k)*. A 2×2 kernel size increases the performance metrics and comparatively reduces the number of parameters. For a coronal planar network with a CA attention module with $k = 3$, the DSC is 0.647 (ET), 0.823 (WT), and 0.734 (TC). Whereas for

CSCA and SSCA, it is 0.614 (ET), 0.798 (WT), 0.710 (TC), and 0.548 (ET), 0.748 (WT), and 0.622 (TC), respectively. These results can be compared with kernel size $k = 2$ as shown in $6-8^{th}$ rows of Table 3.1, showing a significant increase in scores with $k = 2$.

Similarly, we experiment with a reduction ratio of 16 and 8 in the dense layer of the CA network. The segmentation results are better with a $(r = 8)$ reduction ratio. Therefore, we also keep $r = 8$ in all the channel attention modules used in CCSA and SCSA networks. Further, in the SCSA block, for implementing spatial attention, we replace $k = 7 \times 7$ kernel size as shown in Equation 3.7, with $k = 3 \times 3$ to balance parameter numbers, but the model's performance deteriorates to 0.537 (ET), 0.771 (WT) 0.608 (TC). For an axial planar model with $k = 3 \times 3$ in the SCSA attention block, the DSC is 0.670 (ET), 0.866 (WT), and 0.760 (TC). These results can be compared with kernel size $k = 7$ shown in Table 3.1 (5^{th} row), showing a significant increase in scores with $k = 7$.

3.1.5 Limitations

The model exhibits several limitations. Firstly, it faces challenges in effectively eliminating specific false positive tumor lesions. Secondly, the triplanar model does not effectively utilize the depth information of an image, thereby restricting its differentiating capacity. Finally, the considerable counts of trainable parameters in the proposed network pose a challenge for real-time segmentation of brain tumors on edge devices. Advanced edge devices like NVIDIA's Jetson embedded GPUs have memories in the range from 2 GB to 32 GB [Gealy & George (2022)]. However, on these systems, the memory is shared between the CPU cores and the GPU. Therefore, accommodating a larger model for precise and reliable real-time

segmentation on these edge devices is challenging and requires rigorous model optimization (in terms of pruning and bit precision) [O. Ali et al. (2022)].

3.1.6 Conclusions and future scope

This thesis explores a triplanar-ensemble network with a segmenting performance similar to a 3D model. Also, this approach is influenced by the fact that most large publicly available medical datasets consist of 2D images. Therefore, we study 2D networks and optimize them (2.5D) to improve their performance metrics. The proposed network uses three 2D UNet networks to generate axial, coronal, and sagittal slice predictions. These predictions are subsequently integrated into a final multiple-view prediction, which enables partial capturing of spatial information in the depth dimension. Additionally, infusing attention mechanisms into the network causes the inclusion of relevant information from channel and spatial dimensions, thereby suppressing unnecessary information, which improves the discriminating power of the segmentation model.

We can observe from this study that an ensemble of the triplanar network based on UNet provides robust BTS, requires fewer parameters, and thus requires less computational memory. The proposed approach with limited parameters (almost $3\times$ lesser) demonstrates comparable performance to a 3D model, making it suitable for brain tumor segmentation in resource-limited settings. Additionally, we observe that training models across multiple planes enable them to learn and discriminate between different tumor lesions. Models trained using axial and sagittal planar views can segment tumor lesions more robustly than those trained on coronal planar view. Combining ensembles of these models further enhances the

overall segmentation performance. Likewise, we observe that incorporating channel and spatial attention into the network sequentially enables the model to learn significant features from channel and spatial dimensions effectively. Moreover, incorporating channel attention alone into the network also increases the model's discriminating capabilities. In other words, SCSA and CA attention-based models have shown better segmenting performance than CCSA.

In summary, optimizing 2D models using the attention-based triplanar approach can compete with 3D models with limited complexity and computation requirements. These attributes can be extremely useful when implemented in resource-constrained environments or integrated with legacy systems where datasets are in 2D images. The proposed 2D network has shown comparable results to the top-performing BraTS2020 models. However, the performance and parameter numbers can be further optimized.

Future work can focus on further enhancing and refining the proposed methodology. This may involve incorporating additional post-processing methods, such as replacing ET lesions with necrotic tissue according to particular cutoff values determined through experimental analysis. Moreover, investigating the incorporation of diverse 3D, 2.5D, and 2D U-Net architectures shows potential for improving segmentation accuracy. Also, integrating functional modalities, including Magnetic resonance spectroscopic imaging (MRSI), Diffusion-weighted imaging (DWI), and perfusion-weighted MRI, can provide additional biological understanding. Combining structural information from native MRI scans and biological inference from advanced MRI scans can enhance the segmentation precision of brain tumors.

3.2 Methodologies for SD prediction

As previously mentioned, predicting survival in Glioblastoma patients is pivotal for guiding the optimal treatment and care strategies selection. This prediction relies on factors such as BTS, image characteristics, acquisition protocols, qualitative feature extractions, and provided clinical information. Furthermore, as previously discussed, our initial focus involved the development of 3D BTS methods and predicting SD within a resource-constrained environment. This is to be followed by triplanar BTS network development and SD prediction to further progress toward our objective.

We work on four variants of the triplanar BTS network, extract features, perform feature selection, and work on a diverse set of predictor models to predict the SD of GBM patients. The accuracy of predicting SD also relies on both the input features and the particular ML model employed for the task. The dimensionality of the dataset can be reduced by extracting pertinent features from raw data or selecting the most informative ones. This process not only simplifies the model but also mitigates the risk of overfitting and minimizes computational complexity, which can arise due to the curse of dimensionality. The subsequent sections elaborate on the crucial components for predicting SD.

3.2.1 Feature extraction

The feature extraction component comprises both image-based and radiomics-based features [Van Griethuysen et al. (2017)] extracted from input FLAIR and segmentation. as detailed in Table 3.5. Image-based features are directly ex-

tracted from the images. In contrast, radiomics features are extracted from the Python-based radiomics tools [Van Griethuysen et al. (2017)]. Radiomics is a prominently discussed subject in nuclear medicine and general medical imaging. Despite lacking a precise definition, it primarily endeavors to derive quantitative and ideally reproducible information from the imaging scans. This encompasses intricate patterns that prove challenging for human eye recognition or quantification [Mayerhoefer et al. (2020); Gillies et al. (2016)].

Radiomics can be utilized to capture characteristics of tissues and lesions, including shape and heterogeneity. Serial imaging enables observing changes, such as those occurring during treatment or monitoring. Secondly, radiomic data is extractable, indicating that within adequately extensive datasets, it can uncover previously unidentified markers and patterns related to the evolution, progression, and response to treatment in diseases [Mayerhoefer et al. (2020)]. In oncology, the evaluation of tissue heterogeneity holds specific significance. Genomic analyses have revealed that the extent of tumor heterogeneity serves as a prognostic factor for survival and presents a challenge to effective cancer control [Mayerhoefer et al. (2020); F. Yang et al. (2017); Burrell et al. (2013); J. Liu et al. (2018)].

In the initial phase, features are extracted from the ground truth to train the predictor model. These features are subsequently validated using the segmented results obtained from the 3D BTS network, considering that ground truth information is unavailable in the validation2020 dataset. Subsequently, the segmentation results from the triplanar network are utilized for training and validating the SD prediction. We consider four variants of the BTS network and extract four sets of the same features for training the predictor models. Whereas for validation, we validate each variant across all the variants. The variants of feature extraction

from the triplanar networks are:

- Variant 1: No BTS network (used BraTS ground-truth for SD prediction).
- Variant 2: *Super-ensemble* using CA, CCSA, SCSA attention mechanism.
- Variant 3: *Super-ensemble* using CA, CCSAv1, SCSA attention mechanism.
- Variant 4: Variant 3 with postprocessing.

TABLE 3.5
The feature set comprises 1264 features, with 1225 being radiomics-based and 39 being image-based [Rajput et al. (2023a)].

	Image-based features
Shape-based features (27)	Volume of tumor lesions (TLs), the Surface area of TLs, the area-to-volume ratio of TLs, amount of tumor, proportion of TLs, proportion ratio between each TLs.
Location-based features (12)	Centroid of TLs, the distance between the center of TLs and the brain's center.
	Radiomics-based features
Shape features (13)	Elongation, major axis length, minor axis length, mesh volume, maximum diameter row, maximum diameter column, surface area, flatness, sphericity, and surface volume ratio.
First-order features (144)	Energy of an image, maximum intensity, minimum intensity, mean intensity, median intensity, entropy, absolute deviation, interquartile range, variance, skewness, percentile, kurtosis, and uniformity.
Gray-level features (1068)	Neighboring gray-tone difference matrix (NGTDM), Gray-level co-occurrence_matrix (GLCM), Gray-level_size-zone (GLSZ), Gray-level run-length matrix (GLRLM), and Gray-level_dependence_matrix (GLDM).

Note: The parentheses numbers indicate the extracted features' quantity.

In this context, image-based features are derived by analyzing the tumor's location and shape. In contrast, radiomics-based features are obtained by examining both necrotic and non-enhanced tumor lesions by applying wavelet and Laplacian of Gaussian (LoG) filters across varying σ values between 1 to 5. The lower σ value focuses on detailed textures, and the higher σ highlights global textures features. The wavelet filters denoise the images and capture spatial and global signals [Singh & Urooj (2015)]. The LoG filter detects the centers of regions and estimates their size, shape, and location [Kong et al. (2013)]. As a result, we acquire a combined set of 1264 features. Additionally, we incorporate clinical data, specifically patients' ages, as part of our feature set. Hence, a total of 1265 features are considered for performance assessment. Some features may duplicate information or fail to contribute to the prediction, thus potentially reducing the performance of ML models by impacting both the time taken for execution (due to high dimensionality) and the performance [Kira & Rendell (1992)]. Therefore, a feature selection procedure is crucial.

3.2.2 Feature subset selection

The basic aim of feature selection methods is to eradicate irrelevant or duplicated features. In this study, we employ Recursive feature elimination (RFE) [Pedregosa et al. (2012)] and Permutation importance (PI) [MIT & Lopuhin (1965)] as feature selection methods. RFE employs a backward feature selection technique to identify the optimal feature set iteratively. Initially, it includes all features to train the model and then evaluates the importance of each feature. Subsequently, it eliminates the least important features and re-trains the model. This process is re-

peated, ranking features based on their significance and removing the least significant ones until the optimal feature subset is achieved. The details of the dominant features selected by RFE are provided in Table 1 of the Supplementary section.

On the other hand, PI is a model inspection technique that evaluates the impact of individual features on the performance of a trained model. It determines the impact on the model's performance by randomly perturbing the desirable feature's value. A significant reduction in performance following rearrangement indicates the feature's importance, while minimal impact suggests its lower significance. The steps to calculate PI are shown in algorithm 1.

From PI, we obtain the weight of the features based on their significance which ranges from -35 to 1309.95. Typically, dominant features are assigned higher weights compared to others. A zero or negative weight indicates that the feature does not contribute to the prediction. Therefore, we have chosen 100 as the optimal threshold to eliminate small or negative values prompting us to exclude such features and reduce the feature set to 180. Further, these features undergo additional analysis utilizing the SRC, employing an absolute threshold value of 0.5. As a result, we eliminated redundant features, reducing the feature set to 29.

The process for feature selection proceeds as outlined below:

- We filter out features based on their weights generated through PI, representing their influence on the outcome. Features with a weight below the threshold of 100 are excluded, resulting in 180 features.

- Additionally, we further refine the selection by identifying weaker features among the initial 180 using the SRC and a sorting procedure. Here's how it works: (a) We iterate over each feature within the set of 180 features,

beginning with the feature with the lowest PI weight. For each feature, we compute its SRC with the remaining 179 features. (b) Features with an SRC value below 0.5 are selected. (c) Based on this selected feature, we utilize the feature with the highest PI weight value to substitute for the feature with the lowest PI weight chosen in step (a). This iterative procedure persists for every feature within the set of 180 features, executing the loop 179 times. Ultimately, we pinpoint 29 predominant features out of 180 initial features having lesser correlations.

Brief details of the 29 chosen predominant features employing PI are presented in the Supplementary Table 2.

Algorithm 1 PI Algorithm [Rajput et al. (2023a)]:

Input: Trained model denoted as m on the Dataset D
Compute: The metric S of the model m on dataset D, such as the R^2 metric for a regression model.
for each feature j : **do**
  for each repetition r in R: **do**
    Randomly alter column j of Dataset D to create a modified version of the dataset with additional noise, for instance, $\hat{\mathrm{D}}_{r,j}$.

    Calculate the performance measure $s_{r,j}$ of trained-model m on the noisy dataset variant $\hat{\mathrm{D}}_{r,j}$.

  end for
  Measure the significance I of each feature j in the following manner:
  $I_j = s - \frac{1}{r}\sum_{r=1}^{r} s_{r,j}$
end for

3.2.3 Predictor models

As detailed in Section 3.2.1, during the initial phase, the SD regressor model undergoes training utilizing features extracted from the actual segmentation. Sub-

sequently, the validation set is employed to validate the results, utilizing 3D BTS segmentation outcomes. We employed the Random Forest Regressor (RFR) model for this task. RFR [Fernáandez-Delgado et al. (2014)] employs ensemble learning, with Decision trees (DT) serving as the fundamental building blocks. Each decision tree is built using randomly selected samples from the training set, known as "Random Forest". This method is widely used as it has been proven accurate and robust [Fernáandez-Delgado et al. (2014)] across multiple complex problems, including SD prediction [Puybareau et al. (2019); Agravat & Raval (2019)]. The RFR model often outperforms other models because it reduces variability by averaging predictions from each tree. Furthermore, randomization during tree growth and splitting aids in preventing overfitting. [Ishwaran et al. (2008)]. Hence, the RFR model is robust for predicting brain tumor patients' survival [Rajput et al. (2021)]. A five-fold cross-validation technique is employed to train the RFR model. Additionally, the model's hyperparameters undergo fine-tuning using the grid search technique. The model's hyper-tuned parameters are *maximum tree depth, count of trees within the forest, number of features to consider at each split, and the minimum sample count needed at an internal node for a split*.

Additionally, as part of the ablation study, we used other widely used BraTS predictor models such as ANN, Linear regressor, and Gradient boosting regressor (GBR). All these models are employed using the best-performing approach to the BraTS challenges.

Subsequently, in the second phase, as mentioned in section 3.2.1, we extract these features from Variants 1 to 4 (enlisted in section 3.2.1) and employ them as four separate training sets. For the validation set, ground truth is not publicly available. Consequently, we extract features from the segmented results of the

last three variants (referred to as *Val-Variant2, Val-Variant2, and Val-variant3*) and employ them to predict SD. Each of these training variants is assessed across all the variants of the validation feature sets. This approach enables us to pinpoint the model that excels at accurately predicting SD and evaluate if the enhanced segmented results contribute positively to predicting SD days. Additionally, analyzing features set within these variants aids in identifying their behavioral patterns and answering how robust these features are.

We train the RFR and Extra random forest (Extra RF) regression models using this feature set. Both these models are based on ensemble learning principles. However, Extra RF splits nodes randomly and prohibits sample replacement during tree-building. Allowing random splits increases the bias of individual trees but decreases their variance. This can lead to a more stable and reliable model, especially when dealing with noisy or high-dimensional data. Further, to optimize the model's performance, we conduct hyperparameter tuning using grid search techniques. The fine-tuned hyperparameters include *the maximum count of features at each split, maximum depth of the tree, count of trees within the forest, the minimal count of sample needed at a leaf node at a splitting point, and stochastic nature of bootstrapping samples*.

3.2.4 Results and discussions

The BraTS evaluation platform [Spyridon (Spyros) (2021)] evaluates the predictor models. Furthermore, we utilized the top-performing models from the BraTS-2020 challenge as benchmarks to assess our results.

Correlation study of dominant features

In order to understand the relationship among features, we plot the correlation matrix of the feature sets as shown in Figure 3.11 from the initial phase of the study (3D segmentation network) and Figure 3.12 from the second phase of the study (triplanar segmentation network). (Refer to Supplementary Table 3 for the annotation of the features).

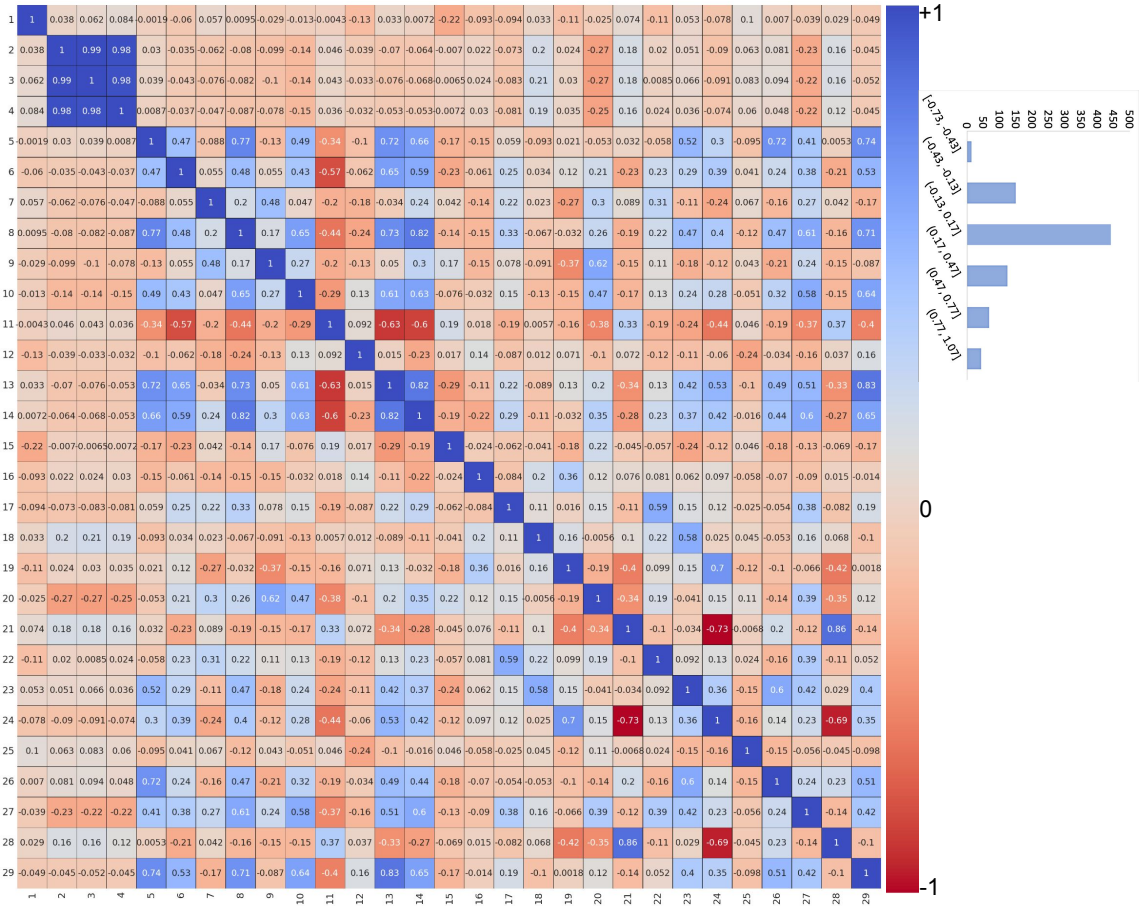


Figure 3.11. Correlation matrix among the features obtained through the PI feature selection technique. The frequency chart on the right side illustrates the frequency and range of correlations in the correlation matrix. (Refer to Supplementary Table 2 for features notation.) [Rajput et al. (2023a)].

The correlation between features serves as an indicator of whether they offer redundant or distinctive information. Higher absolute correlation values imply redundancy within the feature set, potentially hindering the model's performance by promoting overfitting or diminishing its generalization ability. Conversely, lower correlations suggest that features are uncorrelated and carry unique insights. This unique information can enhance the model's performance by providing diverse perspectives, improving its ability to generalize and make accurate predictions. A positive correlation between features means if one feature increases, the other tends to increase, whereas a negative correlation between features means that if one feature increases, the other tends to decrease.

In Figure 3.11, the correlation plot illustrates that most features exhibit low correlation, indicating that they capture distinct phenotypic properties. Moreover, the histogram on the right confirms this finding by illustrating that most selected features exhibit correlation coefficients between -0.13 and +0.17, indicating their uncorrelated nature. This provides further validation for our selected features.

Similarly, Figure 3.12 displays the correlation plot for *Variant 4*, the best-performing model. The plot demonstrates that most feature correlations fall within the -0.25 to +0.25 correlation values range. This indicates that these features exhibit low correlation, suggesting high independence. Such independence among features enables them to capture distinct phenotypic information from tumor lesions effectively. By leveraging these uncorrelated features, our model can provide a more nuanced and comprehensive prediction of SD for glioblastoma patients, potentially improving prognostic accuracy.

Figure 3.12. Correlation map among the features derived from one of the best-performing variants (*Variant4*) (Refer Supplementary Table 3 for features annotation) [Rajput, Kapdi, Raval, et al. (2024)].

Results of the SD predictor models

Table 3.6 illustrates the comparison between our SD prediction results obtained during the initial phase of the study and the top-performing methods of the BraTS2020 challenge.

TABLE 3.6
Performance Comparisons of proposed predictor method with top-ranking methods on the training and validation datasets of BraTS2020. The highlighted method denotes the most effective model from the BraTS2020 challenge. Where UA denotes Unavailable.

Dataset	Method	Accuracy	MSE	medianSE	stdSE	SRC
Training	[McKinley et al. (2020)]	UA	UA	UA	UA	UA
	[Asenjo & Solís (2021)]	0.822	55499.71	11351.02	147319.00	0.833
	[Bommineni (2021)]	UA	UA	UA	UA	UA
	[M. J. Ali et al. (2021b)]	0.641	62305.61	05745.64	200788.00	0.632
	Proposed Method	0.538	60668.61	16037.10	125873.00	0.754
Validation	McKinley et al. (2020)	0.414	098704.66	36100.00	152176.00	0.253
	[Asenjo & Solís (2021)]	0.520	122515.80	70305.26	157674.00	0.130
	[Bommineni (2021)]	0.379	093859.54	67348.26	**102092.00**	0.280
	[M. J. Ali et al. (2021b)]	0.483	105079.40	37004.93	146376.00	0.134
	Proposed Method	**0.552**	**79826.24**	**14148.89**	148288.00	**0.711**

A robust approach should demonstrate proficiency across various performance metrics besides accuracy, as each metric evaluates the models based on different criteria. Hence, we compare the proposed method with best-ranking models [McKinley et al. (2020)] and report the improvement as computed using Equation 3.8. Here the percentage enhancement ϕ for each performance measure x of our proposed method P is determined by:

$$\phi(x) = \frac{Proposed\,model\,(P) - Best\,ranking\,model\,(S)}{Best\,ranking\,model\,(S)} \times 100 \qquad (3.8)$$

As a result, the SD outcomes of the proposed method demonstrate a 33.33% enhancement in accuracy. There is a 19.13% enhancement in MSE, which evaluates the variability around the fitted regressor model and indicates the disparity between the actual and predicted values. However, it is responsive to the outliers [Minoru (2021)]. For medianSE, there is a 60.80% enhancement, which employs the median value of the residuals and remains resilient to outliers. We can observe

a 2.62% enhancement in stdSE and a 181.03% enhancement in the SRC measures often used to measure the relation between the therapy response and the survival days [Molina et al. (2019)]. The results obtained across different metrics demonstrate the robustness of the prediction [Minoru (2021)]. As illustrated in Table 3.6, our model consistently excels across all established measures utilized for survival prediction.

Following the initial phase of the study, the subsequent stage presents outcomes derived from the variants of the triplanar network as detailed in Table 3.7. The results are based on RFR and ExtraRF regressor models from the BraTS2020 training and validation datasets. Further, performance comparisons with various state-of-the-art are shown in Table 3.8. We can observe that the RFR model, which is trained using the feature set generated from *Variant1* (actual truth segmentation), achieves the best accuracy on the training set (0.590) and the validation set (0.607) (specifically, *Val-Variant4*). Likewise, training using *Val-Variant2* features, the RFR model demonstrates strong performance with an accuracy of 0.552 on the validation set (*Variant2*).

Similar observation can be made for the ExtraRF model trained on *Variant1*, which achieves the best accuracy on the training set, whereas trained on *Variant2*, which demonstrates strong performance on validation set *Val-Variant4*. Notably, the ExtraRF model trained on *Variant2* stands out with the lowest medianSE error among all the variants. Further, we observe that *Variant4* feature set exhibited strong performance across all variants of the validation set when trained using both regressor models. Various other performance metrics substantiate this enhancement in performance. This particular feature set has outperformed the *Variant1* feature set, which is derived from ground truth regarding performance

measures, including MSE, medianSE, and stdSE. When considering SRC in a broader context, where its average value is 0.50, it becomes clear that the feature combination exhibits robustness in accurately predicting SD. We can observe that both the regressor models have shown consistent performance. However, the ExtraRF model performs better than the RFR model in terms of medianSE. Further, both the RFR and ExtraRF models trained on *Variant4* obtained the second-best accuracy on all the validation sets.

TABLE 3.7
Performance evaluation on BraTS2020 training and validation datasets. The results of two models (RFR and ExtraRF) with four variants are shown in this figure. Boldface numbers indicate the best outcomes. Where MSE= mean squared error, medianSE = median squared error, stdSE = standard deviation squared error, and SRC = Spearman ranking coefficient.

	RFR-Variant1					RFR-Variant2				
Dataset	Accuracy	MSE	medianSE	stdSE	SRC	Accuracy	MSE	medianSE	stdSE	SRC
Training	**0.590**	59961.29	14329.44	130263.47	0.75	0.494	66597.44	20926.74	154478.65	0.79
Val-Variant2	0.517	83517.04	26053.75	141127.31	0.46	**0.552**	84889.59	27419.79	145913.19	0.38
Val-variant3	0.517	81758.66	33743.34	143279.04	0.50	0.379	82299.18	34042.45	140221.36	0.40
Val-Variant4	**0.607**	84583.28	25863.77	149488.19	0.52	0.536	88256.21	32317.57	152391.84	0.41
	RFR-Variant3					RFR-Variant4				
Training	0.420	70848.61	19095.99	164996.23	0.73	0.540	52490.06	13735.40	110568.72	0.84
Val-Variant2	0.517	89247.45	30042.56	166046.05	0.482	**0.586**	**76529.43**	41402.78	**130978.47**	0.52
Val-Variant3	0.517	88071.62	35003.74	161756.57	0.46	**0.552**	78062.18	34680.22	132088.99	**0.54**
Val-Variant4	0.500	96315.14	29462.49	182386.75	0.38	**0.571**	82070.60	40678.11	138345.10	0.47
	Extra RF-Variant1					Extra RF-Variant2				
Training	**0.573**	72497.92	21034.61	145370.679	0.64	0.513	84720.58	26284.83	164973.83	0.56
Val-Variant2	0.448	88824.51	27677.49	147216.69	0.35	0.517	82763.34	**20467.30**	132469.19	0.38
Val-Variant3	0.379	88509.32	26823.04	147955.13	0.37	0.448	82531.81	24575.59	130747.95	0.36
Val-Variant4	0.483	87686.49	24668.79	150750.72	0.37	**0.571**	87835.71	19673.74	139329.53	0.35
	Extra RF-Variant3					Extra RF-Variant4				
Training	0.509	70212.48	26440.58	130157.39	0.60	0.504	59927.38	20101.86	107835.95	0.72
Val-Variant2	0.517	84665.76	25455.24	133754.20	0.36	**0.586**	**76529.43**	41402.78	**130978.47**	0.52
Val-Variant3	0.483	83751.74	30857.99	131561.07	0.36	**0.552**	78062.18	34680.22	132088.99	**0.54**
Val-Variant4	0.536	87761.17	24263.52	139863.25	0.35	**0.571**	82070.60	40678.11	138345.10	0.47

[a] Validation results from the BraTS2020 challenge online evaluation portal: https://ipp.cbica.upenn.edu/

Lastly, we have compared our best-performing model on all the performance metrics with the top-ranking and leading models in Table 3.8. The bold-faced text

shows the best results. Our proposed method has surpassed the leading approach in both accuracy and MSE. However, in terms of SRC (another vital metric), it ranks as the second-best performer.

TABLE 3.8
Performance comparison between proposed and best-performing methods on training and validation brats2020 datasets. The data was sourced from the validation2020 Leaderboard [Spyridon (Spyros) (2021)]. The bold-faced text shows the best results. NA: Not-Available.

Dataset	Method	Accuracy	MSE	medianSE	stdSE	SRC
Training	(McKinley et al., 2020)	NA	NA	NA	NA	NA
	(Asenjo & Solís, 2021)	0.822	55499.71	11351.02	147319.00	0.833
	(Bommineni, 2021)	NA	NA	NA	NA	NA
	(M. J. Ali et al., 2021b)	0.641	62305.61	05745.64	200788.00	0.632
	(Rajput et al., 2023a)	0.538	60668.61	16037.10	125873.00	0.754
	Proposed (ExtraRF variant4)	0.504	59927.38	20101.86	107835.95	0.725
Validation	(McKinley et al., 2020)	0.414	098704.66	36100.00	152176.00	0.253
	(Asenjo & Solís, 2021)	0.520	122515.80	70305.26	157674.00	0.130
	(Bommineni, 2021)	0.379	093859.54	67348.26	**102092.00**	0.280
	(M. J. Ali et al., 2021b)	0.483	105079.40	37004.93	146376.00	0.134
	(Rajput et al., 2023a)	0.552	79826.24	**14148.89**	148288.00	**0.711**
	Proposed (ExtraRF variant4)	**0.586**	**76529.43**	41402.78	130978.47	0.525

3.2.5 Conclusions and future scope

In this study, we introduce a holistic approach for predicting SD. We determine 29 prominent features crucial for accurate SD prediction through comprehensive analysis. Our validation of these features through correlation matrix and histogram visualization confirms that each carries unique valuable information for precise prediction. Additionally, ensemble-based models with proper fine-tuning have improved SD prediction with limited samples, signifying their robust nature in minimizing overfitting. The resulting trained model demonstrates superior performance across multiple metrics compared to existing methods. Notably, it out-

performs the best-performing methods from the BraTS-2020 competition, providing more accurate predictions. This highlights the effectiveness of our proposed approach and emphasizes its potential for advancing SD prediction in clinical settings.

Further, as future work, clinical data such as age, gender, race, Karnofsky performance score (KPS) [Rajput, Kapdi, Raval, et al. (2024)], and treatment information can be obtained from the TCGA glioblastoma project [2]. This clinical information has proven to be an essential variable in the prognosis of gliomas. Combining this clinical information with our proposed feature set can further validate its effectiveness.

[2]https://www.cancer.gov/tcga

Chapter 4

Interpretability in SD Models

In recent decades, ML models have found widespread use across many fields, including medicine, military operations, social media, education, industry, transportation, trading, and smart devices. Their widespread adoption is due to their exceptional success and effectiveness. These models are adept at handling large volumes of data and discovering complex patterns, leading to significant advancements. However, despite their remarkable success, integrating these models into real-time applications is limited by the challenge of comprehending their internal mechanisms and providing explanations for the decisions made by these ML models. Gaining insight into the decision-making process of AI or ML models is highly important, especially in the medical field, where the transparency of these models is essential for establishing trust between machine and clinical experts. Typically, these models are challenging to understand due to their non-linear nature. Therefore, in this study, the primary focus involves conducting a comprehensive analysis of feature characteristics on a global scale and at the individual sample level for SD prediction. For achieving this objective, post-hoc

interpretable methods, such as SHAP [Lundberg & Lee (2017)], PDP [Friedman (2001)] and ALE [Apley & Zhu (2020)] for generating global visual explanations. Additionally, we utilized SHAP [Lundberg & Lee (2017)], and LIME [Ribeiro et al. (2016a)], were employed to derive local visual explanations.

Post-hoc interpretability methods are approaches utilized following the training of an ML model aimed at aiding in comprehending and interpreting its predictions. These methods clarify the model's decision-making process without modifying its architecture or training procedure [Molnar (2021)]. The employed post-hoc interpretable methods are listed below:

- Deriving global explanation - SHAP, PDP, and ALE
- Deriving local explanation - SHAP and LIME

Moreover, given the widespread use of radiomic features, particularly notable in our study, and their applicability in medical image-based analyses, radiomics often fall short of directly inferring tumor malignancy due to their intricate connection with genetic factors. Addressing this gap involves expanding the interpretation of radiomic features by correlating them with gene expression data. This approach enables validation of the relevance of radiomic features and the discovery of potential biomarkers for predicting tumor behavior, prognosis, or treatment response, thus enhancing clinical decision-making.

4.1 Foundational concepts of employed interpretation techniques

The primary goal of the SHAP method is to assess the influence of individual features on the prediction of a given sample. The SHAP value of any feature represents the average marginal contribution of that feature to the value of the antecedent set, considering all possible arrangements of the feature set. It can be defined in the equation 4.1 [Rozemberczki et al. (2022)].

$$(\Phi_j) = \frac{1}{\mid \Pi(N) \mid} \sum_{\pi\, \varepsilon\, \Pi(N)} \overbrace{(v(\hat{P}_j^{\pi} \cup j) - (v(\hat{P}_j))}^{\substack{\text{marginal contribution of } j^{th} \text{ feature} \\ \textit{in a coalition } \pi}} \tag{4.1}$$

where (Φ_j) represents the SHAP-value of the desirable feature j, $\Pi(N)$ signifies the potential coalitions of all features, π denotes a specific coalition, c stands for the contribution of feature(s), $(\hat{P}_j^{\pi} \cup j)$ denotes the predecessor set of feature j in a particular coalition, which includes the j feature, while $\hat{P}_j$ represents the predecessor set of feature j in a particular coalition, excluding the j feature (Rajput et al., 2023a). For instance, if $\pi = \{X,Y,Z\}$, $j = Y$ and $c\{X\} = 8$, $c\{Y\} = 10$, $c\{Z\} = 9$, $c\{X,Y\} = 18$, $c\{X,Z\} = 20$, $c\{Y,Z\} = 22$ and $c\{A,Y,Z\} = 25$, whereas the feasible predecessor sets in this instance) within a specific coalition $\pi = \{A,Y,Z\} : \{\phi, X\}$ and marginal contribution (MC) of j(=Y) is computed as: $c\{A,Y\} - c\{X\} = 20 - 8 = 12$. Moreover, computing the MC of feature j across all possible coalitions and then averaging will yield the SHAP value (Φ_Y) of feature Y. In essence, this demonstrates the influence of each feature on predicting

SD. It facilitates understanding the overall behavior of the model by integrating each sample's explanation. Refer to Supplementary Table 4 for a more detailed interpretation of this instance. The steps for computing the SHAP value of the desirable feature are presented in Algorithm 2.

The PDP illustrates the overall impact of the desirable feature on the target feature. It analyzes all samples and enables visualization and exploration of the overall relationship between the SD and input variables. The partial dependence function is expressed in Equation 4.2:

$$f(x_s) = E_c[f(x_s,\, x_c)] = \int p_2(x_c) f(x_s, x_c)\, dx_c \tag{4.2}$$

In this context, x_s denotes the feature(s) of interest for which our objective is to visualize partial dependency, and x_c denotes the rest of the feature variables utilized in training the model. $x_c = x_s'$, and $X = x_s + x_c$ constitutes the entire feature set. In PDP, we assume that feature subset x_s and x_c are uncorrelated to each other and hence can be calculated using average interaction effect [Friedman & Popescu (2008)] as shown in Equation 4.3:

$$f(x_s) = \frac{1}{n} \sum_{i=1}^{n} f(x_s,\, x_c) \tag{4.3}$$

Algorithm 3 displays the pseudocode to find the Partial dependency (PD) values of the samples.

Nevertheless, a significant drawback arises in the case of PD plots when predictor variables exhibit substantial interdependence, a frequent occurrence in extensive observational datasets. In contrast, ALE naturally accounts for interaction effects between variables, does not assume independence, provides a more gran-

Algorithm 2 Calculation for the SHAP-value of a feature requires: [Rajput et al. (2023a)]:

Input: The number of features N and their respective real values v, indicating their participation, are considered. The participation vector v for a specific feature is computed by perturbing the feature values within coalition π, where k represents the number of sampling permutations. Further details can be found in the following study [Molnar (2021)].

Output: SHAP value ϕ_j for the desirable feature $j \, \varepsilon \, N$.
for Iteration : 1, 2, ...K: **do**
  Selecting π stochastically from the set of all permutations $\Pi(N)$
  for $j \, \varepsilon \, N$: **do**
    Computing antecedent set $P_i^{\pi} = \{j \, \varepsilon \, N \mid \pi(j) < \pi(i)\}$.
    $\phi_j = \phi_j + \frac{v(\hat{P}_j^{\pi} \cup j) - (v(\hat{P}_j))}{K}$
  end for
end for

Algorithm 3 The steps of obtaining PD value of samples involves [Rajput et al. (2023a)]:

Input: Unique feature values $x_s = x_1, x_2, ...x_n$, where x is desirable feature.

Ouput: PD plot of desirable feature.
Steps:
for i ε (1, 2, . . . , k): **do**
  Replace the original x_1 values with the constant x_{1i} in the training samples.
  Calculating the predicted value vector from the modified version of the training samples.
  Calculate the average of the prediction to find $f'(x_{1i})$.
end for

The PDP for x_1 is generated by plotting the pairs $\{x_{1i}, f'(x_{1i})\} \; for \; i = 1, 2, ..n$

ular view of local effects, offers enhanced interpretability at the local level, and focuses on the local effects within the observed range, and doesn't require extrapolation to derive meaningful insights. ALE [Apley & Zhu (2020)] can be mathematically defined as underneath in Equation 4.4.

$$f_{1,\mathrm{ALE}}(x_1) \equiv \int_{x_{\min,1}}^{x_1} E[f_1(X_1,X_2)\,|\,X_1 = z_1]\,dz_1 - \text{constant} \tag{4.4}$$

$$= \int_{x_{\min,1}}^{x} \int_{p_{2|1}} (x_2|z_1) f_1(z_1,x_2)\,dx_2\,dz_1 - \text{constant}$$

where $f_{1,\mathrm{ALE}}(x_1)$ denotes the accumulated local effects of X_1 on the model's predicted outcome, $\int_{x_{\min,1}}^{x_1}$ is an integral sign, indicating that we are calculating the integral over the range of X_1 from $x_{\min,1}$, and $E[f_1(X_1,X_2)\,|\,X_1 = z_1]$ represents the average prediction when X_1 is fixed at z_1 considering all possible values of X_2. It reflects the accumulation of local effects over this interval. In equation 4.4, we calculate the local effect $f_1(x_1,x_2)$ of x_1 at $(x_1 = z_1, x_2)$, then average this local effect across all values of x_2 with weight $p_{2|1}(x_2|z_1)$, and finally accumulating this averaged local effect overall values of z_1 up to x_1. This avoids the extrapolation required in PD plots.

LIME [Ribeiro et al. (2016b)] aims to construct a model within an interpretable representation, ensuring local fidelity to the classifier's behavior. In other words, this model should accurately capture the local behavior of the classifier it explains. Mathematically, the explanation LIME derives can be defined in Equation4.5:

$$LIME\_model(x) = \mathrm{argmin}_{g \in G}\, L(f,g,\pi_x) + \Omega(g) \tag{4.5}$$

Here, x refers to the specific sample we seek to explain, while f denotes the black box model, $g \in G$ where G is a set of explainable models such as decision trees and linear models, π_x is a proximity measure between an instance z and x to establish the concept of locality around, L serves as an indicator of the extent to which g deviates from an accurate approximation of f within the locality defined by π_x and $\Omega(g)$ defines model complexity. In summary, LIME utilizes an interpretable surrogate model g, which can explain a given instance x with reduced complexity and minimized deviation within its locality. The procedure for training local surrogate models involves the following steps [Molnar (2021)]:

1. Choose the specific instance you are interested in and for which you desire an interpretation of its black-box prediction.

2. Introduce perturbations to your dataset and obtain black-box predictions for these newly created data points.

3. Calculate weights for the new samples based on their closeness to the instance of interest.

4. Train an interpretable model on the dataset, integrating the weighted variations.

5. Interpret the prediction by analyzing the local model.

Employing LIME is straightforward, but the challenge arises in accurately defining the neighborhood, a task that can vary depending on the nature of the problem.

4.2 Interpreting the biological connection between radiomics and gene expressions: Radiogenomics

As elaborated in Chapter 3, section 3.2.1, radiomics encompasses a diverse array of features, encompassing shape and texture features derived from radiological images. These features have demonstrated their capability to capture tumor heterogeneity and malignancy. Notably, this heterogeneity holds significant biological implications, intricately linked to the underlying molecular landscape. Radiomics presents an opportunity to directly interface with various omics data, including proteomics, transcriptomics, and genomics. This integration can predict clinical outcomes such as survival and treatment response, thereby facilitating the development of personalized, patient-tailored medicine (PPM) [Mayerhoefer et al. (2020)].

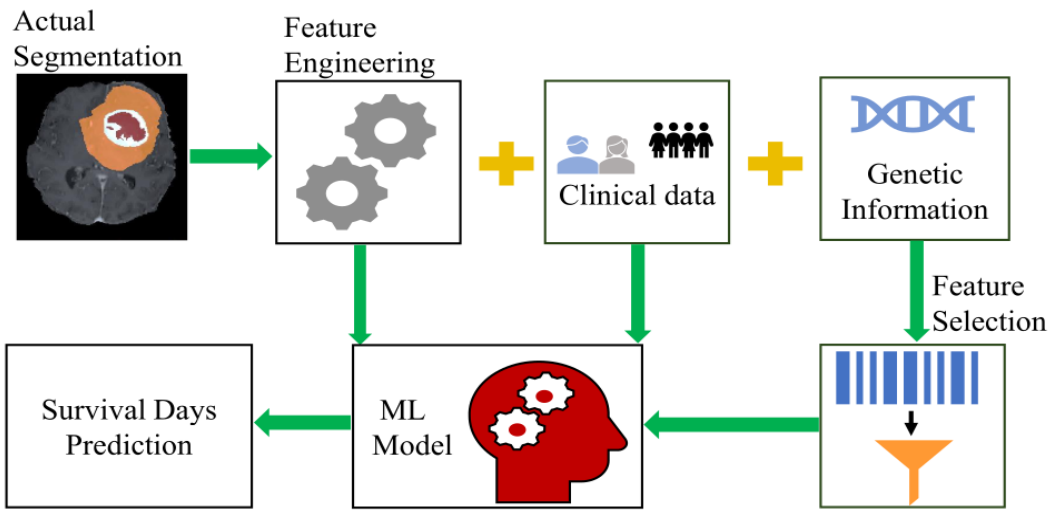


Figure 4.1. Proposed radiogenomics methodology for SD prediction.

Hence, in this study, we have expanded the SD prediction approach to incorporate additional clinical information for predicting survival days and to establish the correlation between radiomic and genomic features in gliomas, as depicted in Figure 4.1.

4.2.1 Data source for radiogenomics

The BraTS2020 dataset [Bakas et al. (2017, 2018); Menze et al. (2014)] was employed to predict BTS and SD. Detailed dataset information is provided in Chapter 1, section 1.6. Additionally, supplementary clinical data such as gender, race, and treatment details are sourced from the TCGA dataset[1], while genetic data is also obtained from the same TCGA dataset. These datasets are utilized to explore correlations between radiomics and genetics. We thoroughly search the TCGA data repository for each patient ID corresponding to the BraTS samples. The search criteria include data categories such as transcriptome profiling, data type for gene expression quantification, and experimental strategy such as RNA-seq. Consequently, each sample in our study comprises gene expression values for 60,560 genes.

4.3 Proposed framework for radiogenomics

As depicted in Figure 4.1, the methodology highlights that BTS is a prerequisite for SD prediction. This approach involves utilizing actual segmentation labels from the BraTS dataset alongside the triplanar network discussed in Chapter 3 for feature extraction. The extracted feature sets include radiomic and location-

[1]https://www.cancer.gov/tcga

based features, elaborated in supplementary Table 2, which are then employed to train various ML models. These models are subsequently validated using the 34 samples standard to the TCGA dataset.

Additionally, we conduct training and validation on TCGA samples, employing a split ratio of 0.25. This process involves testing the samples both with and without clinical details. To perform an ablation study, considering the limited number of TCGA samples, we categorize them into two groups: long-term survivors (SD $<$ 450 days) and short-term survivors (SD $>=$ 450 days). Subsequently, the models are trained and tested using a 0.25 split ratio.

Moreover, to explore the relationship between the extracted features and genes, we leverage genomic and clinical data from TCGA, where each sample contains information on 60,560 genes. Initially, as part of the feature selection process, we remove genes with a value of 0, which typically denotes missing or zero expression values. Subsequently, we utilize Pearson correlation (PC) and SRC with a $p \leq 0.001$ threshold to reduce the feature dimensions to 5830. Additionally, we analyze these 5830 genes to identify those displaying similar behavior to the extracted feature set and those showing uncorrelated or unrelated behavior. It's worth noting that the extracted feature set comprises 29 features outlined in Chapter 3, along with three clinical pieces of information obtained from TCGA.

In this study, we conducted several ablation analyses to evaluate the impact of individual components on the accuracy of SD prediction. Ablation studies are essential, especially in scenarios with limited sample sizes, as they involve identifying and removing unnecessary components or features that could contribute to model complexity. The ablation study encompasses the following:

1. For training, we rely on these segmentation outcomes for feature extraction and validated on TCGA samples:

    - Segmentation from triplanar network (with GTR status (*samples* = 117))
    - Ground-truth segmentation from BraTS2020 dataset(with all resection status (*samples* = 239))

2. For training and validation, we rely on TCGA 34 samples (with and without clinical information).
3. Reframing SD prediction challenge as a binary classification task, given the constraints of a limited dataset.

Our primary approach for prediction and classification tasks involves employing ensemble learning-based ML models, with the only exception being using an LR model for classification. The parameters for all the regressor and classifier models are shown in Supplementary Table 5.

4.4 Results and discussions

We initially discussed interpretations derived from SD prediction using interpretation tools, followed by a discussion of outcomes from radiogenomic analysis in a subsequent section 4.4.3.

During the initial phase of the study, our analysis centers on SHAP summary (both global and local perspectives), SHAP distribution plots (both global and local), SHAP force plots (local), and PDP (global). In contrast, the second phase

delves into more exhaustive and intricate explanations, incorporating additional interpretation methodologies such as LIME (local) and ALE (global), including SHAP-summary, SHAP-waterfall plots, and PDP. Subsequent sections provide a thorough examination of the 29 dominant features.

4.4.1 Global and Local Insights: Exploring SHAP and PDP Explanations

SHAP showcases the share of features in the prediction within all possible combinations of features. The SHAP value reveals how much an individual feature impacts the prediction, with the signs indicating whether the effect on the prediction outcome is positive or negative. Figures 4.2 and 4.4 show the SHAP summary and waterfall plots, respectively, corresponding to the preliminary phase of the study.

The SHAP summary plot enables us to observe the global/general and local (for individual samples) impact of features on the model's prediction. Conversely, the SHAP-waterfall plot allows us to observe the impact of features on a specific sample. This allows us to investigate the contribution of features and their values in a specific prediction, enabling a detailed examination of each feature's behavior for any desired sample.

In the SHAP-summary plot illustrated in Figures 4.2, the X-axis represents the SHAP value, indicating the influence of features on the target variable (in our case, it is *SD*). A higher absolute value indicates a higher effect on the target variable, while the sign (+/-) denotes whether the impact is positive or negative on the SD prediction. On the Y-axis, features are arranged in order of importance, from top to bottom. Each dot on the summary plot corresponds to an instance,

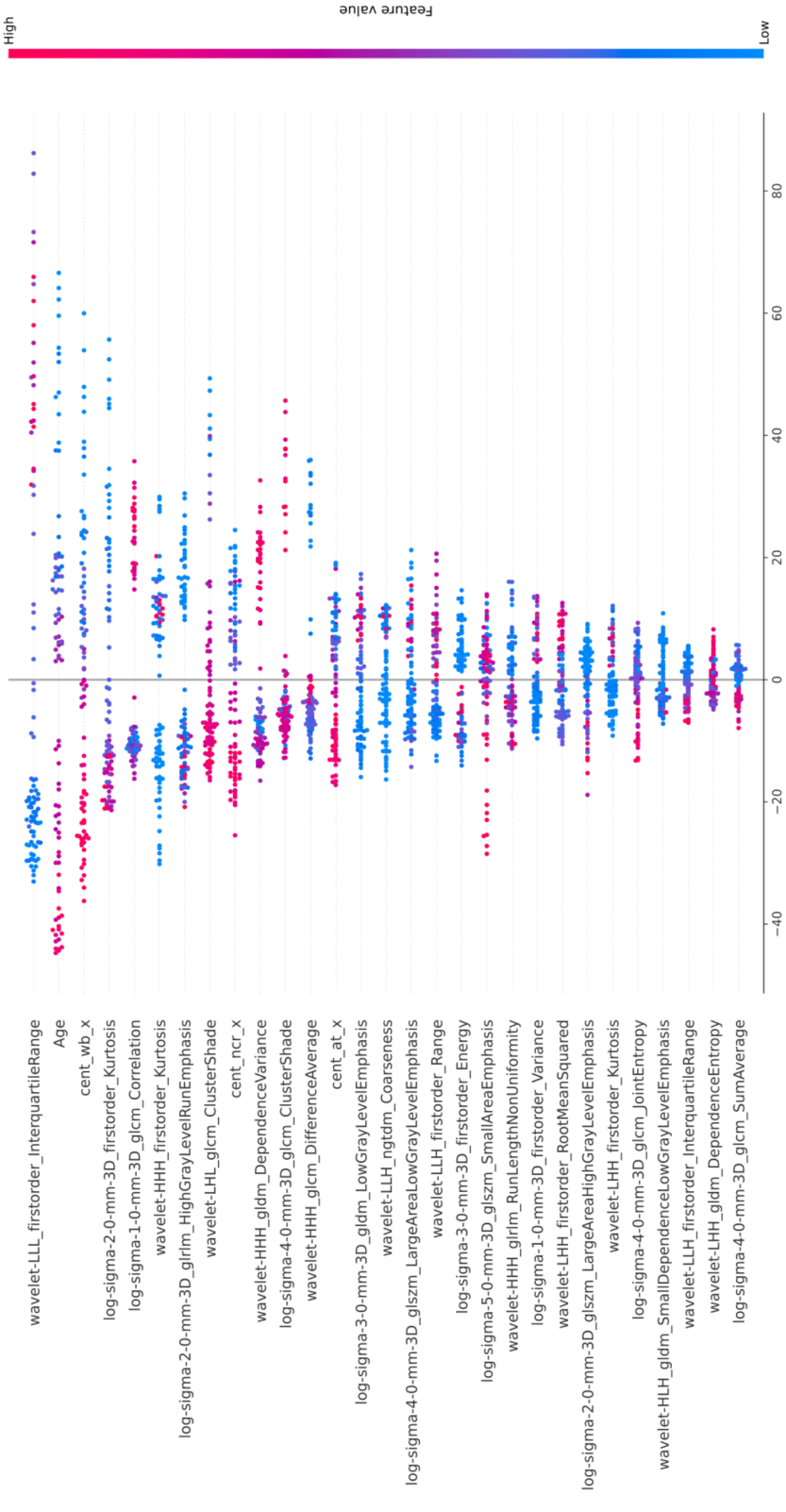


Figure 4.2. The SHAP-summary plot of 29 key features: Each blue point signifies an individual sample. The X-axis displays the Shapley/SHAP value for a particular sample, reflecting the influence of a particular feature on the SD prediction. A higher absolute SHAP value indicates a more significant influence on survival, with the (-) sign indicating a decrease in average survival days and the (+) sign indicating an increase. The Y-axis presents features arranged by importance, with the ordering based on the mean of the absolute SHAP value, from highest to lowest. Additionally, feature values are color-coded to denote high values in red and low values in blue [Rajput et al. (2023a)].

with the color representing the value of the respective instance. Blue indicates a low feature value, while red signifies a high one.

In Figure 4.2, it is evident that the *Wavelet-LLL_firstorder_InterquartileRange (WIR)* feature holds the maximum importance. This feature, derived from a first-order radiomic analysis using the wavelet low-pass filter at each Z, X, and Y direction, characterizes the distribution of pixel intensity values. The *WIR* feature quantifies the pixel intensity within the 25% to 75% percentile range. Examining this plot, we notice that samples with intermediate or high feature values (indicated by purple and red colors) of *WIR* positively contribute to the prediction, exhibiting maximum positive SHAP values. In simpler terms, an intermediate or higher feature value of the *WIR* feature correlates with increased patient survival days. Furthermore, larger samples are concentrated (depicted by blue color) within the SHAP value range of -15 to -25, as observed from the *WIR* feature row on the Y-axis. This suggests that a majority of the samples fall within this SHAP range, and these samples play a role in reducing patients' SD. Additionally, it indicates that tumor intensity (pixel value) information falls within this range, contributing to decreased survival days. This highlights the significant role played by the intensity of tumor pixels in MRI scans. [Aboussaleh et al. (2021); Bae et al. (2018); Tessamma & Ananda Resmi (2010)]. Moreover, exploring spatial heterogeneity can enhance our understanding of the origin and progression of glioblastomas.

The 2^{nd} most significant feature is *Age*. As depicted in Figure 4.2, it is evident that samples with lower Age values exhibit positive SHAP values, indicating increased SD for patients. Conversely, higher Age values are associated with negative SHAP values, suggesting reduced patient SD. This observation corresponds

with medical understanding, wherein the Age of GBM patients plays a critical role in determining survival duration; specifically, a lower Age indicates higher survivability, and an older age of lower survivability [ASCO (2022)].

The 3^{rd} most critical feature is *cent_wb_x*, depicted in Figure 4.2. This feature, a location-based attribute, represents the centroid coordinate of the entire tumor along the X-axis of an MRI image, denoting a physical coordinate. The plot illustrates that intermediate and higher feature values of *cent_wb_x* hurt prediction. Hence, higher feature values are associated with reduced patients' SD. Notably, on the X-axis, which signifies the axial view [Mahmoudzadeh & Kashou (2014)], higher feature values correspond to the physical coordinates of the brain's central region. Thus, the plot indicates that tumors in the brain's central and latter-mid parts will likely decrease patients' survival days [Fyllingen et al. (2021)]. Comparable trends can be noted for the *cent_at_x* and *cent_nec_x* characteristics, denoting the centroids of the active tumor and necrosis areas, respectively.

The 4^{th} most significant feature is *log-sigma-2-0-mm3D_ firstorder_Kurtosis (LFK)*. This feature, a first-order radiomic attribute, is extracted using a LOG filter, indicating the distribution of voxels without considering their spatial relations [Rizzo et al. (2018)]. LFK measures the tailedness or presence of outliers in the data distribution. Observing the plot, it is evident that low kurtosis values correspond to an increase in SD, while higher kurtosis values are associated with a reduction in patients' SD. Most samples fall within the SHAP-value range of -20 to 60.

The 5^{th} most crucial feature is *log-sigma-1-0-mm-3D_glcm_Correlation (Glcm_corr)*. This feature, a second-order radiomic attribute, is extracted using a LOG filter and measures the inter-relationship of intensity between neighboring voxels

[Rizzo et al. (2018)]. According to the plot, higher feature values correspond to an increase in SD, while lower feature values are associated with a reduction in SD. A higher correlation between voxel values increases SD, whereas low correlation values decrease SD.

The 6^{th} most important feature is *wavelet-HHH_firstorder_Kurtosis (FOK)*. This first-order radiomic feature is extracted using a wavelet filter, employing high-pass filters in the series of z, y, and x directions. Similar to the 4^{th} most important feature, the distribution of voxels is independent of their spatial relations. According to the plot, lower feature values are responsible for increasing SD. For further details, please refer to Figure 4.3 for the SHAP distribution plot. In summary, upon comparing all the features, it can be observed that the range of SHAP values for all features falls within -40 to +40 on the X-axis. Additionally, as the importance of features decreases, the range of SHAP values also decreases. This suggests that features with a lower SHAP range have a comparatively lower effect on SD.

Note: The SHAP values of most samples can be further examined by referring to Figure 4.3. This figure illustrates the distributions of SHAP and feature values for the respective features.

Furthermore, the SHAP-waterfall plot illustrated in Figure 4.4 provides a visual interpretation of feature contributions in predicting a specific instance. On the X-axis, the average SD (E[f(x)]) is defined, while on the Y-axis, the features are organized in descending order according to their SHAP values (from top to bottom). From this plot, we can analyze the impact of features, whether negative (blue) or positive (red), and how they contribute to shifting the prediction from the expected outcome $E[f(x)]$. Here, the expected outcome represents the aver-

age of all outcomes considering all the samples. For the specific sample under consideration, the model output is $f(x) = 331.732$, while the expected output is $E[f(x)] = 478.91$. This significant impact can be comprehended by measuring the influences of each feature.

The SHAP value of each feature in Figure 4.4 illustrates its weightage. By summing all the SHAP values from each feature (N) and adding them to the average SD prediction, we can ascertain the individual contributions of these features to the model output. Mathematically, this is represented by $f(x) = E(f(x)) + \sum_N SHAP$, where $\sum_N SHAP$ denotes the summation of SHAP values across all features. This analysis shows that the feature *Age* significantly impacts the model outcome. For instance, in this sample, the Age value of 71.37 decreases the average survival days by 39.33 days (a negative value indicates a reduction in SD). Similarly, the *cent_wb_x* feature with a value of 164.651 also reduces SD by 28.22 days. Mapping these features, *Age* and *cent_wb_x*, to the SHAP-summary Plot (Figure 4.2) or the SHAP-distribution plot (Figure 4.3), which illustrates global impacts, reveals similar observations of SD reduction for these features. For example, visualizing the *Age* and *cent_wb_x* features on the SHAP-summary Plot shows that higher values correspond to a reduction in SD. Similarly, the SHAP distribution plot demonstrates a decrease in SD for these features. This consistency suggests that these features exhibit consistent behavior both globally and locally.

Furthermore, additional insights were gained by combining the SHAP summary (Figure 4.2), SHAP-distribution plot (Figure 4.3), and PDP of the top 6 dominant features (Figure 4.5).

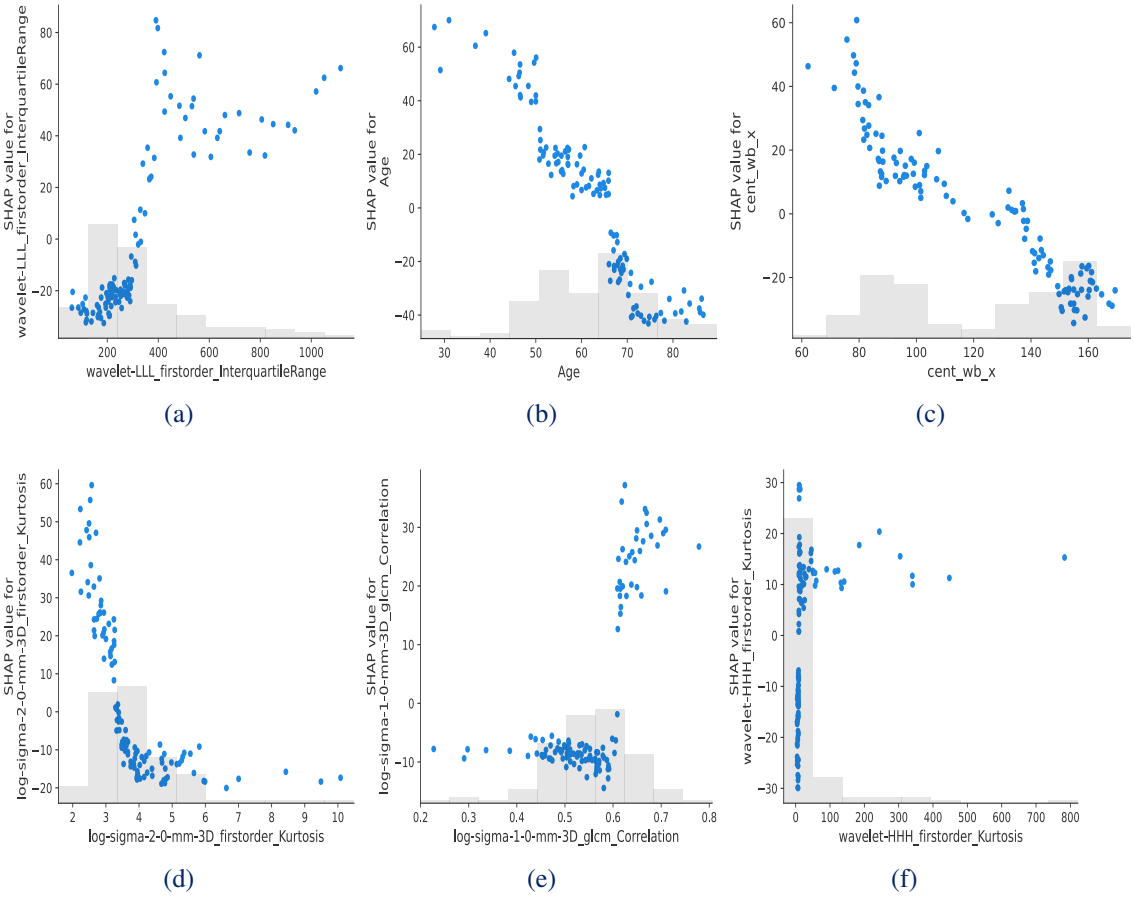


Figure 4.3. Distribution of feature values and its SHAP-values for the topmost six prominent features, namely *(a) WIR feature, (b) Age feature, (c) cent_wb_x feature, (d) LKF, (e) Glcm_Corr feature, (f) FOK feature*. The feature values of corresponding features are shown on the X-axis, and the Y-axis showcases the SHAP value of the corresponding feature. Each point in blue is a sample. The barplot in grey represents histograms of the values of the respective feature. The SHAP-value distribution for the residual features is shown in Supplementary Figure 3 [Rajput et al. (2023a)].

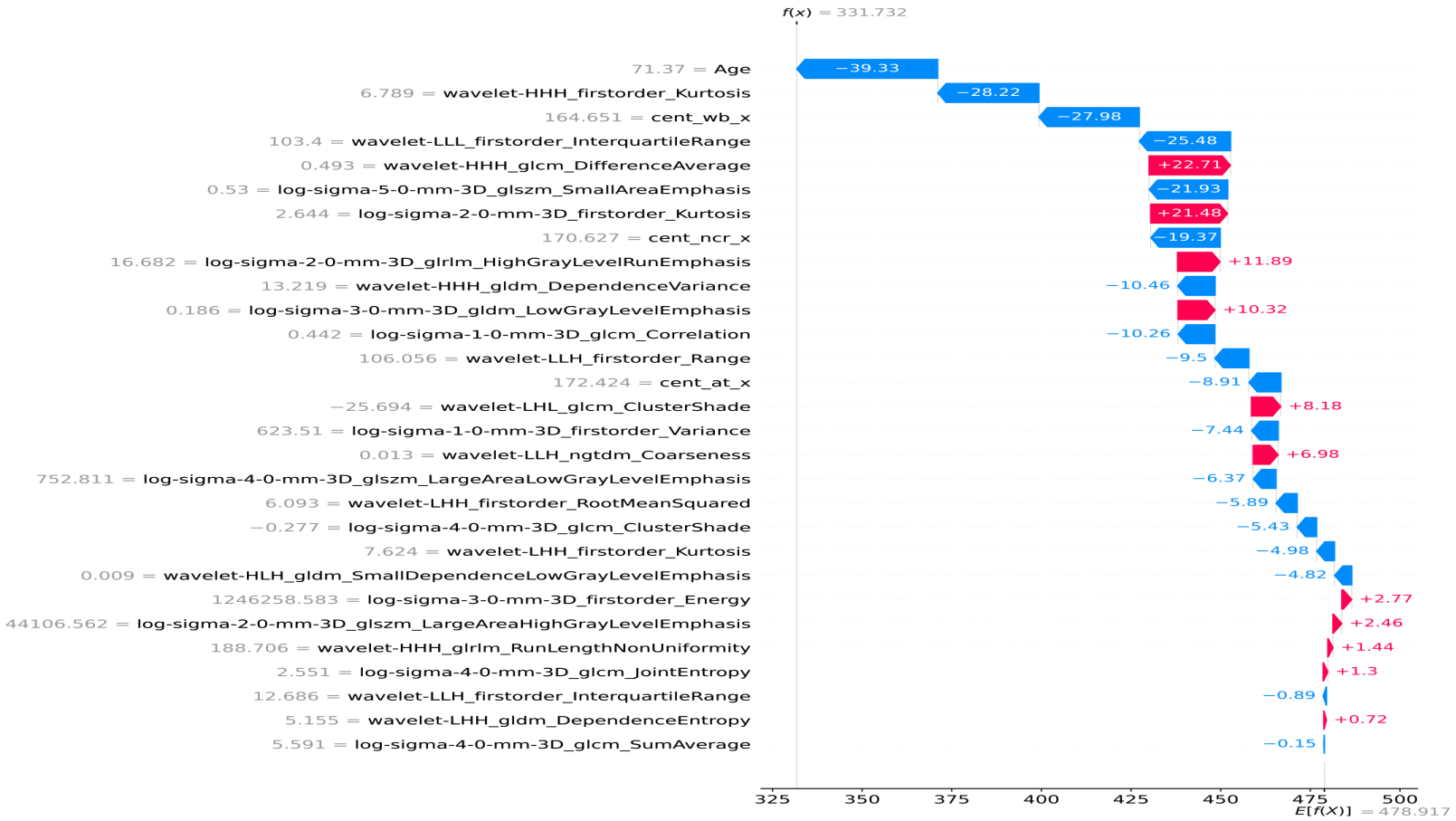


Figure 4.4. The SHAP-waterfall plot depicts the impact of dominant features on a specific sample. Within this visualization, every row demonstrates how each feature contributes to the prediction. Negative contributions are shown in blue, while positive contributions are represented in red. These contributions alter the value from the expected outcomes ($E[f(x)]$) to the model's prediction ($f(x)$). Mathematically, $f(x) = E(f(x))+$ each feature's SHAP value. The value beside each feature row also represents its respective observed value [Rajput et al. (2023a)].

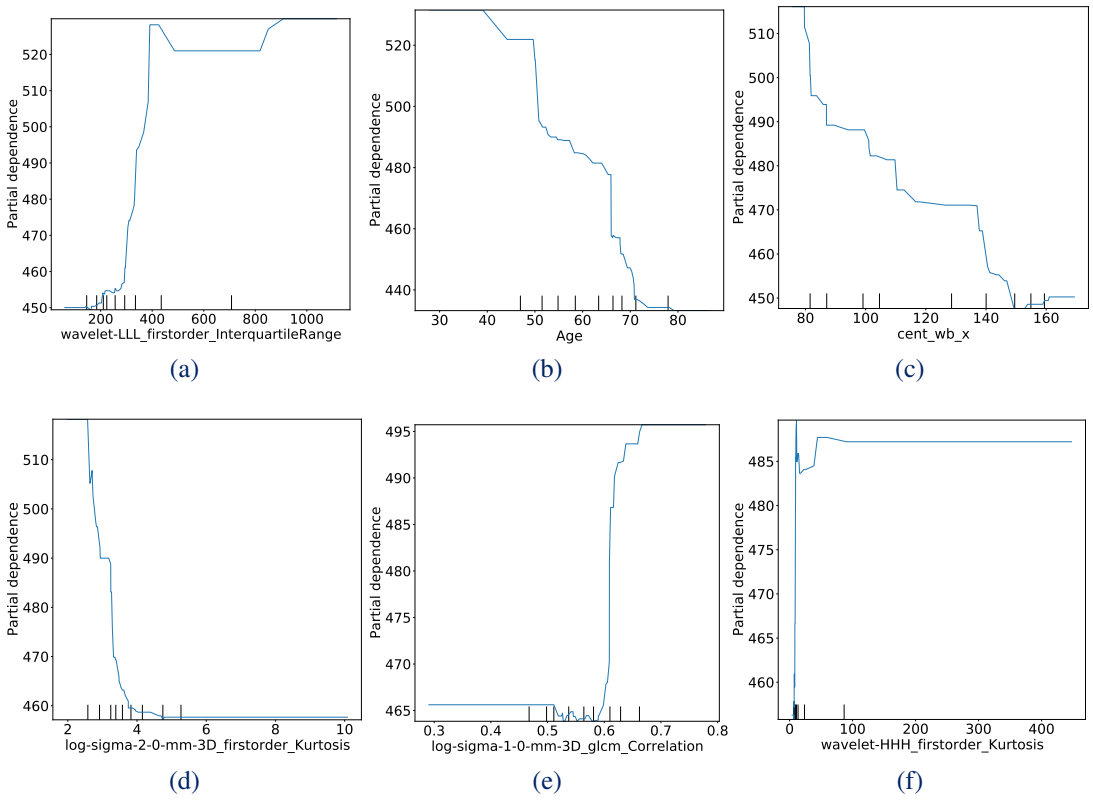


Figure 4.5. The PDP analysis focuses on the six most essential features, illustrating the marginal effects of each feature on SD. Precisely, the plots depict the impact of*(a) WIR feature, (b) Age feature, (c) cent_wb_x feature, (d) LKF feature, (e) Glcm_Corr feature, (f) FOK feature* on SD. The X-axis represents the values of the corresponding feature, while the Y-axis indicates the average rate of change/marginal effect of the feature's value on the SD prediction. Vertical lines on the X-axis illustrate the distribution of most samples. Supplementary Figure 2 contains the PDP for the remaining features [Rajput et al. (2023a)].

The PDP shows a marginal effect between desirable and target features (SD) [Friedman (2001)]. This plot demonstrates how a dependent variable changes as one of the predictor variables varies while all other variables remain constant. Altering the value of a specific feature introduces more variability in the average SD, indicating the feature's significance. In this examination, we focus on the top six features based on their importance, as determined by their absolute SHAP

values. These features include *WIR, Age, cent_wb_x, LFK, Glcm_Corr, and FOK*. The partial dependence plots for these dominant features are displayed in Figure 4.5, while plots for the remaining features can be found in Figure 2 in the Supplementary section. Here, the ordering of the PDP features is derived from the feature importance ranking obtained from the SHAP summary plot. (Figure 4.2).

Moreover, upon visualizing the PDPs, we observe that the marginal impacts align with the order of importance of features identified through the SHAP-value analysis. Here's a comprehensive examination of the top six features: The marginal effect of the *WIR* feature on SD prediction is depicted in Figure 4.5. The trend indicates a sharp increase in value within the range of 100 to 300, followed by a reduction within the range of 300 to 350, and a saturation within the range of 350 to 800 intensity value. This indicates a notable high heterogeneity in intensity within the range of 100 to 300, resulting in a significant increase in marginal impact. Comparing the PDP (Figure 4.5(a)) with SHAP (Figure 4.2) and its distribution plots (Figure 4.3 (a)), it can be inferred that the intensity ranging between 100-300 contributes to a decrease in SD (as demonstrated by the reduction in the SHAP value). Therefore, it can be inferred that tumor pixel intensity within this range negatively impacts a patient's survival. Moreover, as highlighted in this study, increased tumor heterogeneity correlates with increased malignancy [Gupta et al. (2019)]. This finding is consistent with prior studies indicating wavelet filters effectively capture enhanced texture features.

Likewise, in the PDP plot illustrating *Age* feature (Figure 4.5(b)), we can identify the trend of marginal effect, indicating the largest deviation in marginal effects for patients with lower ages, signifying maximum influence on SD. Furthermore, the marginal effect diminishes with the increasing age of patients. Combining

inferences from the SHAP summary and SHAP-distribution plots, it's clear that beyond the age of 60, there is a decline in SD (as indicated by a decrease in SHAP-value beyond this range.

Further, the PD plot for the *cent_wb_x* feature (depicted in Figure 4.5 (c)) highlights that the marginal effect is more noticeable when the centroid falls within the range of values approximately between 75 and 112, and less significant for values ranging from 113 to 160. By comparing these ranges to the SHAP distribution plot, it becomes evident that the former range of values increases the SD while the latter range reduces them. This suggests that tumor lesions located in the central or posterior part can be fatal to patients.

Meanwhile, the *LFK* feature characterizes radiomic first-order statistical attributes, measuring the peakedness of data distribution. A kurtosis value (k) is 3 for a Gaussian distribution. When $k > 3$, the dataset tends to have significant outliers, while $k < 3$ suggests fewer or no outliers. The PD plot (depicted in Figure 4.5(d)) illustrates that for $k = 3$, there is a notable marginal impact on SD. By comparing the PDP with the SHAP and SHAP-distribution plots, it can be inferred that for most samples, $k = 3$, leads to an increase in SD. However, there are also sufficient samples with $k > 3$, which decreases SD. This suggests a considerable number of outliers or intra-heterogeneity among the samples. As mentioned by [Steven et al. (2014)], diffusion kurtosis imaging works on a similar principle of capturing non-normal distribution behavior, which signifies tissue heterogeneity. It is observed that the survival days are positively skewed [Der & Everitt (2005)].

The *Glcm_Corr* is a radiomic feature that evaluates the joint probability of occurrence of specified pixel pairs with designated intensity values. Utilizing the gray-level co-occurrence matrix (GLCM), it examines spatial relationships be-

tween pixels at particular distances and directions. From the PDP (depicted in Figure 4.5(e)), it is clear that when the correlation between pixel pairs equals or exceeds 0.6, it significantly impacts SD. Similarly, when examining the correlation threshold of 0.6 in the SHAP and SHAP-distribution plots, one can observe its positive influence, as evidenced by a positive SHAP value. This finding is consistent with the study by Sanghani et al. [Sanghani et al. (2018)], highlighting the crucial role of texture features in SD prediction.

Furthermore, the *FOK* is another first-order statistical feature, similar to the *LFK* (the 4th most significant feature); but it is computed employing wavelet high pass filters. A comparison of their PDPs (depicted in Figure 4.5(d) and (f)) shows that both capture kurtosis information, however, in distinct dimensions. From the PDP (shown in Figure 4.5(f)), it is apparent that the kurtosis value sharply rises between 0 and 100 and then remains steady for subsequent values. Upon observing the SHAP distribution, it can be inferred that most samples fall within this range, with those near values $1-10$ decreasing SD while the remainder increasing SD. However, some samples are sparsely distributed, indicating their outlier status.

These findings emphasize the significance of these features in predicting SD. Through the analysis above, we recognize that *Age, WIR, cent_wb_x, LFK, Glcm_Corr, and FOK* play pivotal roles in determining a patient's SD. Similar scrutiny can be applied to the remaining features. Notably, the *WIR* feature offers insights into tumor heterogeneity related to an increase in tumor malignancy. Additionally, the Age feature exhibits a clear trend in survivability, indicating diminishing survival prospects with increasing age (this is additionally supported by the Kaplan-Meier (KM) plot, as depicted in Supplementary Figure 4). Meanwhile,

the centroid of tumors facilitates the identification of tumors in the central or posterior regions, which are detrimental to patients.

All these signify the importance of these features in determining the SD. With the above analysis, the *Age, WIR, cent_wb_x, LFK, Glcm_Corr, and FOK* plays a crucial role in determining a patient's SD. Similarly, we can analyze other remaining features. Finally, we agree that the *WIR* feature can tell us about tumor heterogeneity associated with high malignancy. Again, the Age feature shows us the trend of survivability, where the survival chances decrease with the patient's increasing age (this is further validated by the KM [Goel et al. (2010)] plot as shown in the Supplementary Figure 4). Meanwhile, the centroid of tumors enables us to locate tumors in the central or latter-central part, which is detrimental for patients. The alignment of these observations with medical findings and relevant research indicates the features' reliability and validates the effectiveness of explainability methods like SHAP and PDP.

4.4.2 Global and Local Insights: Exploring SHAP, LIME, PDP and ALE Explanations

In the second phase of the study, RFR and ERFR predictor models are applied to predict SD across all feature variants. However, for interpreting feature behavior, our focus is on the best variant of the feature set *(Variant 4)*. We'll first discuss the visual interpretation of feature behavior from the ERFR models, followed by the interpretation from the RFR models.

Figure 4.6 presents the SHAP-summary plots for the *Variant 4* (best-performing) feature set. It is obtained from an ERFR trained model. In the broader context

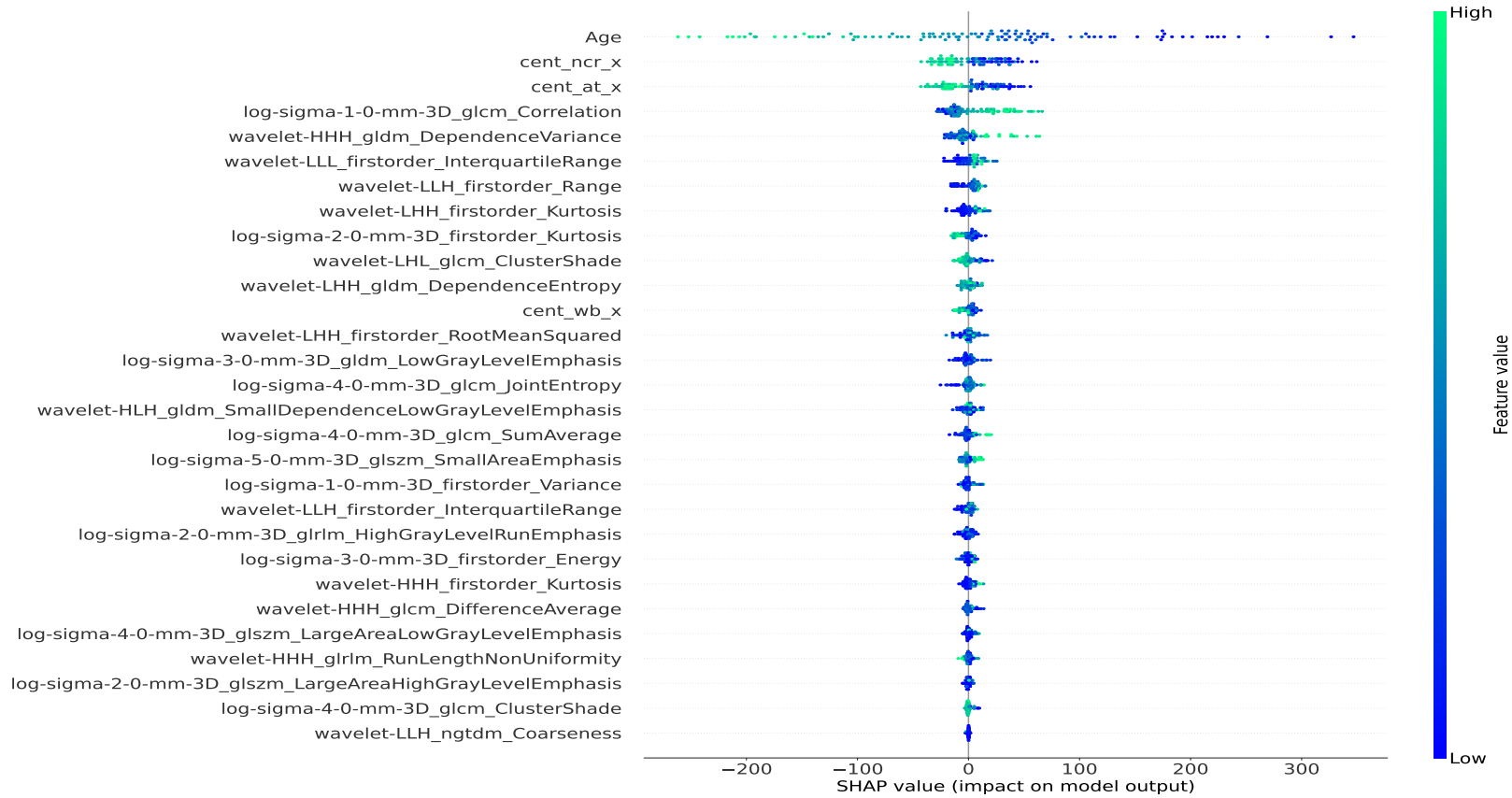


Figure 4.6. The SHAP-summary plots from ERFR model, illustrate the *Variant 4* feature set. Each blue dot on the plot represents an individual patient. The X-axis depicts each sample's Shapley or SHAP value, indicating how a particular feature impacts that patient's survival days. A greater absolute SHAP value suggests a more significant influence on survival days, with the sign indicating whether it contributes to an increase or decrease in average survival days. On the vertical Y-axis, features are arranged by importance and ranked from high to low. This ranking is determined by calculating each feature's average absolute SHAP values. Additionally, the color coding on the right side of the plot distinguishes between high feature values, represented in green, and low feature values, displayed in blue [Rajput, Kapdi, Raval, et al. (2024)]

.

of this study, we consistently observe the persistence of the top dominant features identified in the initial phase. These include *Age, location-based features, Glcm_Corr, WIR, and LFK*. The observation regarding the influence of features or feature values on the increase or decrease in SD prediction shown in Figure 4.6 is comparable across different features, as demonstrated in Figure 4.2.

Furthermore, the SHAP summary plot for all variants is available in the Supplementary section (Figure 5). In this broader context, it becomes evident that the ranking of most features remains consistent across all variants.

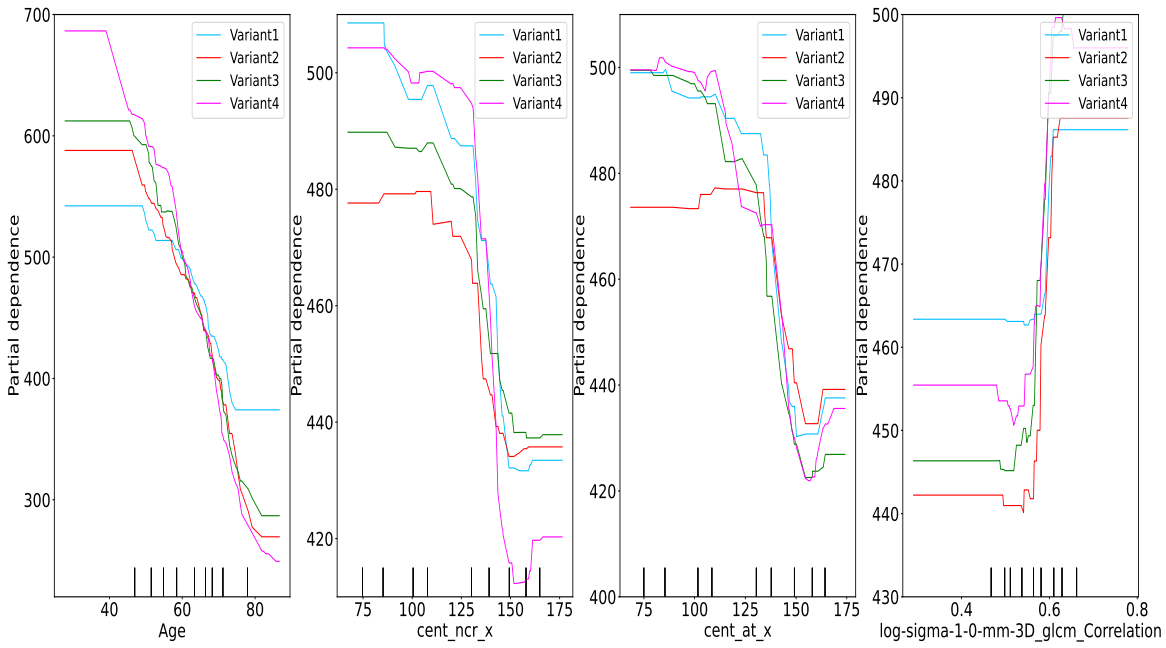


Figure 4.7. PDP plots (from ERFR trained model) of the first four features of all the variants shown in the SHAP summary plot. The X-axis illustrates the feature name including *Age, cent_ncr_x, cent_at_x, Glcm_Corr* and its values. The Y-axis depicts the average changes in SD, corresponding to the values on the X-axis.

Additionally, for more insights, PDP plots were generated for the first four features (*Age, cent_ncr_x, cent_at_x, Glcm_Corr*) depicted in the SHAP summary plot

in Figure 4.6. Here, *Age* refers to the age of patients provided with the dataset, *cent_ncr_x* and *cent_at_x* denote the centroids of necrosis and active tumor lesions, respectively, while *Glcm_corr)* represents the correlation in the context of the GLCM, where it signifies the probability of encountering two specific gray levels (i and j) together in an image at a specific spatial relationship defined by a distance (δ) and an angle (θ). This joint probability distribution offers insights into the relationships between pairs of pixel values in an image, which is valuable for tasks such as texture analysis, pattern recognition, and feature extraction in image analysis and processing. In the context of GLCM, the correlation measures the linear relationship between gray levels in an image at a specific spatial relationship defined by GLCM. Please refer to [Van Griethuysen et al. (2017)] for more detailed information.

The PDP plot showcases the behavior of all variants concerning the four features mentioned in the preceding paragraph. It illustrates how the predicted outcome varies as the selected feature(s) values change while other features remain constant. The *Age* feature exhibits the most significant changes in predicted outcomes, decreasing in importance in descending order. Increasing age has a diminishing effect on SD prediction. A comparable trend is observed for the centroid of necrosis and active tumor. However, in the case of *GLCM_Corr*, the impact is more evident when the correlation exceeds 0.6. Aligning this behavior pattern with the SHAP summary plot indicates that specific feature values contribute to increasing or decreasing SD prediction.

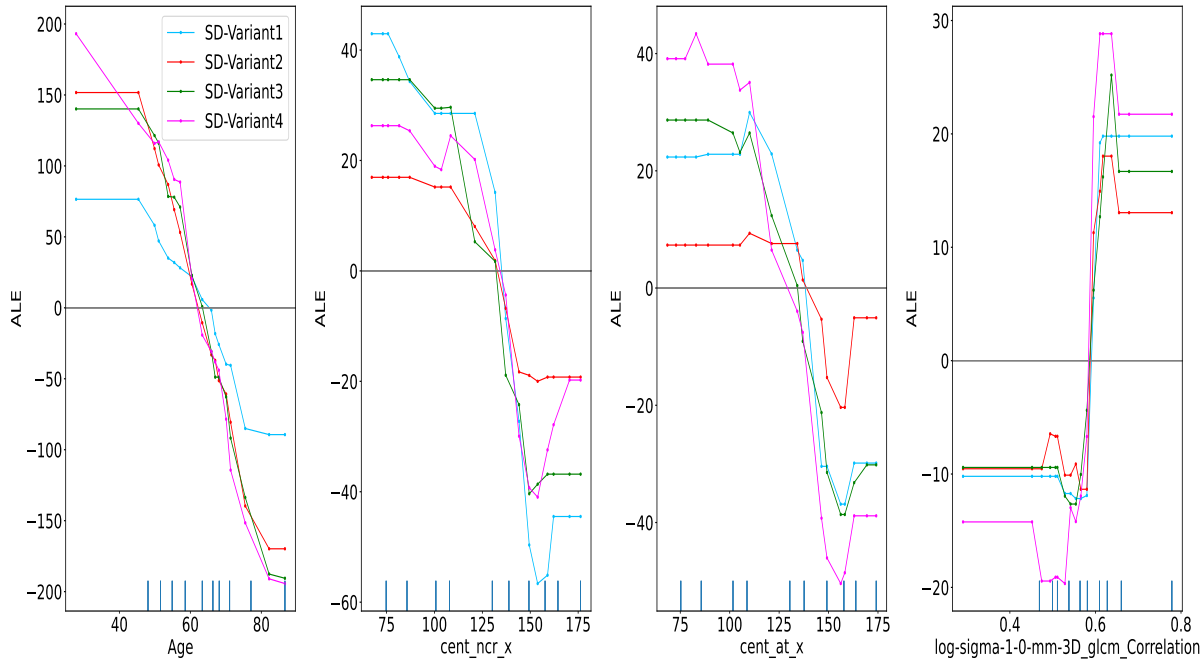
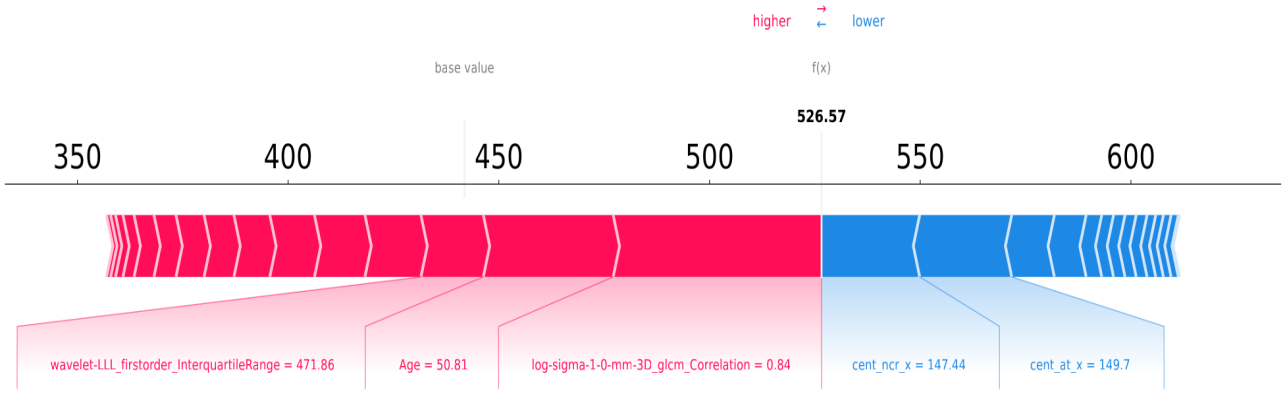


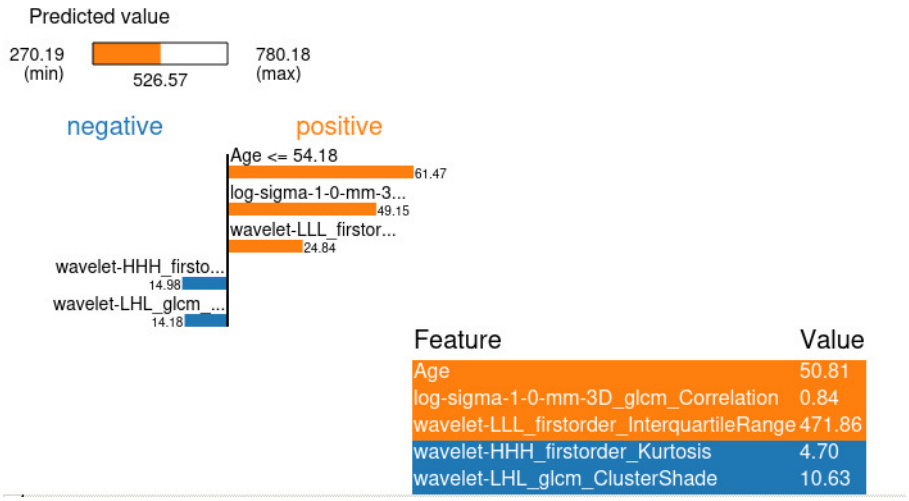
Figure 4.8. ALE plots (from ERFR trained model) of first four features of all the variants shown in SHAP summary plot. The X-axis presents the feature name including *Age, cent_ncr_x, cent_at_x, Glcm_Corr* and its values. The Y-axis depicts the average standard deviation changes corresponding to the values on the X-axis [Rajput, Kapdi, Raval, et al. (2024)].

Additionally, we generate ALE plots presented in Figure 4.8, using the alibi explain tool [Klaise et al. (2021)] for the initial four features of all the variants, as presented in the SHAP summary plot. This plot demonstrates how the model's predictions change as the feature of interest varies. On the X-axis, the feature names and their respective values are depicted. At the same time, the Y-axis illustrates the variations in average SD concerning the values displayed on the X-axis. The overall behavioral pattern remains consistent compared to the PDP plot shown in Figure 4.7. It reveals how the feature values impact the SD prediction negatively or positively. The insights gained from this plot align with those from

the PDP and SHAP summary plots.



(a)



(b)

Figure 4.10. (a) SHAP force plot for a sample illustrating features that contribute beyond an 8% threshold. Where *f(x)* represents the predicted value for this sample from the validation set, a subset of the training set. The base value corresponds to the expected value of the training set. Features with their values highlighted in red indicate an increasing SD, while those in blue signify a decreasing SD. The size of the arrow indicates their contribution to prediction. (b) LIME plot for a sample showcasing the impact of the top five features. The predicted values with minimum and maximum are shown here. The features with their values in orange show their contribution in orange, indicating an increasing SD, while those in blue signify a decreasing SD. The features are arranged in decreasing order of importance.

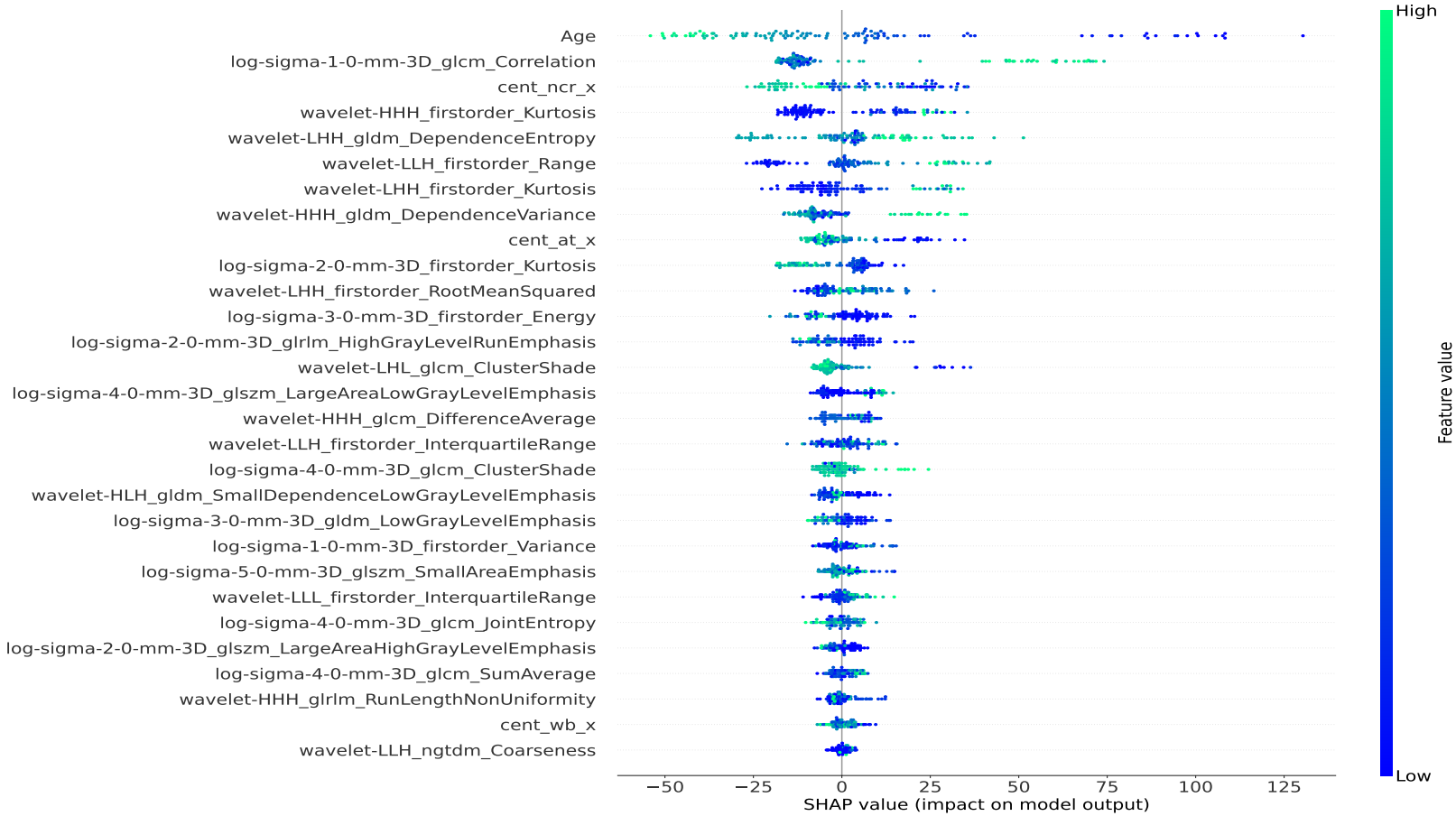


Figure 4.9. The SHAP-summary plot (from the RFR model) for the feature set displays individual patients as blue dots. The X-axis shows each instance's SHAP value, which indicates a particular feature's influence on that patient's survival days. A larger absolute SHAP value signifies a more significant influence on SD prediction, with the sign indicating whether it contributes to an increase or decrease in the average survival days. The Y-axis organizes the features in descending order of significance. Furthermore, the color scheme on the right side of the plot is used to distinguish between high feature values depicted in green and low feature values represented in blue.

Further, for visual interpretation from the RFR model, the SHAP summary plot is presented in Figure 4.9, reaffirming the consistent ranking of top features.

Furthermore, to delve into the behavior at the sample level, we utilized the SHAP force plot and LIME tools to visualize the feature behavior of a sample, as depicted in Figure 4.10 (a) and Figure 4.10 (b), respectively. The Python SHAP (version 0.42.1) and LIME tool (version 0.2.0.1) were employed for this purpose.

In Figure 4.10 (a), we showcased features whose contribution exceeds 8% (meaning, only features with a Shapley value magnitude greater than 8% of the sum of all absolute Shapley values will be displayed). The size and color of the arrow denote the magnitude and direction (increasing/decreasing) of contribution to SD prediction. Red color indicates an increase in SD, whereas blue color signifies a decrease in SD.

In contrast, Figure 4.10 (b) presents the top five contributing features. Here, features and their respective values highlighted in orange indicate an increase in SD, while those in blue signify a decrease in SD. When comparing these two figures, it's apparent that the top three performing features remain consistent and exhibit similar effects on SD prediction. Furthermore, we include a SHAP waterfall plot in Supplementary Figure 6 for the identical sample, indicating a comparable behavior. We employ the Python SHAP tool (version: 0.42.1) for deriving summary, force, and waterfall plots.

For global interpretation, Sklearn (version=1.1.1) [Pedregosa et al. (2011)] PartialDependenceDisplay tool is employed to generate a PD plot, and alibi (version=0.9.4) [Klaise et al. (2021)] ALE tool is used to analyze the behavior of the initial four features from the SHAP summary plot (shown in Figure 4.9). The PDP plot can be observed in Figure 4.11 (a), and the ALE plot in Figure 4.11 (b). The

X-axis in the plot shows features and their respective values, whereas the Y-axis in the PDP plot represents the average predicted SD as the chosen feature varies. In contrast, other features are held constant or averaged. In the ALE plot, the y-axis signifies the accumulated local effect of the desirable feature on the SD integrated over the specified feature range.

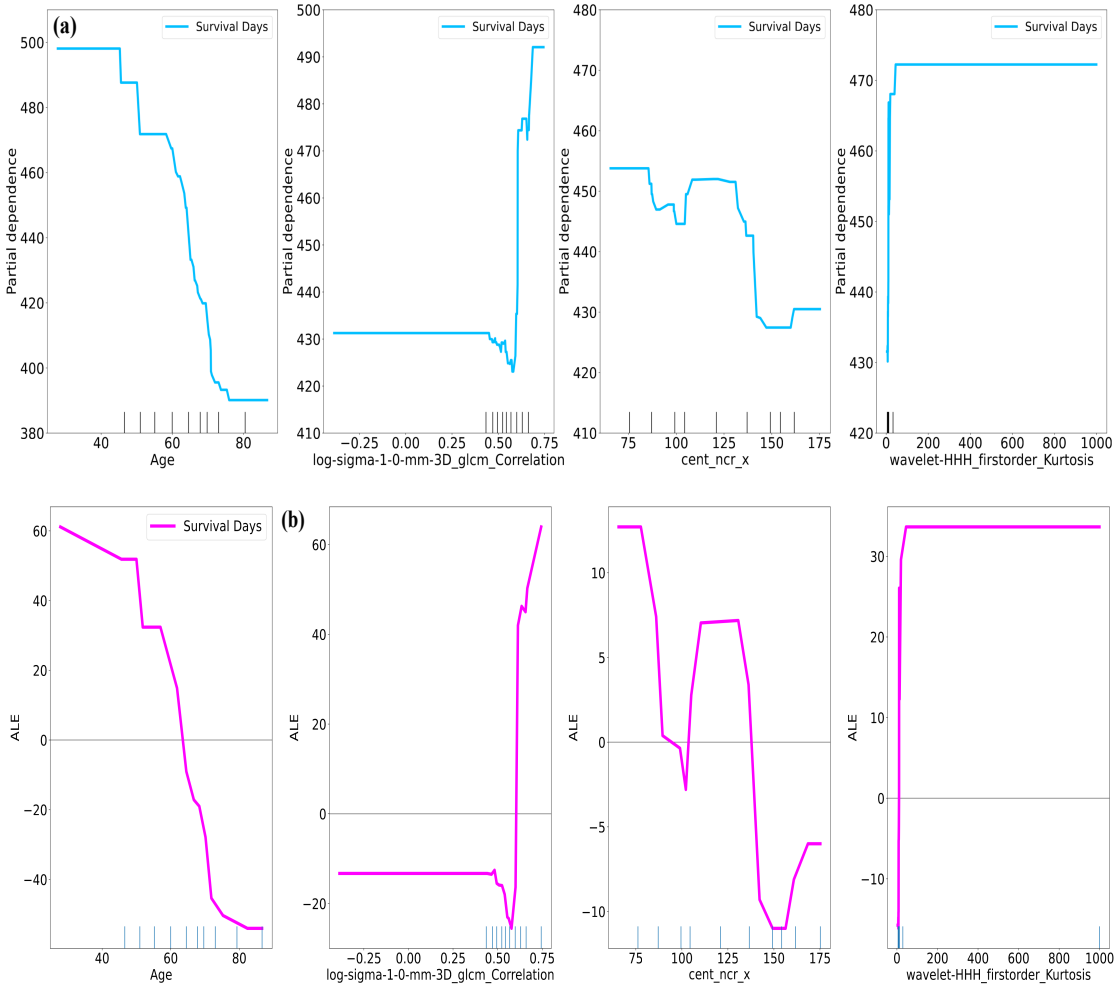


Figure 4.11. Visualizing the first four features using (a) PDP and (b) ALE plots allows for the analysis of the global behavior of these features derived from the SHAP summary plot. Feature values are represented on the X-axis, while the vertical bar on the X-axis indicates data distribution. In PDP plots, the Y-axis denotes the average predicted SD as the selected feature changes. Conversely, in the ALE plot, the Y-axis illustrates changes in the average SD across different intervals of the feature displayed on the X-axis.

Examining the average changes in the SD plotted on the Y-axis for PDP (Figure 4.11 (a)) and ALE (Figure 4.11 (b)) plots for the X-axis reveals a consistent overall behavioral pattern for the desired features. For instance, when analyzing the *Age* feature in a PDP plot, notable changes in the average effect, indicated on the Y-axis (SD:500), are observed within the 40-50 age range (depicted on the X-axis). A larger magnitude of changes in the average effect signifies the importance of the feature in predicting SD. This observation is similarly reflected in the ALE plot for the *Age* feature. The vertical bar on the X-axis reveals the distribution of feature values. For the *Age* feature, the distribution spans evenly from 40 to 80. However, for *Glcm_Corr*, most features exhibit correlation values ranging from 5 to 0.75, with one sample displaying a negative correlation. Similarly, for *cent_ncr_x*, most centroid coordinates of necrosis regions along the X-axis fall between 7-170. Meanwhile, for *FOK*, most values ranging from 3 to 50 indicate high (positive) kurtosis, suggesting a concentration of the distribution towards the tails rather than the mean. In diffusion imaging, kurtosis serves as a metric within diffusion kurtosis imaging (DKI) for evaluating tissue microstructure, providing insights into tissue barrier complexity and cellularity [Steven et al. (2014)]. Deviations from typical kurtosis values may indicate alterations in tissue integrity and potential pathological changes. Studies have found a correlation between positive kurtosis and increased tissue heterogeneity in conditions like Ischemia and Infarction [Steven et al. (2014)]. Cauter et al. [Van Cauter et al. (2012)] noted in their study that high-grade tumors exhibited higher kurtosis values, likely due to greater cellular density, reduced cell size, and increased complexity (heterogeneity) in the tumor microenvironment.

4.4.3 Radiogenomic analysis

The outcomes from radiogenomics analysis are discussed hereafter. Table 4.1 illustrates performance comparisons between regression and classification models for SD prediction. Upon comparing Table 4.1 (a) and (b), we observe that the models, particularly those in Table 4.1(b), perform better when considering all samples (n = 239) from various resection statuses. Table 4.1(c) presents the performance of models with and without clinical information, including gender, race, and treatment type. Models trained with clinical information outperform those without it, highlighting the importance of clinical information as a crucial prognostic factor in SD predictions. Also, various studies have emphasized the correlation between survival prediction and various clinical factors such as age, therapy information, gender, and performance status [Rajput et al. (2023a); Pálsson et al. (2022); Michaelsen et al. (2013); Liang et al. (2020); Li et al. (2021)]. Hence, gender-specific and treatment-specific survival analysis is conducted, as depicted in Figure 4.12 and Supplementary Figure 7.

Table 4.1(d) presents the results obtained by classifying survival days into two categories. Despite the smaller samples, the classifier models have demonstrated superior performance to the regressor models. The best accuracy is 0.727 for the RF classifier. In contrast, the Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) score is highest for LR, followed by XGB with 0.786 and RF with 0.750. The ROC-AUC curve for the LR, RF, and XGB can be seen in Figure 4.13. The ROC-AUC score assesses the classifier's ability to differentiate between positive and negative classes, ranging from 0 to 1. A higher ROC-AUC signifies superior performance. An ideal model achieves a ROC-AUC of 1.0,

while a random model scores 0.5. This suggests that classification may be more effective in scenarios with limited data for survival prediction.

TABLE 4.1
The performance evaluation of SD predictor models using ensemble learning incorporates Random forest (RF), Extra random forest (ExtraRF), Xtreme gradient boosting (XGB), and Logistic regressor (LR). These models are applied to regression and classification tasks, respectively. MSE represents a mean squared error, while ROC-AUC indicates the Receiver Operating Characteristic - Area Under the Curve.

(a) Trained through TN (with GTR status), tested on TCGA dataset.

Regressor Model	Train MSE	Test MSE
RF	098144.45	906973.28
Extra-RF	110966.41	885217.53
XGB	108162.88	886718.95

(b) Trained with all resection samples, tested on TCGA dataset.

Regressor Model	Train MSE	Test MSE
RF	099524.92	**877985.34**
Extra-RF	104315.72	**874357.20**
XGB	101858.18	**884607.68**

(c) TCGA samples with* and without** clinical data were utilized for training and testing, with a split ratio of 0.25

Regressor Model	Train* MSE	Test* MSE	Train** MSE	Test** MSE
RF	425635.98	**141919.67**	417538.62	176197.62
Extra-RF	517174.43	**217368.40**	793569.79	221054.26
XGB	870795.50	**156186.83**	64116.535	198083.56

(d) Binary classification was conducted, dividing samples into two groups: those with SD of less than 450 days and those with a duration of 450 days or more. The training and testing datasets were divided using a ratio of 0.30

Classifier Model	Train Accuracy	Test Accuracy	ROC-AUC score
LR	0.565	0.636	**0.857**
RF	0.913	**0.727**	**0.750**
Extra-RF	0.696	0.636	0.607
XGB	0.565	0.636	**0.786**

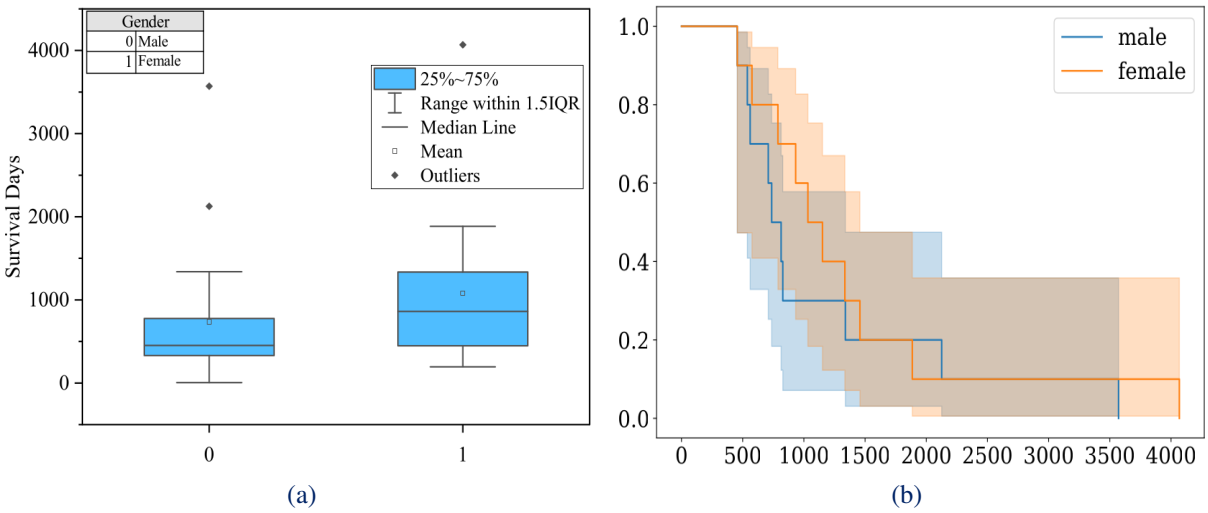


Figure 4.12. Gender-wise boxplots of SD are depicted in (a), while (b) illustrates that survival probability, indicated on the Y-axis, is higher in women compared to men. The X-axis represents SD.

Additionally, framing the SD prediction challenge as a binary classification task results in better performance than regression models, especially when dealing with smaller datasets. The reasons are as follows:

- Simplicity: By dividing survival days into "long-term" and "short-term" categories, it makes it easy for the model to learn patterns and make predictions.
- Interpretability: Categorizing SD into two groups facilitates a more straightforward interpretation of results, distinguishing between patients likely to have long-term or short-term survival and aiding clinical decision-making.
- Imbalanced data handling: In survival prediction tasks, data may be imbalanced and scattered across the population.

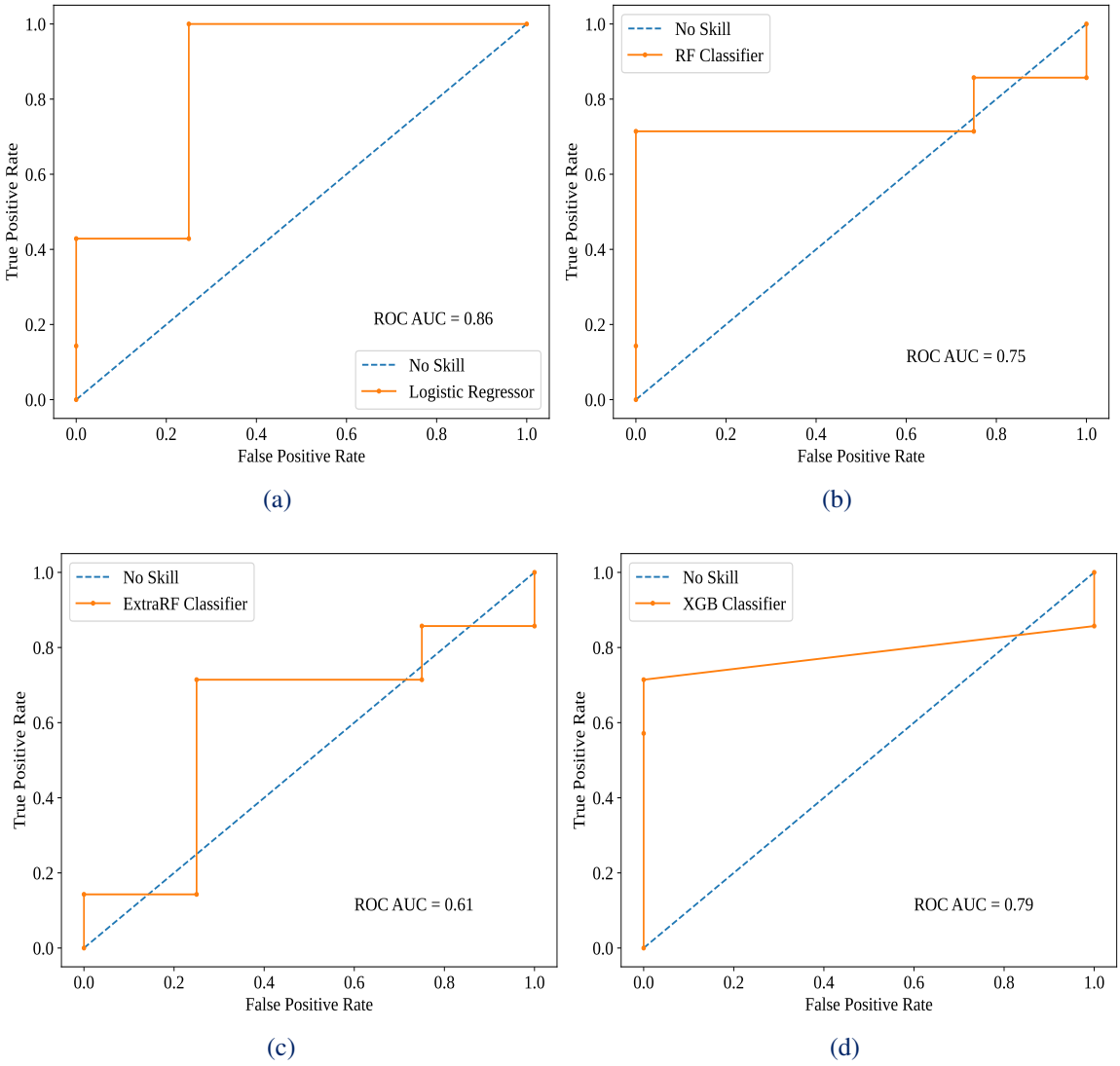


Figure 4.13. Comparison of ROC-AUC curves, evaluates the sensitivity and (1 - specificity) using multiple classifiers, including (a) LR, (b) RF, (c) ExtraRF, and (d) XGB.

In Figure 4.12(a), the genderwise boxplot reveals that females (Class 1) tend to have longer survival days compared to males (Class 0). Specifically, the median survival days for males is 500, whereas for females, it is 1000. Figure 4.12(b) illustrates the survival probability for both genders. The Y-axis depicts survival

probability, while the X-axis depicts SD. Notably, females exhibit a higher survival probability than males. The supplementary Figure 7 also presents survival analysis using the KM statistical method Goel et al. (2010), highlighting the impact of treatment information, encompassing radiation therapy and pharmaceutical therapy.

4.4.4 Interactions in radiomics - genomics

In the subsequent phase of this study, we explore the interaction between the extracted feature sets and genetic information. We reduce the available gene information as outlined in Section 4.3. Further, Figure 4.14 illustrates the correlation-histogram between the initial three extracted feature categories and the reduced gene feature set. The rest of the features' correlation-histograms refer to Supplementary Figure 8.

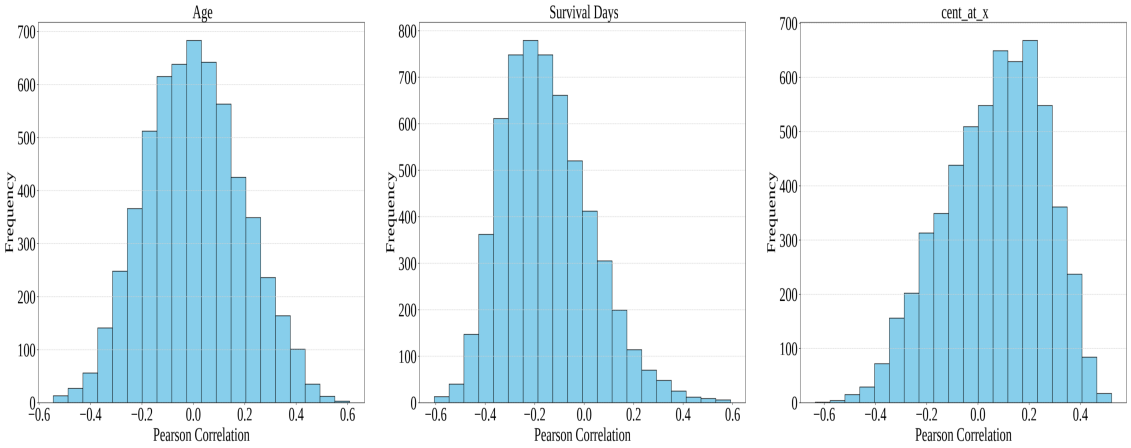


Figure 4.14. The correlation-histogram illustrates the distribution of genes across three feature categories, with their corresponding PC values. Correlation-histogram of the remainder features is illustrated in Supplementary Figure 8.

This correlation-histogram plot illustrates the distribution of genes across var-

ious feature categories and their corresponding PC values. By examining these plots, we gain insight into the skewness and kurtosis within the gene dataset, aiding in identifying asymmetry and tails in the distribution of genes, respectively. For example, the gene distribution in the *Age* plot appears to be normally distributed, while *survival days* exhibits positive skew, and *cent_at_x* displays negative skew. Additionally, the gene distribution in all plots with normal kurtosis indicates minimal outliers in the data distribution.

While PC measures the linear associations among features, 1 signifies a complete positive linear relationship, -1 indicates a complete negative linear relationship, and 0 denotes no linear relationship between the features. Highly correlated features provide redundant information, while less correlated ones provide unique information. We can observe from Figure 4.14 that the feature sets do not fall into extreme values, suggesting that gene information is largely uncorrelated with the feature set. This signifies that it carries unique information useful for SD prediction.

Further, to deepen our understanding of the correlation between the feature set and gene information, we dissect the gene feature set into two categories. In the first category, we examine gene features highly correlated with the feature set to explore their similarities with the PC threshold value $>= 0.7$. Conversely, we focus on gene features exhibiting low correlation with the feature set in the second category. This analysis aims to uncover how these sets could be integrated to provide complementary information, potentially enhancing their utility for SD prediction. Here, we use the PC threshold value $<= 0.006$. The distribution plots of both categories are shown in Supplementary Figure 9 and Figure 10. These

plots aid in identifying the frequency of genes and their corresponding PC values within each feature category. They provide insight into the maximum and minimum genes following each category and their respective PC values.

Following the thresholding process in the first and second categories, we identified 2940 and 188 genes correlating with the extracted feature set. These results are illustrated in Figure 4.15 and Figure 4.16, respectively. In Figure 4.15, extracted feature categories are arranged in descending order, where the maximum gene falls under *WIR*, which measures the variability of intensity values within an image, specifically focusing on the middle 50% of the data. Some of the highlighting genes from Figure 4.15 are shown in Table 4.2, which are related to the poor prognosis of glioma patients. It suggests that specific genes contain complementary information to the features. This may potentially aid in the prognosis of SD and improve the prediction of patient outcomes. However, in this study, we focused on analyzing similar behavior between extracted features and genes. Examining the similarities between radiomic features and genes can offer valuable insights into the fundamental molecular mechanisms of diseases. Researchers deepen their understanding of disease pathogenesis and progression by uncovering the biological pathways and molecular signatures associated with specific radiomic characteristics.

TABLE 4.2
The association between genes and extracted feature set, with genes identified as being linked to poor prognosis of SD in glioma patients. EGFR and EGFR-AS1 are known to have an association with various cancers. However, we found that EGFR-AS1 has a poor prognosis of SD with the used dataset. Details regarding the features are available in the Supplementary Table 2.

Gene	Feature set
PTEN [Hashemi et al. (2023); Han et al. (2016)]	log-sigma-4-0-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis
EGFR-AS1 [Zhu et al. (2023); Dong et al. (2019)]	wavelet-LLH_firstorder_InterquartileRange
EGFR-AS1 [Zhu et al. (2023); Dong et al. (2019)]	wavelet-LHL_glcm_ClusterShade
EGFR-AS1 [Zhu et al. (2023); Dong et al. (2019)]	log-sigma-1-0-mm-3D_firstorder_Variance
EGFR-AS1 [Zhu et al. (2023); Dong et al. (2019)]	log-sigma-4-0-mm-3D_glcm_ClusterShade
EGFR-AS1 [Zhu et al. (2023); Dong et al. (2019)]	log-sigma-2-0-mm-3D_glrlm_HighGrayLevelRunEmphasis
C8orf44 [Bao et al. (2013)]	cent_ncr_x
C8orf44 [Bao et al. (2013)]	wavelet-HHH_glcm_DifferenceAverage
NF1 [Lobbous et al. (2020); D'Angelo et al. (2019)]	log-sigma-4-0-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis
PLK2 [Cao et al. (2021)]	wavelet-HHH_glcm_DifferenceAverage
ATXN2 [Hoelzinger et al. (2005)]	log-sigma-1-0-mm-3D_glcm_Correlation
ATXN2L [Hoelzinger et al. (2005)]	log-sigma-3-0-mm-3D_firstorder_Energy
ATNX10 [Hoelzinger et al. (2005)]	LLH_ngtdm_Coarseness
ZNF554 [Balogh et al. (2020)]	wavelet-LLH_firstorder_Range
RAPGEF2 [S. Long & Li (2019)]	Age, wavelet-HHH_gldm_DependenceVariance
LINC00205 [S. Long & Li (2019)]	log-sigma-4-0-mm-3D_glcm_SumAverage

In contrast, from Figure 4.16, we can observe that most genes positively correlate with the extracted radiomics features. Furthermore, The top three radiomic feature categories, with the highest number of genes falling under each, are *wavelet-HHH_firstorder_Kurtosis (FOK)* (specifically 109 out of 244), *wavelet-LLH_ngtdm_Coarseness (Coarse)* (23 genes), and *waveletHHH_glrlm_Run-LengthNonUniformity (RLNU)* (19 genes). The top 10 genes with the highest absolute correlation values with each *FOK, Coarse,* and *RLNU*, as depicted in Figure 4.16, are listed in Table 4.3, Table 4.4, and Table 4.5 along with their direct or indirect roles in gliomas. The maximum positive correlation value indicates that as one set of features increases, the other set also increases proportionally. The correlation plot between genes and each of the three feature classes is shown in Supplementary Figure 11.

In radiomics, *FOK* represents a first-order statistical radiomic feature extracted from the actual label segmentation map after applying a wavelet high pass filter across all three dimensions. Since these features originate from single-pixel or single-voxel analyses, they are referred to as first-order statistical (FO) texture analysis (TA) features, regardless of the spatial relationships of pixels/voxels. FO represents a histogram representation of image intensities in a predefined region of interest. TA provides a noninvasive approach for quantifying tissue macroscopic heterogeneity, indirectly reflecting microscopic tissue heterogeneity [Soni et al. (2019)]. In this context, kurtosis represents the sharpness or pointedness observed in a histogram, offering insights into the visual contrast and describing the shape of the intensity distribution of data [Mayerhoefer et al. (2020)].

Whereas, *Coarse* extracts high-order TA features which represent the disparity in gray-level values between the central voxel and its neighboring area, thereby capturing the spatial frequency of changes in gray-level intensities [Rajput et al.

Figure 4.15. The distribution of the feature set and genes after thresholding with a PC value of $<= 0.006$. While all genes were included in the plot, only every eightieth gene is represented on the X-axis.

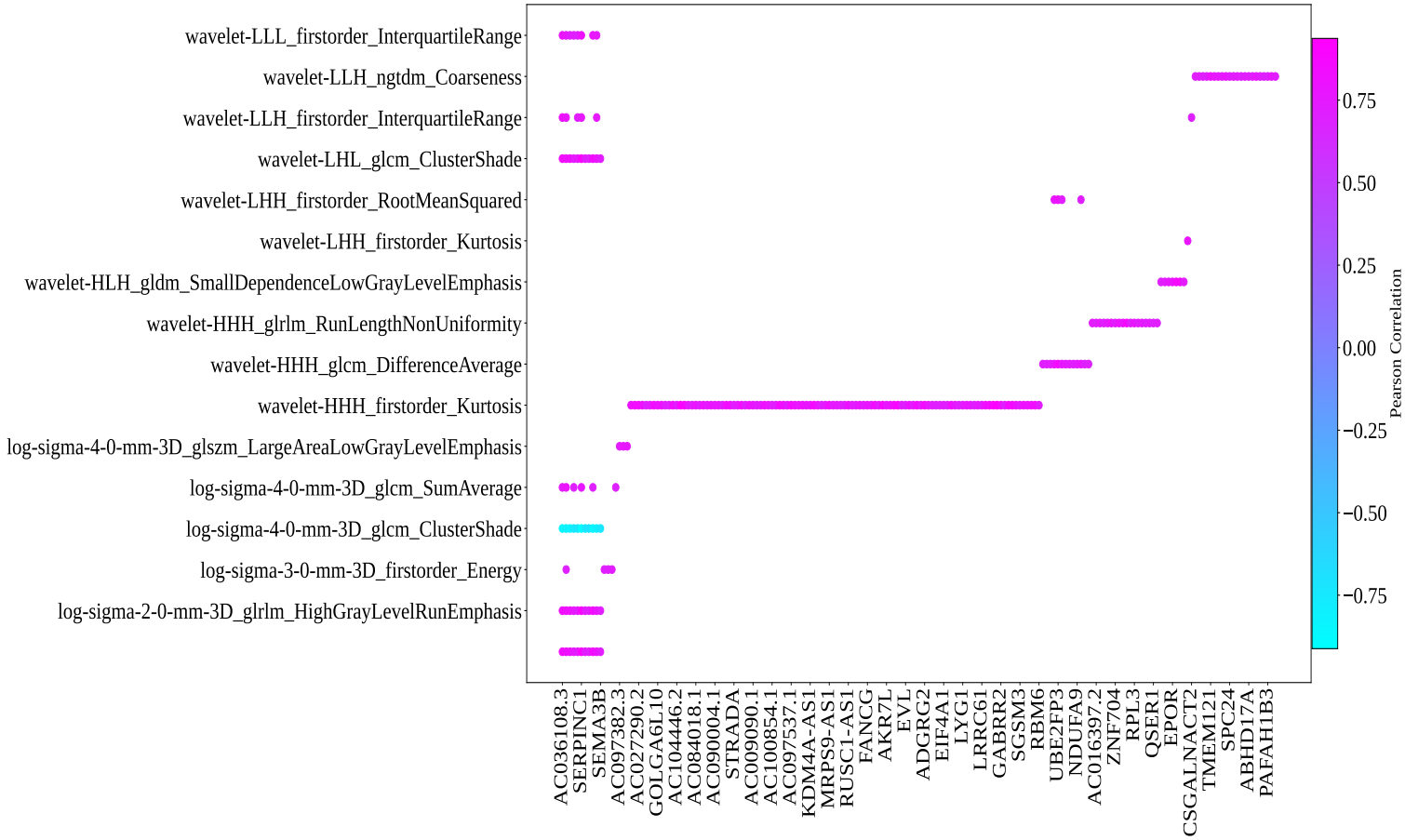


Figure 4.16. The distribution of the extracted feature set is exhibited on the Y-axis and gene expression on the X-axis after thresholding with a PC value of $>= 0.7$. While all genes were included in the plot, only every fifth gene is represented on the X-axis.

(2023a); Mayerhoefer et al. (2020); Papp et al. (2018)]. Similarly, RLNU is also a high-order statistical TA feature that measures the distribution of gray-level intensity values and their spatial relationships in an image regarding the run length of pixels. First and second-order TA are the most commonly employed techniques for survival analysis [Soni et al. (2019); Jang et al. (2020)].

Moreover, wavelet filters have demonstrated significant utility in medical imaging by providing exploration across diverse spatial-frequency scales. This capability facilitates improved correlation of significant features, enhancing computational efficiency and robustness to data variations and noise [Jin et al. (2005)]. Furthermore, heterogeneity plays a crucial role in characterizing malignancy, reflecting adverse tumor biology. Thus, kurtosis has demonstrated its effectiveness in capturing the malignant characteristics of gliomas [Sofie Van Cauter et al. (2012)].

TABLE 4.3
The association between genes and the *FOK* feature class, with genes identified as being linked to poor prognosis in glioma patients.

Gene	PC	Functions
NPR3	0.937	It encodes natriuretic peptide receptors, which regulate blood metabolic and growth processes. This gene is associated with numerous cancers, including osteosarcoma, renal carcinoma, and gliomas [H. Yang et al. (2019)].
AC009318.4 long non-coding RNAs (lncRNAs)	0.916	It is substantial evidence from studies suggesting its role in cancer development and progression [Han Li & Chen (2015)]. However, the potential roles of lncRNAs in the biogenesis and development of gliomas have not been explored [X. Zhang et al. (2012)].

AFDN-DT lncRNAs/ MLLT4 antisense RNA1 (MLLT4-AS1)	0.9	This gene has been found as a tumor suppressor in gastric cancer. Its association between AFDN-DT suppression and DNA hypermethylation and DNA methylation suppression with the restoration of AFDN-DT expressions [Lai et al. (2020)]. Association with brain cancer was established in [Gene Atlas]. [2]
SLC5A9	0.895	Identified as a hypomethylation promoter gene, it also functions as a transporter crucial for tumor cells' survival, relying on external nutrients for growth [Z. Zhang et al. (2011); Nwosu et al. (2023)]. In preclinical studies, transporter inhibition exhibits anti-tumor efficacy. However, many cancer-associated transporters remain poorly characterized and untapped as potential targets for cancer treatment [Nwosu et al. (2023)].
AC005899.6 lncRNAs	0.889	AC005899.7 have been linked to high risk in renal cancer Hao et al. (2023). Similarly, AC005899.6 has been associated with gastric cancer [J. Wei et al. (2021)].
UBOX5-AS1	0.872	lncRNA UBOX5-AS1 is boosted in low oxygen conditions, increasing invasiveness in ovarian endometriosis. Targeting this gene could be a promising treatment approach [H. Liu et al. (2021)]. However, no specific functions of UBOX5-AS1 in the brain haven't been extensively studied or elucidated.

[2] https://www.genecards.org/

TAS2R19	0.866	A bitter-tasting compound. Recent studies have explored how bitter compounds fight cancer and enhance chemotherapy, suggesting their therapeutic potential. TAS2Rs, linked to drug effectiveness, may be crucial in cancer treatment. Understanding how TAS2Rs work with their ligands could lead to better cancer therapies [Costa et al. (2023)]. Further, it is found as the potential gene for analysis in the pseudo and actual progression of Gliomas [Qian et al. (2016)].
AC097537.1 (GPM6A-DT)	0.864	Targeting the GPM6A gene increases the sensitivity of glioblastoma stem cells (GBSC) to radiation treatment. This suggests that inhibiting GPM6A or could be a promising strategy for glioblastoma treatment, as it could effectively target proliferation, invasion, and radioresistance simultaneously [Lacore et al. (2022)].
AC025178.1 lncRNA	0.858	An association between this gene and Bladder cancer [D. Liu et al. (2022)]
ZDHHC11 [Tang et al. (2022)]	0.854	Recent findings indicate that ZDHHCs are linked to the prognosis of tumor patients, although the prognostic roles of these ZDHHCs may vary between different types of tumors. For instance, high expression of ZDHHC11 is associated with favorable prognostic features in glioma patients [Z.-Y. Liu et al. (2023); Tang et al. (2022)].

TABLE 4.4
The association between genes and the *Coarse* feature class, with genes identified as being linked to poor prognosis in glioma patients.

Gene	PC	Functions
TMEM121	0.794	TMEMs, widely expressed in the brain [*National Library of Medicine-National Center for Biotechnology Information - Gene: 80757* (2023)], act as membrane channels for specific substances, yet much of their function remains elusive due to challenges in isolation and purification. Notably, high expression of TMEM140 in gliomas and TMEM97 are implicated in tumor growth of gliomas [Schmit & Michiels (2018)] has been reported.
Compared to normal brain tissues, PRMT1	0.776	PRMT1 expressions in gliomas are increased. Additionally, research suggests that in gliomas with a mutation in the IDH1 gene, there is a decrease in PRMT1 expression [S. Wang et al. (2012); Shen et al. (2024); Lathoria et al. (2023)].
GPX4	0.774	In gliomas, changes in the Xc-system (cellular antioxidant system) and GPX4 activity are detected, indicating their role in tumor advancement [M. Lu et al. (2022)]. GPX4 could emerge as a novel prognostic indicator in glioma, significantly associated with glioma cell proliferation, migration, and apoptosis [Zhao et al. (2017)].

MRPL54	0.767	It contribute to oxidation and cell state regulation. Numerous studies have highlighted the atypical expression of MRPs in different tumors. Associations have been identified between MRPS23, MRPL35, MRPL42, MRPS2, MRPS12, MRPL34 genes and gliomas [Huang et al. (2020)].
RUVBL2	0.759	It is involved in chromatin remodeling, DNA damage repair, and cell cycle regulation, all crucial in cancer. Additionally, expression of RUVBL2 is considerably elevated in tumor tissues compared to normal tissues in lower-grade glioma (LGG)) [Su et al. (2022)].
FAM3A	0.759	Belonging to the FAM3 gene family subgroup, it contributes to the modulation of neuronal apoptosis and holds potential significance in neurodegenerative disorders [J. Liu et al. (2019)].
SAE1	0.742	It facilitates the advancement of glioma cancer [Y. Yang et al. (2019); Q. Guo et al. (2023)].
TTC9B	0.736	It is associated with reduced survival in glioma patients [Gong et al. (2021)].
RPL10	0.731	Ribosomal proteins (RPs) are involved in numerous pathological processes, particularly tumorigenesis and cell transformation. High expression of RPS6 is observed in glioblastoma and has been reported in the study [El Khoury & Nasr (2021)].

FAM98C	0.725	Although the precise function of FAM98C remains unclear, it is believed to be involved in RNA metabolism and gene expression regulation, with high expression in the cerebral cortex [*Protein Atlas - FAM98C Tissue Expression* (2023)].

TABLE 4.5
The association between genes and the *RLNU* feature class, with genes identified as being linked to poor prognosis in glioma patients.

Gene	PC	Functions
BMP2	0.869	Bone morphogenetic proteins (BMPs), specifically BMP2, are recognized for their role in promoting differentiation and inhibiting growth in GBM cells [Persano et al. (2012); C. Liu et al. (2009); Sachdeva et al. (2019)].
GALNT13	0.795	Glycosylation is pivotal in the molecular and cellular mechanisms underlying tumorigenesis. While glycosylation has been associated with diverse biological processes and diseases, its precise role in glioma remains largely unexplored. GALNT13 has been associated with a better prognosis in gliomas according to Yang et al. [Y. Yang et al. (2024)] illustrated in their research that GALNT13 expression was highest in the fetal brain, followed by the adult cerebellum, cerebral cortex, and whole brain. Interestingly, GALNT13 was exclusively detected in neuroblastoma cells and was absent in other cell-lines, including glioma cells.

RPL4	0.79	Its expression is increased in gliomas samples than normal brain [K. Guo et al. (2022)].
TET1	0.773	The gene acts as a tumor suppressor; its downregulation has increased tumor cell proliferation and development [Ji et al. (2021)]. low expression of TET1 is related to poor survival of gliomas patient [Bian et al. (2014)]. Müller et al. highlighted in their study that gliomas with IDH1 mutations exhibited nuclear aggregation of TET1 [Müller et al. (2012)].
RPL3	0.77	RPL3 might have implications in glioma tumorigenesis [B. Wei et al. (2020)].
MAPT-AS1	0.754	It is identified as a lncRNA, has been associated with poor prognosis in glioma patients [X. Wang et al. (2020)].
SEMA4G	0.749	In LGG gliomas, higher gene expression is associated with longer survival [Xiao et al. (2020)]. It has tumor suppressive functions in gliomas [T.-P. Lu et al. (2012)].
TLCD5	0.742	While the gene is widely expressed in the brain [National Center for Biotechnology Information (2024)], its investigation about gliomas remains relatively understudied.
AP003071.5	0.739	This lncRNA is relatively underexplored in gliomas but exhibits high expression in prostate cancer [Che et al. (2022)].

EIF4B	0.733	Lowering EIF4B in cancer cells slows proliferation and enhances their sensitivity to genotoxic stress-induced apoptosis. The research findings highlight EIF4B as a potential target for anti-cancer treatment development [Shahbazian et al. (2010)]. EIF3I and EIF4H have been recognized as favorable candidates for glioma treatment [Krassnig et al. (2021)].

For the genes enumerated in Table 4.3, Table 4.4, and Table 4.5, we attempted to link them to gliomas based on existing studies. In cases where such associations were not identified, we investigated their direct or indirect involvement in other cancers because genes associated with the other cancers might also hold significance in gliomas, given shared or crosstalk biological pathways [H. He et al. (2021)]. Exploring their functions in alternative contexts can offer valuable insights into their potential contributions to glioma development and progression. Moreover, the highly invasive nature of cancer cells constitutes the primary reason for elevated cancer mortality and often facilitates cancer advancement and metastasis [H. He et al. (2021)]. Therefore, understanding the mechanism behind the onset and development of malignant tumors is important. Most of the genes listed in the tables are associated with the poor prognosis of glioma patients, some of these genes are lncRNAs, which are RNA molecules measuring over 200 nucleotides long and do not encode proteins. Studies have shown that dysregulation of lncRNAs is associated with various diseases, including cancer [Jiang et al. (2019)]. It exerts clinical therapeutic effects against tumors by inhibiting mRNA

transcription and binding to protein to block its function [Jiang et al. (2019)].

In gliomas, specific lncRNAs are dysregulated, influencing tumor development and progression. These findings indicate that the dysregulation of lncRNA expression may occur early in tumorigenesis, with these lncRNAs potentially serving crucial functions in glioma initiation [Shi et al. (2017)]. Hence, although lncRNAs may not directly encode proteins that affect gliomas, their dysregulation can still profoundly impact glioma development and progression through indirect mechanisms. However, RNAs have not been thoroughly and systematically studied in gliomas [Z. Wang et al. (2020)].

In this study, we identified genes enumerated in Table 4.3 that exhibit a high positive correlation with the extracted feature class *FOK*, showcasing similar behavior. As discussed, *FOK* is well-known for capturing tumor heterogeneity, thus reflecting malignancy. Moreover, studies have suggested that several genes listed in Table 4.3, Table4.4 and Table 4.5 play pivotal roles in gliomas or other cancers. It's essential to highlight that among these genes are those encoding RNA, which can significantly influence the malignancy of gliomas and contribute to survival prediction. Hence, this study has emphasized ten specific genes (including RNAs) that may be critical in indicating tumor malignancy and predicting patient SD.

4.5 Conclusions

We employed versions of the triplanar model to generate segmentation results consistent with other leading segmentation models. Subsequently, to evaluate the robustness of the feature set proposed in the study by [Rajput et al. (2023a)], we extracted a dominant set of 29 features from all these variants to predict the SD

of the patients. We validated the optimality of these features by examining correlation maps and SRC across all the variants. The trained regressor model demonstrated superior performance across multiple performance metrics. Specifically, the ERFR and RFR models trained on *Variant4* outperformed other state-of-the-art BraTS2020 models across multiple performance metrics, signifying the robust nature of features and predictor models.

The SD results surpassed those of the top-performing models on the BraTS2020 validation set across multiple performance metrics. We analyzed the behavior of the feature set using multiple post-hoc interpretable methods such as SHAP and LIME to derive global explanations and PDP and ALE methods to obtain local explanations. Analyzing the feature set through diverse post-hoc methods such as SHAP, LIME, PDP, and ALE provides a comprehensive understanding of the model's behavior globally and locally. This approach ensures cross-verification, yielding robust and contextually nuanced insights, enhancing confidence, and allowing for model-agnostic interpretations.

The features are arranged in descending order in the SHAP summary plot, aligning with the study's findings by [Rajput et al. (2023a)] when considered within a broader context. Moreover, the features' local and global behavioral patterns are consistent across the respective plots. Notably, first-order statistical features, age, location-based, and texture features play a crucial role in prediction. Additionally, these interpretability methods provide valuable insights into the model, offering human-understandable inferences.

We further validated the robustness of the feature set for SD prediction, emphasizing the importance of robust segmentation results and regressor models. Moreover, employing these interpretability tools enables the extraction of human-

understandable inferences, assisting in the comprehension of ML black-box models.

Our radiogenomic study incorporated extracted radiomics, location-based features, and integrated them with additional clinical information, including age, to enhance SD prediction. The uncorrelated nature of these features enables our model to capture diverse information from various data aspects, thereby enhancing survival prediction performance. Moreover, reframing the problem as binary classification, especially given the limited samples, simplified the problem and improved the model performance. Further, we observed a higher correlation of the extracted features set with gene expressions for a potential link between radiomic features and genes to comprehend malignancy or heterogeneity of tumors at a molecular level. We identified radiomic signatures including *wavelet-HHH_firstorder_Kurtosis*, *wavelet-LLH_ngtdm_Coarseness*, and *wavelet-HHH_glrlm_RunLengthNonUniformity* which are strongly associated with genes. These identified genes are directly or indirectly responsible for tumor malignancy. These radiomic feature classes could be potential biomarkers for predicting tumor behavior, potentially highlighting molecular targets for therapeutic interventions or further investigation.

4.6 Future scope

For SD prediction, clinical data such as age, gender, race, KPS, and treatment information can be obtained from the TCGA glioblastoma project. This clinical information has proven to be an essential variable in the prognosis of glioblastoma. Combining this clinical information with our proposed feature set can further val-

idate its effectiveness. Moreover, leveraging interpretation tools on this combined dataset can yield more nuanced insights into the factors influencing survival outcomes in glioblastoma patients.

Similarly, for radiogenomic analysis, access to a larger dataset can offer a more robust validation of the relationship between radiomic features extracted from medical images and gene expression patterns. With a greater diversity of samples, we can explore a wider range of genomic profiles and their corresponding radiomic signatures. This leads to a deeper understanding of the underlying biological mechanisms driving tumor progression and response to therapy. This expanded dataset will enable more comprehensive analyses and enhance the reliability of our findings in the context of radiogenomics.

Chapter 5

Conclusions and Future Scope

5.1 Conclusions

Accurate and precise brain tumor segmentation and SD prediction of glioblastoma patients are among the most challenging and critical tasks in the medical domain. These tasks are crucial for effective diagnosis and treatment planning. However, efficient and lightweight networks are crucial for seamless incorporation into current medical practices. Therefore, we proposed a triplanar attention-based network that can work in a resource-constrained environment and generate robust segmentation results. The proposed approach with fewer parameters (almost 3$\times$ less) demonstrates comparable performance to a 3D model, making it suitable for brain tumor segmentation in resource-limited settings. We worked on fusing channel and spatial attention, allowing the model to refine the feature map across channels and spatially. We proposed an improved CCSAv1 network that has increased the dice score without much addition to the trainable parameters.

Further, we extracted diverse features for SD prediction encompassing location-

based, geometric-based, first-order statistical, and texture-based radiomics features by applying wavelet and LoG filters. We conducted feature selection and reduced the dimensionality to 29 features. Utilizing a correlation map and SRC, we observed that most features are uncorrelated, validating the capture of unique information from tumor lesions. Subsequently, we employed ensemble learning-based predictor models trained on this feature set to predict SD. The feature set extracted from the 3D and triplanar networks consistently performed well for SD prediction. Notably, the RFR and ERFR models outperformed numerous top-ranking methods in BraTS2020 on the validation set.

In this study, we undertake another essential task: comprehending the decision-making process of black-box ML models, which is crucial for establishing trust between machine learning systems and domain experts. Hence, we utilized various post-hoc interpretable tools (SHAP, PDP, LIME, and ALE) to generate visual explanations, facilitating straightforward and accessible interpretations. We analyzed the behavior of the feature at the generic and local levels (specific to a particular sample). We observed that the behavior of features remained consistent across different feature values, and the ranking of top-contributing features remained the same, albeit with some variations in the order for certain moderately contributing features. This approach ensures cross-verification, yielding robust and contextually nuanced insights, enhancing confidence, and allowing for model-agnostic interpretations. Across all the tools, we found the dominance of *Age, first-order statistical (interquartile range, kurtosis), location-based feature (centroid of necrosis, active tumor), and texture-based features (GLCM_corr)*. We determined the relationship between features, their values, and how they affect the SD prediction. These tools helped to highlight top-ranking features and show how

these features impact SD prediction. The explanations obtained from these analyses were then utilized to extract biological significance from the features. The insights drawn from the visual graphs were consistent with the existing understanding in the medical domain.

However, the feature set under investigation primarily comprises radiomic features renowned for capturing heterogeneity and indicating tumor malignancy through pixel intensity. However, they have limitations in directly correlating with tumor malignancy, which is more closely tied to the molecular landscape. To address this problem, we conducted an analysis linking radiomic features with gene expression in glioma patients, allowing for direct biological inference. In this study, we have included additional clinical information that has improved SD prediction. However, it should be validated further on a larger dataset. The radiogenomic study examined gene expressions with both strong correlations and those exhibiting uncorrelated characteristics. Further, we observed a higher correlation of the extracted features set with gene expressions for a potential link between radiomic features and genes to comprehend malignancy or heterogeneity of tumors at a molecular level. Further, we observed a higher correlation of the extracted features set with gene expressions for a potential link between radiomic features and genes to comprehend malignancy or heterogeneity of tumors at a molecular level. We identified radiomic signatures including *wavelet-HHH_firstorder_Kurtosis*, *wavelet-LLH_ngtdm_Coarseness*, and *wavelet-HHH_glrlm_RunLengthNonUniformity* which are strongly associated with genes. These identified genes are directly or indirectly responsible for tumor malignancy. These radiomic feature classes could be potential biomarkers for predicting tumor behavior, potentially highlighting molecular targets for therapeutic interventions or

further investigation. Moreover, identifying gene expressions that mimic radiomic features may lead to discovering biomarkers for predicting tumor behavior, prognosis, or treatment response.

5.2 Future scope

For BTS, there is scope for further improvement and fine-tuning of the proposed methodology. This could involve refining network architecture and optimizing hyperparameters, such as exploring variations of 2D and/or 3D models and methods based on Graphical Neural Networks. These approaches may offer opportunities to optimize resource consumption, enhance performance, and broaden the scope of clinical applicability.

Additionally, integrating Multimodal medical image fusion techniques, which merge multiple medical modalities into a single image, can enhance the information available. Furthermore, incorporating additional post-processing methods, such as replacing ET lesions with necrotic tissue based on specific cutoff values determined through experimental analysis, could also be beneficial. Incorporating functional-imaging techniques (e.g., PET scans and fMRI scans into conventional MRI can help DNN models learn the physiological, metabolic, and biological details of tumor lesions, enhancing the discriminating potential of the DNN models. Explainable AI in the lesser complex models can be made more accessible due to their simplicity, helping to identify the model's strengths and weaknesses.

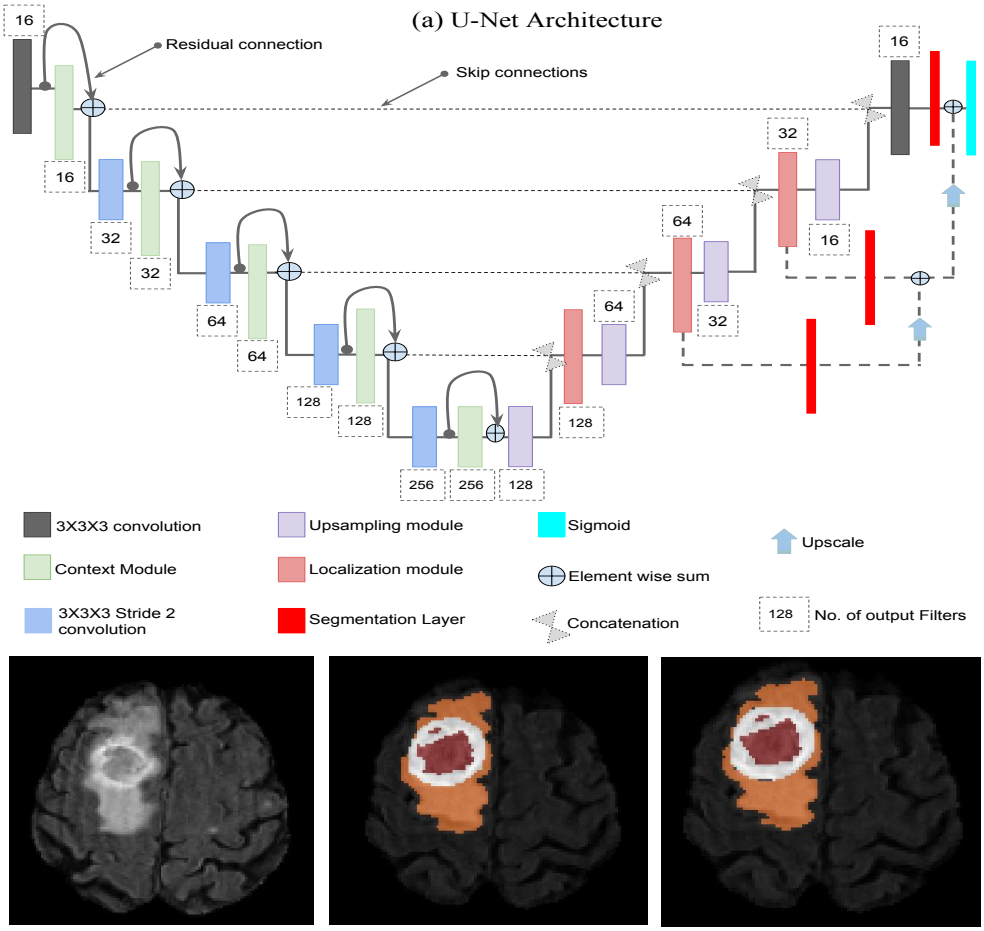
For SD prediction, incorporating clinical data such as age, gender, race, and treatment information has enhanced SD predicting accuracy, necessitating validation on larger datasets. Other clinical factors such as performance status and

comorbidities have also demonstrated their significance in glioblastoma prognosis. When combined with our proposed feature set, these variables can further validate its efficacy. Moreover, including other omics data such as proteomics, genomics, and metabolomics can enhance predictive accuracy. Moreover, interpretation tools can be subsequently applied to enhanced predictor models trained on an enriched feature set using ample datasets to comprehend the decisions of ML models. These will allow for more robust analysis and validation of model interpretations. Moreover, ample datasets facilitate the identification of genomic alterations associated with specific radiomic features and tumor characteristics in radiogenomic analysis.

Appendix

.1 Supplementary:

.1.1 Supplementary Figures:



(b) Input axial-FLAIR (c) Ground truth (d) Segmented result

Figure 1. *No-new network* architecture and comparison of segmentation outcomes obtained, (a) the application of the 3D U-Net for BTS, (b) FLAIR modality, (c) ground truth segmentation, and (d) predicted segmentation, all shown in the axial view.

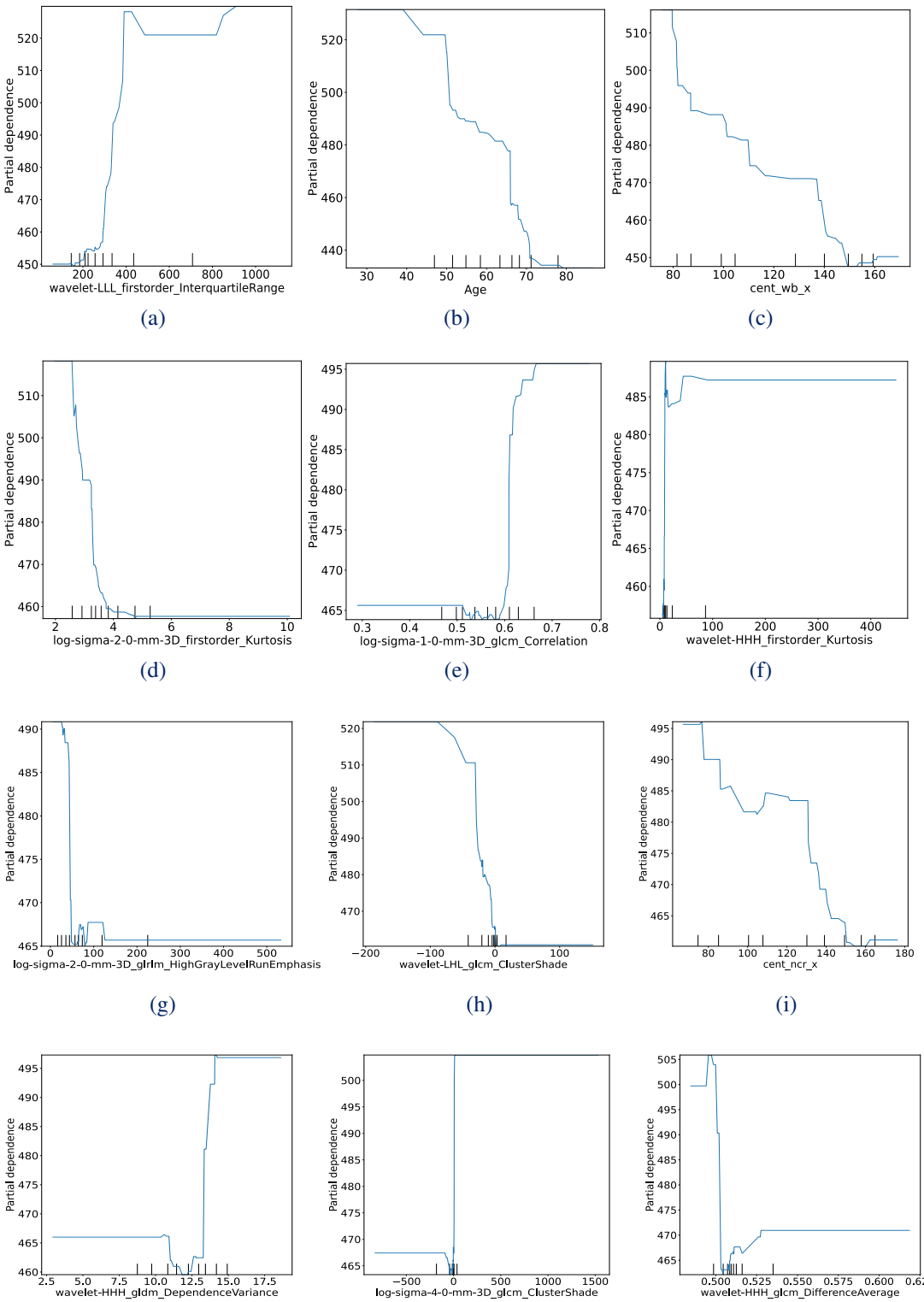
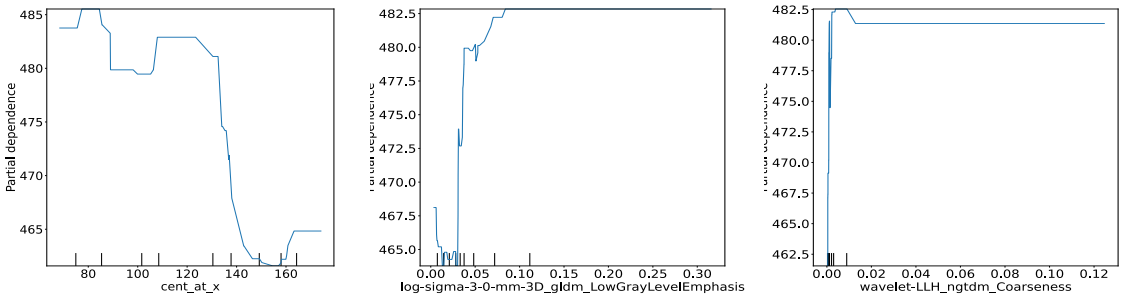
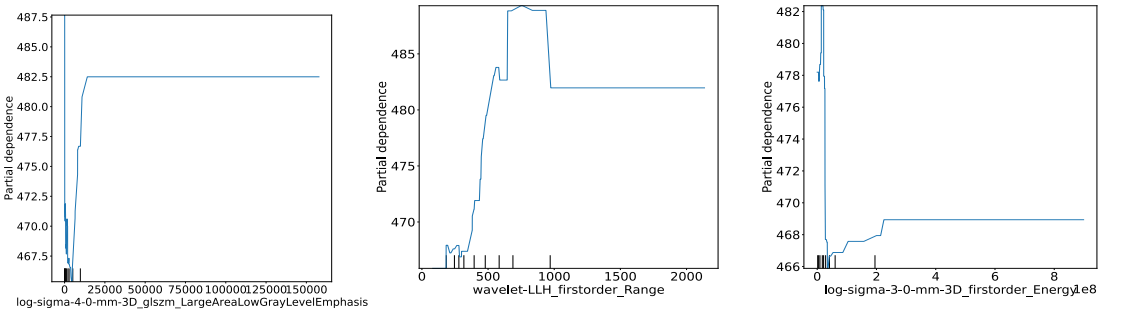


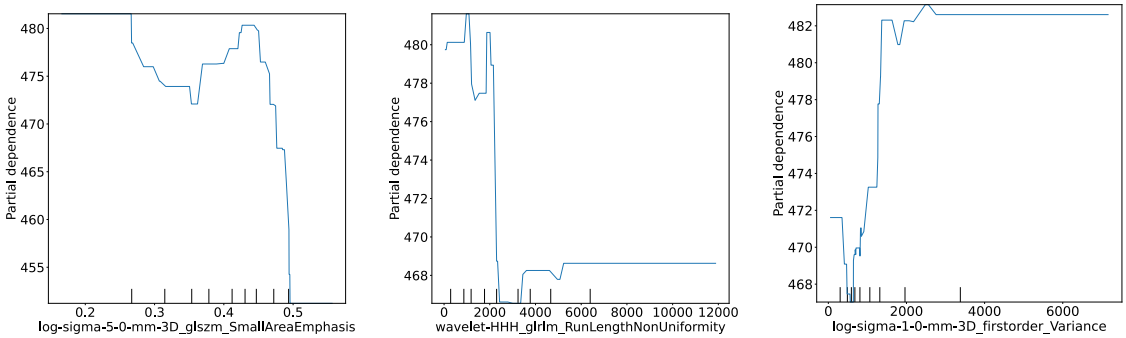
Figure 2. PDP showcasing 29 prominent features.



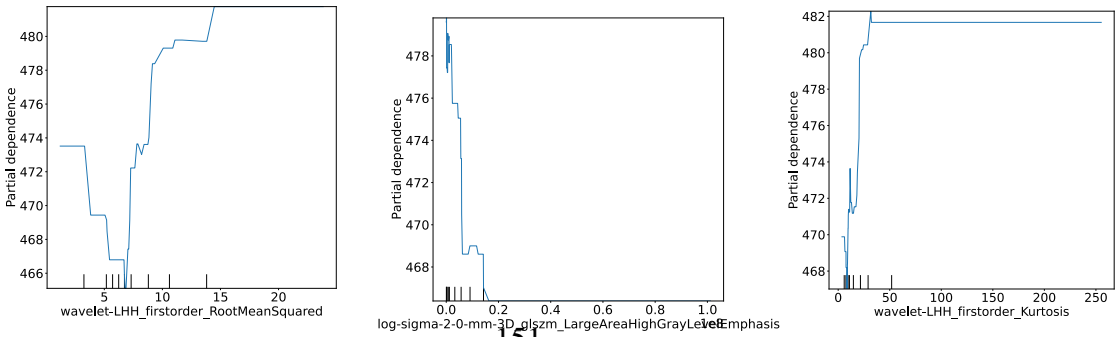
(m) (n) (o)



(p) (q) (r)

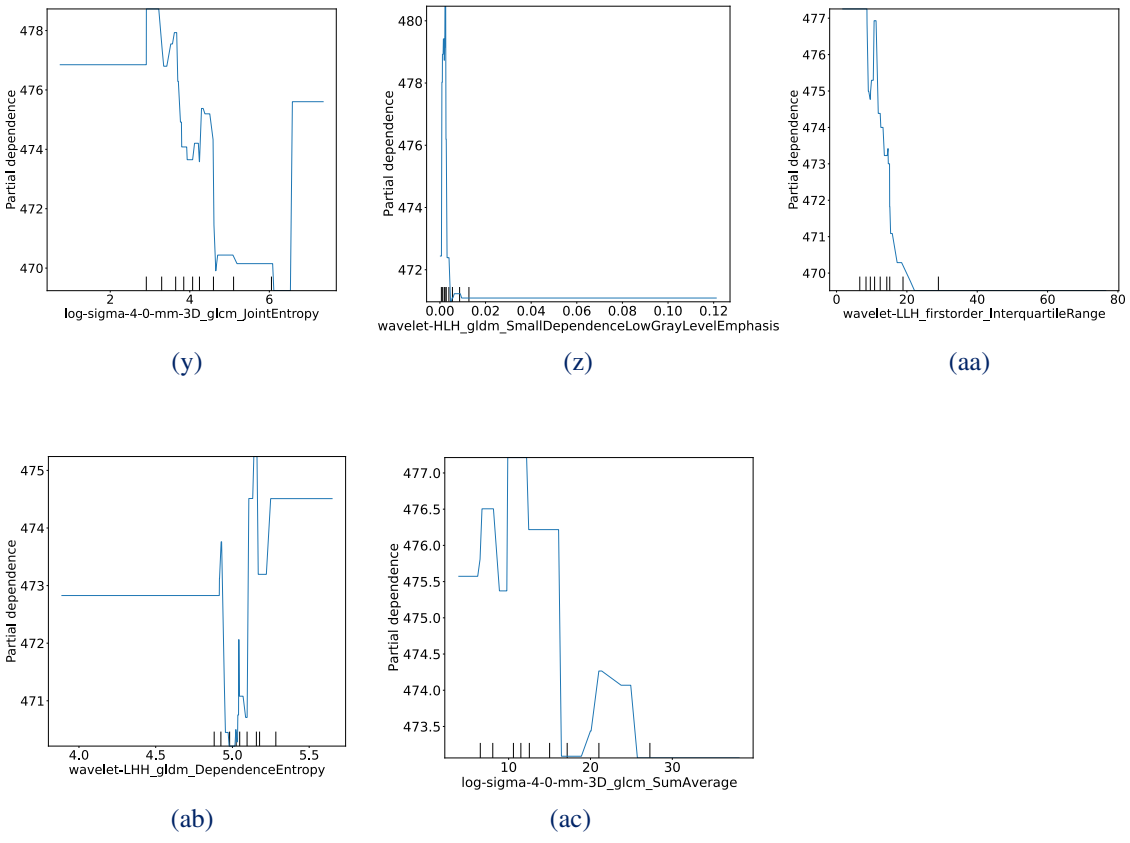


(s) (t) (u)



(v) (w) (x)

Figure 2. PDP showcasing 29 prominent features.



(y) (z) (aa)

(ab) (ac)

Figure 2. PDPs showcasing 29 prominent features: The X-axis represents feature values of the corresponding features, whereas the Y-axis illustrates the average rate of change in the feature's effect on the SD prediction. The vertical bars across the X-axis depict the data distribution. These visualizations offer insights into the general trends of feature(s) of interest, on SD prediction, encompassing all samples.

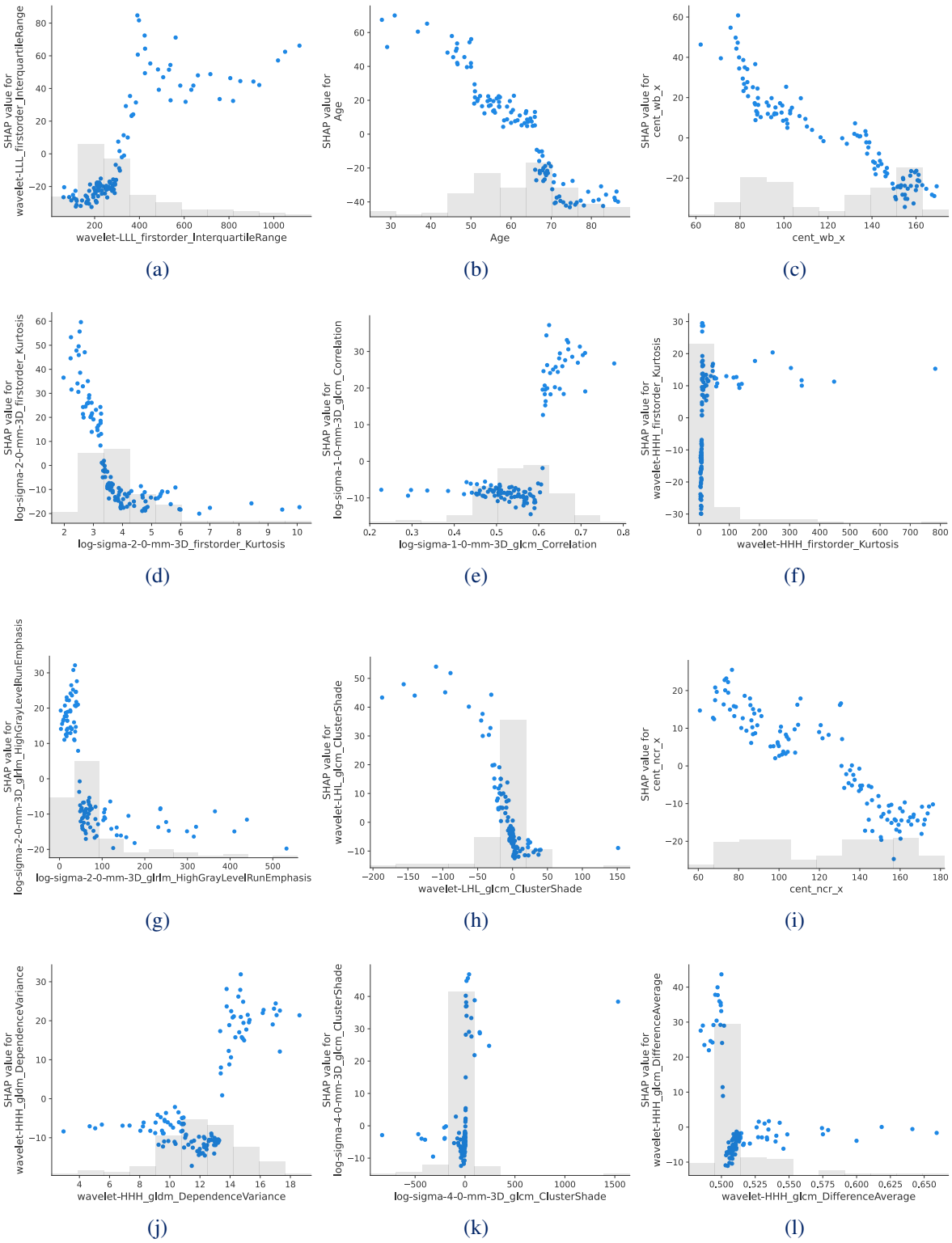


Figure 3. Distribution of SHAP values for the dominant features.

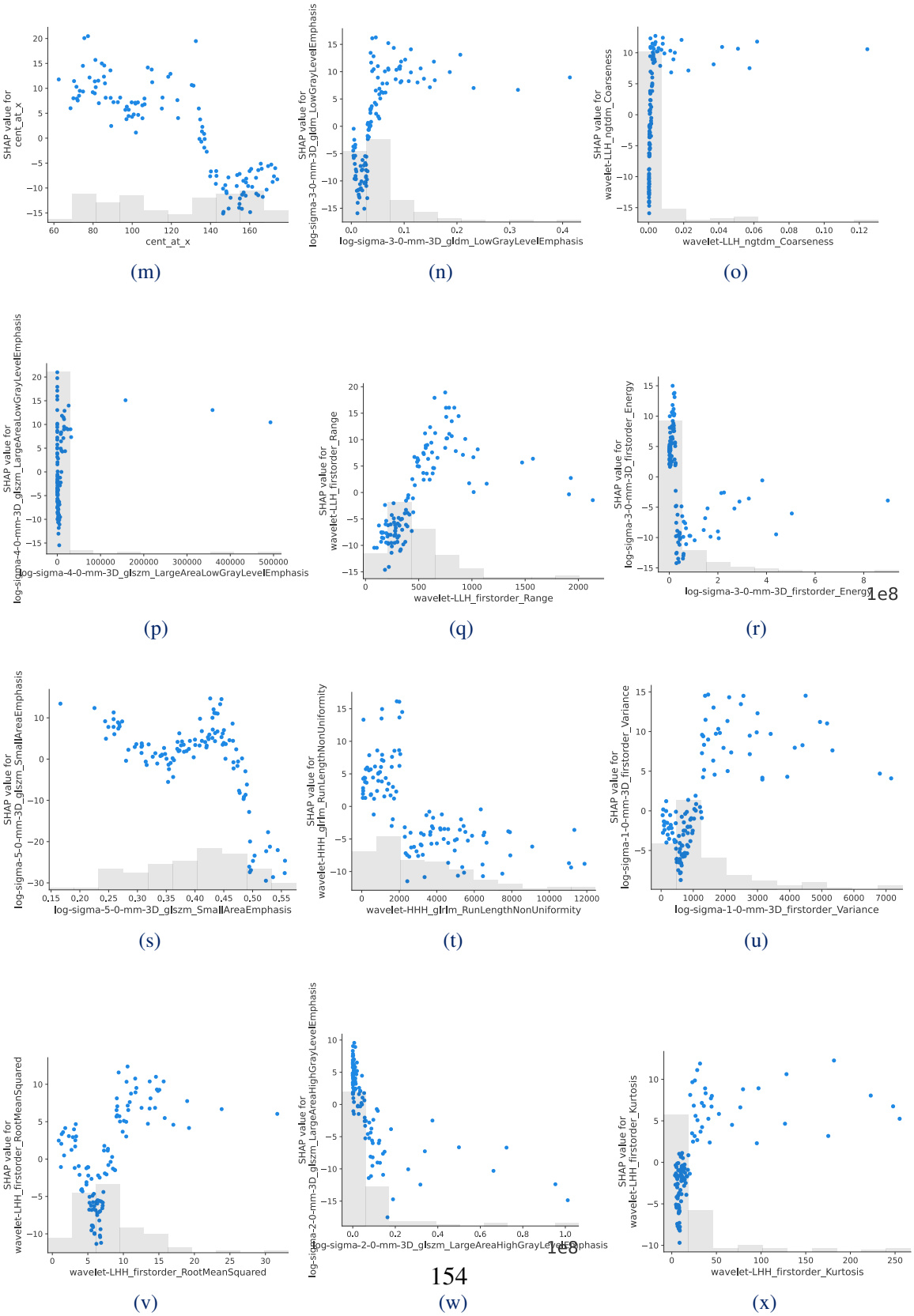


Figure 3. Distribution of SHAP values for the dominant features.

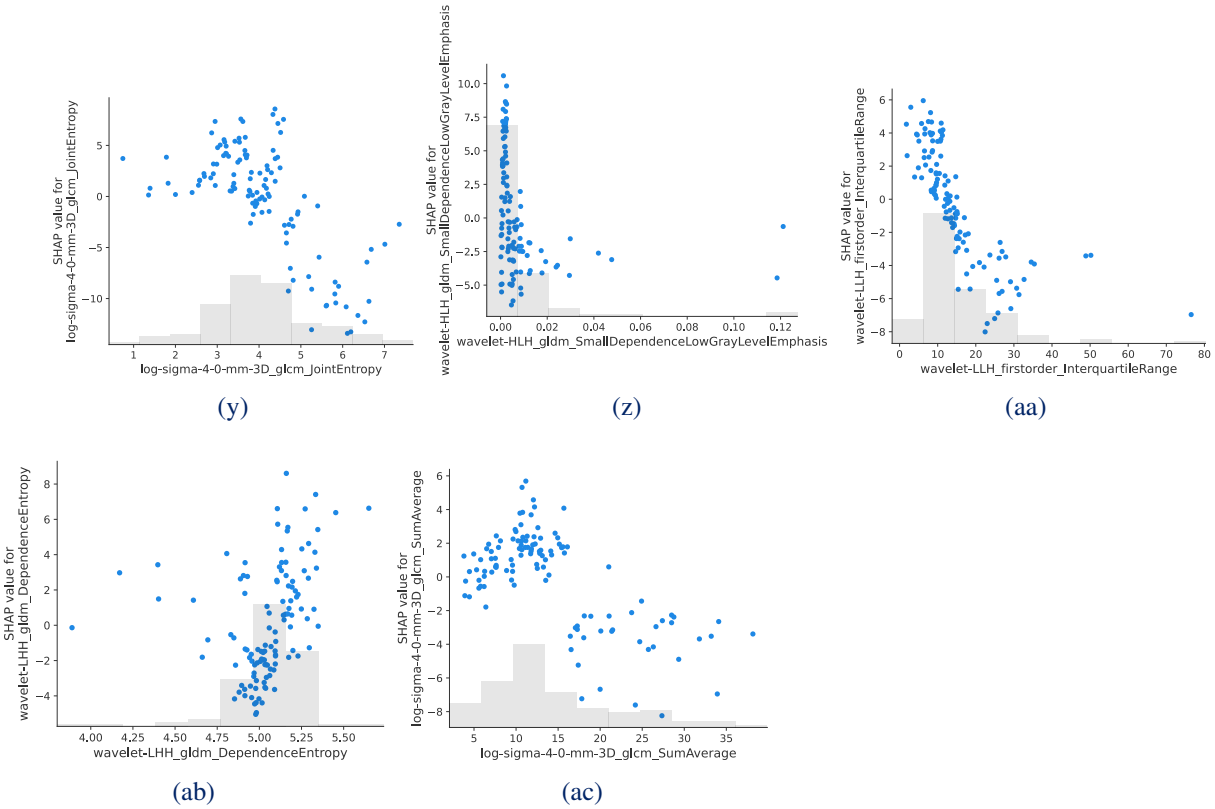


Figure 3. The feature values of each respective feature are displayed on the X-axis, while the SHAP values of corresponding instances are displayed on the Y-axis. The bar plot depicts the distribution of feature values, with each point representing a sample from the training set. This visualization aids in understanding where the majority of SHAP feature values lie, the individual impact of instances on target features, and their distribution. This illustrates the role of these values in defining important features. Across all SHAP plots, it is observable that the magnitude of SHAP values decreases with the declining importance of features, from high to low.

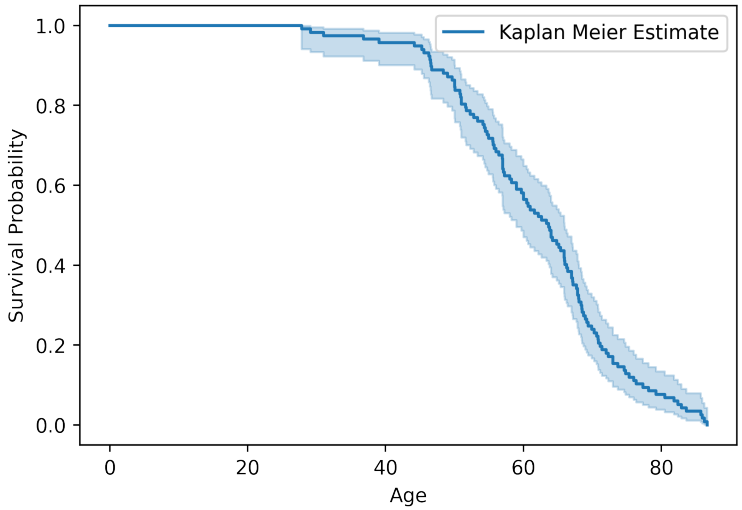


Figure 4. KM survival plot illustrating the probability of survival.

The KM estimator evaluates the proportion of patients who have survived a defined duration after receiving treatment or undergoing surgery. It computes event probabilities over time by dividing the duration into smaller intervals and subsequently recalculating these probabilities to obtain the ultimate estimate. The survival likelihood is defined in Equation 1:

$$S_{t+1} = S_t \times ((C_{t+1} - D_{t+1})/C_{t+1}) \tag{1}$$

In this equation, C represents the count of persons at risk, D denotes the count of deceased persons, and t signifies the time duration. The KM survival plot is depicted in Figure 4. It provides an accumulative measure, where survival probabilities stay consistent until another patient encounters the risk. From examining this graph, it's evident that the chances of survival for older patients are significantly reduced. The likelihood of survival shows a nearly linear decline after the age of 50, followed by an exponential decrease after 70, and significantly diminishes beyond the age of 80. The KM curve, examination conducted on the *age* feature aligns with both the SHAP and PDP assessments, illustrating the expo-

nential decline in survivability starting around the age of 70.

Extra RF - Variant1	Extra RF - Variant2	Extra RF - Variant3	Extra RF - Variant4
Age	Age	Age	Age
log-sigma-1-0-mm-3D_glcm_Correlation	wavelet-LLH_firstorder_Range	cent_ncr_x	cent_ncr_x
cent_ncr_x	cent_at_x	cent_at_x	cent_at_x
cent_at_x	cent_wb_x	log-sigma-1-0-mm-3D_glcm_Correlation	log-sigma-1-0-mm-3D_glcm_Correlation
wavelet-LLL_firstorder_InterquartileRange	log-sigma-1-0-mm-3D_glcm_Correlation	wavelet-LLL_firstorder_InterquartileRange	wavelet-HHH_gldm_DependenceVariance
log-sigma-2-0-mm-3D_firstorder_Kurtosis	cent_ncr_x	wavelet-LLH_firstorder_Range	wavelet-LLL_firstorder_InterquartileRange
cent_wb_x	wavelet-HHH_firstorder_Kurtosis	wavelet-HHH_gldm_DependenceVariance	wavelet-LLH_firstorder_Range
wavelet-LLH_firstorder_Range	log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis	wavelet-LHH_firstorder_Kurtosis	wavelet-LHH_firstorder_Kurtosis
wavelet-HHH_gldm_DependenceVariance	wavelet-LHH_firstorder_Kurtosis	log-sigma-4-0-mm-3D_glcm_SumAverage	log-sigma-2-0-mm-3D_firstorder_Kurtosis
wavelet-HHH_glrlm_RunLengthNonUniformity	log-sigma-3-0-mm-3D_gldm_LowGrayLevelEmphasis	log-sigma-1-0-mm-3D_firstorder_Variance	wavelet-LHL_glcm_ClusterShade
wavelet-HLH_gldm_SmallDependenceLowGrayLevelEmphasis	wavelet-LLL_firstorder_InterquartileRange	cent_wb_x	wavelet-LHH_gldm_DependenceEntropy
log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis	wavelet-HHH_gldm_DependenceVariance	log-sigma-2-0-mm-3D_firstorder_Kurtosis	cent_wb_x
wavelet-LHH_gldm_DependenceEntropy	log-sigma-1-0-mm-3D_firstorder_Variance	wavelet-LHH_gldm_DependenceEntropy	wavelet-LHH_firstorder_RootMeanSquared
log-sigma-3-0-mm-3D_firstorder_Energy	wavelet-HLH_gldm_SmallDependenceLowGrayLevelEmphasis	wavelet-LLH_firstorder_InterquartileRange	log-sigma-3-0-mm-3D_gldm_LowGrayLevelEmphasis
log-sigma-2-0-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis	wavelet-HHH_glrlm_RunLengthNonUniformity	wavelet-HHH_firstorder_Kurtosis	log-sigma-4-0-mm-3D_glcm_JointEntropy
wavelet-LHL_glcm_ClusterShade	log-sigma-4-0-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis	log-sigma-2-0-mm-3D_glrlm_HighGrayLevelRunEmphasis	wavelet-HLH_gldm_SmallDependenceLowGrayLevelEmphasis
log-sigma-1-0-mm-3D_firstorder_Variance	log-sigma-2-0-mm-3D_firstorder_Kurtosis	log-sigma-3-0-mm-3D_firstorder_Energy	log-sigma-4-0-mm-3D_glcm_SumAverage

wavelet-LHH_firstorder_RootMeanSquared	wavelet-LHH_firstorder_RootMeanSquared	log-sigma-4-0-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis	log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis
wavelet-LHH_firstorder_Kurtosis	log-sigma-4-0-mm-3D_glcm_SumAverage	log-sigma-4-0-mm-3D_glcm_JointEntropy	log-sigma-1-0-mm-3D_firstorder_Variance
wavelet-HHH_firstorder_Kurtosis	wavelet-HHH_glcm_DifferenceAverage	wavelet-HHH_glcm_DifferenceAverage	wavelet-LLH_firstorder_InterquartileRange
log-sigma-4-0-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis	log-sigma-4-0-mm-3D_glcm_JointEntropy	wavelet-LHH_firstorder_RootMeanSquared	log-sigma-2-0-mm-3D_glrlm_HighGrayLevelRunEmphasis
log-sigma-4-0-mm-3D_glcm_JointEntropy	wavelet-LHH_gldm_DependenceEntropy	wavelet-HHH_glrlm_RunLengthNonUniformity	log-sigma-3-0-mm-3D_firstorder_Energy
log-sigma-4-0-mm-3D_glcm_SumAverage	log-sigma-2-0-mm-3D_glrlm_HighGrayLevelRunEmphasis	wavelet-LHL_glcm_ClusterShade	wavelet-HHH_firstorder_Kurtosis
wavelet-HHH_glcm_DifferenceAverage	log-sigma-3-0-mm-3D_firstorder_Energy	wavelet-HLH_gldm_SmallDependenceLowGrayLevelEmphasis	wavelet-HHH_glcm_DifferenceAverage
log-sigma-2-0-mm-3D_glrlm_HighGrayLevelRunEmphasis	wavelet-LLH_firstorder_InterquartileRange	log-sigma-3-0-mm-3D_gldm_LowGrayLevelEmphasis	log-sigma-4-0-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis
log-sigma-3-0-mm-3D_gldm_LowGrayLevelEmphasis	log-sigma-4-0-mm-3D_glcm_ClusterShade	log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis	wavelet-HHH_glrlm_RunLengthNonUniformity
wavelet-LLH_firstorder_InterquartileRange	log-sigma-2-0-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis	log-sigma-2-0-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis	log-sigma-2-0-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis
log-sigma-4-0-mm-3D_glcm_ClusterShade	wavelet-LHL_glcm_ClusterShade	log-sigma-4-0-mm-3D_glcm_ClusterShade	log-sigma-4-0-mm-3D_glcm_ClusterShade
wavelet-LLH_ngtdm_Coarseness	wavelet-LLH_ngtdm_Coarseness	wavelet-LLH_ngtdm_Coarseness	wavelet-LLH_ngtdm_Coarseness

Figure 5. The features in all variations have been organized based on their importance, descending as per the SHAP summary plots. Features highlighted in red suggest a consistently similar order across all variations.

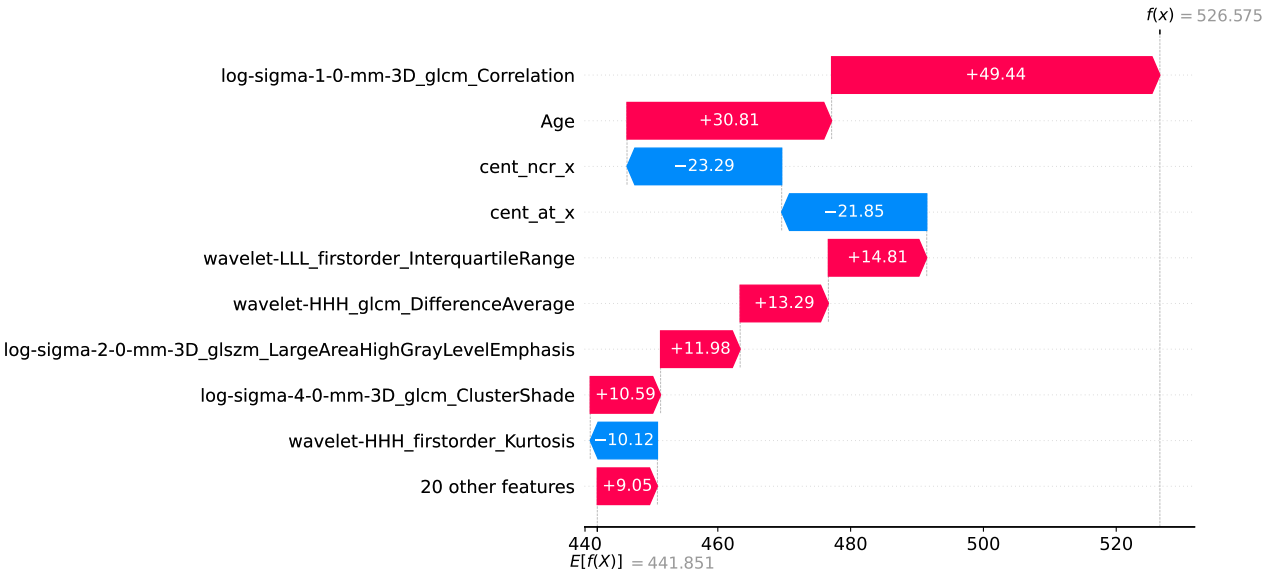


Figure 6. The SHAP waterfall plot for the initial ten features illustrates the behavior of a sample from the validation set, a subset of the training set. Here, $E[f(X)]$ denotes the expected (mean) value, $f(x)$ denotes the predicted value, and the features indicated by the arrow's direction demonstrate their contribution to either increasing or decreasing SD.

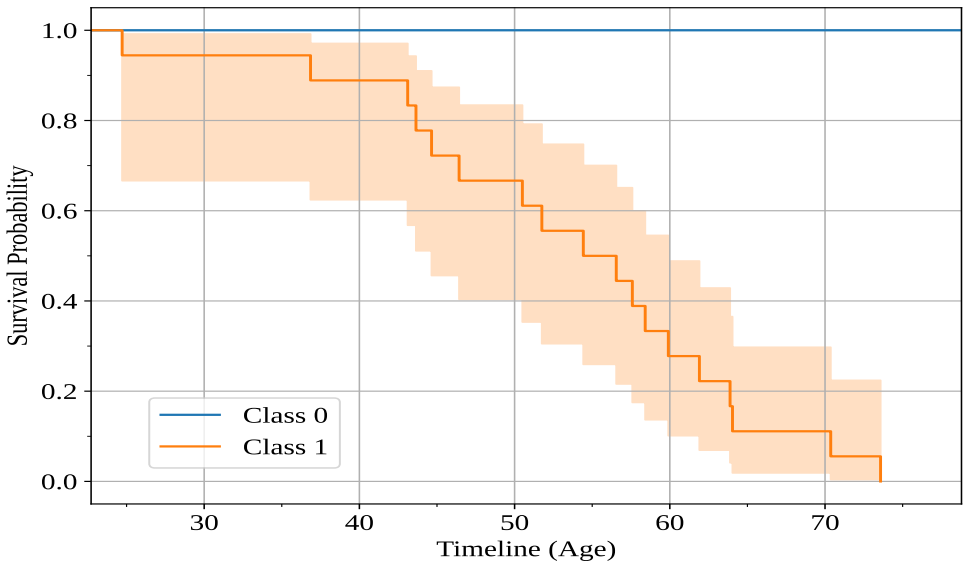
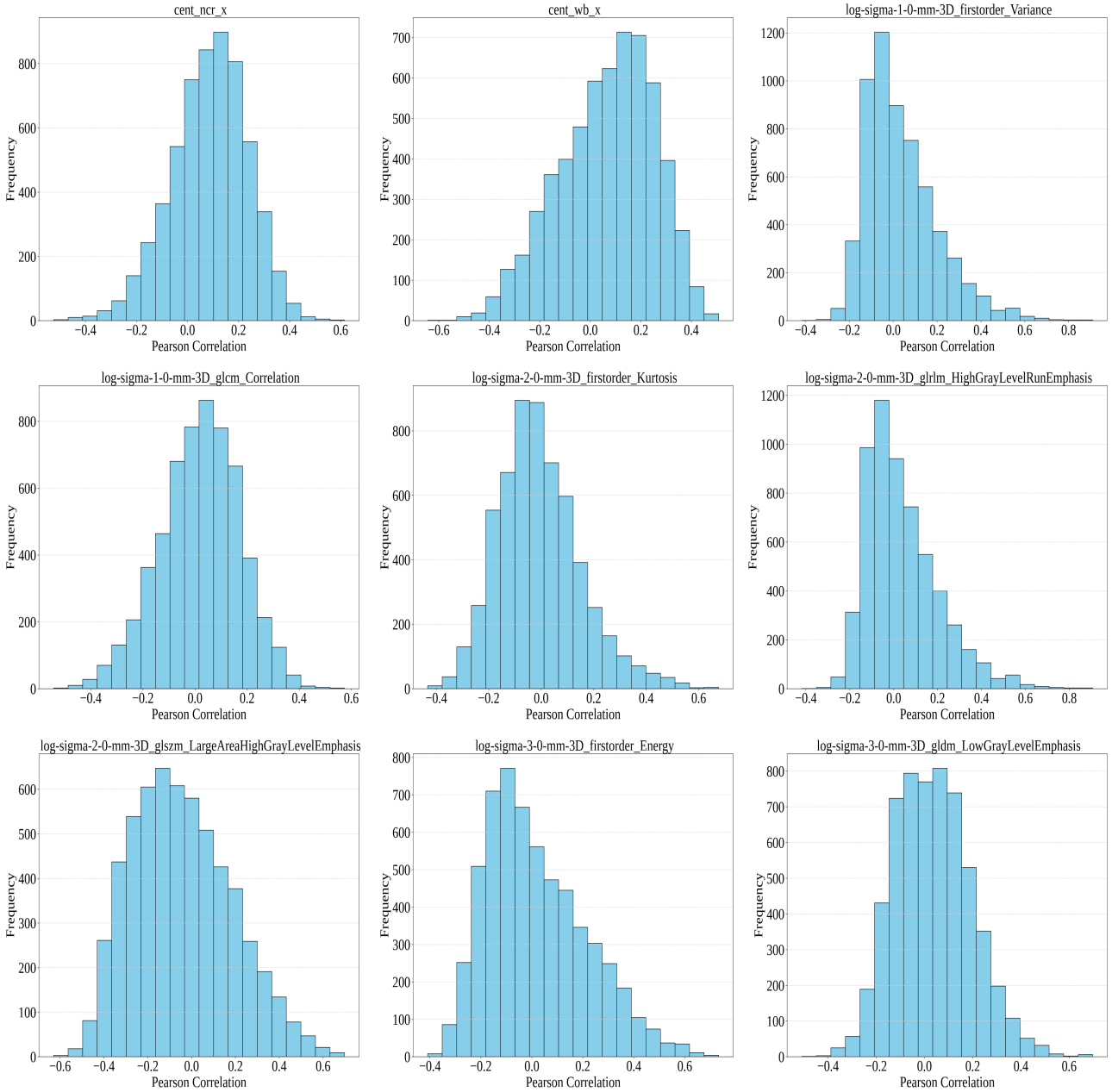
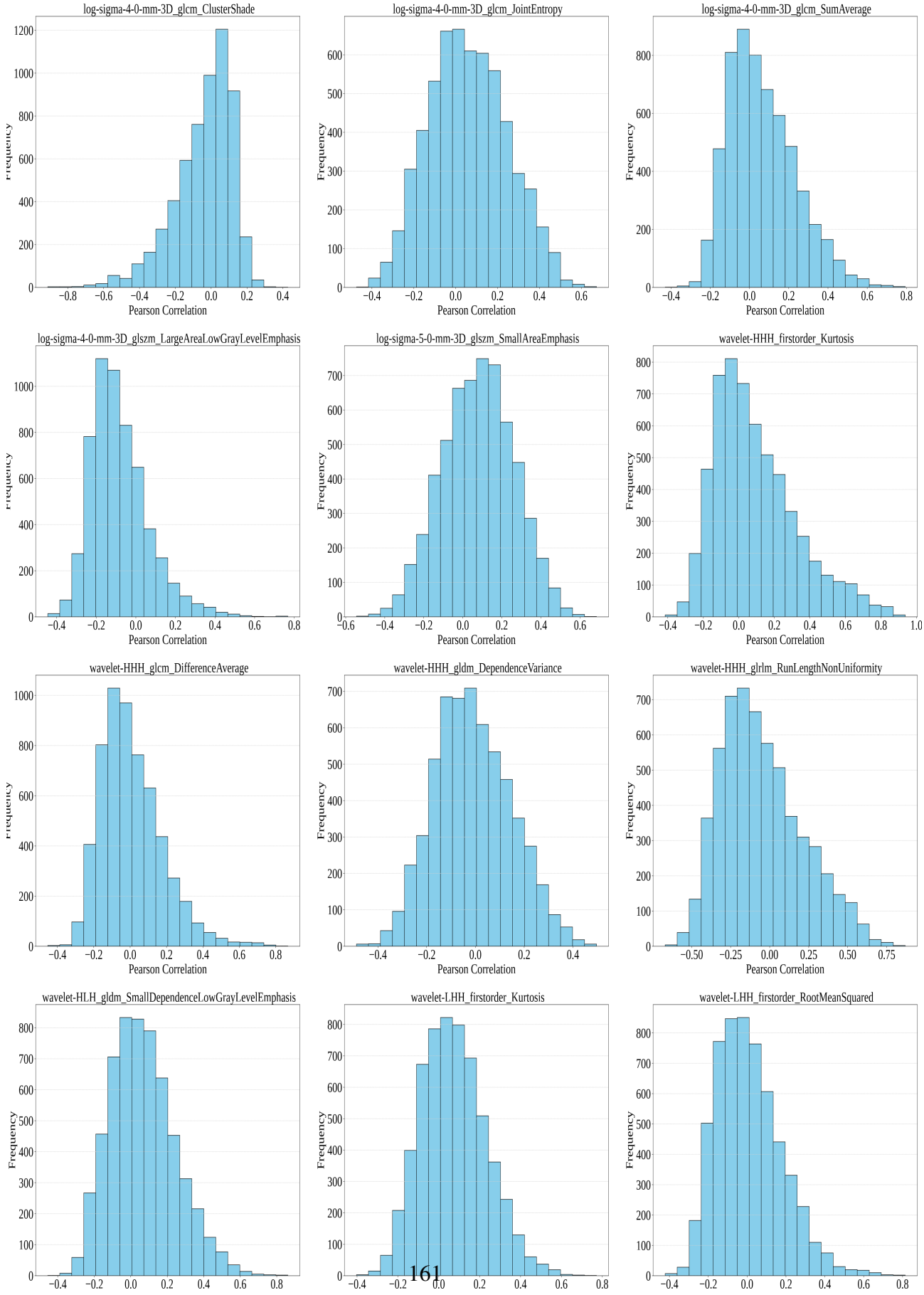


Figure 7. Survival analysis using the KM method for treatment information (clinical information), including both radiation therapy (Class 0) and pharmaceutical therapy (Class 1). The X-axis represents age, and the Y-axis displays the survival probability.


cent_ncr_x
cent_wb_x
log-sigma-1-0-mm-3D_firstorder_Variance
log-sigma-1-0-mm-3D_glcm_Correlation
log-sigma-2-0-mm-3D_firstorder_Kurtosis
log-sigma-2-0-mm-3D_glrlm_HighGrayLevelRunEmphasis
log-sigma-2-0-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis
log-sigma-3-0-mm-3D_firstorder_Energy
log-sigma-3-0-mm-3D_gldm_LowGrayLevelEmphasis
Frequency
Pearson Correlation


log-sigma-4-0-mm-3D_glcm_ClusterShade
log-sigma-4-0-mm-3D_glcm_JointEntropy
log-sigma-4-0-mm-3D_glcm_SumAverage
log-sigma-4-0-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis
log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis
wavelet-HHH_firstorder_Kurtosis
wavelet-HHH_glcm_DifferenceAverage
wavelet-HHH_gldm_DependenceVariance
wavelet-HHH_glrlm_RunLengthNonUniformity
wavelet-HLH_gldm_SmallDependenceLowGrayLevelEmphasis
wavelet-LHH_firstorder_Kurtosis
wavelet-LHH_firstorder_RootMeanSquared
Frequency
Pearson Correlation

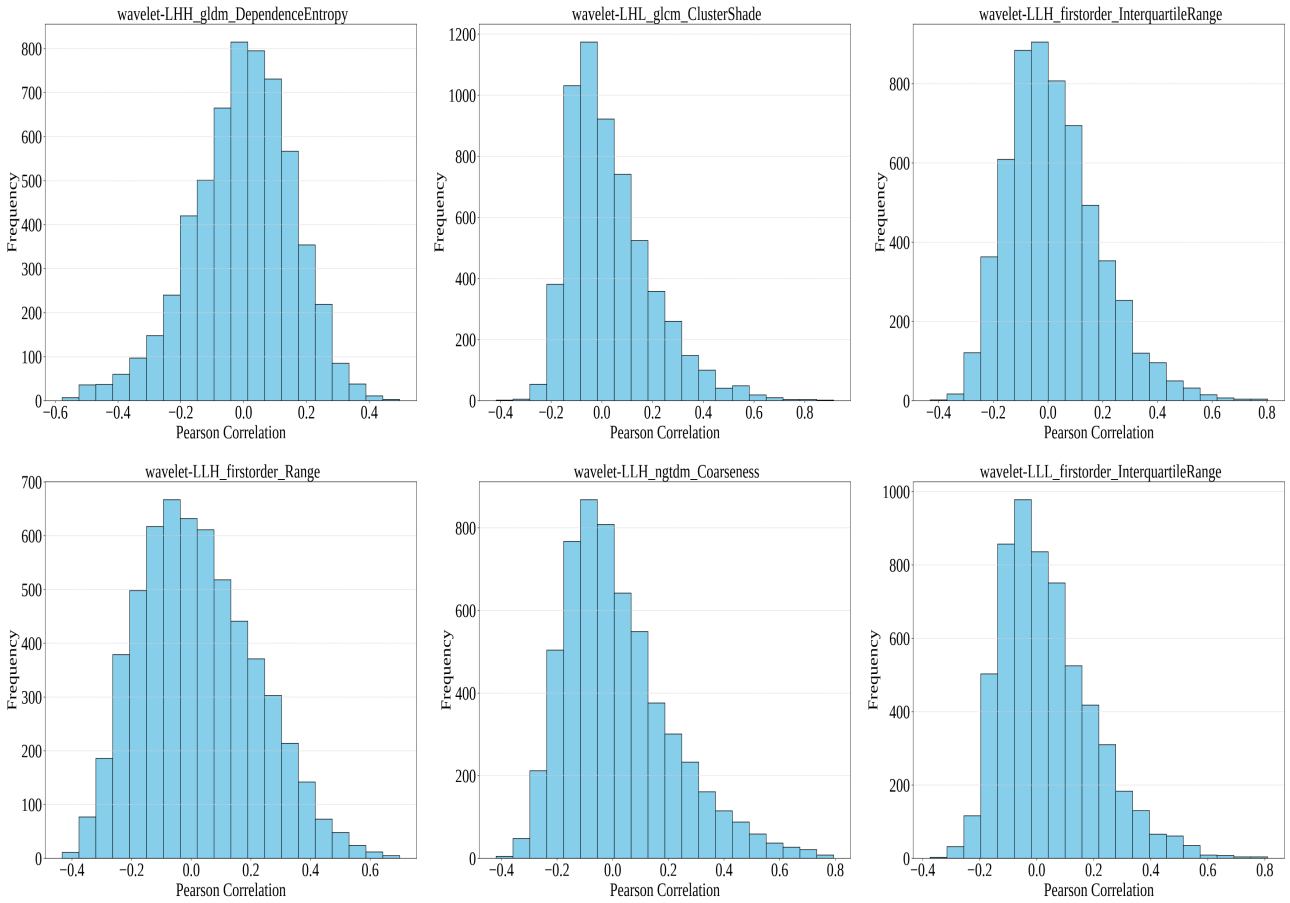
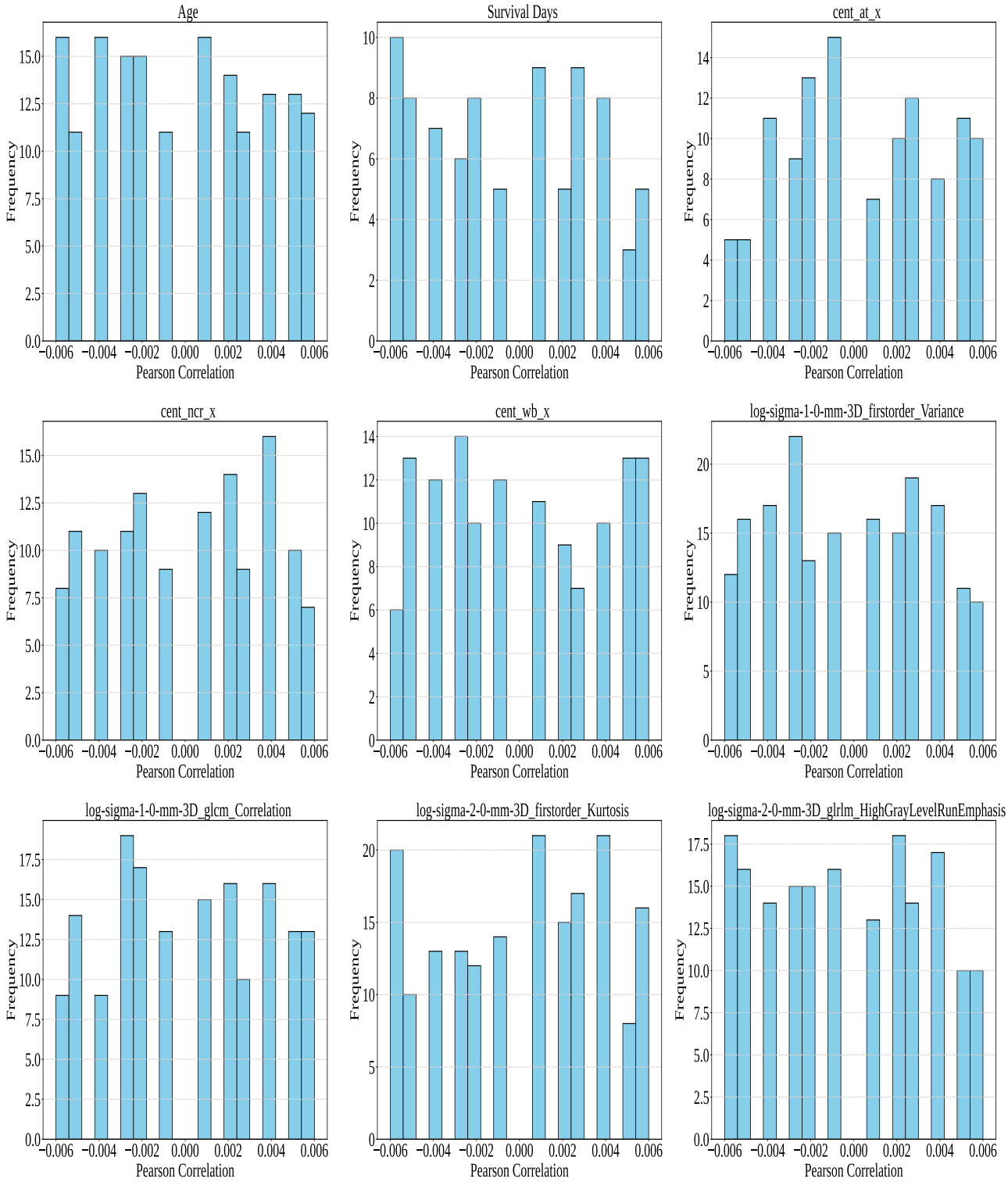
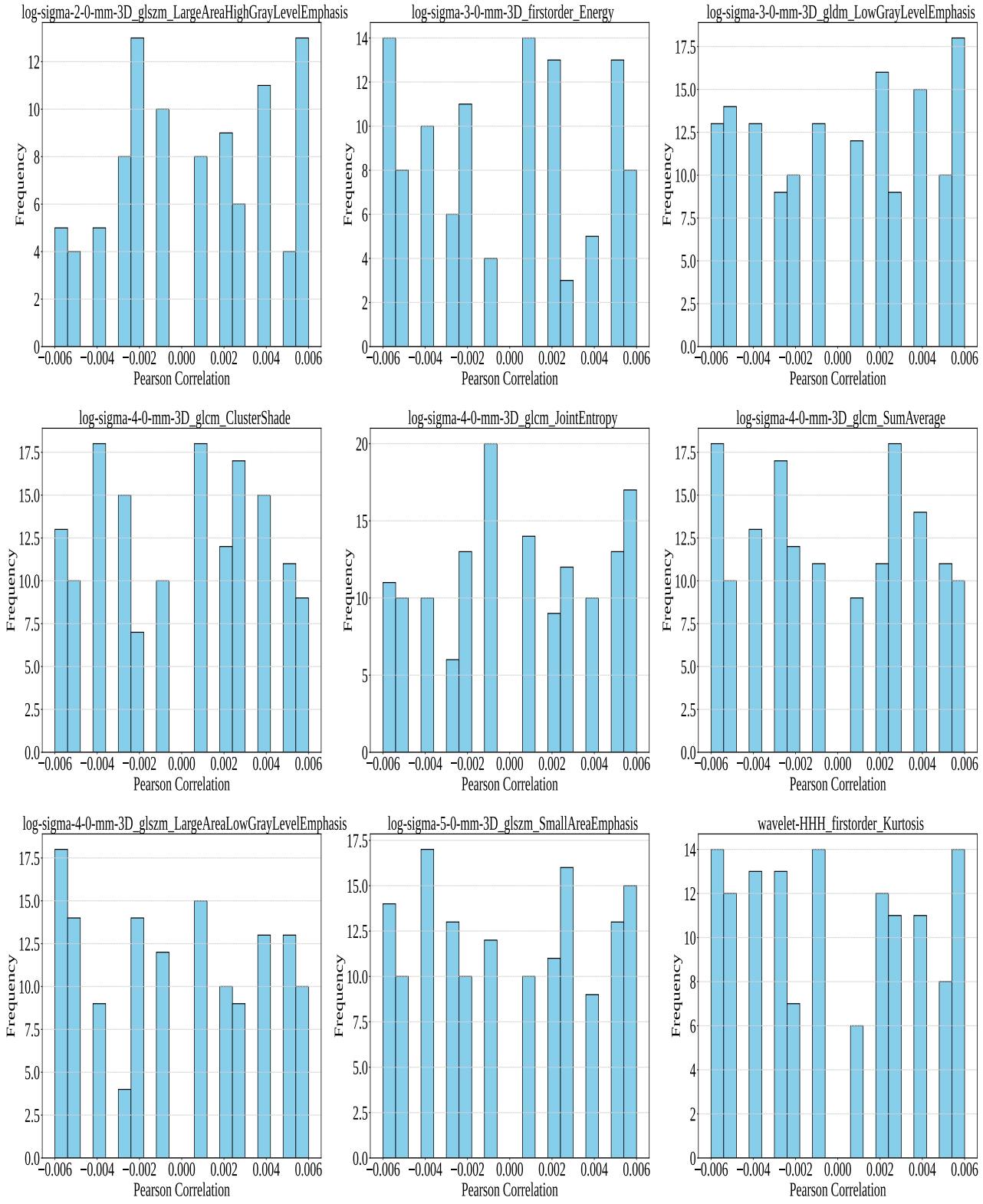


Figure 8. The correlation histogram illustrates the distribution of genes across various extracted feature categories, with their corresponding PC values.


Age
Survival Days
cent_at_x
cent_ncr_x
cent_wb_x
log-sigma-1-0-mm-3D_firstorder_Variance
log-sigma-1-0-mm-3D_glcm_Correlation
log-sigma-2-0-mm-3D_firstorder_Kurtosis
log-sigma-2-0-mm-3D_glrlm_HighGrayLevelRunEmphasis
Frequency
Pearson Correlation


log-sigma-2-0-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis
log-sigma-3-0-mm-3D_firstorder_Energy
log-sigma-3-0-mm-3D_gldm_LowGrayLevelEmphasis
log-sigma-4-0-mm-3D_glcm_ClusterShade
log-sigma-4-0-mm-3D_glcm_JointEntropy
log-sigma-4-0-mm-3D_glcm_SumAverage
log-sigma-4-0-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis
log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis
wavelet-HHH_firstorder_Kurtosis
Frequency
Pearson Correlation
−0.006 −0.004 −0.002 0.000 0.002 0.004 0.006

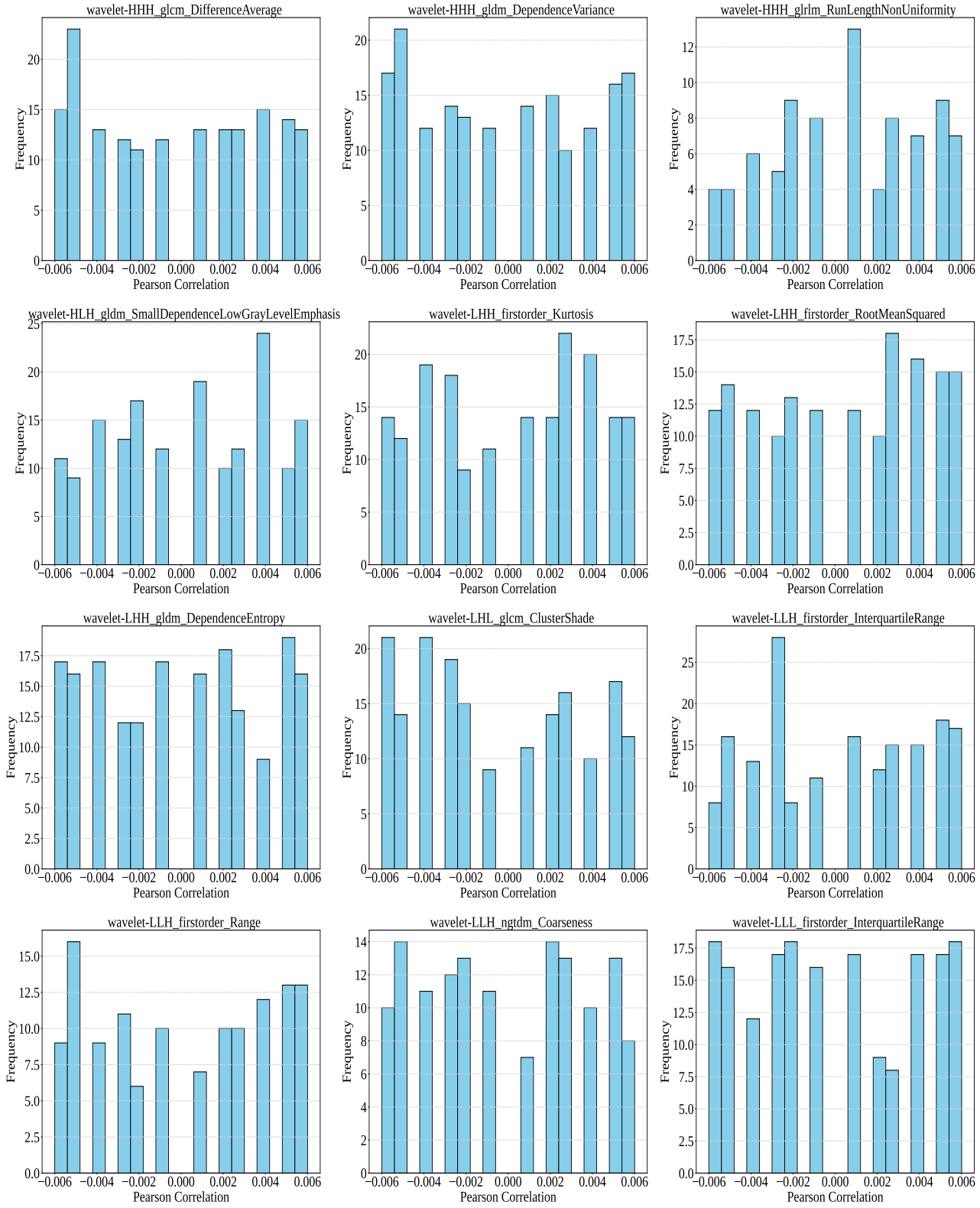
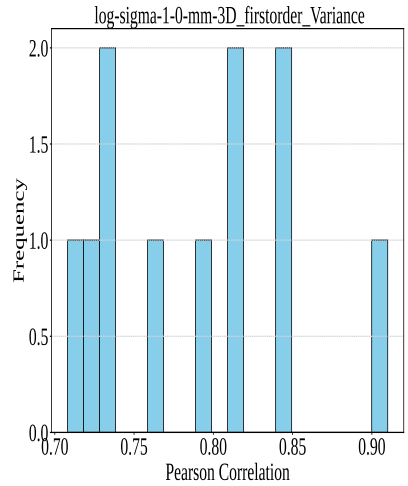
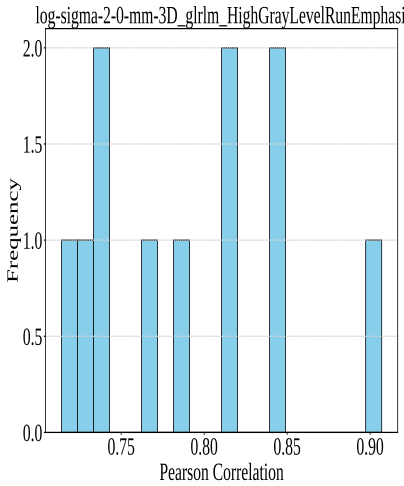
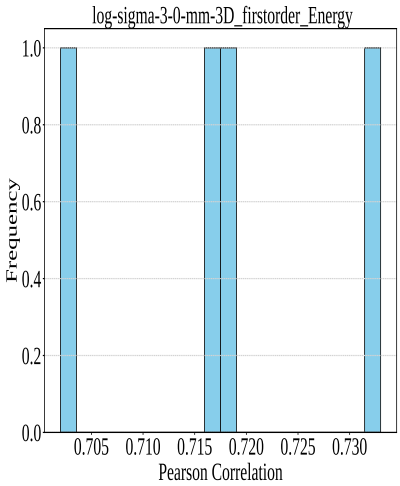
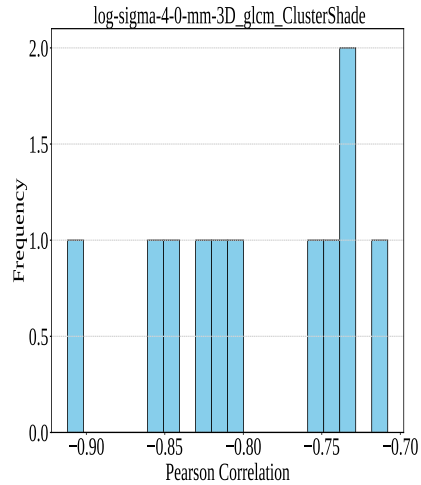


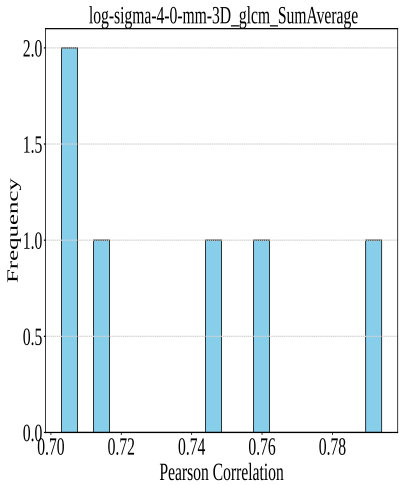
Figure 9. Distribution of genes and radiomics, demonstrating uncorrelated behavior between them after thresholding at $<= 0.006$.

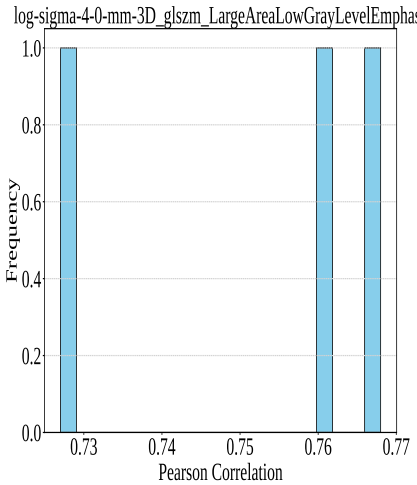

log-sigma-1-0-mm-3D_firstorder_Variance
Frequency
Pearson Correlation

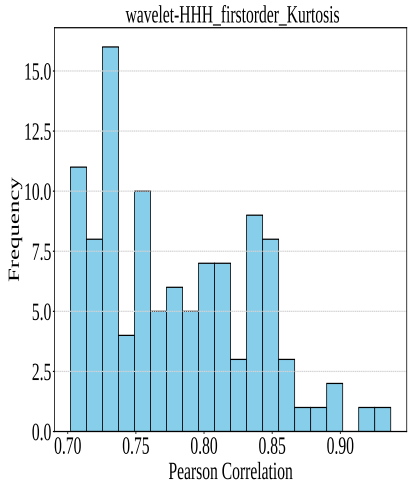

log-sigma-2-0-mm-3D_glrlm_HighGrayLevelRunEmphasis
Frequency
Pearson Correlation

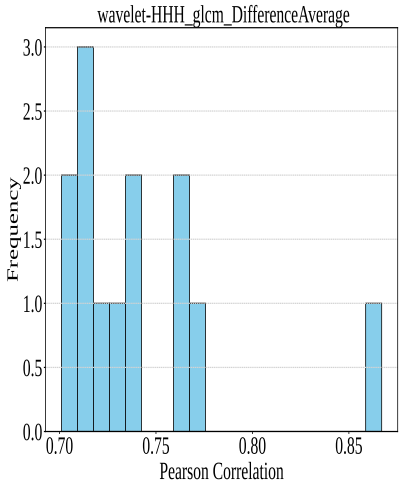

log-sigma-3-0-mm-3D_firstorder_Energy
Frequency
Pearson Correlation

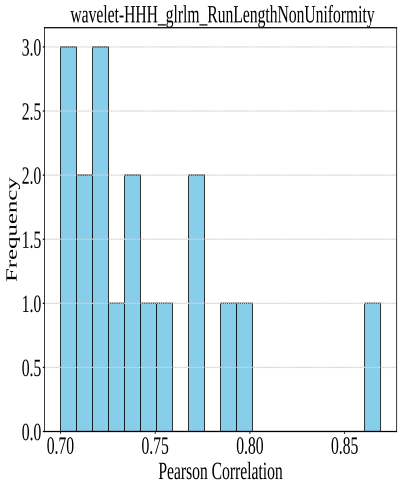

log-sigma-4-0-mm-3D_glcm_ClusterShade
Frequency
Pearson Correlation


log-sigma-4-0-mm-3D_glcm_SumAverage
Frequency
Pearson Correlation


log-sigma-4-0-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis
Frequency
Pearson Correlation


wavelet-HHH_firstorder_Kurtosis
Frequency
Pearson Correlation


wavelet-HHH_glcm_DifferenceAverage
Frequency
Pearson Correlation


wavelet-HHH_glrlm_RunLengthNonUniformity
Frequency
Pearson Correlation

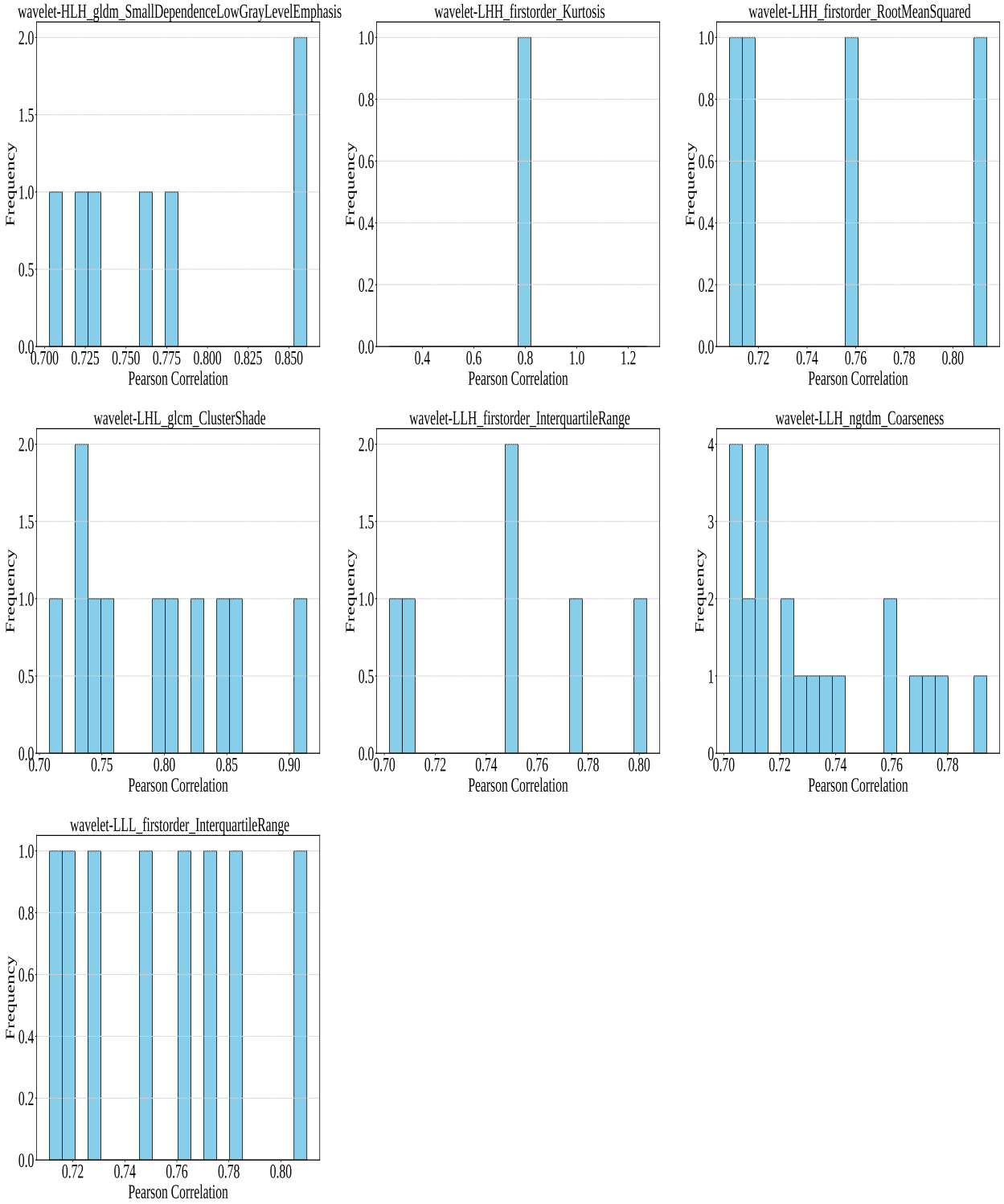
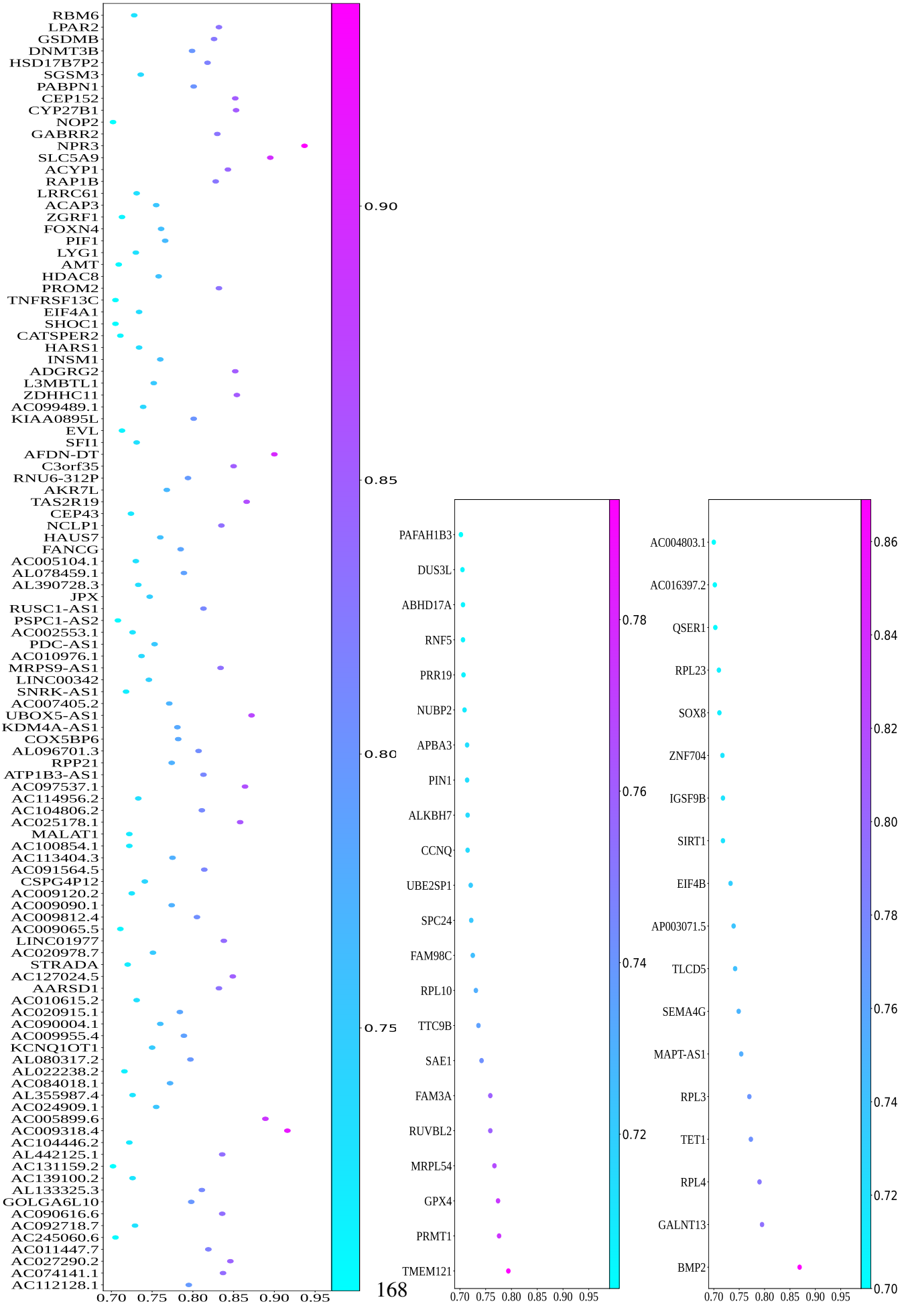


Figure 10. Distribution of genes and radiomics, demonstrating highly correlated behavior between them after thresholding at $>= 0.7$.



(ad) (ae) (af)

Figure 11. Correlations between genes and top three feature classes: (a) FOK, (b) Coarse, and (c) RLNU obtained from Figure 4.16 of the Chapter 4. The top three feature classes were obtained after thresholding at 0.7 where maximum genes fall under these categories.

.1.2 Supplementary Tables:

TABLE 1
Details of the dominant features identified through RFE.

Features with their descriptions
Age: The dataset includes information regarding age.
cent-at-y : The centroid of the enhancing tumor along the y-axis
pos-ed-wb-x: The centroid of the enhancing tumor with respect to the brain centroid along the x-axis
original-shape-LeastAxisLength: It computes the smallest axis length of the ROIs.
Wavelet-LLH-firstorder-Maximum: It assesses the maximum gray intensity within the ROI after applying the Wavelet LLH-band filter.
Wavelet-LLH-Gldm-DependenceVariance: It computes the variance in the image's dependence matrix after applying Wavelet LLH-band filter.
Wavelet-LLH-Glrlm-LongRunLowGrayLevel-Emphasis: It quantifies the distribution of runs of low gray-level intensity values after the application of the Wavelet-LLH filter.
Wavelet-LHL-Glcm-Correlation: It evaluates the correlation between gray-level values and their associated voxels in the gray-level co-occurrence matrix after applying the Wavelet LHL filter.

Wavelet-LHH-Gldm-DependenceNonUniformityNormalized: It measures the variability in the distribution of dependencies among voxel pairs with different gray-level values in an image, after applying Wavelet-LHH band filter.

Wavelet-LHH-Gldm-SmallDependenceHighGrayLevelEmphasis: It highlights small dependence and high gray-level intensity within the image's spatial relationships following the application of the Wavelet-LHH filter.

Wavelet-LHH-Glszm-ZoneEntropy: It measures the variability in the distribution of connected regions of voxels with the same gray-level value in the image, after the application of the Wavelet-LHH band filter.

Wavelet-HLL-Glcm-Imc1: computes the Informational Measure of Correlation 1 (IMC1) from the GLCM, which evaluates the level of information shared among pairs of pixels with varying gray levels, offering insights into both the strength and direction of their correlation within the image, after the application of Wavelet-HLL filter.

Wavelet-HLH-firstorder-Kurtosis: It measures the peakedness of the spread of pixel intensities in the given image after the application Wavelet-HLH filter.

Wavelet-HLH-Gldm-DependenceEntropy: It assesses randomness in the dependencies of an image after the application Wavelet HLH band filter.

Wavelet-HLH-Gldm-SmallDependenceLowGrayLevel-Emphasis: It assesses the combined distribution of small dependence with higher gray-level values after applying Wavelet-HLH filter.

Wavelet-HHH-Glcm-MaximumProbability: It finds the most frequently occurring neighboring pair of intensity values from the grey-level co-occurrence matrix after the application of the Wavelet-HHH filter.

Wavelet-LLL-Glcm-Correlation: It assesses the association between pairs and their corresponding voxel intensity value after applying the Wavelet-LLL filter.

LoG-sigma-1-0-mm-3D-Glcm-Correlation: It evaluates the correlation between pairs and their respective voxel intensity values after applying the LoG filter with a sigma value of 1.

Wavelet-LLH-Ngtdm-Strength: It measures the local variation in gray-level transitions within the neighborhood of the image after applying a Wavelet filter using the LLH band.

LoG-sigma-5-0-mm-3D-Glrlm-RunLengthNonUniformity-Normalized: It assesses the variability in the gray level run lengths in the image, indicating the diversity of consecutive voxels with the same gray-level value and their lengths in different directions within the image after the application of LoG filter with sigma value 5.

LoG-sigma-3-0-mm-3D-Glrlm-RunVariance: It measures variance in the gray-level run lengths in the image, after applying LoG filter with sigma value 3.

LoG-sigma-2-0-mm-3D-Glcm-ClusterShade: It measures the skewness of the GLCM, indicating the asymmetry in the distribution of co-occurring gray-level values within the image, after the application of the LoG filter with sigma value 2.

LoG-sigma-5-0-mm3D-firstorder-TotalEnergy: It quantifies the overall magnitude of image intensity values, after the application of the LoG filter with sigma value 5.

LoG-sigma-3-0-mm-3D-Glcm-MaximumProbability: It indicates the highest probability value within the Gray-Level Co-occurrence Matrix, reflecting the most frequently occurring co-occurring gray-level value pair in the image, after the application of LoG filter with sigma value 5.

LoG-sigma-2-0-mm-3D-firstorder-90Percentile: It evaluates the 90th percentile of intensity values within an image following the application of a LoG filter with a sigma value of 2

LoG-sigma-2-0-mm-3D-firstorder-Skewness: It computes the asymmetry of intensity value distribution relative to the mean intensity value after applying a LoG filter with a sigma value of 2.

LoG-sigma-1-0-mm-3D-Glcm-MCC: It evaluates the texture complexity in the co-occurrence matrix of an image after the application of the LoG filter with sigma value 1.

Wavelet-LLL-Glszm-SmallAreaEmphasis: It evaluates the number of connected voxels with identical gray-level intensity values or the dispersion of smaller-sized zones following the application of the Wavelet-LLL filter.

Wavelet-HLL-Glcm-MCC: It evaluates the texture complexity in the co-occurrence matrix of an image following the application of the Wavelet-HLL filter.

TABLE 2
The prominent features identified leveraging PI and their corresponding weights are provided. A cutoff value of 100 is applied to the weights.

Weight	**Features**
1309.94	***Age*** : The dataset includes information regarding age.

0761.33	***LoG-sigma-1-0-mm-3D-glcm-Correlation***: It evaluates the correlation between pairs and their respective voxel intensity values after the application of LoG filter with sigma value 1.
0722.95	***Wavelet-HHH-Gldm-DependenceVariance***: It computes variance in the image's dependence matrix after applying the Wavelet-HHH filter.
0678.10	***LoG-sigma-4-0-mm-3D-Glcm-JointEntropy***: It computes the randomness in neighborhood intensity values after the application of LoG filter with sigma value 4.
0669.58	***LoG-sigma-2-0-mm-3D-firstorder-Kurtosis***: It computes the peakedness of the intensity distribution of a given image after the application of LoG filter with sigma value 2.
0558.74	***LoG-sigma-2-0-mm-3D-Glrlm-HighGrayLevelRunEmphasis***: It measures the distribution of the higher gray-level values in the image, after the application of the LoG filter with sigma value 2 .
0555.77	***Wavelet-HLH-Gldm-SmallDependenceLowGrayLevelEmphasis***: It measures the spread of small-dependence with lower gray-level values after applying Wavelet HLH band filter.
0509.37	***LoG-sigma-3-0-mm-3D-Gldm-LowGrayLevelEmphasis***: It calculates the dispersion of low gray-level values in the image.
0476.60	***cent-ncr-x***: centroid of necrosis along the X-axis.
0464.70	***Wavelet-LLL-firstorderInterquartileRange***: It calculates the intensity values within the range from the 75^{th} to the 25^{th} percentile of the image, following the application of the Wavelet-LLL filter.

0444.84	***LoG-sigma-4-0-mm-3D-Glcm-ClusterShade***: It measures the skewness of the GLCM, indicating the asymmetry in the distribution of co-occurring gray-level values within the image, after applying LoG filter with sigma value 4.
0438.99	***Wavelet-LHH-firstorder-RootMeanSquared***: It computes the root-mean-square of the image intensity value after applying the Wavelet-LHH filter.
0420.35	***LoG-sigma-4-0-mm-3D-Glcm-SumAverage***: This measure offers understanding regarding the average gray-level value of pixel pairs that co-occur in the image, after applying LoG filter with sigma value 4.
0406.08	***Wavelet-HHH-Glrlm-RunLengthNonUniformity***: It assesses the variability in the gray level run lengths in the image, indicating the diversity of consecutive voxels with the same gray-level value and their lengths in different directions within the image, after the application of the Wavelet-HHH filter.
0395.63	***LoG-sigma-5-0-mm-3D-Glszm-SmallAreaEmphasis***: It evaluates the number of connected voxels with identical gray-level intensity values or the dispersion of smaller-sized zones, after applying LoG filter with sigma value 5.
0357.96	***Wavelet-LLH-Ngtdm-Coarseness***: It assesses the rate of change in the intensity value between neighboring voxels after applying the Wavelet LLH filter.

0357.51	***Wavelet-LLH-firstorder-InterquartileRange***: It calculates the intensity values within the range from the 75^{th} to the 25^{th} percentile of the image, after the application of the Wavelet-LLH filter.
0340.11	***cent-at-x***: centroid of active tumor along the X-axis.
0314.18	***LoG-sigma-4-0-mm-3D-GlszmLargeAreaLowGrayLevelEmphasis***.
0282.22	***Wavelet-HHH-firstorder-Kurtosis***: It assesses the peakedness of the spread of the image's intensity values after applying the Wavelet-HHH filter.
0247.80	***Wavelet-HHH-Glcm-DifferenceAverage***: It evaluates the concurrent relationships between occurrences of pairings with similar intensity values and those with differing intensity values following the application of the Wavelet-HHH filter.
0247.36	***cent-wb-x***: centroid of whole-tumor brain along the X-axis.
0232.56	***LoG-sigma-3-0-mm-3D-firstorder-Energy***: It assesses the magnitude of voxel values in an image after applying LoG filter with sigma value 3.
0229.91	***LoG-sigma-1-0-mm-3D-firstorder-Variance***: It quantifies the spread of intensity values around the mean intensity value following the application of a LoG filter with a sigma value of 1
0226.71	***Wavelet-LHH-firstorder-Kurtosis***: It quantifies the peakedness of the image's intensity distribution after applying the Wavelet-LHH filter.
0217.80	***LoG-sigma-2-0-mm-3D-Glszm-LargeAreaHighGrayLevel-Emphasis***: It quantifies the combined spread of larger size-zones and higher gray level intensity, after the application of the LoG filter with sigma value 1.

0183.47	***Wavelet-LLH-firstorder-Range***: It assesses the distribution of gray-level intensity values of an image.
0131.10	***Wavelet-LHH-Gldm-DependenceEntropy***: It assesses randomness in the dependencies of an image after applying Wavelet LHH band filter.
0118.90	***Wavelet-LHL-Glcm-ClusterShade***: It measures the skewness of the GLCM, indicating the asymmetry in the distribution of co-occurring gray-level values within the image after the application of the Wavelet LHL band filter.

TABLE 3
Annotation of features in the correlation matrix.

Serial No.	Name of feature	Feature Type
1	***Age***	***Meta-Data***
2	***cent-at-x***	***Image***
3	***cent-ncr-x***	***Image***
4	***cent-wb-x***	***Image***
5	***LoG-sigma-1-0-mm-3D-FirstorderVariance***	***Radiomics***
6	***LoG-sigma-1-0-mm-3D-Glcm-Correlation***	***Radiomics***
7	***LoG-sigma-2-0-mm-3D-FirstorderKurtosis***	***Radiomics***
8	***LoG-sigma-2-0-mm-3D-Glrlm-HighGrayLevelRunEmphasis***	***Radiomics***
9	***LoG-sigma-2-0-mm-3D-Glszm-LargeAreaHighGrayLevelEmphasis***	***Radiomics***
10	***LoG-sigma-3-0-mm-3D-Firstorder-Energy***	***Radiomics***
11	***LoG-sigma-3-0-mm-3D-Gldm-LowGrayLevelEmphasis***	***Radiomics***
12	***LoG-sigma-4-0-mm-3D-Glcm-ClusterShade***	***Radiomics***
13	***LoG-sigma-4-0-mm-3D-Glcm-JointEntropy***	***Radiomics***
14	***LoG-sigma-4-0-mm-3D-Glcm-SumAverage***	***Radiomics***
15	***LoG-sigma-4-0-mm-3D-Glszm-LargeAreaLowGrayLevelEmphasis***	***Radiomics***
16	***LoG-sigma-5-0-mm-3D-Glszm-SmallAreaEmphasis***	***Radiomics***
17	***Wavelet-HHH-Firstorder-Kurtosis***	***Radiomics***
18	***Wavelet-HHH-Glcm-DifferenceAverage***	***Radiomics***
19	***Wavelet-HHH-Gldm-DependenceVariance***	***Radiomics***
20	***Wavelet-HHH-Glrlm-RunLengthNonUniformity***	***Radiomics***
21	***Wavelet-HLH-Gldm-SmallDependence-LowGrayLevelEmphasis***	***Radiomics***
22	***Wavelet-LHH-Firstorder-Kurtosis***	***Radiomics***
23	***Wavelet-LHH-Firstorder-RootMeanSquared***	***Radiomics***
24	***Wavelet-LHH-Gldm-DependenceEntropy***	***Radiomics***
25	***Wavelet-LHL-Glcm-ClusterShade***	***Radiomics***
26	***Wavelet-LLH-Firstorder-InterquartileRange***	***Radiomics***
27	***Wavelet-LLH-Firstorder-Range***	***Radiomics***
28	***Wavelet-LLH-Ngtdm-Coarseness***	***Radiomics***
29	***Wavelet-LLL-Firstorder-InterquartileRange***	***Radiomics***

TABLE 4
An illustration of computing SHAP values. Consider a feature set (F) = {X, Y, Z}, and feature's contribution/values are: c{X} = 8, c{Y} = 10, c{Z} = 9, c{X, Y} = 18, c{X, Z} = 20, c{Y, Z} = 22 and c{X, Y, Z} = 25.

Possible combinations of feature	**Marginal Combination**		
	Feature **X**	Feature Y	Feature Z
{**X**, Y, Z}	c{X} -ϕ =8	c{X,Y}-c{X} =10	c{X,Y,Z} -c{X,Y} =7
{**X**, Z, Y}	c{X} -ϕ =8	c{X,Y,Z} -c{X,Z}=5	c{X,Z} -c{X}=12
{Z, Y, **X**}	c{X,Y,Z} -c{Z,Y} = 25-22 = 3	c{Z,Y} -c{Y} =12	c{Z} -ϕ=9
{Y, **X**, Z}	c{X,Y} -c{Y} = 8	c{Y} -ϕ =10	c{X,Y,Z} -c{X,Y} =7
{Z, **X**, Y}	c{X,Z} -c{Z} = 11	c{X,Y,Z} -c{X,Z} =5	c{Z} - ϕ =9
{Y, Z, **X**}	c{X,Y,Z} -c{Y,Z}=3	c{Y} - ϕ= 10	c{Y,Z} -c{Y}=12
SHAP value	$(8+8+8+10+11+3) \mid 6 =$ **6.833**	$(10+5+12+10+5+10) \mid 6 =$ **8.667**	$(7+12+9+7+9+12) \mid 6 =$ **9.334**

TABLE 5
Parameters details of SD predictor models.

(a) Training with IMA ground truth GTR samples

Regression Models	Parameter details
RFR	max depth: 6, max features: log2, min samples leaf: 4, estimators: 45, random state: 0
ExtraRF	max depth: 7, max features: None, min samples leaf': 8, estimators: 50, random state:1
XGBR	alpha: 3, eta: 0.1, lambda: 3, max delta step': 0, 'max depth': 3, 'max leaves': 4, 'min child weight': 12

(b) Training with all resection status

Regression Models	Parameter details
RFR	max depth: 6, max features: None, min samples leaf: 13, estimators: 70, random state: 42
ExtraRF	max depth: 8, max features: None, min samples leaf: 7, estimators: 60, random state: 2
XGBR	alpha: 1, eta: 0.05, lambda: 5, max delta step: 0, max depth: 3, max leaves: 0, min child weight: 3

(c) TGCA samples having clinical information

Regression Models	Parameter details
RFR	max depth: 5, max features: sqrt, min samples leaf: 2, estimators: 15, random state: 0
ExtraRF	max depth: 3, max features: sqrt, min samples leaf: 2, estimators: 10, random state: 5
XGBR	alpha: 1, eta: 0.001, lambda: 1, max delta step: 0, max depth: 6, max leaves: 15, min child weight: 1

(d) TGCA samples without clinical information

Regression Models	Parameter details
RFR	max depth: 3, max features: sqrt, min samples leaf: 2, estimators: 15, random state: 0
ExtraRF	max depth: 2, max features: None, min samples leaf: 7, estimators: 15, random state: 5
XGBR	alpha: 5, eta: 0.05, lambda: 5, max delta step: 0, max depth: 3, max leaves: 0, min child weight: 3

(e) Binary Classification of Survival Days

Classifier	Parameter details
RF	criterion: gini, max depth: 2, max features: log2, min samples leaf: 5, estimators: 25, random state: 1
ExtraRF	criterion: gini, max depth: 2, max features: log2, min samples leaf: 2, estimators: 30, random state=1
XGBR	alpha: 1, eta: 0.001, lambda: 1, max delta step: 0, max depth: 2, max leaves: 0, min child weight: 1
Logistic Regressor	C:1, penalty: l2, tol: 0.001, solver: liblinear

List of Publications

Journal Publications

1. Snehal Rajput, Rupal kapdi, Mehul S. Raval, and Mohendra Roy, "Interpretable machine learning model to predict survival days of malignant brain tumor patients" **Published** in Journal of Machine Learning: Science and Technology, May 2023, (SCI Indexed, IF= 6.8)

   DOI : `https://iopscience.iop.org/article/10.1088/2632-2153/acd5a9` [Chapter 3, Chapter 4].

   No. of citation received: 11

2. Snehal Rajput, Rupal kapdi, Mehul S. Raval, and Mohendra Roy, "Multi-View Brain Tumor Segmentation (MVBTS): An Ensemble of Planar and Triplanar Attention UNets " **Published** in Turkish journal of electrical engineering & computer sciences, 2023, (SCI Indexed, IF= 1.1), Vol. 31: No. 6, Article 2.

   DOI : `https://journals.tubitak.gov.tr/elektrik/vol31/iss6/2/` [Chapter 3].

   No. of citation received: 4

3. Snehal Rajput, Rupal Kapdi, Mehul S. Raval, and Mohendra Roy, "A Triplanar Ensemble Model for Brain Tumor Segmentation with Volumetric Multimodal Magnetic Resonance Images " **Published** in Healthcare Analytics Journal, June 2024, (Scopus Indexed), Vol. 5, Pages: 100307

   DOI : `https://doi.org/10.1016/j.health.2024.100307` [Chapter 3].

   No. of citation received: 3

4. Snehal Rajput, Rupal kapdi, Mehul S. Raval, Mohendra Roy, and Jayendra M. Bhalodiya, "Deriving and interpreting robust features for survival prediction of brain tumor patients " **Published** in International Journal of Imaging Systems and Technology (IMA) (SCI Indexed, IF= 3.3), Vol. 34, Issue 3, Pages: e23105

   DOI : `https://doi.org/10.1002/ima.23105` [Chapter 4].

5. Snehal Rajput, and Mehul S. Raval, "A Review on End-To-End Methods for Brain Tumor Segmentation and Overall Survival Prediction " **Published** in Azerbaijan Journal for High-Performance Computing, 2020.

   DOI : `https://doi.org/10.32010/26166127.2020.3.1.119.138` [Chapter 4].

   https://azjhpc.org/index.php/archives/15- paper/58-a-review-on-end-to-end-methods-for-brain-tumor-segmentation-andoverall-survival-prediction

   No. of citation received: 11

Conference Publications

1. Snehal Rajput, Rupal kapdi, Mohendra Roy, and Mehul S. Raval, “Glioblastoma Multiforme Patient Survival Prediction, ”Proceedings of 2021 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021). Lecture Notes in Electrical Engineering (LNEE, volume 784). Springer Nature. [Chapter 1-2-3]

   DOI : https://link.springer.com/chapter/10.1007/978-981-16-3880-0_6

   No. of citation received: 11

2. Snehal Rajput, Rupal kapdi, Mehul S. Raval, and Mohendra Roy,“Survival prediction of malignant brain tumor patients”, Proceedings of 2022 International Conference on Recent and Future Trends in Smart Electronics Systems and Manufacturing,”River Publishers Series in Proceedings. [Chapter 1,2,3]

   DOI : https://doi.org/10.13052/rp-9788770229852.

Book Chapter

1. Rupal kapdi, Snehal Rajput, Mohendra Roy and Mehul S. Raval, “Interpretability and Explainability of Brain Tumor Segmentation Deep Neural Network Models and Overall Survival Prediction Machine Learning Models ”. Book Chapter on XAI, 2022. CRC publication. ISBN: 978-1-003-33342-5

   DOI : 10.1201/9781003333425 No. of citation received: 1

2. Snehal Rajput, Rupal kapdi, Mohendra Roy and Mehul S. Raval, “U-Net: A Versatile Deep Learning Architecture for Multi-Disease Detection. Book Chapter on Biomedical Imaging ”- Advances in Artificial Intelligence and Machine Learning. Springer Nature Publisher. [Accepted - InPress]

Achievements and Awards

1. Best Paper Award in AI & ML category, 1st International Conference on Recent and Future Trends in Smart Electronics System and Manufacturing (SEISMIC) 2022, organized by Symbiosis International University, Pune.

2. Best Presenter Award in International Research Conclave, “Research pe Charcha- 2023”, organized by Kaushalya – The skilled University, an Affiliated State University.

3. Winner - Third position in Poster Presentation event organized by IEEE Signal Processing Society (SPS), Gujarat Chapter (2024).

Bibliography

AANS. (2023). *Brain tumors.* Retrieved from: https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Brain-Tumors. (Accessed: November 8, 2023)

Aboussaleh, I., Riffi, J., Mahraz, A. M., & Tairi, H. (2021). Brain tumor segmentation based on deep learning's feature representation. *Journal of Imaging*, *7*(12), 269.

Adewole, M., Rudie, J. D., Gbdamosi, A., Toyobo, O., Raymond, C., Zhang, D., ... others (2023). The brain tumor segmentation (brats) challenge 2023: Glioma segmentation in sub-saharan africa patient population (brats-africa). *ArXiv*.

Agravat, R. R., & Raval, M. S. (2019). Brain tumor segmentation and survival prediction. In *International miccai brainlesion workshop* (pp. 338–348).

Agravat, R. R., & Raval, M. S. (2021). 3d semantic segmentation of brain tumor for overall survival prediction. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part ii 6* (pp. 215–227).

Ahmadvand, A., Yousefi, S., & Manzuri Shalmani, M. (2017). A novel markov random field model based on region adjacency graph for t1 magnetic resonance imaging brain segmentation. *International Journal of Imaging Systems and Technology*, *27*(1), 78–88.

Alex, V., Safwan, M., & Krishnamurthi, G. (2017). Brain tumor segmentation from multi modal mr images using fully convolutional neural network. In *Medical image computing and computer assisted intervention-miccai* (pp. 1–8).

Ali, M. J., Akram, M. T., Saleem, H., Raza, B., & Shahid, A. R. (2021a). Glioma segmentation using ensemble of 2d/3d u-nets and survival prediction using multiple features fusion. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part ii 6* (pp. 189–199).

Ali, M. J., Akram, M. T., Saleem, H., Raza, B., & Shahid, A. R. (2021b). Glioma segmentation using ensemble of 2d/3d u-nets and survival prediction using multiple features fusion. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries* (pp. 189–199). Springer International Publishing. doi: 10.1007/978-3-030-72087-2_17

Ali, O., Ali, H., Shah, S. A. A., & Shahzad, A. (2022). Implementation of a modified u-net for medical image segmentation on edge devices. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *69*(11), 4593–4597.

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *82*(4), 1059–1086.

ASCO, A. S. o. C. O. (2022, Feb). *Brain tumor: Statistics.* Retrieved from `https://www.cancer.net/cancer-types/brain-tumor/statistics`

Asenjo, J. M., & Solís, A. M.-L. (2021). MRI brain tumor segmentation using a 2d-3d u-net ensemble. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries* (pp. 354–366). Springer International Publishing. doi: 10.1007/978-3-030-72084-1_32

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, *39*(12), 2481–2495.

Bae, S., Choi, Y. S., Ahn, S. S., Chang, J. H., Kang, S.-G., Kim, E. H., … Lee, S.-K. (2018). Radiomic mri phenotyping of glioblastoma: improving survival prediction. *Radiology*, *289*(3), 797–806.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., … Davatzikos, C. (2017). Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, *4*(1), 1–13.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., ... others (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*.

Balafar, M. (2014). Fuzzy c-mean based brain mri segmentation algorithms. *Artificial intelligence review*, *41*, 441–449.

Ballestar, L. M., & Vilaplana, V. (2021). Mri brain tumor segmentation and uncertainty estimation using 3d-unet architectures. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part i 6* (pp. 376–390).

Balogh, A., Reiniger, L., Hetey, S., Kiraly, P., Toth, E., Karaszi, K., ... others (2020). Decreased expression of znf554 in gliomas is associated with the activation of tumor pathways and shorter patient survival. *International Journal of Molecular Sciences*, *21*(16), 5762.

Bao, Z.-S., Zhang, C.-B., Wang, H.-J., Yan, W., Liu, Y.-W., Li, M.-Y., & Zhang, W. (2013). Whole-genome mrna expression profiling identifies functional and prognostic signatures in patients with mesenchymal glioblastoma multiforme. *CNS neuroscience & therapeutics*, *19*(9), 714–720.

Bhargavi, K., & Jyothi, S. (2014). A survey on threshold based segmentation technique in image processing. *International Journal of Innovative Research and Development*, *3*(12), 234–239.

Bhattacharya, M., & Das, A. (2008). A study on seeded region based improved watershed transformation for brain tumor segmentation. *The XXIX General Assembly of the Int Union of Radio Science*.

Bian, E.-B., Zong, G., Xie, Y.-S., Meng, X.-M., Huang, C., Li, J., & Zhao, B. (2014). Tet family proteins: new players in gliomas. *Journal of neuro-oncology*, *116*, 429–435.

Bommineni, V. L. (2021). PieceNet: A redundant UNet ensemble. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries* (pp. 331–341). Springer International Publishing. doi: 10.1007/978-3-030-72087-2_29

BrainLine. (2023). *Brain imaging: What are the different types?* Retrieved from `https://www.brainline.org/slideshow/brain-imaging-what-are-different-types` (Accessed on: April 11, 2024)

Burrell, R. A., McGranahan, N., Bartek, J., & Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, *501*(7467), 338–345.

Cao, F., Xia, X., Fan, Y., Liu, Q., Song, J., Zhang, Q., ... Yao, S. (2021). Knocking down of polo-like kinase 2 inhibits cell proliferation and induced cell apoptosis in human glioma cells. *Life Sciences*, *270*, 119084.

CEA (French Alternative Energies and Atomic Energy Commission). (April 2024). *A world premiere: the living brain imaged with unrivaled clarity thanks to the world's most powerful mri machine.* Retrieved from `https://www.cea.fr/english/Pages/News/world-premiere-living`

-brain-imaged-with-unrivaled-clarity-thanks-to-world-most-powerful-MRI-machine.aspx (Accessed on: April 11, 2024)

Cepeda, S., Pérez-Nuñez, A., García-García, S., García-Pérez, D., Arrese, I., Jiménez-Roldán, L., ... others (2021). Predicting short-term survival after gross total or near total resection in glioblastomas by machine learning-based radiomic analysis of preoperative mri. *Cancers*, *13*(20), 5047.

Chaddad, A., & Tanougast, C. (2016). Extracted magnetic resonance texture features discriminate between phenotypes and are associated with overall survival in glioblastoma multiforme patients. *Medical & biological engineering & computing*, *54*, 1707–1718.

Chang, H.-H., & Valentino, D. J. (2008). An electrostatic deformable model for medical image segmentation. *Computerized Medical Imaging and Graphics*, *32*(1), 22–35.

Chang, P. D., et al. (2016). Fully convolutional neural networks with hyperlocal features for brain tumor segmentation. In *Proceedings miccai-brats workshop* (pp. 4–9).

Che, P., Jiang, S., Zhang, W., Zhu, H., Hu, D., & Wang, D. (2022). A comprehensive gene expression profile analysis of prostate cancer cells resistant to paclitaxel and the potent target to reverse resistance. *Human & Experimental Toxicology*, *41*, 09603271221129854.

Chen, L., Zhang, H., Xiao, J., Nie, L., Shao, J., Liu, W., & Chua, T.-S. (2017). Sca-cnn: Spatial and channel-wise attention in convolutional networks for im-

age captioning. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 5659–5667).

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, *40*(4), 834–848.

Colman, J., Zhang, L., Duan, W., & Ye, X. (2021). Dr-unet104 for multimodal mri brain tumor segmentation. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part ii 6* (pp. 410–419).

Costa, A. R., Duarte, A. C., Costa-Brito, A. R., Gonçalves, I., & Santos, C. R. (2023). Bitter taste signaling in cancer. *Life Sciences*, *315*, 121363.

Der, G., & Everitt, B. S. (2005). *Statistical analysis of medical data using sas*. Chapman and Hall/CRC.

Dong, Z.-Q., Guo, Z.-Y., & Xie, J. (2019). The lncrna egfr-as1 is linked to migration, invasion and apoptosis in glioma cells by targeting mir-133b/rack1. *Biomedicine & Pharmacotherapy*, *118*, 109292.

Dorjsembe, Z., & Xiao, F. (2023). *Synthetic whole-head mri brain tumor segmentation dataset.* IEEE Dataport. Retrieved from `https://dx.doi.org/10.21227/3ej9-e459` doi: 10.21227/3ej9-e459

D'Angelo, F., Ceccarelli, M., Tala, Garofano, L., Zhang, J., Frattini, V., ... others (2019). The molecular landscape of glioma in patients with neurofibromatosis 1. *Nature medicine*, *25*(1), 176–187.

El Khoury, W., & Nasr, Z. (2021). Deregulation of ribosomal proteins in human cancers. *Bioscience Reports*, *41*(12), BSR20211577.

Fernáandez-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research*, *15*(1), 3133–3181.

Ferreira, A., Solak, N., Li, J., Dammann, P., Kleesiek, J., Alves, V., & Egger, J. (2024). How we won brats 2023 adult glioma challenge? just faking it! enhanced synthetic data augmentation and model ensemble for brain tumour segmentation. *arXiv preprint arXiv:2402.17317*.

Fisher, M., & Aldridge, R. (1999). Hierarchical image segmentation using a watershed scale-space tree.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Friedman, J. H., & Popescu, B. E. (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics*, *2*(3), 916–954.

Fyllingen, E. H., Bø, L. E., Reinertsen, I., Jakola, A. S., Sagberg, L. M., Berntsen, E. M., ... Solheim, O. (2021). Survival of glioblastoma in relation to tumor location: a statistical tumor atlas of a population-based cohort. *Acta neurochirurgica*, *163*(7), 1895–1905.

Ganesh, M., Naresh, M., & Arvind, C. (2017). Mri brain image segmentation using enhanced adaptive fuzzy k-means algorithm. *Intelligent Automation & Soft Computing*, *23*(2), 325–330.

Gealy, C. B., & George, A. D. (2022). Evaluation of parameter-scaling for efficient deep learning on small satellites.

Gies, V., & Bernard, T. M. (2004). Statistical solution to watershed oversegmentation. In *2004 international conference on image processing, 2004. icip'04.* (Vol. 3, pp. 1863–1866).

Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: images are more than pictures, they are data. *Radiology*, *278*(2), 563–577.

Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-meier estimate. *International journal of Ayurveda research*, *1*(4), 274.

Gong, X., Liu, L., Xiong, J., Li, X., Xu, J., Xiao, Y., ... others (2021). Construction of a prognostic gene signature associated with immune infiltration in glioma: a comprehensive analysis based on the cgga. *Journal of Oncology*, *2021*.

Gordillo, N., Montseny, E., & Sobrevilla, P. (2013). State of the art survey on mri brain tumor segmentation. *Magnetic resonance imaging*, *31*(8), 1426–1438.

Guo, K., Duan, X., Zhao, J., Sun, B., Liu, X., & Zhao, Z. (2022). A novel necroptosis-related gene signature for predict prognosis of glioma based on single-cell and bulk rna sequencing. *Frontiers in Molecular Biosciences*, *9*, 984712.

Guo, Q., Fan, Y., Wang, Q., Li, B., Qiu, W., Qi, Y., ... others (2023). Glioblastoma upregulates sumoylation of hnrnp a2/b1 to eliminate the tumor suppressor mir-204-3p, accelerating angiogenesis under hypoxia. *Cell Death & Disease*, *14*(2), 147.

Gupta, M., Rajagopalan, V., & Rao, B. P. (2019). Glioma grade classification using wavelet transform-local binary pattern based statistical texture features and geometric measures extracted from mri. *Journal of Experimental & Theoretical Artificial Intelligence*, *31*(1), 57–76.

Han, F., Hu, R., Yang, H., Liu, J., Sui, J., Xiang, X., ... Song, S. (2016). Pten gene mutations correlate to poor prognosis in glioma patients: a meta-analysis. *OncoTargets and therapy*, 3485–3492.

Hanif, F., Muzaffar, K., Perveen, K., Malhi, S. M., & Simjee, S. U. (2017). Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment. *Asian Pacific journal of cancer prevention: APJCP*, *18*(1), 3.

Han Li, C., & Chen, Y. (2015). Small and long non-coding rnas: novel targets in perspective cancer therapy. *Current genomics*, *16*(5), 319–326.

Hao, C., Li, R., Lu, Z., He, K., & Qiu, T. (2023). Predicting prognosis, immunotherapy and distinguishing cold and hot tumors in clear cell renal cell carcinoma based on anoikis-related lncrnas. *Frontiers in Immunology*, *14*, 1145450.

Hashemi, M., Etemad, S., Rezaei, S., Ziaolhagh, S., Rajabi, R., Rahmanian, P., ... others (2023). Progress in targeting pten/pi3k/akt axis in glioblastoma

therapy: Revisiting molecular interactions. *Biomedicine & Pharmacotherapy*, *158*, 114204.

Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., & Xu, D. (2021). Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International miccai brainlesion workshop* (pp. 272–284).

Hausmann, D., Liu, J., Budjan, J., Reichert, M., Ong, M., Meyer, M., ... others (2018). Image quality assessment of 2d versus 3d t2wi and evaluation of ultra-high b-value (b= 2,000 mm/s2) dwi for response assessment in rectal cancer. *Anticancer research*, *38*(2), 969–978.

Havaei, M., Jodoin, P.-M., & Larochelle, H. (2014). Efficient interactive brain tumor segmentation as within-brain knn classification. In *2014 22nd international conference on pattern recognition* (pp. 556–561).

He, H., Shao, X., Li, Y., Gihu, R., Xie, H., Zhou, J., & Yan, H. (2021). Targeting signaling pathway networks in several malignant tumors: progresses and challenges. *Frontiers in Pharmacology*, *12*, 675675.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the ieee international conference on computer vision* (pp. 1026–1034).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Hirsch, G. V., Bauer, C. M., & Merabet, L. B. (2015). Using structural and functional brain imaging to uncover how the brain adapts to blindness. *Annals of neuroscience and psychology*, *2*.

Hoelzinger, D. B., Mariani, L., Weis, J., Woyke, T., Berens, T. J., McDonough, W., ... Berens, M. E. (2005). Gene expression profile of glioblastoma multiforme invasive phenotype points to new therapeutic targets. *Neoplasia*, *7*(1), 7–16.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 7132–7141).

Huang, G., Li, H., & Zhang, H. (2020). Abnormal expression of mitochondrial ribosomal proteins and their encoding genes with cell apoptosis and diseases. *International journal of molecular sciences*, *21*(22), 8879.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, *195*(1), 215–243.

Isensee, F., Jäger, P. F., Full, P. M., Vollmuth, P., & Maier-Hein, K. H. (2021). nnu-net for brain tumor segmentation. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part ii 6* (pp. 118–132).

Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., & Maier-Hein, K. H. (2017). Brain tumor segmentation and radiomics survival prediction: Contribu-

tion to the brats 2017 challenge. In *International miccai brainlesion workshop* (pp. 287–297).

Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., & Maier-Hein, K. H. (2019). No new-net. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 4th international workshop, brainles 2018, held in conjunction with miccai 2018, granada, spain, september 16, 2018, revised selected papers, part ii 4* (pp. 234–244).

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., Lauer, M. S., et al. (2008). Random survival forests. *Annals of Applied Statistics*, *2*(3), 841–860.

Jang, K., Russo, C., & Di Ieva, A. (2020). Radiomics in gliomas: clinical implications of computational modeling and fractal-based analysis. *Neuroradiology*, *62*(7), 771–790.

Ji, J., You, Q., Zhang, J., Wang, Y., Cheng, J., Huang, X., ... others (2021). Downregulation of tet1 promotes glioma cell proliferation and invasion by targeting wnt/β-catenin pathway. *Analytical Cellular Pathology*, *2021*.

Jiang, M.-C., Ni, J.-J., Cui, W.-Y., Wang, B.-Y., & Zhuo, W. (2019). Emerging roles of lncrna in cancer and therapeutic opportunities. *American journal of cancer research*, *9*(7), 1354.

Jin, Y., Angelini, E., & Laine, A. (2005). Wavelets in medical image processing: denoising, segmentation, and registration. *Handbook of Biomedical Image Analysis: Volume I: Segmentation Models Part A*, 305–358.

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., ... Glocker, B. (2017). Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, *36*, 61–78.

Kaus, M. R., Warfield, S. K., Nabavi, A., Black, P. M., Jolesz, F. A., & Kikinis, R. (2001). Automated segmentation of mr images of brain tumors. *Radiology*, *218*(2), 586–591.

Kazerooni, A. F., Khalili, N., Liu, X., Haldar, D., Jiang, Z., Anwar, S. M., ... others (2023). The brain tumor segmentation (brats) challenge 2023: Focus on pediatrics (cbtn-connect-dipgr-asnr-miccai brats-peds). *ArXiv*.

Kim, H. S., & Lee, D. Y. (2022). Nanomedicine in clinical photodynamic therapy for the treatment of brain tumors. *Biomedicines*, *10*(1), 96.

Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Machine learning proceedings 1992* (pp. 249–256). Elsevier.

Klaise, J., Looveren, A. V., Vacanti, G., & Coca, A. (2021). Alibi explain: Algorithms for explaining machine learning models. *Journal of Machine Learning Research*, *22*(181), 1-7. Retrieved from `http://jmlr.org/papers/v22/21-0017.html`

Koley, S., Sadhu, A. K., Mitra, P., Chakraborty, B., & Chakraborty, C. (2016). Delineation and diagnosis of brain tumors from post contrast t1-weighted mr images using rough granular computing and random forest. *Applied Soft Computing*, *41*, 453–465.

Kong, H., Akakin, H. C., & Sarma, S. E. (2013). A generalized laplacian of gaussian filter for blob detection and its applications. *IEEE transactions on cybernetics*, *43*(6), 1719–1733.

Krassnig, S., Wohlrab, C., Golob-Schwarzl, N., Raicht, A., Schatz, C., Birkl-Toeglhofer, A. M., ... others (2021). A profound basic characterization of eifs in gliomas: Identifying eif3i and 4h as potential novel target candidates in glioma therapy. *Cancers*, *13*(6), 1482.

Lacore, M. G., Delmas, C., Nicaise, Y., Kowalski-Chauvel, A., Cohen-Jonathan-Moyal, E., & Seva, C. (2022). The glycoprotein m6a is associated with invasiveness and radioresistance of glioblastoma stem cells. *Cells*, *11*(14), 2128.

Lai, Y., Xu, P., Wang, J., Xu, K., Wang, L., & Meng, Y. (2020). Tumour suppressive long non-coding rna afdn-dt inhibits gastric cancer invasion via transcriptional regulation. *Journal of Cellular and Molecular Medicine*, *24*(5), 3157–3166.

Lakare, S., & Kaufman, A. (2000). 3d segmentation techniques for medical volumes. *Center for Visual Computing, Department of Computer Science, State University of New York*, *2000*, 59–68.

Lathoria, K., Gowda, P., Umdor, S. B., Patrick, S., Suri, V., & Sen, E. (2023). Prmt1 driven ptx3 regulates ferritinophagy in glioma. *Autophagy*, *19*(7), 1997–2014.

Li, Y., Kim, M. M., Wahl, D. R., Lawrence, T. S., Parmar, H., & Cao, Y. (2021). Survival prediction analysis in glioblastoma with diffusion kurtosis imaging. *Frontiers in Oncology*, *11*, 690036.

Liang, J., Lv, X., Lu, C., Ye, X., Chen, X., Fu, J., ... Zhao, Y. (2020). Prognostic factors of patients with gliomas–an analysis on 335 patients with glioblastoma and other forms of gliomas. *BMC cancer*, *20*, 1–7.

Liu, C., Tian, G., Tu, Y., Fu, J., Lan, C., & Wu, N. (2009). Expression pattern and clinical prognostic relevance of bone morphogenetic protein-2 in human gliomas. *Japanese journal of clinical oncology*, *39*(10), 625–631.

Liu, D., Xu, S., Ma, S., Wang, K., Sun, G., Chen, S., & Zhang, H. (2022). Predicting prognosis and distinguishing cold and hot tumors in bladder urothelial carcinoma based on necroptosis-associated lncrnas. *Frontiers in Immunology*, *13*, 916800.

Liu, H., He, H., Zhang, Z., Wang, L., Zhang, L., Liu, Y., & Xiong, W. (2021). Up-regulation of the long noncoding rna ubox5 antisense rna 1 (ubox5-as1) under hypoxic conditions promotes epithelial-mesenchymal transition in endometriosis. *Annals of Translational Medicine*, *9*(9).

Liu, J., An, P., Xue, Y., Che, D., Liu, X., Zheng, J., ... Yu, B. (2019). Mechanism of snhg8/mir-384/hoxa13/fam3a axis regulating neuronal apoptosis in ischemic mice model. *Cell Death & Disease*, *10*(6), 441.

Liu, J., Dang, H., & Wang, X. W. (2018). The significance of intertumor and intratumor heterogeneity in liver cancer. *Experimental & molecular medicine*, *50*(1), e416–e416.

Liu, J., Li, M., Wang, J., Wu, F., Liu, T., & Pan, Y. (2014). A survey of mri-based brain tumor segmentation methods. *Tsinghua science and technology*, *19*(6), 578–595.

Liu, L., Zhang, H., Rekik, I., Chen, X., Wang, Q., & Shen, D. (2016). Outcome prediction for patient with high-grade gliomas from brain functional and structural networks. In *Medical image computing and computer-assisted intervention–miccai 2016: 19th international conference, athens, greece, october 17-21, 2016, proceedings, part ii 19* (pp. 26–34).

Liu, Z.-Y., Lan, T., Tang, F., He, Y.-Z., Liu, J.-S., Yang, J.-Z., ... Li, Z.-Q. (2023). Zdhhc15 promotes glioma malignancy and acts as a novel prognostic biomarker for patients with glioma. *Bmc Cancer*, *23*(1), 420.

Lobbous, M., Bernstock, J. D., Coffee, E., Friedman, G. K., Metrock, L. K., Chagoya, G., ... others (2020). An update on neurofibromatosis type 1-associated gliomas. *Cancers*, *12*(1), 114.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3431–3440).

Long, S., & Li, G. (2019). Comprehensive analysis of a long non-coding rna-mediated competitive endogenous rna network in glioblastoma multiforme. *Experimental and therapeutic medicine*, *18*(2), 1081–1090.

Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., ... others (2021). The 2021 who classification of tumors of the central nervous system: a summary. *Neuro-oncology*, *23*(8), 1231–1251.

Lu, M., Zhou, Y., Sun, L., Shafi, S., Ahmad, N., Sun, M., & Dong, J. (2022). The molecular mechanisms of ferroptosis and its role in glioma progression and treatment. *Frontiers in Oncology*, *12*, 917537.

Lu, T.-P., Tsai, M.-H., Hsiao, C. K., Lai, L.-C., & Chuang, E. Y. (2012). Expression and functions of semaphorins in cancer. *Translational Cancer Research*, *1*(2).

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Ma, S., Zhang, Z., Ding, J., Li, X., Tang, J., & Guo, F. (2021). A deep supervision cnn network for brain tumor segmentation. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part ii 6* (pp. 158–167).

Mahmoudzadeh, A. P., & Kashou, N. H. (2014). Interpolation-based super-resolution reconstruction: effects of slice thickness. *Journal of Medical Imaging*, *1*(3), 034007.

Marti Asenjo, J., & Martinez-Larraz Solís, A. (2021). Mri brain tumor segmentation using a 2d-3d u-net ensemble. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part i 6* (pp. 354–366).

Mayerhoefer, M. E., Materka, A., Langs, G., Häggström, I., Szczypiński, P., Gibbs, P., & Cook, G. (2020). Introduction to radiomics. *Journal of Nuclear Medicine*, *61*(4), 488–495.

McHugh, H., Talou, G. M., & Wang, A. (2021). 2d dense-unet: a clinically valid approach to automated glioma segmentation. In *Brainlesion: Glioma, multi-*

ple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part ii 6 (pp. 69–80).

McKinley, R., Rebsamen, M., Daetwyler, K., Meier, R., Radojewski, P., & Wiest, R. (2020). Uncertainty-driven refinement of tumor-core segmentation using 3d-to-2d networks with label uncertainty. *arXiv preprint arXiv:2012.06436*.

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

Messaoudi, H., Belaid, A., Allaoui, M. L., Zetout, A., Allili, M. S., Tliba, S., … Conze, P.-H. (2021). Efficient embedding network for 3d brain tumor segmentation. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part i 6* (pp. 252–262).

Michaelsen, S. R., Christensen, I. J., Grunnet, K., Stockhausen, M.-T., Broholm, H., Kosteljanetz, M., & Poulsen, H. S. (2013). Clinical variables serve as prognostic factors in a model for survival from glioblastoma multiforme: an observational study of a cohort of consecutive non-selected patients from a single institution. *BMC cancer*, *13*(1), 1–11.

Minoru. (2021). *regression - what does the median absolute error metric say about the models?* Cross Validated.

(URL:https://stats.stackexchange.com/q/253892 (version: 2017-04-13) accessed: 2021-06-12)

Mishra, R. (2010). Mri based brain tumor detection using wavelet packet feature and artificial neural networks. In *Proceedings of the international conference and workshop on emerging trends in technology* (pp. 656–659).

MIT, M. K., & Lopuhin, K. (1965, Aug). *permutation_importance.* Retrieved from `https://eli5.readthedocs.io/en/latest/blackbox/permutation_importance.html`

Mohammed, Y. M., El Garouani, S., & Jellouli, I. (2023). A survey of methods for brain tumor segmentation-based mri images. *Journal of Computational Design and Engineering*, *10*(1), 266–293.

Molina, G., Chawla, A., Clancy, T. E., & Wang, J. (2019). *The correlation between the proportion of patients with pancreatic ductal adenocarcinoma who received neoadjuvant therapy and overall survival between 2004 and 2015.* American Society of Clinical Oncology.

Molnar, C. (2021). *Interpretable machine learning a guide for making black box models explainable*. Leanpub.

Müller, T., Gessi, M., Waha, A., Isselstein, L. J., Luxen, D., Freihoff, D., … others (2012). Nuclear exclusion of tet1 is associated with loss of 5-hydroxymethylcytosine in idh1 wild-type gliomas. *The American journal of pathology*, *181*(2), 675–683.

Myronenko, A. (2019). 3d mri brain tumor segmentation using autoencoder regularization. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 4th international workshop, brainles 2018, held in conjunction with miccai 2018, granada, spain, september 16, 2018, revised selected papers, part ii 4* (pp. 311–320).

National Center for Biotechnology Information. (2024). *Ncbi gene: 219902.* https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=219902. (Accessed: April 15, 2024)

National brain tumor society - brain tumor facts. (n.d.). https://braintumor.org/brain-tumors/about-brain-tumors/brain-tumor-facts/. ((Accessed on October 27, 2023))

National library of medicine-national center for biotechnology information - gene: 80757. (2023). Online. Retrieved from https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=80757 (Accessed on: April 15, 2024)

Noori, M., Bahri, A., & Mohammadi, K. (2019). Attention-guided version of 2d unet for automatic brain tumor segmentation. In *2019 9th international conference on computer and knowledge engineering (iccke)* (pp. 269–275).

Nwosu, Z. C., Song, M. G., di Magliano, M. P., Lyssiotis, C. A., & Kim, S. E. (2023). Nutrient transporters: connecting cancer metabolism to therapeutic opportunities. *Oncogene*, *42*(10), 711–724.

of Encyclopaedia Britannica, T. E. (2024). *tumour.* Retrieved from https://www.britannica.com/science/tumor (Accessed: 1 February 2024)

Ostrom, Q. T., Price, M., Neff, C., Cioffi, G., Waite, K. A., Kruchko, C., & Barnholtz-Sloan, J. S. (2022). Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2015–2019. *Neuro-oncology*, *24*(Supplement_5), v1–v95.

Pálsson, S., Cerri, S., Poulsen, H. S., Urup, T., Law, I., & Van Leemput, K. (2022). Predicting survival of glioblastoma from automatic whole-brain and tumor segmentation of mr images. *Scientific Reports*, *12*(1), 19744.

Papp, L., Pötsch, N., Grahovac, M., Schmidbauer, V., Woehrer, A., Preusser, M., ... others (2018). Glioma survival prediction with combined analysis of in vivo 11c-met pet features, ex vivo features, and patient features by supervised machine learning. *Journal of Nuclear Medicine*, *59*(6), 892–899.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, *12*, 2825–2830.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2012). Scikit-learn: Machine learning in python. *arXiv:1201.090*.

Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Deep convolutional neural networks for the segmentation of gliomas in multi-sequence mri. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: First international workshop, brainles 2015, held in conjunction with miccai 2015, munich, germany, october 5, 2015, revised selected papers 1* (pp. 131–143).

Persano, L., Pistollato, F., Rampazzo, E., Della Puppa, A., Abbadi, S., Frasson, C., ... Basso, G. (2012). Bmp2 sensitizes glioblastoma stem-like cells to temozolomide by affecting hif-1α stability and mgmt expression. *Cell death & disease*, *3*(10), e412–e412.

Phophalia, A., & Maji, P. (2018). Multimodal brain tumor segmentation using ensemble of forest method. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: Third international workshop, brainles 2017, held in conjunction with miccai 2017, quebec city, qc, canada, september 14, 2017, revised selected papers 3* (pp. 159–168).

Protein atlas - fam98c tissue expression. (2023). Online. Retrieved from `https://www.proteinatlas.org/ENSG00000130244-FAM98C/tissue` (Accessed on: April 15, 2024)

Puybareau, E., Tochon, G., Chazalon, J., & Fabrizio, J. (2019). Segmentation of gliomas and prediction of patient overall survival: A simple and fast procedure. In A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, & T. van Walsum (Eds.), *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries* (pp. 199–209). Cham: Springer International Publishing.

Qian, X., Tan, H., Zhang, J., Liu, K., Yang, T., Wang, M., ... Zhou, X. (2016). Identification of biomarkers for pseudo and true progression of gbm based on radiogenomics study. *Oncotarget*, *7*(34), 55377.

Rajput, S., Agravat, R., Roy, M., & Raval, M. S. (2021). Glioblastoma multiforme patient survival prediction. *arXiv preprint arXiv:2101.10589*.

Rajput, S., Agravat, R., Roy, M., & Raval, M. S. (2022). Glioblastoma multiforme patient survival prediction. In *Proceedings of 2021 international conference on medical imaging and computer-aided diagnosis (micad 2021) medical imaging and computer-aided diagnosis* (pp. 47–58).

Rajput, S., Kapdi, R., Raval, M., & Roy, M. (2023b). Multi-view brain tumor segmentation (mvbts): An ensemble of planar and triplanar attention unets. *Turkish Journal of Electrical Engineering and Computer Sciences*, *31*(6), 908–927.

Rajput, S., Kapdi, R., Roy, M., & Raval, M. S. (2024). A triplanar ensemble model for brain tumor segmentation with volumetric multiparametric magnetic resonance images. *Healthcare Analytics*, 100307.

Rajput, S., Kapdi, R. A., Raval, M. S., & Roy, M. (2023a). Interpretable machine learning model to predict survival days of malignant brain tumor patients. *Machine Learning: Science and Technology*, *4*(2), 025025.

Rajput, S., Kapdi, R. A., Raval, M. S., Roy, M., & Bhalodiya, J. M. (2024). Deriving and interpreting robust features for survival prediction of brain tumor patients. *International Journal of Imaging Systems and Technology*, *34*(3), e23105.

Ramteke, R., & Monali, K. Y. (2012). Automatic medical image classification and abnormality detection using k-nearest neighbour. *International Journal of Advanced Computer Research*, *2*(4), 190.

Ratan, R., Sharma, S., & Sharma, S. (2009). Multiparameter segmentation and

quantization of brain tumor from mri images. *Indian Journal of Science and Technology*, *2*(2), 11–15.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016a). "why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, san francisco, ca, usa, august 13-17, 2016* (pp. 1135–1144).

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016b). " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1135–1144).

Rizzo, S., Botta, F., Raimondi, S., Origgi, D., Fanciullo, C., Morganti, A. G., & Bellomi, M. (2018). Radiomics: the facts and the challenges of image analysis. *European radiology experimental*, *2*(1), 1–8.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–miccai 2015: 18th international conference, munich, germany, october 5-9, 2015, proceedings, part iii 18* (pp. 234–241).

Roy, A. G., Navab, N., & Wachinger, C. (2018). Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks. In *Medical image computing and computer assisted intervention–miccai 2018: 21st international conference, granada, spain, september 16-20, 2018, proceedings, part i* (pp. 421–429).

Rozemberczki, B., Watson, L., Bayer, P., Yang, H.-T., Kiss, O., Nilsson, S., & Sarkar, R. (2022). The shapley value in machine learning. *arXiv preprint arXiv:2202.05594*.

Sachdeva, R., Wu, M., Johnson, K., Kim, H., Celebre, A., Shahzad, U., ... others (2019). Bmp signaling mediates glioma stem cell quiescence and confers treatment resistance in glioblastoma. *Scientific reports*, *9*(1), 14569.

Salman, Y., Assal, M., Badawi, A., Alian, S., & El-Bayome, M. E.-M. (2006). Validation techniques for quantitative brain tumors measurements. In *2005 ieee engineering in medicine and biology 27th annual conference* (pp. 7048–7051).

Sandabad, S., Benba, A., Tahri, Y. S., & Hammouch, A. (2016). Novel extraction and tumour detection method using histogram study and svm classification. *International Journal of Signal and Imaging Systems Engineering*, *9*(4-5), 202–208.

Sanghani, P., Ang, B. T., King, N. K. K., & Ren, H. (2018). Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning. *Surgical oncology*, *27*(4), 709–714.

Schmit, K., & Michiels, C. (2018). Tmem proteins in cancer: a review. *Frontiers in pharmacology*, *9*, 408621.

Shahbazian, D., Parsyan, A., Petroulakis, E., Hershey, J. W., & Sonenberg, N. (2010). eif4b controls survival and proliferation and is regulated by proto-oncogenic signaling pathways. *Cell cycle*, *9*(20), 4106–4109.

Shen, S., Zhou, H., Xiao, Z., Zhan, S., Tuo, Y., Chen, D., ... Wang, J. (2024).

Prmt1 in human neoplasm: cancer biology and potential therapeutic target. *Cell Communication and Signaling*, *22*(1), 102.

Shi, J., Dong, B., Cao, J., Mao, Y., Guan, W., Peng, Y., & Wang, S. (2017). Long non-coding rna in glioma: signaling pathways. *Oncotarget*, *8*(16), 27582.

Singh, S. P., & Urooj, S. (2015). Wavelets: biomedical applications. *International Journal of Biomedical Engineering and Technology*, *19*(1), 1–25.

Sofie Van Cauter, M., Veraart, J., Sijbers, J., Peeters, R. R., Himmelreich, U., De Keyzer, F., ... others (2012). Gliomas: Diffusion kurtosis mr imaging in grading1. *Radiology*, *263*, 492–501.

Soltaninejad, M., Pridmore, T., & Pound, M. (2021). Efficient mri brain tumor segmentation using multi-resolution encoder-decoder networks. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part ii 6* (pp. 30–39).

Soni, N., Priya, S., & Bathla, G. (2019). Texture analysis in cerebral gliomas: a review of the literature. *American Journal of Neuroradiology*, *40*(6), 928–934.

Spyridon (Spyros), B. C. S. (2021). *Validation survival leaderboard 2020.* `https://www.cbica.upenn.edu/BraTS20//lboardValidationSurvival.html`. (accessed: 2021-06-12)

Steven, A. J., Zhuo, J., & Melhem, E. R. (2014). Diffusion kurtosis imaging: an emerging technique for evaluating the microstructural environment of the brain. *American journal of roentgenology*, *202*(1), W26–W33.

Strudel, R., Garcia, R., Laptev, I., & Schmid, C. (2021). Segmenter: Transformer for semantic segmentation. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 7262–7272).

Su, X., Zheng, G., Gui, Z., Yang, X., Zhang, L., & Pan, F. (2022). A systematic analysis reveals the prognostic and immunological role of reptin/ruvbl2 in human tumors. *Frontiers in Genetics*, *13*, 911223.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support: Third international workshop, dlmia 2017, and 7th international workshop, ml-cds 2017, held in conjunction with miccai 2017, québec city, qc, canada, september 14, proceedings 3* (pp. 240–248).

Sun, H., & Wang, T. (2008). Lcg-mrf-based segmentation of mri brain images. In *2008 international conference on computer science and information technology* (pp. 375–378).

Tang, F., Yang, C., Li, F.-P., Yu, D.-H., Pan, Z.-Y., Wang, Z.-F., & Li, Z.-Q. (2022). Palmitoyl transferases act as potential regulators of tumor-infiltrating immune cells and glioma progression. *Molecular Therapy-Nucleic Acids*, *28*, 716–731.

Tarasiewicz, T., Kawulok, M., & Nalepa, J. (2021). Lightweight u-nets for brain tumor segmentation. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in*

conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part ii 6 (pp. 3–14).

Tessamma, T., & Ananda Resmi, S. (2010). Texture description of low grade and high grade glioma using statistical features in brain mris..

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4itk: improved n3 bias correction. *IEEE transactions on medical imaging*, *29*(6), 1310–1320.

Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J., et al. (2014). Multi-modal brain tumor segmentation using deep convolutional neural networks. *MICCAI BraTS (brain tumor segmentation) challenge. Proceedings, winning contribution*, 31–35.

Van Cauter, S., Veraart, J., Sijbers, J., Peeters, R. R., Himmelreich, U., De Keyzer, F., ... others (2012). Gliomas: diffusion kurtosis mr imaging in grading. *Radiology*, *263*(2), 492–501.

Van Griethuysen, J. J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., ... Aerts, H. J. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer research*, *77*(21), e104–e107.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wacker, J., Ladeira, M., & Nascimento, J. E. V. (2019). Transfer learning for brain tumor segmentation. *arXiv preprint arXiv:1912.12452*.

Wang, S., Tan, X., Yang, B., Yin, B., Yuan, J., Qiang, B., & Peng, X. (2012). The role of protein arginine-methyltransferase 1 in gliomagenesis. *BMB reports*, *45*(8), 470–475.

Wang, S., Zhang, Y., Dong, Z., Du, S., Ji, G., Yan, J., ... Phillips, P. (2015). Feed-forward neural network optimized by hybridization of pso and abc for abnormal brain detection. *International Journal of Imaging Systems and Technology*, *25*(2), 153–164.

Wang, X., Gao, M., Ye, J., Jiang, Q., Yang, Q., Zhang, C., ... others (2020). An immune gene-related five-lncrna signature for to predict glioma prognosis. *Frontiers in genetics*, *11*, 612037.

Wang, Z., Tang, W., Yuan, J., Qiang, B., Han, W., & Peng, X. (2020). Integrated analysis of rna-binding proteins in glioma. *Cancers*, *12*(4), 892.

Wei, B., Wang, R., Wang, L., & Du, C. (2020). Prognostic factor identification by analysis of the gene expression and dna methylation data in glioma. *Mathematical Biosciences and Engineering: MBE*, *17*(4), 3909–3924.

Wei, J., Zeng, Y., Gao, X., & Liu, T. (2021). A novel ferroptosis-related lncrna signature for prognosis prediction in gastric cancer. *BMC cancer*, *21*, 1–12.

Wong, K.-P. (2005). Medical image segmentation: methods and applications in functional imaging. In *Handbook of biomedical image analysis: Volume ii: Segmentation models part b* (pp. 111–182). Springer.

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: Convolutional block

attention module. In *Proceedings of the european conference on computer vision (eccv)* (pp. 3–19).

Wu, G., Zhang, L., Chen, X., Lin, J., Wang, Y., & Yu, J. (2021). Convolutional neural network with asymmetric encoding and decoding structure for brain vessel segmentation on computed tomographic angiography. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part i 6* (pp. 51–59).

Wu, M.-N., Lin, C.-C., & Chang, C.-C. (2007). Brain tumor detection using color-based k-means clustering segmentation. In *Third international conference on intelligent information hiding and multimedia signal processing (iih-msp 2007)* (Vol. 2, pp. 245–250).

Wu, Z., Wei, J., Wang, J., & Li, R. (2022). Slice imputation: Multiple intermediate slices interpolation for anisotropic 3d medical image segmentation. *Computers in Biology and Medicine*, *147*, 105667.

Xiao, K., Liu, Q., Peng, G., Su, J., Qin, C.-Y., & Wang, X.-Y. (2020). Identification and validation of a three-gene signature as a candidate prognostic biomarker for lower grade glioma. *PeerJ*, *8*, e8312.

Xu, C., & Prince, J. L. (1997). Gradient vector flow: A new external force for snakes. In *Proceedings of ieee computer society conference on computer vision and pattern recognition* (pp. 66–71).

Xu, C., & Prince, J. L. (1998). Generalized gradient vector flow external forces for active contours. *Signal processing*, *71*(2), 131–139.

Xu, J. H., Teng, W. P. K., Wang, X. J., & Nürnberger, A. (2021). A deep supervised u-attention net for pixel-wise brain tumor segmentation. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part ii 6* (pp. 278–289).

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, *9*, 611–629.

Yang, F., Wang, Y., Li, Q., Cao, L., Sun, Z., Jin, J., ... others (2017). Intratumor heterogeneity predicts metastasis of triple-negative breast cancer. *Carcinogenesis*, *38*(9), 900–909.

Yang, H., Jin, L., & Sun, X. (2019). A thirteen-gene set efficiently predicts the prognosis of glioblastoma. *Molecular Medicine Reports*, *19*(3), 1613–1621.

Yang, X., & Fei, B. (2011). A multiscale and multiblock fuzzy c-means classification method for brain mr images. *Medical physics*, *38*(6Part1), 2879–2891.

Yang, Y., Liang, Z., Xia, Z., Wang, X., Ma, Y., Sheng, Z., ... others (2019). Sae1 promotes human glioma progression through activating akt sumoylation-mediated signaling pathways. *Cell Communication and Signaling*, *17*, 1–14.

Yang, Y., Teng, H., Zhang, Y., Wang, F., Tang, L., Zhang, C., ... others (2024). A glycosylation-related gene signature predicts prognosis, immune microenvironment infiltration, and drug sensitivity in glioma. *Frontiers in Pharmacology*, *14*, 1259051.

Yousefi, S., Azmi, R., & Zahedi, M. (2011). Brain tumor segmentation in 3d mris using an improved markov random field model. In *International conference on graphic and image processing (icgip 2011)* (Vol. 8285, pp. 744–750).

Yuan, Y. (2021). Automatic brain tumor segmentation with scale attention network. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 6th international workshop, brainles 2020, held in conjunction with miccai 2020, lima, peru, october 4, 2020, revised selected papers, part i 6* (pp. 285–294).

Yushkevich, P. A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, *31*(3), 1116–1128.

Zhang, J., Lv, X., Zhang, H., & Liu, B. (2020). Aresu-net: Attention residual u-net for brain tumor segmentation. *Symmetry*, *12*(5), 721.

Zhang, X., Sun, S., Pu, J. K. S., Tsang, A. C. O., Lee, D., Man, V. O. Y., ... Leung, G. K. K. (2012). Long non-coding rna expression profiles predict clinical phenotypes in glioma. *Neurobiology of disease*, *48*(1), 1–8.

Zhang, Z., Tang, H., Wang, Z., Zhang, B., Liu, W., Lu, H., ... others (2011). Mir-185 targets the dna methyltransferases 1 and regulates global dna methylation in human glioma. *Molecular cancer*, *10*(1), 1–16.

Zhao, H., Ji, B., Chen, J., Huang, Q., & Lu, X. (2017). Gpx 4 is involved in the proliferation, migration and apoptosis of glioma cells. *Pathology-Research and Practice*, *213*(6), 626–633.

Zhu, D., Ouyang, X., Zhang, Y., Yu, X., Su, K., & Li, L. (2023). A promising new cancer marker: Long noncoding rna egfr-as1. *Frontiers in Oncology*, *13*, 1130472.

Zikic, D., Ioannou, Y., Brown, M., & Criminisi, A. (2014). Segmentation of brain tumor tissues with convolutional neural networks. *Proceedings MICCAI-BRATS*, *36*(2014), 36–39.