Survival Prediction of Glioblastoma Patients Using Pre-Operative Brain MRI Scans

A Thesis

submitted for the award of the degree of

Doctor of Philosophy

in

Computer Science and Engineering Department

Submitted by

Gurinderjeet Kaur
(Reg no: 901803017)

Under the Guidance of

Dr. Prashant Singh Rana
Associate Professor

Dr. Vinay Arora
Associate Professor



THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Thapar Institute of Engineering and Technology,
Patiala, Punjab - 147004, India
January 2023

Certificate

I hereby certify that the work, which is being presented in the thesis, entitled "Survival Prediction of Glioblastoma Patients Using Pre-Operative Brain MRI Scans", in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy and submitted to the institution is an authentic record of my work carried out under the supervision of Dr. Prashant Singh Rana and Dr. Vinay Arora. I have included the citation for the text(s), figure(s), or table(s) from which it was derived.
The matter presented in this thesis has not been submitted either in-part or full to any other University/Institute for the award of any other degree.

Gurinderjeet Kaur

(Gurinderjeet Kaur)
Registration No. 901803017

This is to certify that the above statements made by the candidate are correct and true to the best of my knowledge.

Verified by:

(Dr. Prashant Singh Rana)
Supervisor
Computer Science and Engineering Department
Thapar Institute of Engineering and Technology, Patiala, Punjab, India

(Dr. Vinay Arora)
Supervisor
Computer Science and Engineering Department
Thapar Institute of Engineering and Technology, Patiala, Punjab, India

To my father, mother, and siblings for their unconditional love and to my supervisors for their selfless and exemplary role in my life.

Acknowledgements

I would like to thank my supervisors, Dr. Prashant Singh Rana and Dr. Vinay Arora, for their guidance and support during my PhD tenure. This thesis would not have been feasible without their unwavering support and faith in me. Their commitment to my thesis extends well beyond their duty as academic supervisors and includes ongoing personal support without which this path would not have been possible.

I would like to give thanks to Dr. Rana and Dr. Arora for letting me be one of their students. I am grateful to them for their unconditional support and for allowing me to pursue my academic curiosity without hesitation. Their words of wisdom and encouragement always guided me and served as helpful reference points.

My sincere thanks to Head of the Computer Science and Engineering Department Dr. Shalini Batra and my research committee members: Dr. Husanbir Singh Pannu, Dr. Shreelekha Pandey, and Dr. Neeru Jindal. They graciously devoted their time to provide informative feedback on my work. I would also like to thank Dr. Vishal Srivastava, who helped me throughout my PhD research work.

I would like to acknowledge and give my warmest thanks to Dean of Academic Affairs Dr. Maninder Singh for providing access to the required computing resources for my research work.

I owe this thesis to my parents, my in-laws and other family members who always stood by me and provided strength in pursuing this work.

Abstract

Glioblastoma is identified as highly invasive and aggressive grade IV type glioma. Grade IV gliomas are malignant brain tumors that are cancerous by nature and cause life threatening consequences for the patients. The life span of glioblastoma patients is less than two years due to its abnormal growth and outspread to other parts of the patient's body. Therefore, it is very important to predict the overall survival time of glioblastoma patients on prior basis for appropriate treatment planning. Early detection and precise segmentation of tumor region are two pre-requisites for accurate survival estimation of the patient. In this research work, retrospective study is done using Multimodal Brain Tumor Segmentation (BraTS) data of glioma patients made publicly available by the University of Pennsylvania.

In *first* technique, the radiomic features were extracted using PyRadiomics Python library from pre-operative raw structural multi-parametric Magnetic Resonance Imaging (mpMRI) scans and, scans segmented using three-dimensional (3D) deep-supervised U-shaped Convolutional Neural Network (CNN) inspired encoder-decoder architecture after doing their necessary pre-processing. After removing irrelevant features, regression models based on Machine Learning (ML) were developed by considering selected radiomic features and clinical data to predict the Overall Survival (OS) time of Glioblastoma Multiforme (GBM) patients within a period of days only.

In *second* technique, the features were extracted using a pre-trained two-dimensional (2D) Residual Neural Network (RNN) pre-operative raw structural mpMRI scans and, scans segmented using 3D deep-supervised UNet model architecture after doing their required pre-processing. The dimension-reduced principal components were integrated with clinical data and the handcrafted features of tumor subregions to compare the performance of automated ML-based regression techniques.

In *third* technique, a lightweight 2D methodology was proposed to predict the survival time of GBM patients based on the pre-operative raw 3D mpMRI scans and clinical data provided in the publicly available BraTS 2020 dataset. Firstly, a 2D Residual UNet for Segmentation (ResUNet-SEG) model was trained to perform semantic segmentation on brain tumour subregions. Then, the raw and segmented mpMRI volumes was used alongwith clinical data to train the 2D CNN for Survival Prediction (CNN-SP) model to predict the survival time in days.

Keywords: Glioblastoma, Glioma Segmentation, Survival Prediction, Deep Learning, Machine Learning, Magnetic resonance imaging, Convolutional neural network, Radiomics.

Table of Contents

Title **Page No.**

Abstract . . . . . **vii**

Table of Contents . . . . . **ix**

List of Figures . . . . . **xiv**

List of Tables . . . . . **xviii**

List of Abbreviations . . . . . **xxi**

Chapter 1 Introduction **1**
1.1 Brain Tumor 2
1.1.1 Glioma 3
1.1.2 Grades of Gliomas 3
1.1.3 Glioblastoma Multiforme 5
1.2 Medical Imaging and Treatment Planning 5
1.2.1 Manual and Automated Glioma Delineation 7
1.2.2 Difficulties in Automated Glioma Delineation 9
1.2.3 Survival Prediction 10
1.3 Current Research Trends 11
1.4 Organisation of the Thesis 13

Chapter 2 Review of Literature **17**
2.1 Review Methodology 17
2.1.1 Dataset 18
2.1.2 Brain Tumor Segmentation 19
2.1.3 Survival Prediction 20
2.2 Generic Workflow for Brain Tumor Segmentation and Survival Prediction . . . 21
2.2.1 Pre-processing and Data Augmentation 21
2.2.1.1 Pre-processing 22
2.2.1.2 Data Augmentation 25
2.2.2 Glioma Segmentation Architectures 27
2.2.2.1 Single Networks 28

2.2.2.2 Cascaded Networks . . . 32
2.2.2.3 Self-ensemble Networks . . . 34
2.2.2.4 Parallel Networks . . . 34
2.2.2.5 Ensemble Networks . . . 34
2.2.3 Post-processing . . . 39
2.2.3.1 Connected Component Analysis . . . 39
2.2.3.2 Conditional Random Field . . . 40
2.2.3.3 Morphological Operations . . . 40
2.2.3.4 Relabelling the Output Label . . . 40
2.2.4 Feature Extraction and Selection . . . 41
2.2.5 Overall Survival Prediction Models . . . 46
2.2.5.1 Machine Learning Models . . . 46
2.2.5.2 Deep Learning Models . . . 47
2.3 Research Gaps and Findings . . . 52
2.4 Problem Formulation . . . 55
2.5 Objectives . . . 56
2.6 Proposed Methodology . . . 56

Chapter 3 Radiomic Features-based Survival Prognosis of Glioblastoma Multiforme Patients . . . 57
3.1 Contribution of the Proposed Technique . . . 57
3.2 Materials and Methods . . . 58
3.2.1 Dataset and Proposed Methodology . . . 58
3.2.2 Pre-processing . . . 60
3.2.3 Tumor Segmentation Using Deep Learning . . . 63
3.2.4 Radiomic Feature Extraction Using PyRadiomics . . . 65
3.2.5 Feature Selection Using Machine Learning . . . 66
3.2.6 Machine Learning Algorithms for Survival Prediction . . . 67
3.3 Experiments . . . 77
3.3.1 Experimental Set-up . . . 77
3.3.2 Evaluation Parameters . . . 78
3.4 Results and Discussion . . . 80
3.4.1 Comparison with State-of-the-Art Techniques . . . 84
3.4.2 Strengths and Limitations of the Proposed Methodology . . . 86
3.5 Inference and Findings . . . 87

Chapter 4 Deep Learning Features-based Survival Prognosis of Glioblastoma Multiforme Patients . . . 89

4.1 Contribution of the Proposed Technique . . . 89
4.2 Materials and Methods . . . 90
4.2.1 Dataset and Proposed Methodology . . . 90
4.2.2 Pre-processing . . . 92
4.2.3 Segmentation Architecture . . . 92
4.2.4 Deep Learning and Handcrafted Feature Extraction . . . 93
4.2.5 Feature Selection . . . 94
4.2.6 Automated Model Selection . . . 96
4.2.7 Survival Prediction . . . 104
4.3 Experiments . . . 104
4.3.1 Experimental Set-up . . . 104
4.3.2 Evaluation Parameters . . . 105
4.4 Results and Discussion . . . 106
4.4.1 Comparison with State-of-the-Art Techniques . . . 108
4.4.2 Strengths and Limitations of the Proposed Methodology . . . 112
4.5 Inference and Findings . . . 113

Chapter 5 Automated Neural Network-based Survival Prognosis of Glioblastoma Multiforme Patients . . . 115
5.1 Contribution of the Proposed Technique . . . 115
5.2 Materials and Methods . . . 116
5.2.1 Dataset and Proposed Methodology . . . 116
5.2.2 Pre-processing and Data Augmentation . . . 117
5.2.3 Segmentation Module . . . 118
5.2.4 Post-processing . . . 124
5.2.5 Survival Prediction Module . . . 124
5.3 Experiments . . . 126
5.3.1 Experimental Set-up . . . 127
5.3.2 Evaluation Parameters . . . 127
5.4 Results and Discussion . . . 127
5.4.1 Comparison with State-of-the-Art Techniques . . . 129
5.4.2 Strengths and Limitations of the Proposed Methodology . . . 131
5.5 Inference and Findings . . . 131

Chapter 6 Conclusions and Scope for further research . . . 133
6.1 Conclusion . . . 133
6.2 Scope for further research . . . 134

References **137**

List of Publications **155**

Appendix **157**

List of Figures

Figure No.	Title	Page No.
1.1	Brain Tumor	1
1.2	Categories of Brain Tumor Based on Aggressiveness and Origin Location	2
1.3	Grades of Glioma – LGG (Top row), HGG (Bottom row) [1]	4
1.4	Segmentation of Brain Tumor Sub-regions [2]	5
1.5	Different Multimodal MRI Scans- T1, T1CE, T2, T2-FLAIR	6
2.1	Different Multimodal MRI Scans (a) FLAIR (b) T1 (c) T1CE (d) T2 [1]	19
2.2	MRI Scans with Segmentation Labels in Multiple Views [1]	20
2.3	Overview of Steps for Survival Prognosis [1]	21
2.4	The Generic Workflow for Brain Tumor Segmentation and Overall Survival Prediction	22
2.5	Types of Pre-processing Used in BraTS 2020 SP Techniques	23
2.6	Different Forms of Data Augmentation [3, 4]	26
2.7	Different Types of Networks Used for Glioma Segmentation	29
2.8	Post-processing Techniques Used by SP Techniques Submitted in BraTS Challenge	39
3.1	Proposed Methodology	60
3.2	Steps of Pre-processing Performed on Multimodal 3D MRI Volumes	61
3.3	Illustration of 3D Deep Supervised UNet Segmentation Architecture	64
3.4	Radiomic Feature Extraction Performed on FLAIR and Segmented MRI Scans	65
3.5	Terminology Used for Determining Sensitivity and Specificity	78
3.6	Illustration of Segmentations Generated in Comparison to the Ground-truth for Four Different Patients using Different MRI Sequences (FLAIR, T1CE, T1, and T2)	81
3.7	Feature Importance Plot of Ridge Regressor Model	82
4.1	The Proposed Methodology for Early SP of Glioblastoma Patients	91
4.2	Illustration of 3D UNet Encoder-Decoder based Segmentation Architecture which takes four Modalities as Input and Generates a Probability-based Segmentation Map	93
4.3	Illustration of Pre-Trained 2D Resnet50 Architecture used for Extraction of Deep Features from Raw and Segmented 3D MRI Scans	95

4.4 (a, b, c, d) Plots for the Cumulative Proportion of Variance for Principal Components Generated using Matplotlib And Sklearn Libraries, (e) Pairplot for Different Handcrafted and Clinical Features Generated Using Seaborn Library 98
4.5 Feature Importance Plots for Top Six Regression Models 101
4.6 Overview of Complete Survival Prediction Pipeline 104
4.7 Zoomed Segmentation Maps Generated Using 3D UNet from the Brats Validation Data for Subject Ids: 2, 7, 12, 15, 23. The Generated Segmentation Maps were Overlayed On The T1CE Modality Of Individual Patients With Labels 1 For NCR And NET (Red Colour), 2 For ED (Green Colour), 4 For ET (Yellow Colour) 107
4.8 (a) Age Groups of 236 Patients from the BraTS 2020 Training Set, (b) Age Groups of 29 Patients from the BraTS 2020 Validation Set 110

5.1 Top-level Schematic Diagram of the Proposed Methodology 116
5.2 Different PNG Slices Generated from NIfTI After Pre-processing 118
5.3 Different PNG Slices Generated After Data Augmentation 118
5.4 The Architecture of the 2D ResUNet-SEG Model 119
5.5 Different Blocks Used in Encoding and Decoding Pathways of Segmentation Architecture 120
5.6 The Architecture of the 2D CNN-SP Model 125
5.7 Segmentation Maps Generated Using 2D ResUNet-SEG from the BraTS Validation Dataset for Patient IDs: 2, 8, 15, 22, 34. The Generated Segmentation Maps are Overlayed on the FLAIR Modality of Individual Patients with Labels 1 for NCR and NET (Red colour), 2 for ED (Green colour), 4 for ET (Yellow colour). 128
5.8 (a) Age Groups of 236 Patients from the BraTS 2020 Training Set, (b) Survival Day Groups of 236 Patients from the BraTS Training Set, (c) Age Groups of 29 Patients from the BraTS 2020 Validation Set, (d) Survival Day Groups of 29 Patients from the BraTS Validation Set 130

6.1 Future Research Directions for BTS and OS Prediction 136
F.1 Shapiro ranking of selected shape features 161
F.2 Shapiro ranking of selected features: age, diagnostics and tumor volume . . . 161
F.3 Shapiro ranking of selected NGTDM features 162
F.4 Shapiro ranking of selected GLDM features 162
F.5 Shapiro ranking of selected GLSZM features 163
F.6 Shapiro ranking of selected first order features 164
F.7 Shapiro ranking of selected GLCM features 165

F.8 Shapiro ranking of selected GLRLM features 166

List of Tables

Table No.	Title	Page No.
1.1	Class Imbalances in the Dataset [5]	10
2.1	BraTS 2020 Data for Brain Tumor Segmentation	19
2.2	BraTS 2020 Data for Overall Survival Prediction	19
2.3	Summary of Segmentation Architectures for BraTS 2020 SP Problem (A: Network Type, B: Input Dimension, C: Segmentation backbone, D: Depth/Levels/Layers, E: Modifications, F: Normalization, G: Uniqueness, H: Sampling)	36
2.4	Summary of SP Model Details Used for BraTS 2020 Dataset	49
2.5	Summary of Model Training Hyperparameters for Integrated BTS and SP Techniques Using BraTS 2020 Dataset (A: Loss Functions, B: Epochs, C: Patch Size, D: Optimizer, E: Learning Rate, F: Learning Rate Decay (LRD)/ Weight Rate Decay (WRD), G: Batch Size)	50
2.6	Summary of Implementation Framework Details	52
3.1	Patient Demographics	59
3.2	Description of Machine Learning Algorithms Trained for Survival Prediction	68
3.3	Description of Error Metrics Used to Compare Model Performance on the Training Dataset	75
3.4	Performance Of Different ML Models on the Training Dataset Before Hyperparameter Tuning (n=119 GBM Cases)	76
3.5	Hyperparameter Tuning on the Training Dataset	76
3.6	List of Positively and Negatively Related Features with OS time of GBM Patients	83
3.7	Performance of Segmentation Model on the Validation Dataset (N=125 Cases)	84
3.8	Performance of Survival Model on the Validation Dataset (N=29 GBM Cases)	85
4.1	Summary of Hyperparameters, Weights, MAE and MSE Metrics for Six Best Performing Models on Training Data generated using Automated ML Pycaret Package	99
4.2	Range of Feature Vectors Used for Training and Validating Different Regression-based Models	102
4.3	Performance of Proposed Segmentation Model on the Validation Dataset (N=125 Cases)	108

4.4 Performance of Proposed Survival Model on the Validation Dataset (N=29 GBM Cases) . . . 109
4.5 Summary of Features Extracted by Previous Studies and the Current Study To Predict The Survival Time Of GBM Patients using the Same Set Of Patients . . 111

5.1 The Details of Feature Maps Generated at Each Stage of Segmentation Architecture . . . 122
5.2 The Details of Feature Maps Generated at Each Stage of Survival Prediction Architecture . . . 125
5.3 Comparison of the Proposed Methodology with State-of-the-art Techniques Evaluated Using CBICA Online Portal for the Validation Set (n=29 Patients) . . 129

A1 Feature selection using ML-based models . . . 157
A2 Cross-validation MAE of different regression models on training dataset before hyperparameter tuning . . . 160
A3 Cross-validation MAE of different regression models on training dataset after hyperparameter tuning . . . 160

List of Abbreviations

AB	Accumulated Block
AE	Accumulated Encoder
ADA	AdaBoost
ANNs	Artificial Neural Networks
AUNet	Asymmetric UNet
ATT	Attention
ARD	Automatic Relevance Determination Regression
BN	Batch Normalization
BR	Bayesian Ridge
BiFPN	Bi-directional Feature Pyramid Network
BCE	Binary classification error
BTS	Brain Tumor Segmentation
CCE	Categorical Cross-Entropy
CNS	Central Nervous System
CBICA	Centre for Biomedical Image Computing and Analytics
CSF	Cerebrospinal Fluid
CN	Channel Normalization
CHSQ	Chi-square Score
CSV	Comma Separated Value
CT	Computed Tomography
cindex	Concordance index
CIFE	Conditional Infomax Feature Extraction
CMIM	Conditional Mutual Information Maximization
CRF	Conditional Random Fields
CCA	Connected Component Analysis
CNN	Convolutional Neural Network
CV	Cross Validation
DT	Decision Tree
DCNNs	Deep CNNs
DL	Deep learning
DNNs	Deep Neural Networks
DSC	Dice Similarity Coefficient
DTI	Diffusion Tensor Imaging

DMFNet	Dilated Multi-Fiber Network
DISR	Double Input Symmetric Relevance
ED	Edema
EffUNet	Efficient UNet
EN	Elastic Net
ET	Enhancing Tumor
EOR	Extent of Resection
ET	Extra Trees
XGB	Extreme Gradient Boosting
FMUs	Feature Mining Units
FFNN	Feedforward Neural Network
FSCR	Fischer Score
FCNNs	Fully CNNs
GDL	Generalized Dice Loss
GLM	Generalized Linear Model
GANs	Generative Adversarial Networks
GBM	Glioblastoma Multiforme
GBR	Gradient Boosting Regressor
GPU	Graphical Processing Unit
GLCM	Gray Level Co-Occurrence Matrix
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
GM	Grey Matter
GTR	Gross Total Resection
GTV	Gross Tumour Volume
GN	Group Normalization
HD	Hausdorff Distance
HGGs	High-Grade Gliomas
IPP	Image Processing Portal
IN	Instance Normalization
ICAP	Interaction Capping
ICC	Intraclass Correlation Coefficient
IDH	Isocitrate dehydrogenase
JMI	Joint Mutual Information
KNN	K-nearest neighbors

LoG	Laplacian of Gaussian
LLAR	Lasso Least Angle Regression
LRD	Learning Rate Decay
LASSO	Least Absolute Shrinkage and Selection Operator
LAR	Least Angle Regression
LE	Lesion Encoder
LightGBM	Light Gradient Boosting Machine
LR	Linear Regression
LGGs	Low-Grade Gliomas
ML	Machine Learning
MRI	Magnetic Resonance Imaging
MRF	Markov Random Fields
MAE	Mean Absolute Error
MSE	Mean Square Error
MSD	Mean Squared Deviation
mSE	median Squared Error
MICCAI	Medical Image Computing and Computer-Assisted Intervention
MRMR	Minimum Redundancy Maximum Relevance
MF	MultiFiber
MLP	Multi-Layer Perceptron
BraTS	Multimodal Brain Tumour Segmentation Challenge
mpMRI	Multi-parametric MRI
MIFS	Mutual Information Feature Selection
MIM	Mutual Information Maximization
NCR	Necrosis
NGTDM	Neighboring Gray Tone Difference Matrix
NIfTI	Neuroimaging Informatics Technology Initiative
NET	Non-Enhancing Tumor
NA	Not Applicable
NN	Neural Network
MGMT	O(6)-Methylguanine-DNA-methyltransferase
OARs	Organs-At-Risk
OMP	Orthogonal Matching Pursuit
OS	Overall Survival
PAR	Passive Aggressive Regressor
PWI	Perfusion-Weighted Imaging

PNG	Portable Network Graphics
PCA	Principal Component Analysis
RF	Random Forest
RFR	Random Forest Regressor
ReLU	Rectified Linear Unit
RFE	Recursive Feature Elimination
RGBA	Red-Green-Blue-Alpha
ROI	Region Of Interest
RIC	Relative Invasiveness Coefficient
RNC	Relative Necrosis Coefficient
RELF	ReliefF
ResNet	Residual Network
RNN	Residual Neural Network
ResUNet-SEG	Residual UNet for Segmentation
RMSE	Root Mean Square Error
STP	Segmentation Then Prediction
SK	Selective Kernel
srUNet	Self-ensemble ResUNet
std	Standard Deviation
stdSE	standard deviation Squared Error
SWA	Stochastic Weight Averaging
STR	Subtotal Resection
SVR	Support Vector Regression
SP	Survival Prediction
SyN	Symmetric Normalization
T1CE	T1 Contrast Enhanced
T1	T1-weighted
T2-FLAIR	T2 Fluid Attenuation Inversion Recovery
T2	T2-weighted
3D	Three-Dimensional
TNR	True Negative Rate
TPR	True Positive Rate
TC	Tumor Core
T-EDet	Tumour-EfficientDet
2D	Two-Dimensional
UNet-AT	UNet Attention

VAE	Variational Auto Encoder
VOI	Volume Of Interest
WRD	Weight Rate Decay
WCEL	Weighted Cross-Entropy Loss
WM	White Matter
WT	Whole Tumor
WHO	World Health Organization

Chapter 1

Introduction

The brain and spinal cord together form the Central Nervous System (CNS) of the human body. The adult brain weighs approximately three pounds and is firmly wrapped in the skull. It is a complex organ made up of nerve cells (neurons) and supporting (glial) tissues. The brain contains four primary parts: the meninges are the membranes that enclose the brain; the cerebrum is the predominant portion of the brain and is separated into two hemispheres; the cerebellum is the rear section of the brain, situated under the cerebral hemispheres; and the brain stem interacts with the spinal cord. The brain regulates all the functions of our body: our senses, perception, logic, memory, sentiment, reaction, breathing, *etc.* It is the most complicated component of the human body and is made up of billions of cells. The spinal cord is a thin tube of nerve tissue which stretches into the pelvis from the brain down the back. It is covered with bony vertebrae. The spinal cord transmits neural impulses from the brain and the remainder of the body, back and forth. Tumors in the brain and spinal cord are referred to as CNS tumors. Such tumors arise when irregular cells are developed in some portion of the brain or spine [6] as shown in Figure 1.1.



Figure 1.1: Brain Tumor

1.1 Brain Tumor

Uncontrollable growth of abnormal cells in the brain is termed as brain tumor [7]. According to a study conducted in the United States, 23 people out of every 100,000 diagnosed yearly were found to have brain tumors related to the CNS [8]. The categorisation of brain tumors is shown in Figure 1.2.

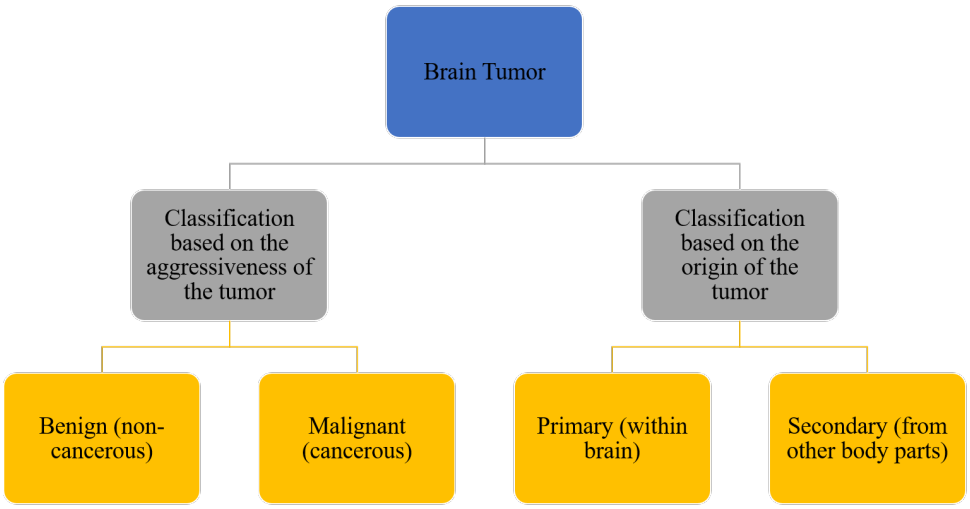


Figure 1.2: Categories of Brain Tumor Based on Aggressiveness and Origin Location

Based on aggressiveness and malignancy, tumors are categorised as benign or malignant in the medical field. A cyst without cancer cells is known to be benign. They could be completely dysfunctional and will not replicate once they have been removed. Benign tumors are less aggressive and do not interfere with the tissues surrounding them. A cancer cell tumor is referred to as malignant, if it becomes more severe and grows quickly with added head pressure. They penetrate surrounding brain tissue as well as other areas of the body, like the spinal cord, beyond where it has already evolved [9]. Brain tumors can also be categorized as primary or secondary on the basis of origin. Primary brain tumors are those tumors that develop from the same brain tissue or nearby underlying tissues. Primary tumors can be either benign or malignant. Secondary or metastatic tumors are generally malignant tumors that originate elsewhere and rapidly migrate towards the brain [10]. There are about 120 types of brain tumors [11].

The exact reason for the development of brain tumors is not known. Doctors give their multiple explanations for the development of brain tumors. Old or destroyed natural cells should die inside the human brain and new cells must substitute them. Perhaps sometimes

the technique can operate in the wrong way. New cells are generated in excess even if they are not needed by the body. In addition, the weak cells sometimes donot vanish. Thus, unchecked cell growth triggers the tumors [12]. Such irregular cell development abrupts the working of regular brain functions and have harmful consequences on the well-being of patients [13]. The brain tumor disorder can result in damage of the CNS. That would either impair the patient's intellectual ability or inflict damage to the patient's psychological health. This contributes to psychological problems and eventually degrades the level of patient's health. Brain tumors can, therefore, contribute towards life-threatening conditions, if not treated properly [14].

1.1.1 Glioma

It is a form of malignant brain tumor which grows from brain or spine glial cells. In the CNS, glial cells are the building block cells of the supporting tissue [15]. Over 60% of brain tumors found in adults are gliomas. Astrocytomas, ependymomas, glioblastoma, medulloblastomas, and oligodendrogliomas are all various types of gliomas [16]. Gliomas are known to be deadliest and aggressively growing brain tumors with the highest risk and incidence of mortality [17]. Gliomas are perhaps the most prevalent brain tumors in adulthood, responsible for a significant percentage among all primary brain tumors, around 40%, and 78% of all instances of malignant brain tumors [18].

For infants, brain tumor mortality levels are about one-fourth relative to other tumors. The estimates claim that the brain tumor is, indeed, the fifth main cause of female's death between twenty to thirty-nine years age group. It was reported that in 2015 alone in the United States, 23,000 additional cases of brain cancer were identified; and 87,240 new cases of major brain tumors were predicted to be identified by 2020 [19]. The medical community primarily emphasizes on glioma, as it mainly impacts the adolescents. Despite advancements in the healthcare sector such as surgical techniques, chemotherapy and radiation therapy, there are still incurable cases of malignant brain tumors [20].

1.1.2 Grades of Gliomas

The World Health Organization (WHO) program has classified CNS tumors for diagnostic and research purposes. The tumors are differentiated on the basis of biological activity and histological form. More than 100 forms of predominant, histologically distinct CNS tumors are distinguished by their own band of clinical observations, therapies, and outcomes [21]. The first version of the CNS's WHO tumor classification was published in 1979. It has been updated four times, most recently in 2016, and is used as the world standard for CNS tumor identification. This classification method serves as a focal point for interactions between

fundamental science researchers and clinicians around the world [22]. The recent version allowed use of modern diagnostic approaches focused solely on microscopy by using molecular criteria to identify CNS tumor entities. Certain cancer forms are staged while the brain tumors are graded as per the WHO tumor classification. The WHO has categorised gliomas into four grades based on the tumor's malignancy, aggressiveness, infiltration, recurrence, and other histology-based characteristics. Low-Grade Gliomas (LGGs) are classified as grade I or grade II gliomas, whereas High-Grade Gliomas (HGGs) are classified as grade III or grade IV gliomas. Their grade level defines the aggressive nature of gliomas. Gliomas of grade I are benign and slow-growing tumors. It is more likely that gliomas of grade II may regrow and expand over time. Grades III and IV, on the other hand, are extraordinarily lethal; and the survival rate is poor [21].

Based on the appearance of gliomas, the whole cancerous area is divided into subcomponents such as Necrosis (NCR), Enhancing Tumor (ET), Non-Enhancing Tumor (NET), and Peritumoral Edema (ED). The Tumor Core (TC) of glioma is comprised of NCR, ET, and NET. The Whole Tumor (WT) of glioma is comprised of NCR, ET, NET and ED. It is important to note from the Magnetic Resonance Imaging (MRI) that LGG does not contain ET in most situations, but HGG does, as shown in Figure 1.3.

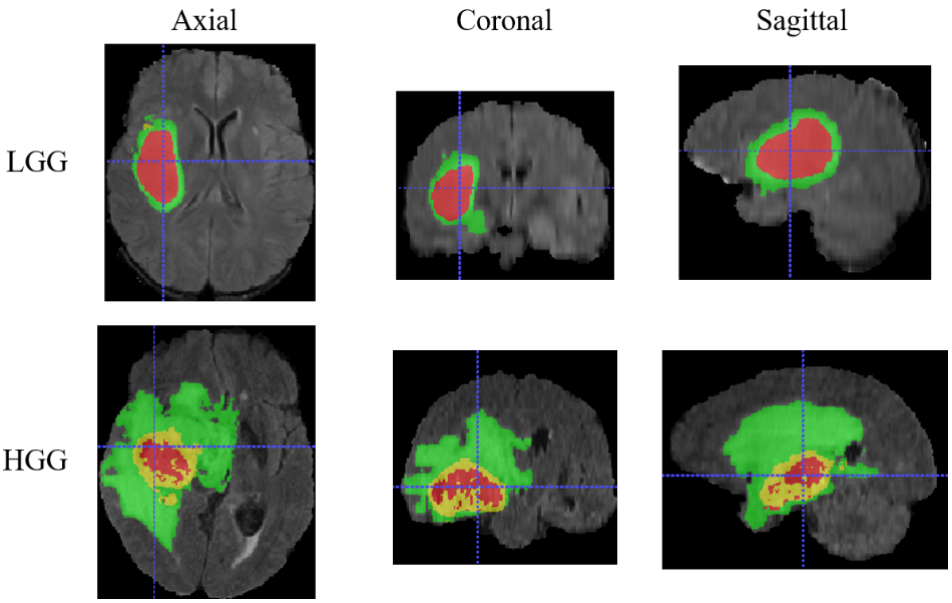


Figure 1.3: Grades of Glioma – LGG (Top row), HGG (Bottom row) [1]

The survival duration of glioma patients is significantly dependent on the tumor grade [23]. The projected five-year and ten-year relative survival rates for patients with malignant brain tumors are 35.0% and 29.3% respectively [8]. HGG patients have shorter survival times than LGG patients. Astrocytomas, oligodendrogliomas, and ependymomas can be graded from grades I to III due to their nature, whereas glioblastoma becomes grade IV tumor. Glioblas-

toma Multiforme (GBM) is the most widespread and lethal glioma cancer in young people [24].

1.1.3 Glioblastoma Multiforme

Glioblastoma is by far the most dangerous and deadliest of all primary CNS tumors, with substantial radiographic heterogeneity represented by numerous sub-regions, including ET, NET, and peritumoral ED as shown in Figure 1.4 [2].

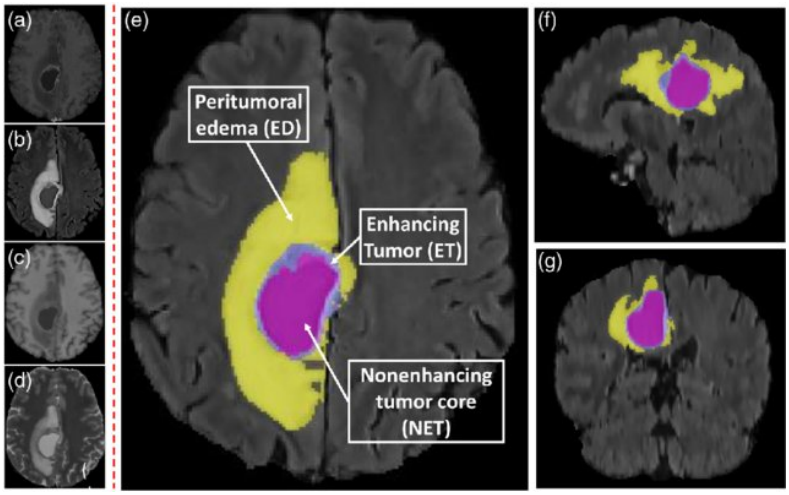


Figure 1.4: Segmentation of Brain Tumor Sub-regions [2]

The above figure contains multiple MRI sequences such as (a) T1 Contrast Enhanced (T1CE), (b) T2 Fluid Attenuation Inversion Recovery (T2-FLAIR), (c) T1-weighted (T1), and (d) T2-weighted (T2) sequences and the segmented peritumoral ED infiltration, ET, and NET core are seen overlaid on (e) axial, (f) sagittal, and (g) coronal view of T1CE sequence. Although there have been several improvements in diagnosis, the median lifespan of GBM-diagnosed patients is still 12–16 months [25]. The cases for GBM are reported to be the largest among all cancerous tumors, about 12,150 cases in 2016 and 12,390 in 2017. GBM accounts for 46.6% of primary malignant brain tumors, and 14.9% of all primary brain and other CNS tumors. The frequency of GBM rises with aging, and with maximum occurrence in age group between 75 to 84 years [24].

1.2 Medical Imaging and Treatment Planning

MRI is frequently used in clinical diagnosis and is an efficient approach for depicting the internal heterogeneity of gliomas utilizing a variety of radiographic traits. Multiple images and appearances of some subjects can be generated simply by modifying the sequence of MRI scanning, which enables the depiction of valuable views of subregions of gliomas using

different modalities. MRI modalities may give supplementary information on the anatomic structures. It has been chiefly employed in clinical uses, such as segmentation of the brain, heart, and intervertebral disc. Each modality is represented by grey-scale illustrations highlighting various types of tissue. Native T1 and T1CE provide a clear contrast between grey and white matter regions, emphasizing the tumor, *i.e.*, NCR, ET and NET regions without ED. T2 and T2-FLAIR images improve the contrast of the ED in its totality [26], as shown in Figure 1.5. As a result, the utilization of multi-parametric MRI (mpMRI) for the segmentation of brain tumors has gained significant attention.

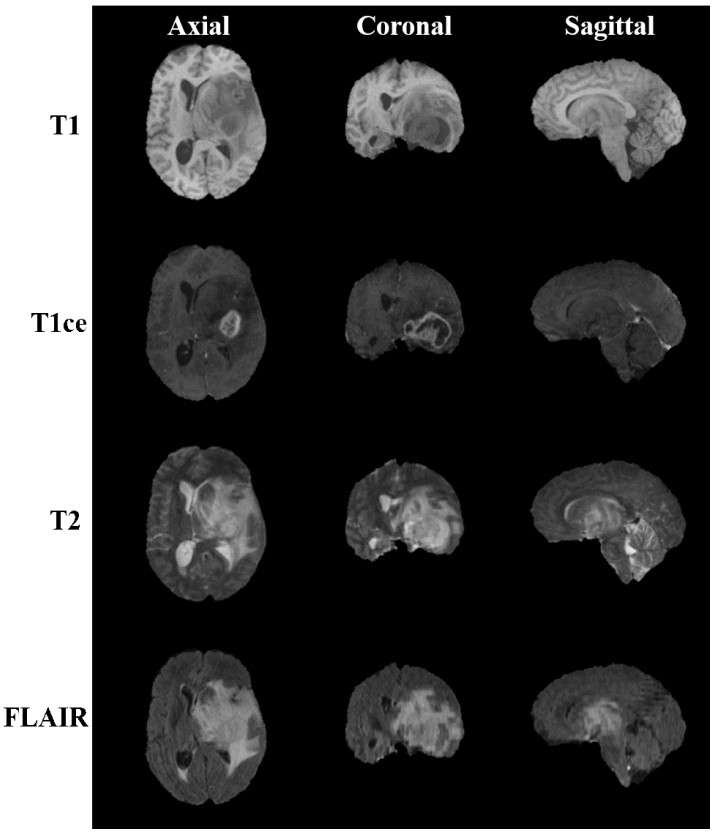


Figure 1.5: Different Multimodal MRI Scans- T1, T1CE, T2, T2-FLAIR

Brain tumor occurrence has progressed over time, varies depending on age, sex, ethnic origin and region. Glioma diagnosis is a very challenging job because of the different form, size and location of gliomas. The use of the new medical technologies will support the clinical population during the diagnostic assessment phase in the early diagnosis and evaluation of gliomas. With the emergence of advanced techniques for diagnostic imaging, the need of neurosurgeons, medical services, and different treatment methods is also intensified. In addition, the histological advancements made in the identification of different brain tumors have contributed a lot towards the treatment procedure of brain tumor disease [16].

Brain imaging is quite useful to provide a clear picture about the physiological conditions

by generating detailed representations of the human body's internal organs or tissues. A well-organized diagnostic evaluation and qualified diagnosis are important for determining the position and extent of brain abnormality in order to schedule an effective treatment [27]. The MRI scan is widely used to diagnose gliomas in clinical practice owing to high definition multiplanar anatomical details and significantly enhanced tissue characterization. When the origin and existence of the glioma is detected, then the appropriate treatment plan can be followed to heal the abnormality. The oncologist should schedule one of the following medical treatments upon finding the tumor; (i) radiotherapy, (ii) chemotherapy, and (iii) tumor resection. Radiation and chemotherapy are prescribed to delay the tumor development and the tumor resection can be used to eliminate the tumor area entirely [21, 28]. The patient is exposed to radiation beams during radiation therapy, usually from various directions and with specific intensity levels, with the aim to optimize the radiation exposure transmitted to the intended tumor while reducing the exposure to susceptible healthy parts called as Organs-At-Risk (OARs). In order to conduct a radiation therapy session, OARs need to be identified with the help of Computed Tomography (CT) or MRI of the patient's brain [29].

Understanding the size and scale of a brain tumor is of utmost importance in determining its progression, treatment response and surgery preparation. The physicists are currently calculating the large tumor diameter and resembling it as an ellipsoid, a highly inaccurate calculation that appears to exaggerate the size of the tumor. The segmentation technique is used to remove the infected area using the brain MRI scan. Many methods have been used to separate pathological cells from regular cells (such as grey matter, white matter, and cerebrospinal fluid). From high-resolution MRI scans, it is possible to obtain detailed information about brain anatomy and cell organization in the brain, from where defects in tissue formation can be quickly identified [30].

1.2.1 Manual and Automated Glioma Delineation

Expert radiologists gain full details on the brain tumor by multi-modal MR analysis in medical practice, and accomplish the delineation process manually. A non-automated segmentation procedure suffers from large inconsistencies, not just between separate radiologists as well as by the same specialist. For example, Porz *et al.* [31] reported an aggregate deviation of 16% to 40% among individuals who conduct the same brain tumor delineation and an average deviation of 5% to 35% among individuals who conduct the delineation thrice in one-month intervals.

In fact, the neoplastic tumor can differ greatly in anatomical properties, position, and extent of progression among patients, making segmentation and diagnosis of brain anomalies

a very complex, complicated, and time-consuming job. In specific, certain neoplasms alter the underlying healthy structures or occur in combination with edema, triggering variations in the intensity values of the nearby structures. MRI scans have inherent acquisition issues about inhomogeneity and heterogeneity of the intensity spectrum between sequences and scanners. Clinical specialists also describe the intrinsic nature of the tumor tissue on the basis of visual analysis of 2D organ slices with minimal understanding of the detailed three-dimensional (3D) brain layout [32]. Consequently, the automated segmentation of tissue anomalies from medical artifacts is, indeed, a crucial part of computer-aided diagnostic imaging developments of the next decade.

Depending on the extent of human participation, Brain Tumor Segmentation (BTS) may be divided into three categories, *viz. manual* segmentation, *semi-automated* segmentation, and *fully automatic* segmentation. A human operator employs specialized tools to precisely sketch or paint around tumor locations during **manual segmentation**. The accuracy of segmentation findings is strongly dependent on the human operator's training and expertise, as well as understanding of brain anatomy. Manual segmentation is commonly utilized as a gold standard for semi-automated and completely automatic segmentation, despite being laborious and time demanding. **Semi-automated segmentation** blends human and computer knowledge. User engagement is required for the segmentation process to begin, as well as for feedback and assessment of segmentation outcomes. Although semi-automatic segmentation approaches save time over manual segmentation, the outcomes are still dependent on the operator. There is no need for human intervention in **fully autonomous** BTS. To handle segmentation difficulties, artificial intelligence and previous knowledge are blended [33].

Current automated and semi-automatic glioma segmentation methods may be generally classified as either generative or discriminative methods [1]. Generative-based glioma segmentation approaches usually involve detailed knowledge regarding the tumor and its tissue and morphology, that could be collected from probabilistic image atlases [34]. The generative model has been well explored on the unseen records, but it is still challenging to directly turn prior information into a suitable probable model, as such models allow many parameters of tumor characteristics to be reported [35]. The discriminative models learn tumor characteristics explicitly from radiologist's manually annotated MRI scans. The discriminative model-based techniques address the issue of tumor segmentation as a pattern of identification problem, *i.e.*, classification of image voxels as tumors or regular tissues based on image features [36].

In case of brain tumors, image segmentation can help to assess tumor volume rapidly and accurately, and can also recognise patient-specific characteristics that help diagnose and

schedule treatment. Scientists have developed automatic methods to solve three major problems to help both radiologists and oncologists; (i) to minimize uncertainty during the diagnosis of the glioma; (ii) to include non-invasive analysis of suspicious cells and also to examine their impact on the patient's body; and (iii) to analyse a vast amount of medical data at shorter time and to promote minimal human intervention in the processing of medical images [37]. Because of noise disturbance, bias field distortion, and tissue movement during MRI acquisition, the image becomes distorted, and the grayscale intensities are not the standard. In addition, the number of images produced by MRI is enormous; the brain tumor seems to have an infiltrative expansion without any apparent boundaries or other features; and several automated segmentation mechanisms are not reasonably competent [38].

1.2.2 Difficulties in Automated Glioma Delineation

Automatic segmentation in gliomas is, indeed, a difficult task. A wide range of explanations exist on this issue. One concern is that the forms, compositions, and positions of these brain tumors are extremely unpredictable [21]. The influence of tumor density affects the composition of the underlying healthy tissues [1]. Certain concerns, such as lack of homogeneity in intensities [39], or varying intensity levels between the same modalities and acquisition scanners, may also influence the precision in segmentation. Furthermore, when viewed in various facilities, the same tumor cells can end up with significantly different high contrast values [28]. Hence, accurate and automated segmentation of the brain tumor is still a great concern.

Instead of one MRI modality, multiple MRI modalities are required to resolve certain difficulties. Various MRI sequences display specific image detail from various sides and illustrate different objects. A single modality is challenging to have adequate distinctive details on lesion tissues, but all modalities such as Flair, T1, T1CE, and T2 are typically used as a diagnostic basis for BTS [40]. For building efficient computer-aided programs, the learning of a well-generalizing classifier includes large manually annotated datasets. Unfortunately, the minimal accessibility of massive image databases manually annotated by radiologists is typically a big obstacle in the realistic implementation of such approaches [41].

The automated segmentation of brain tumors is a tough process owing to tumor tissue heterogeneity. It also suffers from class imbalance, for example, the tumor occupies a small volume in comparison to other areas of the brain. This is evident from Table 1.1 which demonstrates that normal tissues take up 98% of the volume of the brain. In addition to this, there is a lack of homogeneity across the different datasets, as well as the absence of large-scale annotated multimodal datasets [5]. Therefore, it may be difficult to identify mi-

nor changes in molecular pictures by eye examination as the target-to-background ratio in these images is not that much significant. In addition, visual interpretation by physicians is not only time-consuming, but also often results in substantial differences across interpreters owing to their varying levels of competence. The developing intelligent approaches have significant promise for overcoming these issues by automating picture interpretation.

Table 1.1: Class Imbalances in the Dataset [5]

Labels	**% of Brain Volume**
Normal Tissues	98.00
Necrotic	00.18
Peritumoral Edema	01.10
Non-Enhancing Tumor	00.12
Active Tumor	00.18

In the field of medical imaging analysis, Machine Learning (ML) based image processing has been extensively used. Conventional ML algorithms need the artificial intervention of feature extraction and selection, and are thus, subjective to some degree. Moreover, subtle and dispersed changes may be disregarded using artificial feature computation and selection. It is anticipated that fully automated approaches would incorporate local and global information for more precise interpretation. Deep Learning (DL), as a state-of-the-art ML approach, may be able to overcome the aforementioned obstacles by abstracting higher-level information and enhancing predictions from data using deep and sophisticated Neural Network (NN) topologies [42].

1.2.3 Survival Prediction

Despite rigorous treatment including radiation, temozolomide, and surgery, almost all patients with GBM will have a recurrence. It may be difficult to resect these tumors, and in certain situations, they may even be inoperable due to their spatial and temporal intratumor heterogeneity, magnitude, and location. Lack of effective therapy and poor prognosis are significantly influenced by the inability to execute full surgical tumor excision and inadequate medication transport to the brain. In addition to some host characteristics like age and gender, it has been demonstrated that certain biological variables, such as the state of the O(6)-Methylguanine-DNA-methyltransferase (MGMT) promoter methylation, the 1p/19q deletion, and the IDH1 gene mutation status, might explain to some degree this observed difference to some extent. GBM's widely varied clinical behavior suggests that the present concept of GBM as a single disease entity is inadequate. This is further confirmed by the fact that several efforts have been made, using methods such as molecular subtyping, to separate GBM into a greater number of subgroups that are more different. Due to

the complication of the tests and the high cost, however, these procedures are difficult to repeat in ordinary clinical practice. Also, tumors can change from one subtype to another in a complicated way as the disease gets worse. Automatic estimation of the ultimate survival of patients with glioblastoma is an important yet unresolved problem, with no proven procedure in clinical practice. Predicting overall survival in brain tumor patients using multimodal MRI is a tough challenge.

MRI scan plays a vital role in neuro-oncology for initial diagnosis and assessment of treatment response and is increasingly used as a powerful non-invasive predictive tool. Researchers have identified that MRI provides distinct information that can predict survival independently of pathologic and clinical data. The process that extracts various quantitative features on the basis of intensity, volume, shape, and textural variations from radiographic images and design predictive algorithms to find the association of these vast features to the survival and outcome of the patient is known as radiomics. Radiomics incorporates several essential disciplines, including radiology for imaging interpretation, computer vision for quantitative feature extraction, and ML for classifier evaluation and regression.

In particular, for the survival forecast among the specialist physicians, the precision of categorical predictions ranged from 23 to 78%. A great variation in the estimation of survival is found; and even experienced physicians are surprisingly weak in predicting the survival time of the patient. Whether the patient's eventual survival can be correctly estimated on the basis of pre-operative scans utilizing information variables such as radiomics, tumor position, and tumor form remains an open issue [43].

1.3 Current Research Trends

Automated ML-powered programs eliminate the tedious, and lengthy delineation process performed by professionals manually. It is reproducible, as well as much quicker than professional delineation. The most significant benefit of such a program is that it will ultimately assess the right diagnosis, procedure or recovery strategy and track the disease progression. Promising development is achieved utilizing the conventional methods of ML, which measured handcrafted, radiological features at low rates to classify images, and trained a classifier or regressor for glioma delineation and prediction of Overall Survival (OS). All such handcrafted characteristics were typically described by professional neuro-radiologists based on advanced awareness of the precise histological treatment of glioma which may had been a possible source of prejudice. This quick, direct approach of evaluating features often theoretically overlooked a lot of valuable details hidden inside the MR images, preventing the maximum efficient use of sub-region delineation for predicting overall survival

[36, 44].

DL has advanced the state-of-the-art in several areas of computer vision [45–52] and pattern recognition in different source-based images [53–56], including medical image analysis [57–60]. To develop high-quality models, such strategies automatically find the underlying data representation. Deep features may capture significantly discriminative properties and explanatory variables of the data that were overlooked or unknown during manual feature building [3]. Advancements in DL have greatly increased the efficiency of automatic medical image delineation. Convolutional Neural Network (CNN) approaches, instead of identifying handcrafted attributes, collectively train feature extractor and classifier to dynamically extract high-yield knowledge and boost model efficiency. Influenced by the superior findings of this approach, clinicians are gradually utilizing CNN for the segmentation of brain tumors, and also for the estimation of patient's overall survival. A very well-trained CNN framework will complete the segmentation process with reasonable precision in minutes. There are several difficulties, including restricted professionally-annotated training datasets, differences in procedures for image acquisition and MRI scanning *etc* [61, 62]. Despite the fairly successful performance of automated glioma segmentation, the findings of OS estimates are still far from acceptance in clinical practice.

Due to their inadequacy to analyze unprocessed natural data, the capacity of standard ML approaches is restricted. To turn raw data (such as pixels in the case of an image) into a feature vector, domain specialists must create a feature extractor. Then, a learning model, also known as a classifier, may be applied to the input data to find or categorize patterns. In contrast, DL algorithms may automatically extract from raw data the necessary features or representations for detection or classification. DL approaches learn representations via many layers, where each layer performs a basic non-linear modification on the representation of the previous layer to produce a more abstract representation for the subsequent layer. The top layer retrieves those characteristics from the lower layer which are more significant and useful for discriminating and suppressing changes that are not significant. Consider the challenge of recognizing an item in a picture consisting of an array of pixels, for instance. The first layer of the DL model discovers the presence or absence of edges in certain directions and picture locations. Typically, the second layer finds patterns by detecting certain edge configurations and disregarding minor edge location differences. The third layer may build patterns into enormous combinations that correspond to a subset of recognizable items. Objects will be detected by following layers as a mixture of these components [63].

The primary distinction between conventional ML and DL techniques is how features are retrieved from input data. Traditional ML approaches use manually produced features col-

lected through the application of feature selection and extraction algorithms. DL approaches extract characteristics from various layers hierarchically. Therefore, researchers no longer need to manually discover and extract key characteristics from the supplied data. With adequate training data, classical ML approaches provide better outcomes. But after a certain threshold, when the quantity of training data rises, the performance of ML algorithms no longer improves and becomes stagnant, while in the case of DL algorithms, their performance often improves as training data grows [63].

To effectively construct well-generalizing deep models, a massive quantity of ground-truth data is required to prevent overfitting of such large-capacity learners and "memorization" of training sets. In the context of brain-tumor delineation from MRI, gathering high-quality ground-truth data is time-consuming, costly, and highly human-dependent, posing a substantial challenge for the use of Deep Neural Networks (DNNs) in the area of medical image processing. Also, most of the manually annotated image sets aren't balanced, which means that examples from some classes are often under-represented. Data augmentation approaches, which provide synthetic training examples, are currently being developed in the literature to address the issue of small medical training sets [3].

Automatic segmentation of glioma sub-regions (NCR, ED, and tumor enhancement) and reliable estimation of survival duration from multiple MRI modalities provide significant clinical importance in glioma detection, prognosis, and treatment. Nonetheless, glioma segmentation and OS estimation are quite difficult due to the high degree differences in heterogeneous presentation and human physical condition. Therefore, the creation of a robust and quickly optimized system ultimately results in a considerably more precise diagnosis, allowing significant long-term progress in patient's treatment planning. This system is more emphasized as the computer has a better interpretation of 3D images, instead of a traditional two-dimensional (2D) perspective of a radiologist.

1.4 Organisation of the Thesis

The thesis is organized into six chapters. A brief outline of these chapters is given below:

Chapter 1: Introduction
This chapter provides the relevant information about the various aspects and issues of current research work. Here, we can have an overview of different grades of gliomas, followed by significance of medical imaging in treatment planning and Survival Prediction (SP) of GBM patients. The challenges of manual and automated glioma delineation have also been discussed. Further, the current research trends in the field of BTS and SP have been deeply studied.

Chapter 2: Review of Literature

This chapter reviews the available literature on the existing automated techniques for SP of GBM patients using 3D MRI scans and some clinical information. A detailed explanation of available data relating to the steps required for OS prediction is given. A generic workflow is presented in the chapter with different approaches followed by different researchers at each step of the complete process. Further, the research gaps found after reviewing the previous studies, findings followed by problem formulation, and set objectives of this research work have also been explained. Finally, it briefly describes the research methodology followed to achieve these objectives.

Chapter 3: Radiomic Features-based Survival Prognosis of Glioblastoma Multiforme Patients

In this chapter, a retrospective analysis has been made using data of glioma patients made publicly available by the University of Pennsylvania. The radiomic characteristics were extracted from pre-operative structural mpMRI sequences after pre-processing and segmentation using 3D UNet DL architecture. After the removal of irrelevant features, regression models based on ML were trained by considering the selected features to predict the OS time of GBM patients within a period of days only. Finally, the results were analyzed through different performance metrics such as Accuracy, Mean Square Error (MSE), median Squared Error (mSE), standard deviation Squared Error (stdSE) and Spearman Rank Correlation (ρ).

Chapter 4: Deep Learning Features-based Survival Prognosis of Glioblastoma Multiforme Patients

This chapter proposes DL-based 3D U-shaped CNN based on encoder-decoder architecture to segment the brain tumor. Further, feature extraction was performed on these segmented and raw MRI scans using a pre-trained 2D Residual Neural Network (RNN). The dimension reduced principal components were integrated with clinical data and the handcrafted features of tumor subregions to compare the performance of regression-based automated ML techniques.

Chapter 5: Automated Neural Network-based Survival Prognosis of Glioblastoma Multiforme Patients

This chapter proposes a methodology based on 2D ResUNet-SEG plus 2D CNN-SP architectures; and the results have been compared with the existing 2D models using different performance metrics. It is possible to convert the 3D MRI volumes into 2D images to train 2D DL-based CNN models. The lightweight 2D models help in reducing the memory consumption and processing time.

Chapter 6: Conclusion and Scope for Further Research

This chapter draws the conclusions from the findings of current work. It also provides the

necessary directions for undertaking further research on the topic under investigation.

Chapter 2

Review of Literature

Glioblastoma is a grade IV brain tumor with very low life expectancy. Physicians and oncologists urgently require automated techniques in clinics for BTS and SP of GBM patients to perform precise surgery, followed by chemotherapy treatment. This chapter aims at examining the recent methodologies which used automated learning and radiomics to automate the process of SP. Automated techniques use pre-operative raw MRI scans and clinical data related to GBM patients. All the SP methods submitted for the Multimodal Brain Tumor Segmentation (BraTS) Challenge in different years have been examined to extract the generic workflow for SP. The maximum accuracy achieved by twenty-one state-of-the-art different latest SP techniques on the biggest version of BraTS 2020 dataset ranges between 61.7% and 65.5% using the testing and validation subsets, respectively [64, 65]. The comparisons have been made on the basis of segmentation architectures, SP models, training parameters and hardware configurations. The limited accuracies achieved in the literature led us to review the various automated methodologies and evaluation metrics to find out the research gaps, and undertake further research related to the survival prognosis of GBM patients for improving these accuracies in future.

2.1 Review Methodology

The literature review is focused mainly on studying the proposed latest techniques using the 2017, 2018, 2019, and 2020 versions of BraTS dataset. The dataset is available online on https://www.med.upenn.edu/cbica/brats2020/data.html. The google scholar was used to find the required articles for the purpose of literature review. These research papers were obtained by conducting the search with keywords “brats”, “brain tumor”, “segmentation”, “survival” and “prediction”. Then, the obtained papers were selected on the basis of survival data, abstract, basic methodology techniques and evaluation metrics.

The use of 3D MRI scans to segment gliomas helped in diagnosing and treating the glioma patients. The BraTS challenge has been held yearly since 2012 as part of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) to assess current segmentation methods for pre-operative MRI-based BTS. The challenge

makes medical data of GBM patients freely accessible for academic research purposes. This dataset contains a vast collection of pre-operative mpMRI scans and clinical data with corresponding expert-derived ground-truth annotations across four MRI modalities [1, 66–69]. A considerable amount of research in BTS has been performed using the BraTS dataset [21, 28, 70]. The data is gathered from various sources, organisations, and scanners, including both LGGs and HGGs. The dataset contains training, validation, and testing data. For training data, segmentation ground-truth, and patient' s survival information is provided. Validation and testing data, on the other hand, are devoid of segmentation indicators. Researchers can submit their results to the Centre for Biomedical Image Computing and Analytics (CBICA) Online Image Processing Portal (IPP) to evaluate validation and test data performance.

The BraTS challenge 2020 consists of two main tasks: segmentation of glioma brain tumor sub-regions, and prediction of OS. The objective of the BTS task is to define the glioma and its internal components, including the WT, ET and TC. The SP task predicts the patient's survival time using the characteristics derived from MR imaging and clinical data. In recent years, the BraTS challenge has encouraged researchers to submit automated BTS and SP methods for evaluation and discussion using a publicly accessible mpMRI dataset [1]. In the computer-aided diagnosis of gliomas, accurate BTS and SP are two crucial yet difficult objectives. Traditionally, these two activities were carried out independently, with little consideration given to their relationship. In order to extract robust quantitative imaging characteristics from a glioma image, it is necessary to accurately segment the three main sub-regions of the tumor image, *viz.*, ET, TC and WT. Researchers think that these operations should be carried out in a cohesive framework to maximise the benefit.

2.1.1 Dataset

The objective of the BraTS 2020 challenge was to evaluate novel BTS and SP algorithms using pre-operative mpMRI scans and clinical data. In each patient's MRI scans, *i.e.*, T1, T2, T1CE, and FLAIR available in Neuroimaging Informatics Technology Initiative (NIfTI) format are co-registered to the anatomical template of the T1 modality of the same patient as shown in Figure 2.1.

Various clinical techniques and scanners were used to collect data from a range of institutions [1, 66, 67]. All patients in the training set are labeled for the three tumor tissues, *i.e.*, ET, ED, and NCR/NET. All patient scans are skull-stripped and interpolated to 1 mm isotropic resolution. One to four evaluators segmented the images using the same labelling technique to prepare the ground-truth annotations. The annotations were validated by experienced neuroradiologists [68, 69]. The details are summarized in Table 2.1.

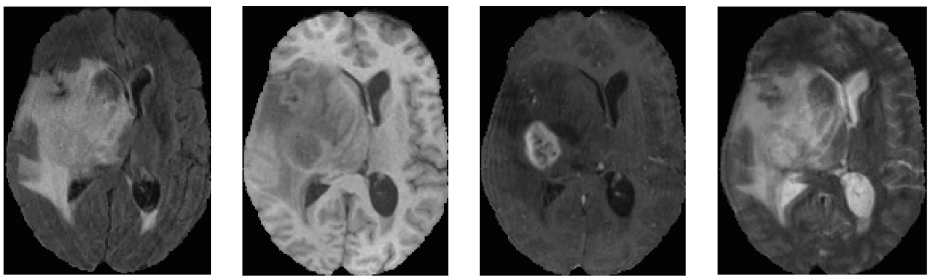


Figure 2.1: Different Multimodal MRI Scans (a) FLAIR (b) T1 (c) T1CE (d) T2 [1]

Table 2.1: BraTS 2020 Data for Brain Tumor Segmentation

Data	**Training**	**Validation**	**Testing**	**MRI voxel spacing**	**MRI dimensions**
BraTS pre-operative MRI scans	369 (293 HGG, 76 LGG)	125	166	$1 \times 1 \times 1$	$240 \times 240 \times 155$

There were 236 patient's survival statistics, including their IDs, age, OS (in days), and surgery status. Gross Total Resection (GTR) surgery was performed on 119 of the 236 patient's data, while the remainder underwent Subtotal Resection (STR) or Not Applicable (NA) operation. Out of 119 data with GTR status, one patient is alive; therefore, the data used for survival training is of 118 patients by most of the researchers. Based on this decreased number of observations, researchers created models to predict patient' s OS. The probability of survival days is divided into three categories: *short-term*, *mid-term*, and *long-term*. However, only patients with resection status GTR are considered for evaluation. The performance of proposed models can be assessed using validation and testing data. The details are highlighted in Table 2.2.

Table 2.2: BraTS 2020 Data for Overall Survival Prediction

Data	**Training**	**Validation**	**Testing**
Clinical data (.csv data)	236 (119 GTR, 89 short, 60 mid and 87 long term survivors)	29	107

2.1.2 Brain Tumor Segmentation

Researchers solved the BTS task using clinically obtained training data and generating segmentation labels of the different glioma sub-regions using DL and ML models. ET, TC, and WT are the three sub-regions that are being examined in BraTS 2020 BTS. As illustrated in

Figure 2.2, the provided segmentation labels have values of 1 for NCR and NET (red colour), 2 for ED (green colour), 4 for ET (yellow colour), and 0 for background (black colour). Researchers submit their segmentation labels for assessment as a single multi-label file in NIfTI format to CBICA IPP.

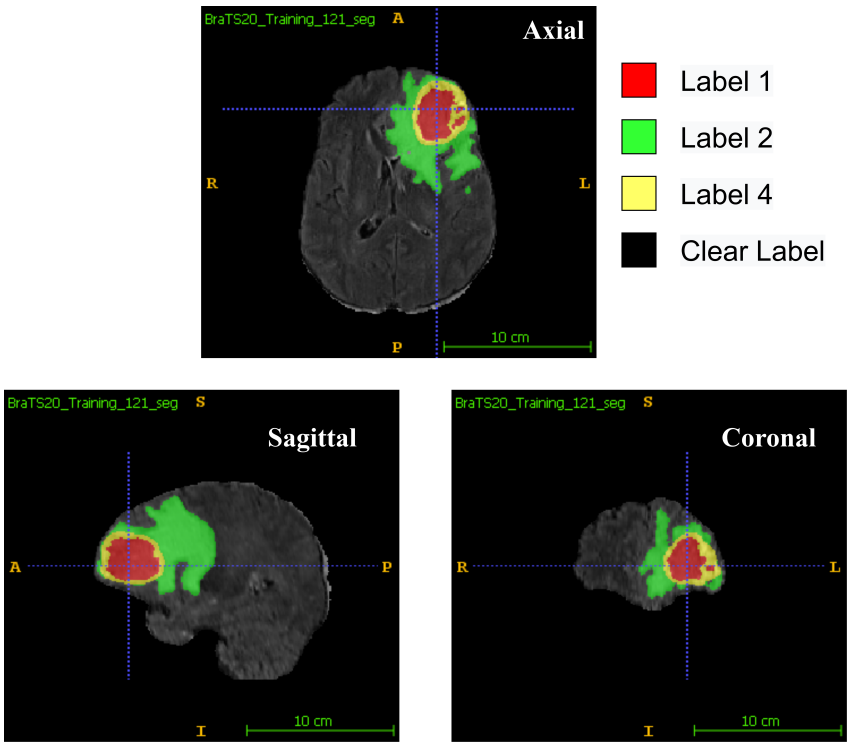


Figure 2.2: MRI Scans with Segmentation Labels in Multiple Views [1]

2.1.3 Survival Prediction

The segmentation labels generated after performing BTS can be used to extract imaging or radiomic features which can be used further to train ML and DL models to forecast the OS of GBM patients (in days), as illustrated in Figure 2.3. It is not essential to limit the parameters to just volumetric features only. Researchers can even use intensity, histological, textural, spatial, glioma diffusion attributes and histogram-based features along with clinical features like patient' s age and resection status. The SP task requires a prediction of OS for patients with GTR. The prediction should be the number of days, and validation of the algorithm is dependent on the accuracy of categorization into three categories, *viz.*, long-survivors (greater than fifteen months), short-survivors (less than ten months), and mid-survivors (within ten and fifteen months). The prediction results obtained by researchers are submitted as a Comma Separated Value (CSV) file, including the subject ids to CBICA's IPP.

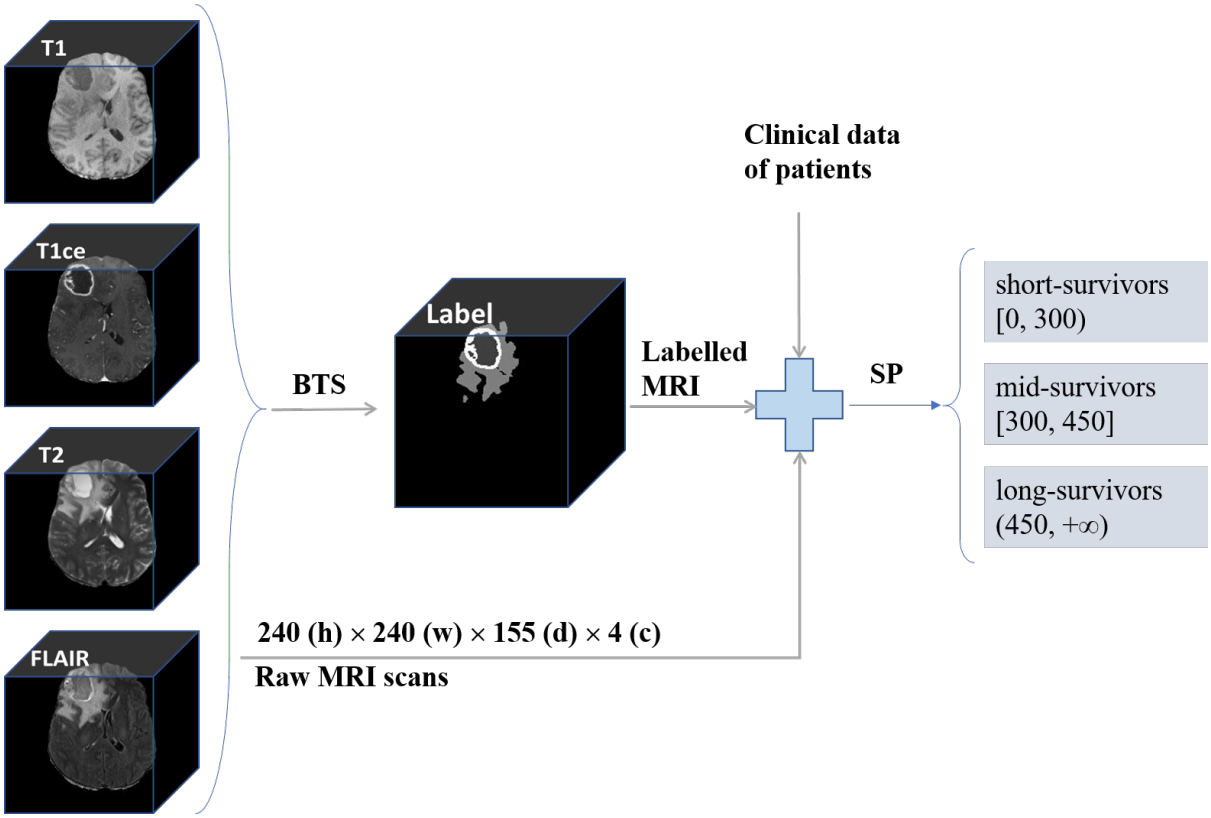


Figure 2.3: Overview of Steps for Survival Prognosis [1]

2.2 Generic Workflow for Brain Tumor Segmentation and Survival Prediction

Various end-to-end approaches for BTS and SP have been proposed in the literature. In some manner, all of these approaches assert their supremacy and usefulness above the others. It has been observed that the automated approaches accomplish the objectives of BTS and SP by following a similar set of steps. This section explains the generic workflow followed to predict the OS of patients, as shown in Figure 2.4.

2.2.1 Pre-processing and Data Augmentation

Deep CNNs (DCNNs) are data operating algorithms. These algorithms require vast data in order to come up with accurate conclusions. Pre-processing techniques can enhance critical data inside the original input scan, such as cropping the area of interest when it includes redundant or misleading information alongwith standardising input data, .

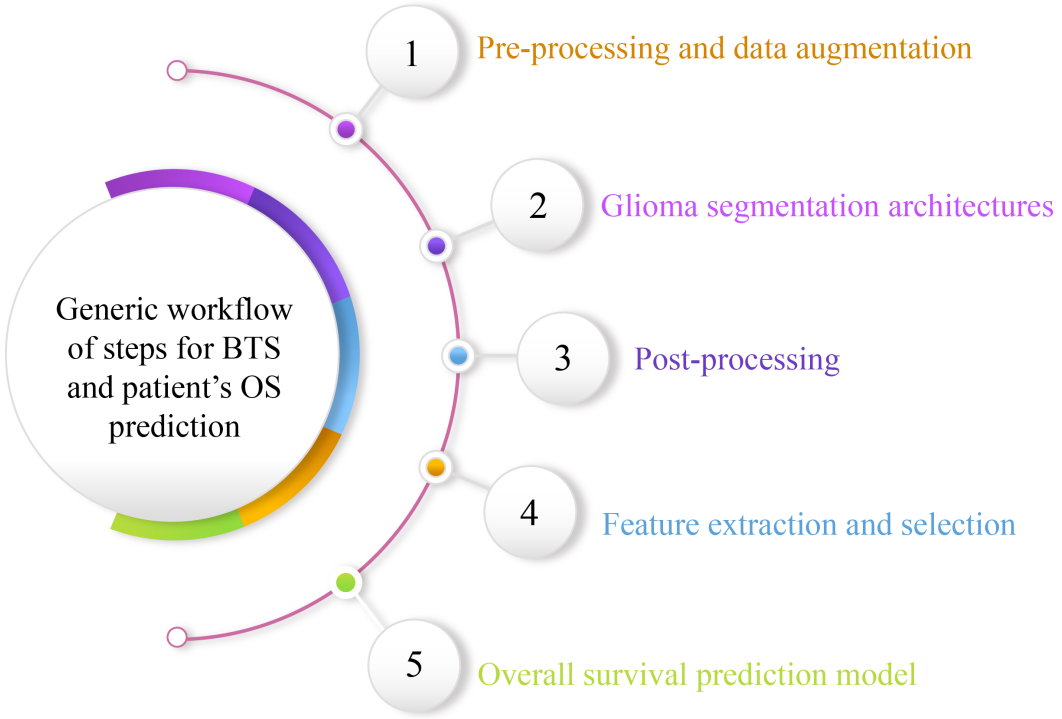


Figure 2.4: The Generic Workflow for Brain Tumor Segmentation and Overall Survival Prediction

2.2.1.1 Pre-processing

The BraTS 2020 challenge organisers provided pre-operative MRI scans and ground-truth annotations for the training set. To minimise variability in patient scans induced by imaging procedures, the challenge coordinators supplied data that had been co-registered to the T1 anatomical template of individual patients, normalised to 1mm isotropic resolution, then skull-stripped. Pre-processing improves network performance and training. As shown in Figure 2.5, the authors [39, 71–87] employed the following pre-processing methods to account for intensity inhomogeneity throughout the dataset.

(i) ***Min-max normalization:*** One of the most frequent methods for normalising data is min-max normalisation. '0' and '1' are the minimum and maximum values for each characteristic, respectively, and all other values are between 0 and 1; and is calculated using *Eq.* 2.1.

$$new_{image} = \frac{value(image) - minimum}{maximum - minimum} \tag{2.1}$$

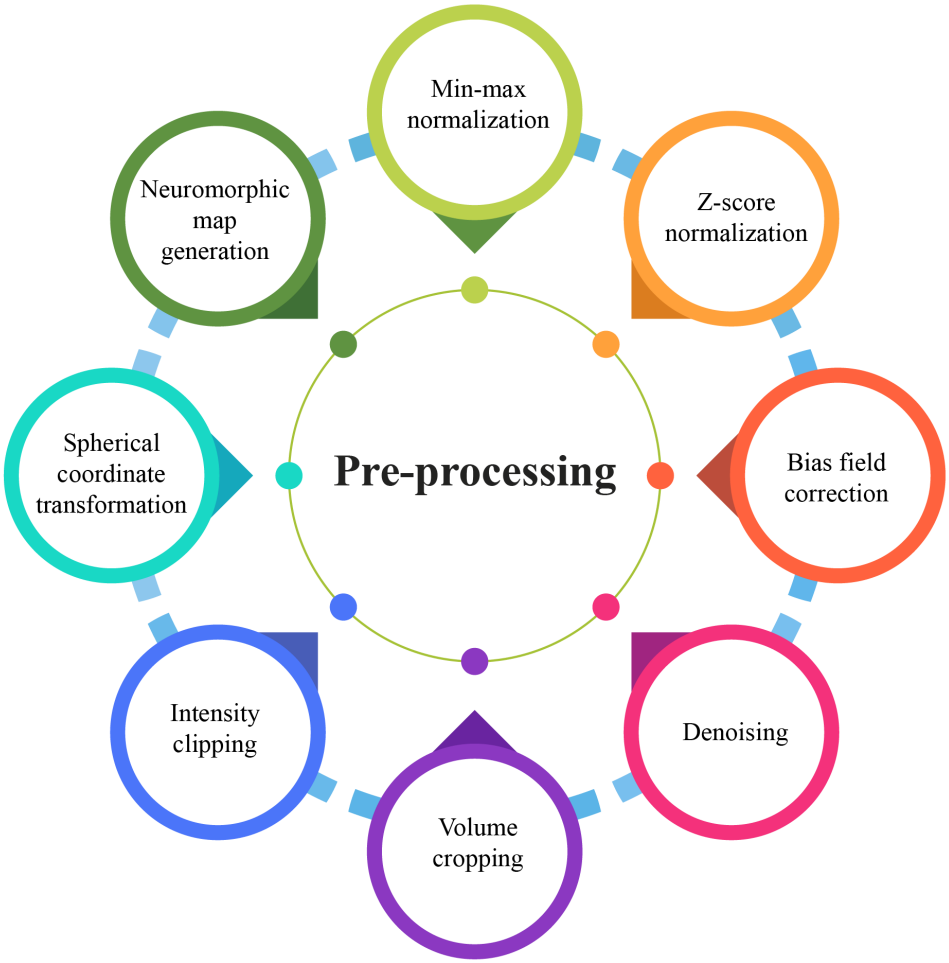


Figure 2.5: Types of Pre-processing Used in BraTS 2020 SP Techniques

where: $value(image)$ = only the brain volume's area,
$minimum$ = minimum intensity value,
$maximum$ = maximum intensity value.

Although min-max normalisation ensures that all features have the same scale, it does not handle outliers well. Before feeding the MRIs as input for model training, González *et al.* [71], Parmar *et al.* [72], and Carmo *et al.* [73] normalised each MRI sequence in the range [0,1].

(ii) ***Z-score normalization:*** Researchers use z-score intensity normalisation to minimise intensity variation within MRIs of different patients and variations among multiple modalities of the same patient. Z-score normalisation is a data normalisation method that eliminates the outlier problem. According to the *Eq.* 2.2, z-score normalisation ensures intensity with a zero mean (mean) and a unit standard deviation (std) for each volume (considering only non-zero voxels):

$$image = \frac{(value(image) - mean(image))}{std(image)} \tag{2.2}$$

where: $value(image)$ = only the brain volume's area,
$mean(image)$ = mean of the brain volume,
$std(image)$ = standard deviation of the brain volume.

If the parameter *'value(image)'* is the same as the mean of all feature values, it is normalised to '0' ; if it is less than the mean, it is a negative number; and if it is greater than the mean, it is a positive number. The original feature's standard deviation decides the magnitude of the negative and positive values. If the standard deviation of abnormal data is big, the normalised values will be closer to '0' . It deals with outliers but does not generate normalised data on the same scale. Researchers used z-score normalization to set each MRI volume's mean and unit standard deviation before these volumes were fed into training models [71–84].

(iii) ***Bias field correction:*** In MRI scans, low-frequency bias field signals are very smooth that cause MRI corruption, especially when the MRI scanners are obsolete. There are differences in the magnetic field strength of MRI scans captured with different procedures and scanners across various institutions. Segmentation architectures, textural analysis, and classification techniques that rely on the grey-level values of image pixels will not provide satisfactory results. It is necessary to correct the bias field signals before submitting raw MRI data to automated algorithms. Agravat *et al.* [75], Patel *et al.* [77], and Soltaninejad *et al.* [82] used N4ITK Insight Toolkit [39] to reduce bias field in all available structural MRIs.

(iv) ***Denoising:*** MRI scans are often corrupted by gaussian noise produced by the random thermal motion of electrical components and decreases image quality and reliability. Noise reduction can be achieved using a variety of noise filtering methods in order to improve image analysis. Agravat *et al.* [75] used denoising as a pre-processing step before training the segmentation architecture.

(v) ***Volume cropping:*** Patel *et al.* [77] firmly cropped all MRI scans to eliminate empty voxels outside the brain area. The MRI data used for training and generating inference has been pre-truncated to a size of $160 \times 192 \times 128$ from the original image's centre point. MRI volumes were similarly cropped by Akbar *et al.* [84] and González *et al.* [71] from the central point.

(vi) ***Intensity clipping:*** Pang *et al.* [85] reduced the amount of background region by using intensity trimming. Carmo *et al.* [73] also clipped MRI sequences within the interval

[-5,5].

(vii) ***Spherical coordinate transformation:*** Russo *et al.* [86] proposed spherical coordinate transformation as a pre-processing stage to improve segmentation results. Pre-training weights for future training stages were achievable due to the utilisation of spherical transformation in the first cascade pass. Extreme augmentation was also added via the spherical coordinate transformation. It was a practical step since it provides invariance for rotation and scaling. However, such invariance has disadvantages, mainly when dealing with WT segmentation; it introduces many false-positive areas. As a result, the researchers employed a Cartesian model to filter away false-positive areas discovered via spherical pre-processing.

(viii) ***Neuromorphic map generation:*** Han *et al.* [87] used a neuromorphic inspired pre-processing technique for tumor segmentation. It followed the concept of attention-based learner combined with CNN to generate saliency maps which aided in better image analysis. This pre-processing technique mimicked the brain's visual cortex and enabled attention-based learning as a pre-determined method for predicting target object regions. During the BraTS 2019 challenge, the neuromorphic attention module demonstrated the effectiveness of training 3D objects with a 2D UNet constraint. The previous 2D UNet with neuromorphic attention module used three channels of incoming image data instead of the initial four channels [88]. Based on the generated saliency maps, this study analysed four-channel MRI data as three-channel input scans. The goal was to include neuromorphic characteristics into the input MRI data.

2.2.1.2 Data Augmentation

Data augmentation, also known as artificial data generation, is a common approach used for improving the generalisation capabilities of DNNs. Data of good quality is hard to collect, and acquiring new examples is also very time-consuming and expensive. Medical image analysis, especially tumor delineation, is affected by this problem [3]. Several data augmentation techniques were used by researchers to increase the number of training samples which included random axis mirror flips [71, 72, 75–80, 89], intensity scaling and intensity shifting [73, 76, 79], isotropic scaling, per-channel gamma corrections [77], rotations [77–79, 85, 89], random scaling [80], elastic transformations, and zoom-in and zoom-out [89]. Different ways of augmenting MRI scans are illustrated in Figure 2.6.

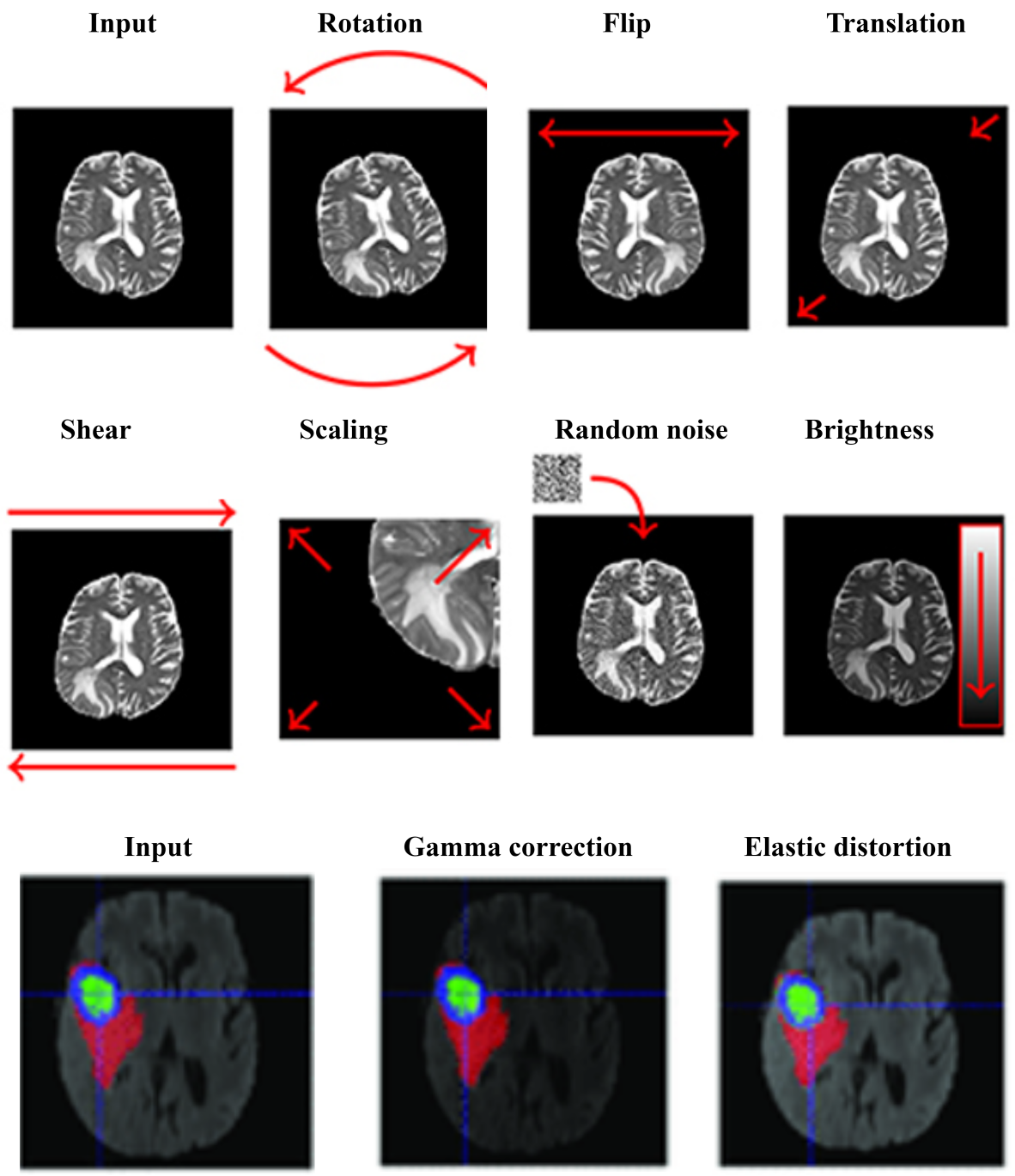


Figure 2.6: Different Forms of Data Augmentation [3, 4]

2.2.2 Glioma Segmentation Architectures

ML approaches, particularly DL architectures, have been applied recently in the medical domain to improve diagnostic and treatment procedures by building automated systems to categorise patients, regression-based predictions, and semantic segmentation applications [90]. However, due to data availability constraints, automation in medical applications is a big challenge. An unbalanced dataset is a common problem in medical imaging. For example, the cancer area is small relative to a healthy brain and MRI scan background. Creating a good DL model, in general, is challenging, but it is now essential to automatically process data of the vast number of patients. Significantly, a deep architecture can generate a wide range of features without the researchers' input. ML models are suitable for small datasets, but researchers need significant characteristics to train such models. ML/DL limitations have been addressed in many ways. Standard techniques for overcoming the imbalanced data problem in segmentation problems include the Weighted Cross-Entropy Loss (WCEL) function, the Generalized Dice Loss (GDL) function, and balanced patch selection [91].

The accuracy of automated diagnostic systems has been improved by using several different models [92–96]. 3D MRI with DNNs is now widely used to diagnose brain tumors because of their valuable and precise performance. Artificial Neural Networks (ANNs) such as Fully CNNs (FCNNs) and their ensembles contribute to the medical field by automatically segmenting the tumor's sub-regions. Accurate BTS algorithms are needed more when the machine observes volumetric MRI scans in three dimensions rather than the actual 2D perspective of a human interpretation. Even though mpMRI provides complete information, it is still difficult to differentiate all sub-regions due to false image features. DL-based approaches often outperform standard ML methods for image semantic segmentation [67].

In their work [97], the authors covered the fundamental, generative, and discriminative approaches for BTS. DNN has recently gained popularity for the segmentation of radiological scans. DeepMedic [70], UNet [62], V-Net [98], SegNet [99], ResNet [100], and DenseNet [101] are the examples of CNN that generate semantic segmentation maps. UNet is generally acknowledged as a standard backbone architecture to perform semantic segmentation in medical imaging. UNet is an encoder-decoder architecture that reduces feature maps to half on the encoder path and doubles feature maps on the decoder path. The skip connections between UNet's parallel stages aid in feature reconstruction.

Due to the volumetric structure of MRIs, organs are often scanned like 3D entities and subsequently segmented using 3D CNN-based architectures. Myronenko *et al.* [95] (BraTS

2018 winner) proposed a variational autoencoder-based 3D encoder-decoder architecture for BTS. Isenee *et al.* [102] (second in BraTS 2018 challenge) utilised a modified version of the basic UNet [62] using diligent training and data augmentation throughout the model training and testing phase. The performance improvement was observed when Leaky Rectified Linear Unit (ReLU) was utilised instead of the ReLU activation function. The BraTS 2019 winner, Jiang *et al.* [94], employed a two-stage cascaded UNet architecture to segment brain tumors. Different tricks for 3D BTS such as sampling of patient' s input data, random patch size-based training, semi-supervised learning, architecture depth, learning warm-up rate, ensembling, and multiproblem learning were used.

Model training and hyperparameter tuning are essential issues related to the high resource consumption by these models. 2D ANNs have fewer parameters than 3D models and can be trained quickly, which allows for better hyperparameter tuning. Due to the absence of depth information, the 2D UNet performs poorly in 3D segmentation. As a result of class imbalance and increasing computing expenses, 3D UNet-based models often face many difficulties. Many recent studies have utilised ensemble models that average the output probabilities of separate models to enhance generalisation power and segmentation performance by providing uncertainty estimate for each voxel in the vision [94, 95, 102–105]. Due to different weights and hyperparameters used for model optimisation, individually trained models vary in one way or the other. As a result, ensembles have certain disadvantages like training time, memory constraints, and increased model complexity. New versions of the UNet were developed by Isensee *et al.* [93, 102] to show that a well-trained UNet can outperform complex ensemble approaches. Many later-year submissions that attempted to utilise the modified 3D UNets or ensembles of 3D UNets [94, 95, 106] seem to be inspired by Isensee's work. The glioma segmentation architectures used in BraTS 2020 SP task to segment tumor sub-regions are classified in Figure 2.7.

2.2.2.1 Single Networks

Using a three-layer deep 3D UNet encoder-decoder framework, Agravat *et al.* [75] presented a tumor segmentation architecture with deep connections to increase network depth. It enabled the gradient to flow straight to the previous layers, allowing the classification layer to watch the preceding levels closely. The layers' extensive connections further enhanced the difficulty in identifying patterns. An unbalanced dataset caused problems with BTS with a higher proportion of non-tumorous slices than tumorous slices, which reduced network accuracy. The 3D patch-based input ensured that the network did not overlearn the background voxels. The network was trained using a mix of dice loss and focal loss functions to enhance its performance to address this class imbalance. A smaller subcomponent size

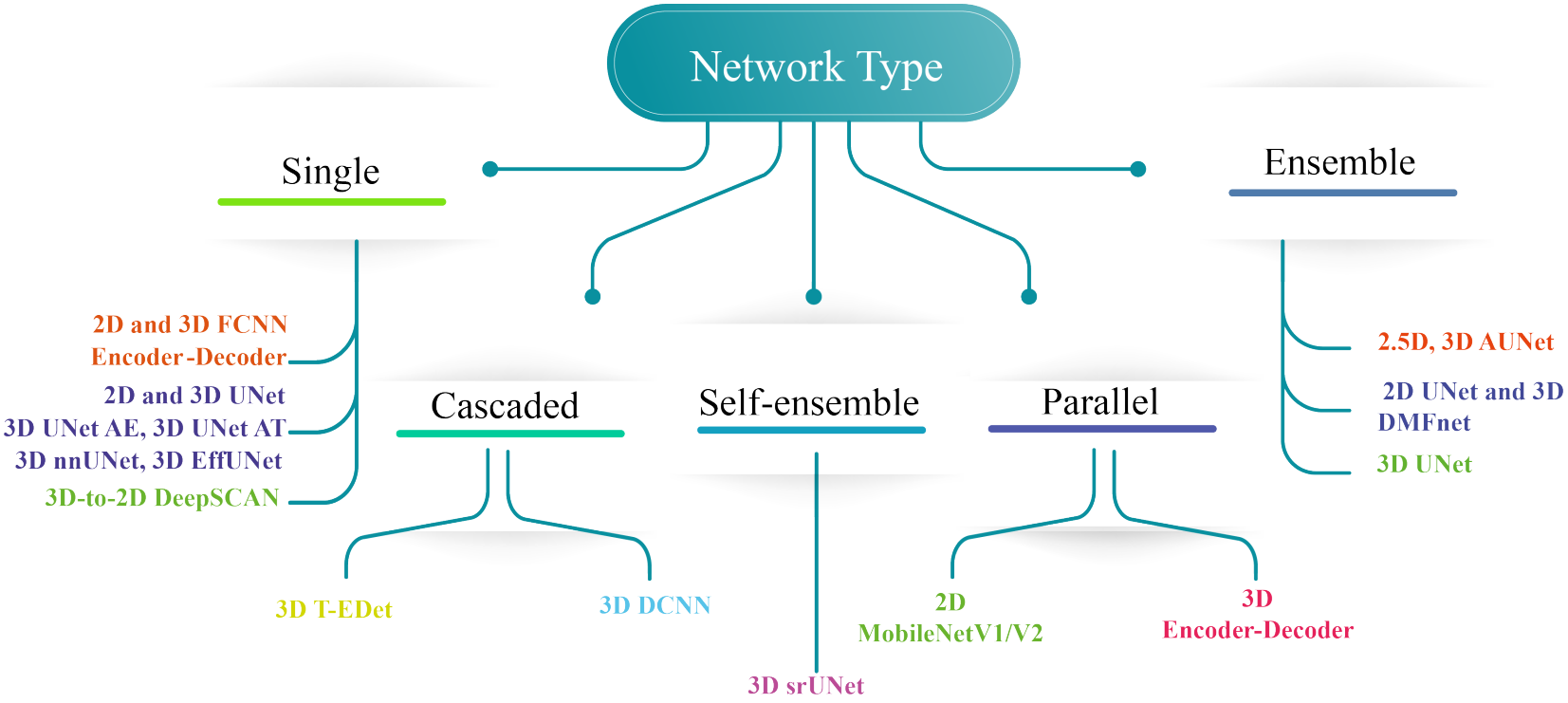


Figure 2.7: Different Types of Networks Used for Glioma Segmentation

(NCR and ET) contributed to the failure of the network, especially in LGG cases.

Using an Accumulated Encoder (AE), Chato *et al.* [107] developed a modified 3D UNet consisting of three layers of encoder/decoder modules with concatenation links. AE included ReLU, Accumulated Block (AB), and max-pooling. To enhance the quality of low-level features, the AB developed two feature maps depending on element-wise addition. After a last max-pooling layer, two branches of convolutions generated these feature maps. After clipping the training patch, Channel Normalization (CN) [108] was applied to prevent overfitting and solve the class imbalance problem. In order to trigger the training procedure with network convergence and performance improvement, the mini-batch method was employed. After two convolutional layers, each encoder unit had Batch Normalization (BN) [109], a ReLU, and a max-pooling layer.

Anand *et al.* [74] suggested 3D FCNN with encoding and decoding pathways. Dense blocks and Transition Down blocks were used in the encoding route. A limited number of output feature maps were chosen for each convolutional layer to avoid the parameter explosion. The spatial dimension of the feature maps was decreased using the Transition Down blocks in the network. Dense blocks and Transition Up blocks make up the network's decoding or up-sampling route. The reversed convolution layers were used in the Transition Up blocks to move sample feature maps up in the hierarchy. The features from the encoding portion were combined with the upsampled features as input for the dense blocks in the decoding phase. With complicated cases, the suggested network did not perform well. The trained network was fine-tuned, and hard mining was conducted on tough instances [110]. With the help of DSC, a threshold-based selection of complicated instances was made.

Parmar *et al.* [72] showed that an increase in batch size might lead to smaller patches with less contextual information. Contrarily, a bigger patch size may provide more contextual information, resulting in smaller batch size, raising the stochastic gradient variance, and reducing the amount of optimization achieved. Batch pools with varying patch sizes were created. The model could learn global information from the most prominent patch and relevant texture from the minor patch with the same parameters by employing various cropping and padding stages between the convolution layers. Patching would also enable a less powerful Graphical Processing Unit (GPU) to handle the big image. The down block of the UNet framework [93] lowered patch size and increased channel length. Overfitting was eliminated by using the down block, which transmitted data from front to end. Reconstruct the position information using the down block outputs, which were then combined in the up block. Based on the probabilistic matrix, the patch size and number of channels were determined. Setting appropriate threshold values for each class helped to accomplish this goal.

Glioma segmentation was solved by Pang *et al.* [85] by using a 2D encoder-decoder arrange-

ment. The encoder portion utilised five Feature Mining Units (FMUs) and four downsample modules to encode the input data. With the use of this structure, the input data may be cleaned of the noise while retaining the crucial features that made segmentation possible. Each of the convolution kernels utilised in the FMU interacted via skip connections; the characteristic channel information collected from these processes was spliced together [111–113]. All convolution operations were combined with BN operations [109] and ReLU activation functions. The decoder's FMU had the same structure as the encoder's. At the same time, the output data of each FMU in the decoder was upsampled to the original size of the image to maintain fundamental patterns. As a result, incorrect information was eliminated, and only appropriate data persisted [114, 115].

Suter *et al.* [116] used the usual ground-truth labelling for Cerebrospinal Fluid (CSF), White Matter (WM) and Grey Matter (GM) [117] to train a nnUNet-framework [118] for locating healthy areas. All images were subjected to an MRI modality-specific piece-wise linear intensity transformation based on these healthy tissue labelling to complement the healthy tissue intensity and map to a set intensity range for consistent grey-value binning. Only the WM and GM labelling were utilised, and CSF segmentation was frequently compromised by insufficient skull-stripping. The N4 correction method [39] was used to correct the bias field in all of the images.

To extract the local characteristics of the tumor, Zhao *et al.* [80] developed a Segmentation Then Prediction (STP) architecture trained on patches. A global branch retrieved the WT's global characteristics to forecast survival days. An encoder and a decoder were used in the segmentation module. Eight convolutional blocks and three downsampling layers comprised the encoder part. There was a downsampling layer between every two convolutional blocks. Because of the limited GPU memory, the Group Normalization (GN) [119] was substituted with the IN. An output layer followed three pairs of upsampling stages and convolutional blocks in the segmentation decoder. Convolutions were used in an upsampling layer, followed by trilinear interpolation and skip connections between the encoder-decoder paths. The number of channels in each convolutional block was decreased to half of the channels in the preceding block. The output layer included convolutional kernels and three channels to generate the ET, WT, and TC segmentation masks.

Patel *et al.* [77] hypothesised that networks using conventional convolutions were less capable of producing correct segmentation labels for glioma sub-regions. As a result, the authors suggested inserting modified Selective Kernel (SK) blocks [120] into the recommended UNet. An attention mechanism allowed the network to automatically change its receptive field to accommodate spatial information collected at various scales. Due to the limitations of GPU memory, the suggested framework learned both anatomical context and com-

plex representations. Downsampling and upsampling of feature maps were achieved via max-pooling with trilinear interpolation, which further helped in decreasing the network's memory footprint. Deep supervision [105, 121] was used to promote quicker network convergence. The proposed SK blocks may be broken down into three operations: (i) Split, (ii) Channel Selection, and (iii) Spatial Attention. A channel attention technique was used to enable the network to alter the receptive field of the SK block adaptively. A primary spatial attention mechanism, comparable to the convolutional block attention module [122], was used to make the network concentrate more on prominent image regions. The channel data was aggregated using the max and average pooling methods across the channel dimension.

Dai *et al.* [78] used the UNet Attention (UNet-AT) framework for BTS, which made use of attention-gated blocks' capacity to focus more on semantic information to improve segmentation performance. Further, there was no need to increase effort for fine-tuning the model parameters and pre-process the data. IN was selected at random for the UNet-AT. The activation function is Leaky-ReLU with a leakiness of 0.01.

McKinley *et al.* [79] used DeepSCAN, a 3D-to-2D FCNN that performed well in the BraTS 2019 challenge [104] and was trained using uncertainty-aware loss. The instances were classified based on confidently segmented core or a poorly segmented or missing core. It was assumed that every tumor had a core. Thus, the criterion for classifying core tissue was lowered when the core, as delineated by the classifier, was poorly defined or absent. The network started with 3D convolutions to compress a non-isotropic 3D patch to 2D. In the bottleneck, a shallow encoder/decoder system utilised densely linked dilated convolutions. This architecture was extremely similar to the authors' BraTS 2018 proposal. The main changes were that the authors utilised IN rather than BN and included a basic local attention block across dilated dense blocks.

2.2.2.2 Cascaded Networks

Using the DeepLabv3+ [123, 124] model, Miron *et al.* [76] developed a version of a two-stage cascaded asymmetric UNet in which decoders kept track of the segmentation region contour to add more regularisation. Large dimensions were used to protect global information. Due to memory constraints, a two-branch decoder with interpolation and deconvolution was employed on the first level of the cascaded model. Model inputs were initial volumes concatenated along with segmentation maps produced from the first step. Two different convolutions were applied to the segmentation and outline of the segmentation region as a regularisation step. Shorter decoders were able to retain more information throughout the upsampling process. When needed, residual additive connectors were added. In order to

extract features at different resolutions, the encoder was changed by adding atrous convolutions having varying dilation rates. A Spatial Pooling unit was introduced in the encoder to make use of the dilated convolutions. After combining the five feature maps, the decoder received them as a single data stream. There were two output tuples at this point; one acquired via upsampling and the other through transposed convolutions.

In contrast to conventional Cartesian space imagery and volumes, Russo *et al.* [86] presented a new technique for feeding DCNN with spherical dimension converted input data to improve feature learning. The DCNN used as a baseline for the proposed approach was developed from the work of Myronenko [95], built on a Variational Auto Encoder (VAE) UNet with modified input shape and loss function depending on the kind of transformation employed in the pre-processing phase. The VAE consisted of a UNet comprising two decoder divisions: a segmentation decoding branch used to get the ultimate segmentation and an extra decoder used to regularise the shared encoder. For COVID-19 severity evaluation and progression prediction [125], the Lesion Encoder (LE) architecture was suggested. The original LE used the UNet structure [62], which included an encoder and decoder inspired by the EfficientNet [126]. The decoder translated the lesion feature maps to the input MRI size and produced the segmentation maps while the encoder learnt and recorded the lesion features in the input images.

Tumor-EfficientDet (T-EDet), a variation of EfficientNet [126], was suggested by Carmo *et al.* [73]. Using a pre-trained EfficientNet [126] with D4 variant, the calculations of backbone weights and features were the starting point for EfficientDet. Because the initial network was developed with a 3-channel input, a significant issue emerged with varying input channels. The 4-channel (4 modalities) input was converted into a 3-channel input using an adaptation convolution. The Bi-directional Feature Pyramid Network (BiFPN) altered the characteristics obtained from the pre-trained EfficientNet to get its initial features. A transposed convolution accompanied by BN and the same swish activation [127] was employed to adapt this representation to the size of the input MRI data. This new representation was given to the segmentation architecture. It comprised three blocks of depth-wise convolutions [128], batch norm, and swish, coupled with a final convolution to reduce the number of classes. In isolation and then as the backbone of the planned multiproblem experiment, the UNet 3D architecture based on the work of Carmo *et al.* [129] was utilised. In terms of training technique, several hyperparameters comparable to Isensee's work [93] were used. The primary experiment included the end-to-end refining of segmentations generated by a solid pre-trained UNet 3D backbone model, as well as the inclusion of an attention-based SP branch.

2.2.2.3 Self-ensemble Networks

Pei *et al.* [81] presented a technique based on 3D reUNet DL. The reUNet was made up of a standard residual ResUNet and a self-ensemble model. The problem of gradient vanishing stopped with the use of the self-ensemble model and the ResUNet model. For OS prediction, the tumor segmentation probabilities derived from 3D Self-ensemble ResUNet (srUNet) were utilised. Even though DL-based techniques performed well in semantic segmentation, a large number of training samples were needed.

2.2.2.4 Parallel Networks

Soltaninejad *et al.* [82] presented a segmentation method that consisted of two major branches with varying degrees of resolution. Based on the technique described in Newell *et al.* [130], the native resolution branch was an encoder-decoder architecture with residual blocks. In the last layer, the activation function was changed with a sigmoid function. The native branch dealt with better resolution small-sized MRI patches. For segmentation, another branch with a broader view but a poorer resolution was added. In order to match features from the larger receptive field with the native branch, these features were trimmed. As a final step, the characteristics from both branches were combined and presented to the final layer for segmentation. In order to create the delineation heatmaps from the features, two convolution layers were utilised for each path. In order to train the network, several loss functions were employed, such that each branch learned relevant characteristics together with the final layer. The output was sent through a Sigmoid function which used stochastic gradient descent to learn the network's topology to train the network. For each class, a separate model was trained against all other labels to identify patterns.

2.2.2.5 Ensemble Networks

When it comes to the BTS task, González *et al.* [71] presented an ensemble of seven models with 3D and 2D input methods, all based on Asymmetric UNet (AUNet). According to the 2.5D model, each modality's low-level characteristics were extracted separately using a multi-input approach. Multi-view 2.5D Inception block was designed to combine features from various perspectives of a 3D visual and accumulate multi-scale features into a single 3D image. Five layers were used in the enhanced 3D UNet to extract additional semantic information, with the encoding route having twice the number of convolutional blocks as the decoding path. There were four 2.5D Multi-View Inception blocks and a modified 3D UNet architecture that included residual blocks, Instance Normalization (IN) [131], transpose convolutions for performing upsampling, and additive skip connections within encoding and decoding paths in the multi-input module.

Ali *et al.* [83] suggested a combination of 2D and 3D approaches to maximise their advantages. The authors also showed that if hyperparameter tuning was done effectively, 2D models might attain comparable or even better performance than 3D models. Several axial, sagittal, and coronal views were used to train the model to compensate for the absence of depth information in 2D models. A combination of 2D and lightweight 3D models was used to train the ensemble model introduced by Asra *et al.* [132] to extract both 2D and 3D benefits. The 3D Dilated Multi-Fiber Network (DMFNet) [92] model was selected because fewer parameters could be learnt more rapidly than other 3D complicated models. MultiFiber (MF) and dilated convolutions were used with group convolutions to decrease parameters and preserve prediction precision. The information was sent using a multiplexer between the fibres by combining two convolutions with the four-channel input. The spatial information of brain tumors was then captured using dilated convolutions. Convolutions were performed with multiple dilation rates to check how many pixels were missed throughout the process. An ensemble of these two models was constructed in order to make use of both.

Zhang *et al.* [133] focused on utilising the brain parcellation-based knowledge, coarse-to-fine approach, and ensembling to enhance UNet segmentation performance. Four UNets were trained independently to perform BTS. UNet1 investigated how distinct tumor areas were more apparent in various MRI modalities. UNet2 focused on developing a brain parcellation prototype using a T1 MRI as input. Brain parcellations plus mpMRI images were used to train UNet3. Utilizing cropped MRI as well as brain parcellations as input, UNet4 resulted in coarse-to-fine segmentation. The final output was achieved via a majority vote of both coarse and fine segmentation. Table 2.3 summarises the specifics of segmentation architecture used to segment tumor sub-regions before performing feature extraction.

Table 2.3: Summary of Segmentation Architectures for BraTS 2020 SP Problem (A: Network Type, B: Input Dimension, C: Segmentation backbone, D: Depth/Levels/Layers, E: Modifications, F: Normalization, G: Uniqueness, H: Sampling)

Authors	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
González *et al.* [71]	Ensemble	2.5D, 3D	AUNet	5	Inception block (2.5D), residual block (3D)	IN	Multi-input strategy	5-fold cross-validation (CV)
Agravat *et al.* [75]	Single	3D	Encoder-Decoder FCNN	3	Dense module	BN	Deep supervision	-
Miron *et al.* [76]	Cascaded	3D	Two-stage cascaded asym-metrical UNet	3	Residual additive connections, spatial pooling module, Atrous convolutions with dilations	GN	Regularization by contouring	-
Chato *et al.* [107]	Single	3D	UNet AE	3	Accumulated block	CN, BN	Information fusion strategy	-
Anand *et al.* [74]	Single	3D	FCNN	3	Dense blocks, transition up/down blocks	-	Hard mining on complex cases	stratified sampling
Parmar *et al.* [72]	Single	3D	UNet	5	Downblock or Upblock	IN	Patch selection strategies	stratified sampling
Soltaninejad *et al.* [82]	Parallel	3D	Encoder-Decoder	2	Residual block	-	Multi-resolution method	-

Table 2.3 continued from previous page

Authors	A	B	C	D	E	F	G	H
Pang *et al.* [85]	Single	2D	Encoder-Decoder	4	Residual units	BN	More abundant features	-
Ali *et al.* [83]	Ensemble	2D, 3D	2D UNet and 3D DMFnet	3	Group convolution with MF, dilated convolutions	-	Utilized 2D and 3D benefits	-
Russo *et al.* [86]	Cascaded	3D	DCNN	3	ResNet block in variational autoencoder	-	Spherical coordinate transformed input adds extreme augmentation, variational autoencoder-based UNet	-
Akbar *et al.* [84]	Parallel	2D	MobileNet V1/V2	3	Removed classification part, additional layers to produce the ratio of tumor sub-regions	BN	Focused on DL-based feature extraction rather than segmentation	stratified sampling, 5-fold CV
Marti Asenjo *et al.* [89]	Single	3D	EffUNet	5	EfficientNet block in encoder	-	Deep supervision	-
Carmo *et al.* [73]	Cascaded	3D	T-EDet	5	D4 variation of EfficientNet, BiFPN, multiATTUNet	BN, GN	Modification of 2D semantic segmentation EfficientNet model, MultiATTUNet architecture	-
Han *et al.* [87]	Single	2D	UNet	4	Neuromorphic attention-based learner	BN	Neuromorphic saliency map	-

Table 2.3 continued from previous page

Authors	A	B	C	D	E	F	G	H
Pei *et al.* [81]	Self-ensemble	3D	srUNet	4	Regular ResUNet and srUNet model	-	Self-ensemble avoids gradient disappearance and regularises training	-
Suter *et al.* [116]	Single	3D	nnUNet-framework	5	Piece-wise linear intensity transform, grey-level binning	IN	Identified healthy region labels for CSF, WM, GM	5-fold CV
Zhao *et al.* [80]	Single	3D	Encoder-Decoder	4	Residual connection	IN	STP with local and global branch	5-fold CV, self-training
Patel *et al.* [77]	Single	3D	UNet	5	Selective kernel with attention mechanism, residual bottleneck block	GN	The network adjusts its receptive field via an attention mechanism, deep supervision	-
Dai *et al.* [78]	Single	3D	UNet-AT	5	Attention-gated block	IN	Self-training model again with the training set and previously generated pseudo annotations for the validation set	-
McKinley *et al.* [79]	Single	3D-to-2D	DeepSCAN	-	Dilated dense block with attention	IN	Separate cases based on confidently segmented/ vaguely segmented/ missing core	-
Zhang *et al.* [133]	Ensemble	3D	UNet	4	Brain parcellation model using T1 MRI	-	Use of brain parcellation, coarse-to-fine strategy	-

2.2.3 Post-processing

Several post-processing techniques for eliminating false positives and improving segmentation results have been suggested. Traditional post-processing techniques, like threshold or region-growing methods, focus on isolated regions or pixels by utilising manually defined thresholds. Recently, post-processing techniques such as Connected Component Analysis (CCA), Conditional Random Fields (CRF), and Gaussian Markov Random Fields (MRF) have been utilised to infer pixel pairings given previous information such as pixel intensity distributions and geographic distance. Recently, the researchers coupled CRF to NN models in an end-to-end learning manner for image segmentation issues [134]. The post-processing shown in Figure 2.8 increased the DSC and HD95 of the ET, WT, and TC classes considerably.

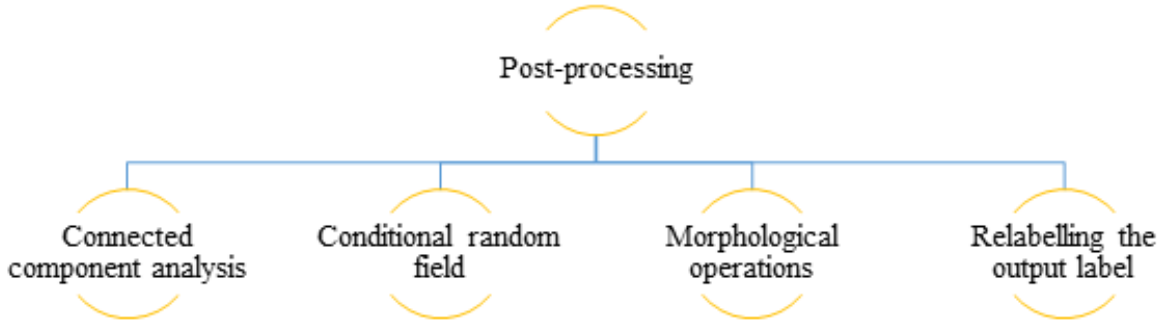


Figure 2.8: Post-processing Techniques Used by SP Techniques Submitted in BraTS Challenge

2.2.3.1 Connected Component Analysis

CCA cluster voxels depending on connectivity and voxel intensity levels are comparable. The highly small linked components are removed from the result since they are false positives due to misleading segmentation findings. Agravat *et al.* [75] chose CCA to eliminate the tumor, which had a volume of less than one thousand voxels. To post-process the segmented labels, Anand *et al.* [74] utilised class-wise 3D CCA. Marti *et al.* [89] defined the primary tumor volume as the most considerable linked (solid) volume. Over the training dataset, the distance from each smaller area (secondary tumor volumes) that was not linked to the primary tumor was computed. After tumor segmentation, the secondary volumes situated at a distance more significant than the average range plus 1.5 times the normal deviation were deemed false-positive and; were therefore eliminated and categorised as background. Patel *et al.* [77] used CCA to minimise the frequency of false positives in any class with fewer than ten voxels. Zhao *et al.* [80] used CCA to enhance segmentation performance as well.

Dai *et al.* [78] eliminated a small isolated WT/ET (just under ten voxels) out from prediction, modifying the TC size to match the size of the ET.

2.2.3.2 Conditional Random Field

The algorithm assumes the voxel category based on characteristics linked to a specific voxel, independent of the voxel's connection to other adjacent voxels. CRF analyses this connection and creates a graphical model to handle prediction dependencies. CRF was employed by Anand *et al.* [74] to smoothen the segmentation maps.

2.2.3.3 Morphological Operations

Morphological operations were used to modify voxel values depending on their neighbourhood, taking into account the size and form of the voxel. After screening the spherical segmentation using the cartesian filter, Russo *et al.* [86] utilised post-processing to improve ET segmentation. The scientists discovered that isolated voxels were responsible for numerous erroneous positive ET segmentations. The scientists used a binary opening operation to separate thin branches across ET sites before filtering out those with less than 30 voxels. Whenever the ET segmentation remained after these filters, the original was recovered and utilised as the final one. Alternatively, the ET partition was deleted, implying that there was no ET in the current volume.

2.2.3.4 Relabelling the Output Label

González *et al.* [71] established a threshold value that represents the smallest size of the ET area. When the overall amount of predicted ET voxels was less than the threshold, the label of all voxels in the ET area was substituted for one of the NCR regions. In the validation dataset, the threshold value was calculated to obtain maximum performance in this region. If the segmentation model did not identify a TC, the authors concluded that the identified WT area matched the TC and relabeled it as TC. ET was produced in the vicinity of the NCR, according to Agravat *et al.* [75]. Furthermore, in HGG, its size cannot be minimised. Because of its small size, ET was changed to NCR. Three hundred was the experimentally determined threshold for relabelling. As a post-processing step, Carmo *et al.* [73] used a particular threshold in ET segmentations, where segmented regions with less than 300 voxels were utterly wiped out. It was done to prevent false-positive predictions of the ET region, which may cause DSC to be zero in individuals without ET. According to Patel *et al.* [77], most LGG patients have no ET portion, and the inclusion of even a single misdiagnosed ET voxel in estimated label maps leads to a DSC of 0. To mitigate the severity of the penalty, the authors substituted all ET voxels by TC if the number of anticipated voxels was

less than 50. Zhao *et al.* [80] discovered that if the segmented ET was less than 350 voxels, the authors suspected no ET in LGG instances and removed those identified voxels from its segmentation result.

2.2.4 Feature Extraction and Selection

Radiomics is studied to investigate the relationship between medical imaging aspects and underlying biological properties. It is a technique of extracting efficient quantitative characteristics from radiographic scans and developing prediction architectures that link image attributes to clinical outcomes. The fundamental concept behind radiomics is that intra-tumor imaging heterogeneity can be retrieved from the MRI scans. The tumorous areas estimated from the patient's MRI volume were then utilised to predict OS time.

González *et al.* [71] used four MRI modalities to derive radiomic characteristics from the whole brain and five sub-regions. WT, TC, ED, NCR, and ET are the feature extraction sub-regions. The Python library PyRadiomics [135] was used to extract all 2,568 handcrafted features. The Recursive Feature Elimination (RFE) method was employed to reduce features to prevent overfitting. Radiomics was also used to estimate two additional characteristics: the Relative Invasiveness Coefficient (RIC) and Relative Necrosis Coefficient (RNC). Agravat *et al.* [75] collected the patient's age, statistical data, and NCR shape for the SP model.

Miron *et al.* [76] retrieved radiomic characteristics from segmentation labels, which primarily defined the area of interest's 3D shape, size, and centre of mass. Each tumor sub-regions ratio to the total brain size and the WT were included. A new binary characteristic called 'High Risk' was set, if the tumor was in the lobe with a higher risk probability or in the left hemisphere. A greater risk was assessed if the centre of mass was located within 128-150 slides (sagittal plane), 60-150 slides (coronal plane), or 0-41 slides (transversal plane).

Chato *et al.* [107] utilised a feature-level and decision-level information fusion approach for creating the OS prediction model. The intersection of two brain midplanes splits a brain volume into four or eight small sub-volumes to retrieve volumetric characteristics. Five volumetric features indicating the volume of a brain region/WT/ET/NCR-NET/ED were gathered from each small volume, plus the volume of the whole brain and the WT. The centroid, volume, diameter, orientation, principal axis length, surface area extent, and solidity of the tumor site were among the twenty-eight shape and location characteristics. Only GTR resection status was considered for evaluation. There were 119 samples with GTR resection status out of 236 samples in the BraTS 2020 training dataset. A dataset of 119 samples was insufficient to build a reliable OS prediction model. Thus, resection status was utilised as an additional non-imaging characteristic (*i.e.*, GTR as '1' , STR or NA as '0'). The patient's age

was utilised as another non-imaging characteristic since it was one of the variables considered by physicians to select an appropriate treatment strategy for glioma patients.

Anand *et al.* [74] extracted radiomic features such as mean, median, and entropy, as well as second-order features such as Gray Level Co-Occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Dependence Matrix (GLDM), Gray Level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM). A total of 1022 distinct characteristics were retrieved using a different mix of segmentation maps. Each characteristic's significant value was given using a forest of trees [136, 137].

Volume and shape characteristics were derived from the whole brain and tumor substructures by Parmar *et al.* [72]. The 2D/3D diameter, major axis, minor axis, elongation, spherical index, whole brain' s flatness, and tumor sub-regions were retrieved to identify interrelationships as shape characteristics. For matching segmentation labels, a clinical characteristic (age) was assigned.

For SP, Soltaninejad *et al.* [82] used statistical characteristics of already segmented tumor areas from the previous segmentation step. The volume sizes of the WT, TC, and ET that were normalised *w.r.t.* the total brain volume were considered. The mean intensities for each tumor tissue type, *i.e.*, the WT, TC, and ET, was another set of characteristics. As a result, eight characteristics were utilised for OS prediction.

Pang *et al.* [85] extracted characteristics from MRI data by masking the ED, ET, and NCR-NET regions. The particular feature selection procedures included Cox univariate analysis, Cox multivariate analysis, and the ph hypothesis test. Minor axis length, sphericity derived from the segmentation result using the ED mask, and busyness extracted from the segmentation result using the ET area mask were among the selected attributes. In addition, the patient' s age was included in the final feature set. According to a bivariate study, the patient's age and survival time were substantially linked at 0.01. The Pearson correlation value was -0.353, which suggested that the patient's age can affect OS.

To estimate the OS time, Ali *et al.* [83] collected radiomic and image-based characteristics. The image-based characteristics were volume, surface area and ratios, and the WT and ET centre positions. The Laplacian of Gaussian (LoG) filters was used to extract radiomic characteristics. The input image's texture and intensity-based characteristics were retrieved and filtered using LoG filters with sigma values 1, 2, and 3. Fourteen shape-based characteristics (tumor volume, diameter, surface area), first-order features, and GLCM features were recovered from segmented tumors. The characteristics for the tumorous areas ET, ED, TC, and WT were extracted. Because of tumor overlapping, some features were common. The top-performing features were chosen using Random Forest (RF) RFE. The patients' age was in-

cluded in the feature vector since it provided crucial information about their survival.

Individual lesion encoders were utilised by Russo *et al.* [86] to extract the latent variables of the input MRI scans, which were further used as input parameters to predict patient OS. A high-dimensional feature vector (d = 256) was generated for each MRI scan. Principal Component Analysis (PCA) regulated the feature dimensionality by selecting various principal components for subsequent analysis since the high-dimensional feature space tended to overfit. The age distribution seems normal; however, the OS distribution appears skewed, with most instances having an OS of fewer than 400 days. The tailed distribution of the OS values was modelled using a Tweedie distribution.

Akbar *et al.* [84] suggested modifying MobileNet V1 and MobileNet V2 as interchangeable feature extractors. The predictive architecture was fed with extracted features and patients' age data. The MobileNet V1 design was changed by eliminating a portion of the categorization and linking it to many additional layers, resulting in three characteristics (the ratio of NEC, ED, and ET area to the overall area of the tumor). MobileNet V1 received a 2D image with three channels that were a multiple of 32 in size. In contrast, the accessible dataset consisted of four cropped 3D images from the image's centre. T1CE, T2, and FLAIR images were utilised to modify. Axially sliced 3D MRI scans were obtained, yielding 128 2D image data of 160×192. The three images were stacked as three channels, and the 128 slices were converted into batch numbers for a single epoch. A dropout with a factor of 0.1 was employed to avoid overfitting [138]. Similar changes were made to the MobileNet V2 design, except that the output feature was used instead of the categorization layer. The same approach was used to input 3D MRI images to this architecture, as in the MobileNet V1 design. MobileNet V2-based modified network was selected based on the test findings.

Marti Asenjo *et al.* [89] created three distinct datasets for each patient's 2D anatomical plane, *i.e.*, axial, sagittal, and coronal. The identification of low-level characteristics was enhanced by using a pre-trained EfficientNet network [126] based on ImageNet [139] as the UNet encoder Efficient UNet (EffUNet). This idea was to expand the number of layers and channels. The 3-channel decoder input layer was pre-trained on ImageNet, and further upgraded to a 4-channel configuration to accommodate all MRI modalities for each patient. The fourth layer was seeded with random weights drawn from a Gaussian distribution. The Laplacian and tumor gradient were vital in glioma tumor growth models [140, 141]. For all MRI sequences and the segmentation matrix, the Laplacian and the vector module of the 3D gradient were derived. Transformations were performed using MRI images (pixel values) and segmentation masks for a particular area, resulting in a binary matrix with '1' within and '0' outside the specified region. A complete tumor contains the whole of all three groups. NCR/NET volume, Gross Tumor Volume (GTV), and WT were also included in three ring-

shaped structures. Following extraction, dimensionality reduction was required to select those characteristics with a higher connection to patient survival. Geometrical, statistical, locational, textural, and other characteristics were retrieved. Three variables (GTR, STR, and NA) were added as categorical and binary characteristics. Three distinct feature selection techniques were used: *fscchi2* (ranking for classification using chi-square tests), *fscmrmr* (ranking for classification using Minimum Redundancy Maximum Relevance (MRMR) algorithm) and *fsrftest* (ranking for regression using F-tests). Each algorithm was then assigned the significance of each feature.

Han *et al.* [87] developed the neuromorphic saliency map to improve image analysis. An attention and saliency map was created to enhance tumor segmentation by simulating the visual brain and applying neuromorphic pre-processing. Several orientation-selective characteristics were used to build the neuromorphic NN. The orientation selectivity inspired by the visual cortex's "simple cell" enabled the resilience of abstract feature extraction [142]. The authors were able to generate a saliency map using the down-up resizing network. The saliency map proved successful at either eliminating noise or attracting attention. A new abstract image was created using UNet and a neuromorphic NN mixed network for BraTS 2019 [88]. It demonstrates the possibility of using a neuromorphic NN for image pre-processing or image post-processing.

Pei *et al.* [81] retrieved shape characteristics from segmentations produced and included non-radiomic variables like age. In the training phase, features were chosen based on their significance using an RF classifier based on survival risk. Data were randomly divided as training and validation in an 8:2 ratio in the training phase to determine the feature significance. The number of NC, ED, ET, shape elongation, flatness, most petite axis length, and surface area were among the 34 characteristics retrieved.

Suter *et al.* [116] retrieved features using the PyRadiomics library [143] for all four MRI modalities and the two segmentation labels WM and GM. The LoG filtered, intensity-based features were extracted for the original wavelet image. For tumor location characteristics, the images were registered using the MNI152 atlas [144]. Because the tumor's mass impact often altered the ventricles' location and shape, the authors used the Symmetric Normalization (SyN) [145] to reshape the afflicted regions appropriately. In the first stage, the authors eliminated any characteristics having a concordance index (cindex) less than 0.55. For all feature pairings having a correlation coefficient of 0.95 or above, the authors progressively eliminated the feature with the lowest c-index. The use of two priors was investigated. When the sequence prior was used, the characteristics were restricted to the T1CE and FLAIR MRI images. Features having an Intraclass Correlation Coefficient (ICC) (2,1) of 0.85 or above were evaluated for robustness prior. These priors were evaluated both in-

dividually and together. Many ML techniques need or benefit from regularly distributed target variables. All sixteen feature selection and eleven ML technique combinations were examined. ReliefF (RELF), Fischer Score (FSCR), Chi-square Score (CHSQ), Joint Mutual Information (JMI), Conditional Infomax Feature Extraction (CIFE), Double Input Symmetric Relevance (DISR), Mutual Information Maximization (MIM), Conditional Mutual Information Maximization (CMIM), Interaction Capping (ICAP), MRMR, and Mutual Information Feature Selection (MIFS) techniques were used for feature selection.

In their suggested STP model, Zhao *et al.* [80] used both local and global branches. The predictions from these two branches were combined to create the outcome. The segmentation module's output contains three local branch channels representing the ET, WT, and TC probabilities. As a result, the authors got three masks for the various tumor sub-regions. The authors were then given three bounding boxes, which were minimal cubes containing various sub-regions. The Region Of Interest (ROI) Alignment [146] was applied to the encoder outputs to get the features. The local tumor characteristics were obtained using 3D Global Average Pooling.

Patel *et al.* [77] fed one whole image through the proposed network and retrieved 2048 characteristics from the bottleneck layer's end. This procedure was done for each of the eight mirror axis-flipped copies of the input and each of the five unique models in the proposed ensemble, yielding 40 different versions of these 2048 characteristics. The characteristics were average global pooled over batch and geographic dimensions to merge these 40 versions. Using this full feature set to train a model would significantly overfit, requiring dimensionality reduction using PCA. The authors also incorporated the age and volumetric features of tumor sub-regions in the suggested model. Patients with GTR were assessed; the authors chose to train the suggested cox model on a collection of 118 cases. The authors discovered experimentally that a model with ten main components outperformed all others on the validation set.

Based on the segmentation findings, Dai *et al.* [78] proposed a biophysics-guided architecture for SP. The authors created a map based on each patient' s tumor' s structure with four discrete values for different classes. Tumor heterogeneity and invasiveness was revealed by the tumor' s spatial distribution [147]. Instead of data-driven radiomic characteristics, the authors explored tumor growth biophysics modelling [148]. The RIC, defined as the ratio between the hypoxic TC and the infiltration front [149], was of particular interest. The authors used the biophysics-guided invasiveness feature [150], which used TC and WT to calculate RIC. In predictive modelling, this biophysics-guided feature outperformed the data-driven radiomics method. The tumor invasiveness was described using a RIC generated from the tumor structural map. The WT and TC first semi-axis length ratio was computed as the RIC

in this research.

Zhang *et al.* [133] utilised the numbers 1, 2, and 3 to denote three resection statuses, *viz.*, NA, GTR, and STR. In addition, three new techniques were used in the suggested model: (i) using brain parcellation to obtain more features, (ii) integrating classifier and regressor models, and (iii) selecting characteristics in addition to those obtained by Xue *et al.* [151]. The authors utilised brain parcellation to separate tissue labels, particularly CSF, WM, and GM. The surface area of each overlapping region between two brain structures was then computed (*i.e.*, overlapped regions between ET and CSF, ET and WM, ET and GM, ED and CSF, ED and WM, and NCR/NET and GM). The technique for calculating surface area was the same as in Xue's study [151].

2.2.5 Overall Survival Prediction Models

Various OS prediction techniques submitted in BraTS 2020 Challenge are briefly explained in this section under sub-sections of ML and DL models.

2.2.5.1 Machine Learning Models

Miron *et al.* [76] demonstrated that using ExtraTrees [152] on derived radiomic characteristics was highly sensitive to minor changes in the parameters, owing to the limited size of the training sample. The CSV file's age proved to be an essential characteristic. The most significant results were achieved when the depth of the trees was limited to just seven levels. Chato *et al.* [107] trained an OS time classification model using a basic NN ML technique. The Random Forest Regressor (RFR) was trained by Anand *et al.* [74] using the thirty-two most significant characteristics out of 1022. Parmar *et al.* [72] utilised age, volumetric, shape features for training the multi-fold RF classifier. Pang *et al.* [85] also utilised RF regression to estimate patient survival time. After identifying key characteristics, Ali *et al.* [83] utilised RFR with Grid search for SP. Patel *et al.* [77] used a cox proportional hazards model trained on in-depth characteristics derived from the proposed segmentation network to predict OS. A linear regressor with RIC and age as factors was employed by Dai *et al.* [78]. Agravat *et al.* [75] used five-fold Cross-Validation (CV) for training an RFR for OS prediction in instances with GTR. The RFR outperforms other cutting-edge methods that utilise linear regression and ANNs.

Soltaninejad *et al.* [82] used an RF model to extract the feature set from each Volume Of Interest (VOI). RF parameters (*i.e.*, tree depth, the number of trees) were adjusted using five-fold CV on the training dataset. The matching parameters, 50 trees of depth 10, gave the best generalisation and accuracy. Using the RF in the regression model, the predictions as the

number of survival days were produced. To forecast OS values, Russo *et al.* [86] developed a Tweedie distribution-based Generalized Linear Model (GLM) [153], *i.e.*, Tweedie Regressor. The resection status and age were essential OS predictors, and these were combined with the LesionEncoder DL features to feed into the Tweedie Regressor for OS prediction.

Suter *et al.* [116] investigated eleven ML methods, including Automatic Relevance Determination Regression (ARD), AdaBoost, Decision Tree (DT), Extra Tree(s), Gaussian processes, Multi-Layer Perceptron (MLP), Nearest Neighbors, radius neighbours, passive-aggressive, RANSAC, RFs, stochastic gradient descent, Support Vector Regression (SVR), linear, and Theil-Sen regression. On the training set, the effectiveness of all permutations was assessed using ten-fold stratified CV. McKinley *et al.* [79] used a combination of linear regression and RF classification to predict the survival of HGG patients based on age, the number of different tumor components, and TCs. The authors developed a least-squares regression model that predicted OS based on age, distinct cores, and various tumors. The age-based RF classifier obtained comparable accuracy in CV. The authors permitted the RF classifier to overrule the linear model's prediction. To improve robustness, the authors only permitted this when the RF was confident in its prediction.

Due to the small sample size, Zhang *et al.* [133] used both classifier and regressor to estimate survival. Firstly, the authors used a linear classifier for allocating individuals to one of three groups and compute the likelihood of each class. A linear regressor was utilised to determine the connection between the likelihood and survival days. Short survivors -$(0,300)$, mid survivors -$(300,450)$, and long survivors -$(450,+\infty)$ should be the connection between the three groups and the survival days. The survival days D predicted from the regressor may differ from the classifier's allocated class C. The authors created a discriminator to verify the consistency between C and D to address this issue, *i.e.*, they utilised a classification model to lead the regression model. The correlation between features and survival days was determined using univariate linear regression. Then, features with the maximum correlations were chosen. Experiments using leave-one-out CV led to the selection of $K = 5$.

2.2.5.2 Deep Learning Models

González *et al.* [71] used handcrafted radiomic characteristics to compare a 3D DenseNet CNN model *w.r.t.* conventional ML techniques for the OS prediction. To operate best with small batch sizes, the authors developed an enhanced version of the 2D model proposed by Huang *et al.* [135], *i.e.*, the 3D DenseNet model, by replacing all 2D convolution filters with 3D convolutions and BN operations with IN operations. For categorization, OS time was divided into three survival groups. Because the forecast is needed in days, a scalar value matching each group's median number of surviving days was assigned to each class (*i.e.*, 150

for short, 363 for mid and 615 days for long survivors).

Akbar *et al.* [84] presented an SP model based on the linear architecture that takes the patient' s age as input and features extracted using MobileNetV1/MobileNetV2. The regression-based activation layer was used for predicting the number of days of survival. Because of its capacity to tolerate overfitting, Pei *et al.* [81] developed a risk-guided standard ML RFR approach for OS prediction. To find the optimum parameters, a grid search was undertaken. The optimum number of features were seven with 31 estimators and 50 depth for the RF model.

Carmo *et al.* [73] used a CNN3DAtt branch to conduct SP for multi problems. Seven input channels were obtained by combining all input MRI modalities and three channels from the backbone architecture. CNN3DAtt was inspired by Gorriz *et al.* [154]. Attention-based CNN was initially a 2D network. However, with the addition of GN and 3D convolutions, it was transformed into a 3D network. The network generated Attention (ATT) maps by making use of the sigmoid activation function. The attention map's 3D features were converted to 1D for the fully-linked layers using Average Global Pooling per channel. At this point, the age value was inserted as an additional neuron. The result was a single neuron that was utilised as a direct activation for survival days. Only when age and survival statistics for the patient were available, survival loss was applied before applying the segmentation loss so that the survival gradients would impact the segmentation loss. The MultiATTUNet architecture was named after the Multi problem approach, while the MultiUNet architecture was named after the same design but without the surviving branch.

Han *et al.* [87] created a Feedforward Neural Network (FFNN) with four inputs for OS prediction. The FFNN topology took three survival time outputs in three categories: short, medium, and long. The segmentation result of WT was added as input after training different input combinations. There were many suggestions on the OS forecast based on previous BraTS challenges. The top-ranking techniques varied in input feature sets, ranging from a simple 'age' feature to a complex 16-feature set. Model creation without any common guideline to select feature set and NN layer structure was very difficult. The feasible assessment was obtained from the confusion matrix using the given training dataset. As a result, the scenario with four inputs was chosen, even though its impact on post-processing was still unknown. The FFNN was overfitted as a result of the network training and network implementation.

Zhao *et al.* [80] merged three sub-tumor local features, shaped into a single-dimensional matrix, and scaled by age. The global branch of the SP module based on 3D ResNet50 was used to predict lifespan days. It consisted of a convolutional layer, a max-pooling layer, four bottleneck subgroups, an average global pooling layer, and a fully connected layer. There

were skip connections, and three convolutional, BN, and ReLU layers existed in each bottleneck. In the first and third combinations, the kernel size of the convolutional layer was $1 \times 1 \times 1$. In the second combination, the kernel size of convolution was $3 \times 3 \times 3$. The authors then utilized the global average pooling layer to extract global tumor characteristics from the WT, subsequently multiplied by patient's age. The outputs of the two branches were then combined to provide SP findings. The SP model details are summarized in Table 2.4.

Table 2.4: Summary of SP Model Details Used for BraTS 2020 Dataset

Authors	**SP model**
González *et al.* [71]	SVR, 3D DenseNet CNN
Agravat *et al.* [75]	RFR with 5-fold CV
Miron *et al.* [76]	Extra Trees with a depth of 7 levels
Chato *et al.* [107]	Simple NN
Anand *et al.* [74]	RFR
Parmar *et al.* [72]	RFC with multi-fold CV
Soltaninejad *et al.* [82]	RFC with 50 trees having a depth of 10, 5-fold CV
Pang *et al.* [85]	RFR with 5-fold CV
Ali *et al.* [83]	RFR with Grid search
Russo *et al.* [86]	GLM with Tweedie distribution
Akbar *et al.* [84]	MobileNetV1/MobileNetV2 with Droput 0.1
Marti Asenjo *et al.* [89]	DT (ensembled by RUSBoost method), SVM with quadratic kernel function, ensembled of regression trees (Matlab ML models)
Carmo *et al.* [73]	MultiATTUNet
Han *et al.* [87]	FFNN
Pei *et al.* [81]	RFR with grid search
Suter *et al.* [116]	ARD, AdaBoost, DT, Extra Tree(s), Gaussian processes, linear, MLP, Nearest Neighbors, passive-aggressive, radius neighbours, RANSAC, RF, stochastic gradient descent, SVR and Theil-Sen regression
Zhao *et al.* [80]	3D ResNet50
Patel *et al.* [77]	Cox proportional hazards model
Dai *et al.* [78]	Linear regression
McKinley *et al.* [79]	Linear regression model and the RF classification model
Zhang *et al.* [133]	Linear classification, linear regression models

The training details of all the SP pipelines reviewed in this study are summarized in Table 2.5.

Table 2.5: Summary of Model Training Hyperparameters for Integrated BTS and SP Techniques Using BraTS 2020 Dataset (A: Loss Functions, B: Epochs, C: Patch Size, D: Optimizer, E: Learning Rate, F: Learning Rate Decay (LRD) / Weight Rate Decay (WRD), G: Batch Size)

Authors	**A**	**B**	**C**	**D**	**E**	**F**	**G**
González *et al.* [71]	dice loss(BTS), cross-entropy loss(SP)	100(BTS), 5(SP)	$128 \times 128 \times 128$	adam	0.0001(BTS), 0.00005(SP)	-	1(BTS), 2(SP)
Agravat *et al.* [75]	dice & focal losses	610	$64 \times 64 \times 64$	adam	-	-	1
Miron *et al.* [76]	dice loss	250	$128 \times 240 \times 240$	-	$1e-4$	first 5 epochs warm-up LR	1
Chato *et al.* [107]	GDL	100	$64 \times 64 \times 64$	adam	$10e-4$	0.95 in LR with a period the drop of 5 epochs	Mini-batches 4, 8
Anand *et al.* [74]	WCEL	-	$64 \times 64 \times 64$	-	0.0001	10% LR factor based on validtion loss	4
Parmar *et al.* [72]	weighted multiclass dice loss	200+500	$128 \times 128 \times 128$	adam	$5 \times 10e-4$	$5 \times 10e-7$ (WRD)	4
Soltaninejad *et al.* [82]	Binary classification error (BCE)	100	$128 \times 128 \times 128$	rmsprop	$2.5 \times 10e-4$	-	5 mini-batches
Ali *et al.* [83]	-	250	$128 \times 128 \times 128$	-	$1e-05$	-	2
Marti Asenjo *et al.* [89]	cross-entropy, dice, hausdorff losses	100	$80 \times 80 \times 77$, 12 parts	adam	$1e-04$	divide the LR by 2 every 5 epochs	20, 1

Table 2.5 continued from previous page

Authors	A	B	C	D	E	F	G
Carmo *et al.* [73]	dice loss(BTS), smooth L1 Loss(SP)	300	128 × 128 × 128, 160 × 160 × 128	RAdam	$5e-04$	$1e-5$ WD, exponential LR decay by 0.985	1
Pei *et al.* [81]	dice loss	300	-	adam	0.001	$lr_i = lr_0 * \left(1 - \frac{i}{N}\right)^{0.9}$	1
Zhao *et al.* [80]	dice, binary cross-entropy losses(BTS), mean square error(SP)	200, 3000	80 × 160 × 160, 56 × 224 × 224	adam(BTS), stochastic gradient descent	$1e-4$, $1e-6$	$lr = lr \times \left(1 - \frac{epoch}{epochs}\right)^{0.9}$	1, 6
Patel *et al.* [77]	the unweighted sum of dice soft cross-entropy loss & boundary-WCEL	330	128 × 128 × 128	adam	$1 \times 10e-4$ to $1 \times 10e-7$	decoupled WD, cosine decay schedule for LRD	1
Dai *et al.* [78]	cross-entropy, soft dice loss	-	random	adam	$10e-4$	WD set to $10e-5$	2, 4
McKinley *et al.* [79]	cross-entropy loss, soft dice loss	60,000/5 iterations, 10 epochs	5 × 196 × 196	adam	($10e-4$ to $10e-7$), ($10e-5$ to $10e-8$)	weight decay $10e-5$, cosine annealing LR (20,000 gradient steps)	2

The implementation frameworks used for the techniques reviewed in this study are summarized in Table 2.6.

Table 2.6: Summary of Implementation Framework Details

Authors	**GPU Type and RAM**	**Platform/Packages**
González *et al.* [71]	NVIDIA Titan Xp 12 GB	Python, PyRadiomics
Agravat *et al.* [75]	NVIDIA Quadro K5200 and Quadro P5000	PyRadiomics
Miron *et al.* [76]	11 GB	PyTorch
Chato *et al.* [107]	NVIDIA Titan RTX 24 GB	Matlab 2019b
Anand *et al.* [74]	NVIDIA GeForce RTX 2080 Ti	PyTorch
Parmar *et al.* [72]	NVIDIA Quadro P5000	Keras with Tensorflow
Soltaninejad *et al.* [82]	NVIDIA Titan X 12 GB	PyTorch
Ali *et al.* [83]	NVIDIA Titan Pascal 12 GB	Keras with Tensorflow, PyRadiomics, OpenCV
Akbar *et al.* [84]	-	Keras with Tensorflow
Marti Asenjo *et al.* [89]	-	Matlab
Carmo *et al.* [73]	NVIDIA Titan X	-
Suter *et al.* [116]	NVIDIA Titan Xp	PyRadiomics
Zhao *et al.* [80]	NVIDIA GTX 1080Ti	PyTorch
Patel *et al.* [77]	NVIDIA Tesla V100 32GB	DeepNeuro with Tensorflow 2.2 backend
Dai *et al.* [78]	NVIDIA P100	-

2.3 Research Gaps and Findings

Even the most advanced algorithms have certain flaws, including the inability to investigate the effect of alternative scanners, protocols, sequences, and dataset sizes, resulting in limited usefulness in regular clinical practice. The segmentation of gliomas is a prerequisite for predicting the survival of the patient. Due to the histological complexity of gliomas, automatic segmentation is not so precise, which affects the SP task as well. From the literature review of SP based on glioma segmentation, it is evident that several research challenges need to be more focused on.

The research gaps and findings of the existing techniques are explained below:

(i) **Limited annotated data:** A large amount of well-annotated training data is required to build deep generalized models successfully. However, it is a tough job to acquire high-quality, labeled ground-truth data in medical imaging. The patient' s medical information is not publicly available due to ethical barriers and privacy protection. To com-

bat the limited size of publicly available medical datasets, techniques like data augmentation can generate synthetic training examples. Thus, the generation of artificial medical images is still an unexplored research pathway in the literature [151, 155–158]. To understand the tumor heterogeneity more clearly, advanced imaging techniques, such as Perfusion-Weighted Imaging (PWI) and Diffusion Tensor Imaging (DTI), may be employed [78].

(ii) **Architecture failure:** The existing network architectures failed to segregate the tumor for several HGG and LGG scans. The networks failed because of the small size of the whole tumor, the small size of the NCR, and the absence/small volume of ET. If NCR was not recognised from the raw MRI scans, then during feature extraction model marks all features zero except age due to network failure. In LGG situations, a smaller subcomponent size was observed [75].

(iii) **Difficult model training:** Researchers have encountered optimization difficulties, such as overfitting, vanishing or exploding gradients, and slow convergence speed while training 3D deep models. High-dimensional 3D data requires many parameters, massive memory utilization, and GPU processing to capture more characteristic features. There is a necessity for proper hyperparameter tuning to obtain the best model performance [159–162]. The number of network parameters grows in proportion to the network's depth, require more memory and time to fine-tune them on each epoch [65].

(iv) **Overfitting:** The technique is prone to overfitting because the training set is minimal [76]. The number of OS individuals examined is less than those of segmentation issues. With a smaller training dataset, the methods cannot extract meaningful information [72]. The trained model's poor performance on unseen patient data demonstrated its overfits to the training dataset [75]. Overfitting is a frequent issue with DNN-based approaches. It may arise owing to a lack of a sufficient quantity of labeled training data for BTS. Overfitting may be addressed by decreasing network complexity (network layers and parameters) or by producing extensive training data utilising image augmentation methods. Augmentation methods generate new images by manipulating input data and their associated ground-truth labels [65].

(v) **Worst tumor sub-regions to segment:** Miron *et al.* [76] found that the worst score was achieved on the ET segmentation, which appears to still suffer from false positives, consistent with other researchers' findings in the challenge. When researchers compared the mean outcomes of subregion segment scores for datasets, they discovered that the ET segment performed poorly [72]. While some segmentation methods performed well in ET and TC, they performed poorly in WT. It means that the models did

not do well when it came to categorising ED labels. It was also apparent from the DSC median values, where the median DSC for TC was more significant than the WT [83].

(vi) **LGG cases:** There are certain outlier instances whereby performance has declined [72]. The overall performance of suggested solutions was outstanding in more than half of the instances, and researchers believed that the LGG examples decreased the mean values. The increased ET sensitivity values demonstrated that the models had missed a small number of ET occurrences [83]. Training LGGs and HGGs separately have been found to improve performance. When it comes to testing, though, this information is not readily accessible. Instead of a distinct boundary between the TC and the surrounding ED, LGGs often exhibit a gradual reduction in tumor-cell density [163].

(vii) **Feature selection:** As a result of utilising radiomic characteristics, the authors had achieved unsatisfactory performance. Many characteristics were initially retrieved, resulting in a tough time choosing the most important ones, which may explain the problem. The robustness and repeatability of the feature selection method must be evaluated in future research [71]. Due to a variety of network characteristics, normalisation has a minimal impact [72]. Due to its reduced fluctuation inaccuracy, PCA was selected for processing high-dimensional data [86]. Adding more characteristics to the radiomics model may improve its performance on unknown sets [78]. The maximum reported accuracy for the SP task is less than 63% [65]. The dependency of the SP feature extraction step on segmented tumor labels can affect the complete survival prognosis outcome. The biological significance of the retrieved characteristics is also significant. The SP cannot be accurate if the significance of the characteristics is not adequately understood.

(viii) **Class biasness:** MRI volumes can be segmented through DL techniques, but this is becoming complicated due to highly imbalanced data. In the case of LGG, this class imbalance is more serious. The segmentation networks do not perform well with LGG volumes. Data with large-class variance or inadequate variation contribute to poor performance in the processing of medical images. The segmentation classes are highly asymmetric. For a sample brain tumor MRI slice, about 98% of the voxels belong to either healthy tissue or the surrounding black region, 1.02% belong to the ED class, 0.29% belong to the ET class, and 0.23% belong to the NCR/NET class [132, 164]. The class imbalance may be dealt with via appropriate training data sampling, better loss functions, and augmentation methods. There are only a few survival models that can forecast mid-survivors. It indicates that separating among three classes is more complicated than just separating between two.

(ix) **Model generalization:** CV could not generalise the networks even after considering very few parameters [72]. When comparing official validation and testing data, it is essential to note that OS MSE increased significantly, while accuracy did not decrease as much. Further research on the OS multiproblem is required [73]. The model' s generalizability can be enhanced by identifying architectural improvements [77]. The use of traditional linear regression models for OS prediction provides superior results due to their robustness and generalizability [78].

(x) **Non-consistent performance:** There is a more considerable performance difference in OS prediction than BTS [87]. The performance obtained on the training set was not preserved on the testing set. The researchers found that normalisation based on normal healthy brain sub-regions reduced the DSC values for the segmentation model when no prior knowledge was used. Ultimately, such normalization approaches did not result in consistency on unseen data [116].

2.4 Problem Formulation

According to the WHO, glioblastoma is identified as highly invasive and aggressive grade IV type glioma. Grade IV gliomas are malignant brain tumors that are cancerous by nature and cause life threatening consequences for the patients. The life span of GBM patients is less than two years due to its abnormal growth and outspread to other parts of the patient's body. Therefore, it is very important to predict the overall survival time of GBM patients on prior basis for appropriate treatment planning. Early detection and precise segmentation of tumor region are two pre-requisites for accurate survival estimation of the patient. Tumor segmentation is extremely valuable for estimating the volume of the cancerous brain tissues prior to surgery or radiation therapy. Tumor segmentation is performed by radiologists in clinics by using pre-operative brain MRI scans. Manual delineation of gliomas from MRI modalities can be a very challenging and time-consuming task for expert radiologists. The manual delineation of MRI scans suffer from inter-observer and intra-observer variability. Also, this task becomes very laborious and impractical with high-dimensional MRI scans. The low contrast of imaging modalities and complicated appearance of gliomas make it exceptionally difficult for the radiologist to create a distinction between a healthy tissue and an abnormal tissue. Large variations are observed in manual glioma segmentations and survival time predictions made by experienced clinicians. Moreover, manual work performed by radiologists distract their attention and time away from other treatment related tasks. A growing number of researchers have suggested various ML and DL computational literature approaches to achieve automated glioma segmentation and survival prognosis. However, the proposed radiomic solutions need potential improvements before actual acceptance in

clinical care. Therefore, developing a robust BTS and SP model is still very much in demand.

2.5 Objectives

The following objectives are set for the research proposal:

(i) To propose and implement machine learning-based model for segmenting glioma sub-regions in pre-operative brain MRI scans.

(ii) To perform feature extraction and selection for the pre-operative and segmented brain MRI scans.

(iii) To predict the survival time of glioblastoma patients through proposed model using selected feature set.

2.6 Proposed Methodology

To achieve the set objectives, the following methodology will be adopted:

(i) ***To achieve the first objective:*** DL-based models are proposed to produce segmentation labels of the different glioma sub-regions. The sub-regions considered for evaluation are ET, TC and WT.

(ii) ***To achieve the second objective:*** Feature extraction is perfomed with python packages like PyRadiomics or automated feature extraction using pre-trained CNNs or trained from scratch. Feature selection is performed using PCA or ML regression models to select the best features related to SP of glioblastoma patients.

(iii) ***To achieve the third objective:*** ML-based models are trained using PyCaret package or DL models based on selected features to predict the survival time of the patient. The proposed techniques are evaluated using BraTS dataset of pre-operative brain MRI scans, which is existing benchmark repository for BTS and SP of glioblastoma patients.

Chapter 3

Radiomic Features-based Survival Prognosis of Glioblastoma Multiforme Patients

The combination of radiomics and artificial intelligence has emerged as a strong technique for building predictive models in radiology. This study aims to address the clinically important issue of whether a radiomic profile can predict the OS time of GBM patients having GTR through pre-operative structural MRI scans. A retrospective analysis was made using data of glioma patients made publicly available by the University of Pennsylvania. The radiomic characteristics were extracted from pre-operative structural mpMRI sequences after pre-processing and segmentation using 3D deep-supervised UNet-based encoder-decoder architecture. After removing irrelevant features, regression models based on ML were developed by considering selected features to predict the OS time of GBM patients within a period of days only. The patients were divided into three survivor groups depending on their projected survival time. To validate the significance of the selected feature set, statistical analysis was performed. As many as 494 patients were considered to improve SP by using more effective feature extraction and selection techniques. The ridge regressor acquired the highest Spearman rank correlation of 0.635 with an accuracy of 69%, the greatest of all the previous works for categorical predictions of such patients. The researchers in the past who used radiomic characteristics for the OS prognosis of GBM patients could yield limited results only. However, the current research work recorded an enhanced accuracy and Spearman rank correlation for the three survivor classes of GBM patients using ML, feature selection, and radiomics. The significance of this work lies in the selection of patients with GTR and the extraction of characteristics through the use of radiomics and artificial intelligence.

3.1 Contribution of the Proposed Technique

"Radiomics" is the process of extracting quantitative information from radiological images, such as T1, T2, T1CE, and T2-FLAIR images and building algorithms that correlate these factors with patient survival and outcome [165]. Radiomics expose pathophysiology, particularly intratumor heterogeneity, by turning medical scans into high-dimensional data.

Unseen by the human eye, this radiomic extraction procedure enhances clinical visual perception [166]. Further, clinical data such as the patient's age and the Extent of Resection (EOR) might be useful in predicting the course of the disease. Predicting a patient's prognosis is quite important when making decisions about treating and managing them. As a prognostic factor, the pre-operative tumor volume is also considered. As some of these traits are interpretation-dependent, there is an urgent need for further quantifiable radiological examination [167].

The proposed technique contributes in the following manner:

- The tumor and its subregions were segmented using 3D deep-supervised UNet-based encoder-decoder architecture.
- The segmented and raw FLAIR MRI scans were used for radiomic feature extraction.
- The feature space was reduced from 1112 to 186 features using ML-based feature selection algorithms.
- The selected feature set was used to train automated ML regression-based models.
- The best performing ridge regressor model was used to predict patients' survival time in days.
- The obtained results were compared with state-of-the-art techniques to validate the performance of the proposed methodology.

This chapter has been organized into various sections. Section 3.1 provides an overview about the proposed technique's contribution in the field under investigation. Section 3.2 deals with the dataset and explains the proposed methodology. The experimental set-up and evaluation parameters have been described in Section 3.3. The results and discussion are presented in Section 3.4. Finally, Section 3.5 carries the findings of the proposed technique.

3.2 Materials and Methods

3.2.1 Dataset and Proposed Methodology

A retrospective collection of glioma patients titled 'BraTS' data made publicly accessible by the Perelman School of Medicine, the University of Pennsylvania, was used for this investigation. Data was collected from 19 different institutions using a variety of clinical procedures and scanners [1, 66–69]. The inclusion criteria were the availability of skull-stripped pre-operative structural mpMRI scans such as T1, T1CE, T2, FLAIR, and clinical informa-

tion such as the patient's age and degree of resection. Only patients with GTR were used for SP training and validation. Because this publicly accessible data contained no patient identities, no institutional review board permission was necessary. Table 3.1 illustrates the detailed demographics of the patients' data. This patient population-based cohort research aimed to investigate the relationship between radiomic characteristics collected from pre-operative MRI images and GBM patients' OS prognosis.

Table 3.1: Patient Demographics

Data	**Input**	**Training Cohort**	**Validation Cohort**	**Output**
Segmentation	Pre-operative MRI sequences	N=369	N=125	Labeled scan
Survival	Surgery status, patient's age	N=236	N=29	Survival days
	Surgery (GTR)	N=119	N=29	
	Age mean (in years)	61.98	57.27	

MRI sequences available in NIfTI format were co-registered to the anatomical template of the same patient's T1 MRI sequence. All patients in the training set were labeled (seg NIfTI file) for the different tumor tissues, *i.e.*, NCR/NET (label 1), ED (label 2), and ET (label 4). Skull-stripping and interpolation to 1mm isotropic resolution were performed on all patient images. One to four experienced neuroradiologists segmented the scans using the same labeling approach to create ground-truth annotations. The given segmentation label values are represented with colours as given below for the four unique labels:

- Label 0: Background (black colour)
- Label 1: NCR and NET (red colour)
- Label 2: ED (green colour)
- Label 4: ET (yellow colour)

Following BTS, the segmentation labels may be utilized to extract imaging/radiomic characteristics, which can subsequently be used to train ML and DL models for predicting the survival of GBM patients. It is not necessary to limit the parameters to solely volumetric features. Clinical factors such as the patient's age and resection status, intensity, histological, textural, spatial, glioma diffusion properties, and histogram-based information may also be used by the researchers. Individuals who had GTR surgery must provide a prognosis for the SP job. The performance of the method is verified using either classification accuracy or the regression-based MSE measure. The researchers submit their prediction results as a CSV file with subject ids to CBICA's IPP (https://ipp.cbica.upenn.edu/).

In this retrospective work, we improved survival prognosis based on MRI-based radiomic

features generated from delineated tumor subregions and clinical data. Figure 3.1 depicts the proposed architecture for BTS and SP utilizing mpMRI.

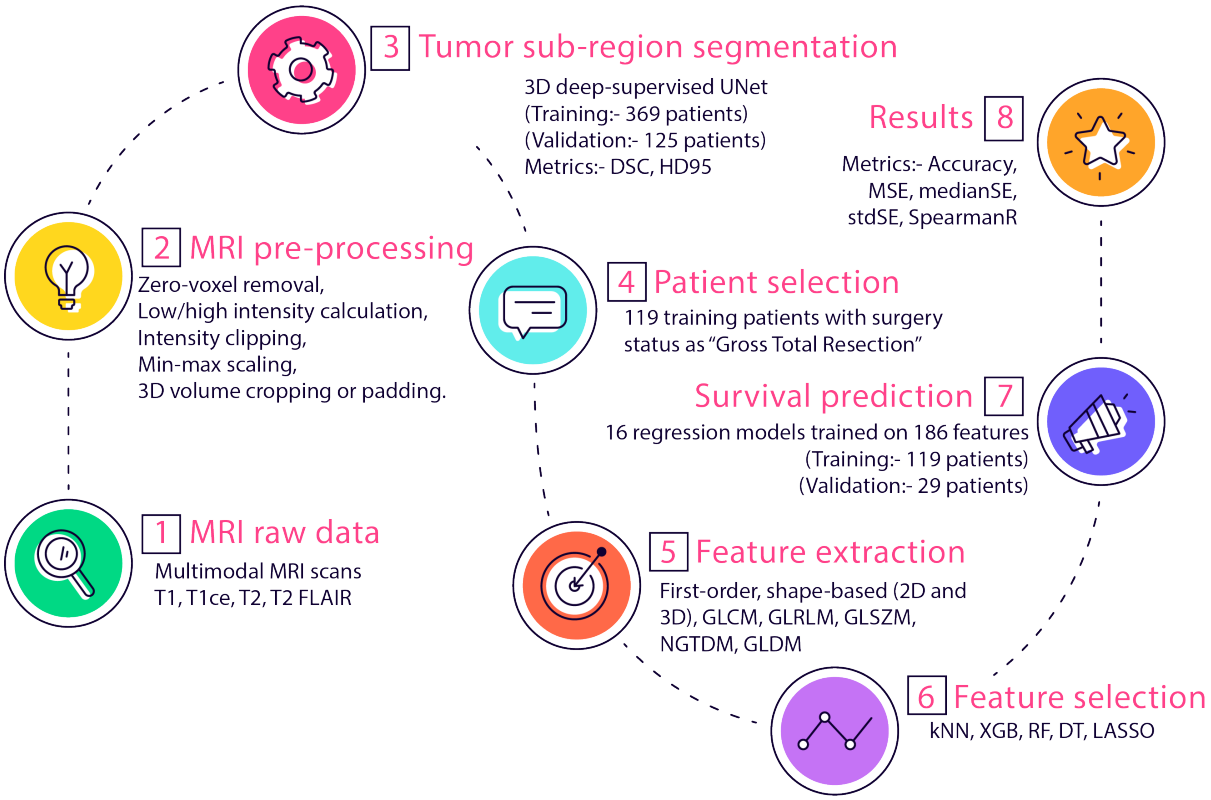


Figure 3.1: Proposed Methodology

Firstly, structural mpMRI images were pre-processed; then, a 3D UNet model with deep supervision and Stochastic Weight Averaging (SWA) was used to separate tumor subregions. An independent set of glioma patients was used to validate the segmentation technique. Afterward, radiomic characteristics were collected from tumor subregions and overall tumor volume. In order to prioritize the relevance of features, four ML-based regression models with CV and grid search were utilized. On the basis of their features, at least two models were considered to be the best and were included in the final feature set. Statistical metrics were used to select the best ML algorithm, *i.e.*, the Ridge regressor. Its effectiveness was tested on previously unidentified GBM patients in the GTR surgery group. The validation set was evaluated using the CBICA IPP. 'Overall survival' refers to the time between the first brain MRI and the final clinical follow-up or death. Patients were no longer followed after five years (approximately) for this data. No censored data has been used in this study.

3.2.2 Pre-processing

The authors used five pre-processing approaches to account for intensity inhomogeneity across the dataset, as illustrated in Figure 3.2. Standardizing the intensity of the raw MRI

volumes was necessary since the data came from many institutions with different acquisition equipment. The BraTS 2020 challenge organizers provided pre-operative MRI images and ground-truth annotations for the training set. To ensure that imaging procedures did not cause too much variation in patient scans, the coordinators provided data that had been co-registered to each patient's T1 anatomical template, normalized to $1mm$ isotropic resolution, and then skull-stripped. The performance of the network, as well as its training, was improved by pre-processing.

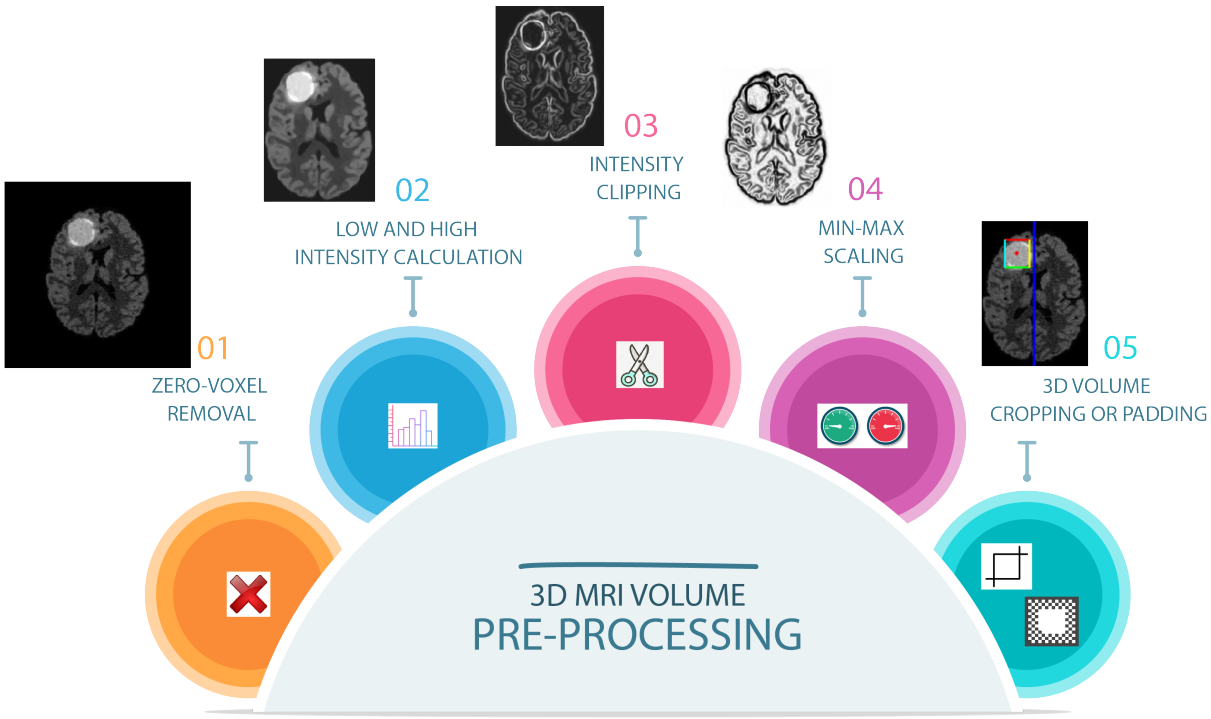


Figure 3.2: Steps of Pre-processing Performed on Multimodal 3D MRI Volumes

The 3D brain MRI NifTI scans were first processed by removing zero voxels, *i.e.*, the extra background area in the scan other than brain region, as per the formula given in *Eq.* 3.1.

$$image[non_{zeros}] = image_{intensity} > 0 \tag{3.1}$$

where: $image[non_{zeros}]$ = brain's volume area other than background,
$image_{intensity}$ = voxel's intensity value of the MRI scan.

The non-zeros-based scan was used to calculate the low and high intensities to be clipped as per *Eq.* 3.2.

$$low, high = percentile\Big(image[non_{zeros}], \Big[low_{perc}, high_{perc}\Big]\Big) \tag{3.2}$$

where: low = calculated lowest intensity outlier value to be clipped,
$high$ = calculated highest intensity outlier value to be clipped,
$image[non_{zeros}]$ = brain's volume area other than background,
low_{perc} = lowest intensity percentile to be clipped is set as '1',
$high_{perc}$ = highest intensity percentile to be clipped is set as '99'.

The low and high percentiles considered were 1 and 99 for clipping raw MRI scans as given in *Eq.* 3.3.

$$image_{clipped} = clip(image[non_{zeros}], low, high) \tag{3.3}$$

where: $image[non_{zeros}]$ = brain's volume area other than background,
low = calculated lowest intensity outlier value to be clipped,
$high$ = calculated highest intensity outlier value to be clipped,
$image_{clipped}$ = intensity clipped brain's volume area other than background.

The percentile and clipping functions have been taken from the python-based NumPy library to perform mathematical operations. The intensity clipped scan was further used to perform the min-max scaling for intensity standardization as per the formula mentioned in *Eq.* 3.4.

$$image_{min-max} = \frac{image_{clipped} - \min(image)}{\max(image) - \min(image)} \tag{3.4}$$

where: $image_{min-max}$ = intensity clipped brain's volume area other than background,
$min(image)$ = minimum intensity value of the image,
$max(image)$ = maximum intensity value of the image,
$image_{clipped}$ = intensity clipped brain's volume area other than background.

As the original MRI scans consume high memory and processing time, thus, the raw scans with dimensions $240 \times 240 \times 155$ have been randomly cropped to remove all the zero background. As a model input, the available four modalities' 3D voxels were clipped to get only the brain area. In their research work [95], the researchers used $128 \times 128 \times 128$ 3D data to suit the GPU RAM and produced the best results in the BraTS 2018 competition. So, taking a cue from the researchers, the useless black region around the actual volume of interest,

i.e., the brain region was also cropped in the proposed methodology. Further, processing of larger patches might have temporarily made the memory consumption to spike higher. Images were then randomly re-cropped to a fixed patch size of $128 \times 128 \times 128$ after being cropped to a variable size using the smallest bounding box enclosing the whole brain. This enabled us to eliminate a major part of the unnecessary background from the original volume and provided a comprehensive view of each brain tumor. To ensure that the network learns enough anatomic context, large input patch sizes of at least $128 \times 128 \times 128$ voxels have become the standard for training 3D networks. Cropping was done mainly for raw modalities; and padding was done usually on segmentation label scans. Model training was done only on original data; and no data augmentation technique was used to generate the synthetic data.

3.2.3 Tumor Segmentation Using Deep Learning

Based on 3D MRI data, techniques for 3D volumetric segmentation of brain tumors are becoming more accessible. These techniques provide more exact and accurate volume and location estimations of tumors and the extraction of radiomic characteristics. The 3D U-shaped encoder-decoder architecture was employed to segment tumor subregions, *viz.*, WT, TC, and ET. The TC area is comprised of the NCR, ET, and NET. The WT is formed by the TC and ED regions. The 3D design preserved global spatial relationships between voxels from various human brain regions. Figure 3.3 illustrates the segmentation architecture. It consists of a contracting path (encoder), a connecting path (bottleneck), and an expanding path (decoder).

Encoder: The 3D UNet encoder used four distinct 3D mpMRI sequences with an input dimension of $128 \times 128 \times 128 \times 4$. It included the repetitive application of two 3D convolutions, followed by a non-linear activation function based on GN [119] and ReLU [168]. Further, downsampling was accomplished via a 3D max-pooling procedure. The preceding procedures were performed four times to build 32, 64, 128, and 256 features.

Bottleneck: The bottleneck repeatedly applied two dilated 3D convolutions, GN and ReLU, to yield 256 feature maps identical to the input. Dilated convolutions boosted network depth without increasing the size of feature maps or memory usage. Finally, a single 3D convolution operation converted 512-feature maps to 256-feature maps. The first input was 256 feature maps from dilated convolutions, and the second came from the encoder stage 4 output. Before convolution, these two 256-feature maps were concatenated, followed by GN and ReLU.

Decoder: Each step along the decoder's extended route included spatial upsampling of the

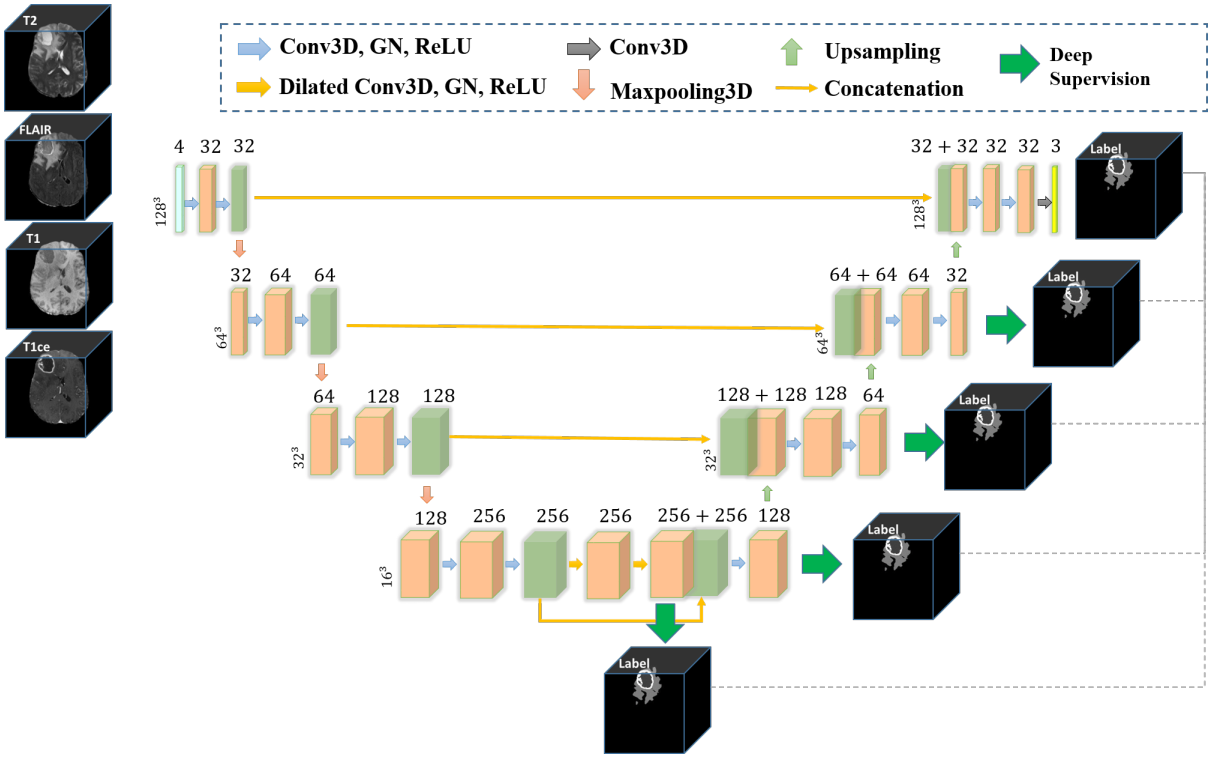


Figure 3.3: Illustration of 3D Deep Supervised UNet Segmentation Architecture

feature map using trilinear interpolation with a scale factor of two. Following each upsampling layer, two convolutions were used to decrease the number of feature maps by half, concatenated with the feature maps collected from the parallel contracting path, and then GN and ReLU operations were performed. A 3D convolution operation assigned the appropriate number of volume-wise tumor class labels to each 48-component feature vector (labeled as 1, 2, and 4). Deep supervision was used at multiple decoder stages to avoid the gradient vanishing problem. Multiple segmentation maps at various resolution levels were developed and integrated to improve resolution-based segmentation maps. To construct secondary segmentation maps, $1 \times 1 \times 1$ convolutions transposed the feature maps from each network level.

Loss Function: Because of the extremely imbalanced nature of this data, the segmentation models are trained using the Dice Loss function in conjunction with a Ranger optimizer. The 3D UNet segmentation model was trained with the 'dice loss' as a metric to calculate the overlap of 'predictions' and 'targets' tumor subregions. The loss function calculated batch-wise and channel-wise is given in *Eq.* 3.5:

$$L_{dice} = 1 - \left\{ \frac{2 \times \Omega + \epsilon}{\sum_n p^2 + \sum_n t^2 + \epsilon} \right\} \tag{3.5}$$

where, 'L_{dice}' denotes dice loss function, 'Ω' denotes the intersection operation calculated as given in *Eq.* 3.6, 'p' stands for prediction, and 't' represents targets, 'n' stands for the channel numbers for four MRI modalities, and 'ϵ' represents the smoothing factor ($\epsilon = 1$).

$$\Omega = \Omega(p, t) \tag{3.6}$$

The dice loss function considers the projected probabilities in their natural form without first transforming them into binary output. The numerator provides individual accurate predictions between the input and the goal, whereas the denominator provides standard correct predictions between the input and the target. This ratio normalizes the loss according to the target mask and permits learning even from the smallest spatial representation of the mask, which is accomplished by dividing the loss by the target mask.

3.2.4 Radiomic Feature Extraction Using PyRadiomics

MRI scans may provide quantitative phenotypic aspects of brain tumors. Using the PyRadiomics Python Package [143], we extracted radiomics characteristics from ED, NET, NCR, and the WT region. Radiomic feature extraction was performed on label 1, *i.e.*, NCR/NET region of segmented and complete raw FLAIR MRI data. The NCR/NET region was selected for radiomic feature extraction because of its solid visible boundaries compared to blurred ED and ET regions. The characteristics retrieved for each GBM case were divided into three groups. The extracted four different feature categories have been illustrated in Figure 3.4.

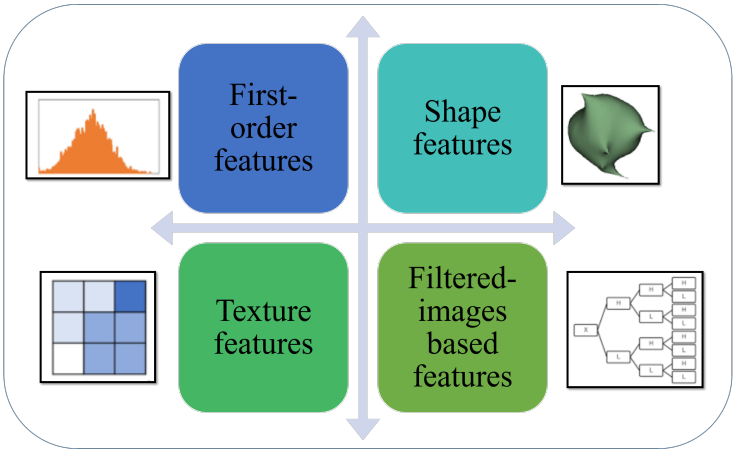


Figure 3.4: Radiomic Feature Extraction Performed on FLAIR and Segmented MRI Scans

The first group, first-order statistics, included 18 features such as mean, median, range, variance, uniformity, skewness, kurtosis, energy, entropy, total energy, minimum, interquartile

range, maximum, 10th percentile, robust mean absolute deviation, 90th percentile, mean absolute deviation, and root mean squared. These characteristics defined the tumor region's gray level intensity.

The second group, shape, included 14 features, with surface area to volume ratio, mesh volume, voxel volume, maximum 2D diameter slice, maximum 3D diameter, maximum 2D diameter column, surface area, maximum 2D diameter row, sphericity, major axis length, elongation, minor axis length, least axis length, and flatness. These characteristics defined the tumor region's shape.

The third group, texture, included five NGTDM, fourteen GLDM, sixteen GLSZM, sixteen GLRLM, and twenty-two GLCM. These characteristics defined the tumor region's texture.

For the purpose of this work, original MRI scans, images filtered through the LoG, and images produced using low-pass and high-pass wavelet decomposition were utilized to extract radiomic features. As LoG filtering helps in improving image edges and perhaps improve tumor boundaries, the wavelet decomposition technique was used to decompose images into many levels of detailed elements.

With a bin width of 5 and a voxel array shift of 300, more than a thousand radiomic characteristics were extracted. These are detailed as below:

- Original image-based features: 105
- LoG filtered image-based features (sigma: 1, 3, 5): 273
- Wavelet decomposed image-based features: 364
- Square computed of image intensities-based features: 91
- Square root computed of absolute image intensities-based features: 91
- Logarithm computed of absolute image intensities-based features: 91
- Exponential computed of original image-based features: 91
- Diagnostic image-based features: 4

These 1110 radiomic features were coupled with patient's age data and tumor volume (non-zero voxels) with a total of 1112 features for predicting the OS time of GBM patients.

3.2.5 Feature Selection Using Machine Learning

For data analysis, feature selection is a critical component of data mining applications [169, 170]. A fraction of the characteristics collected was redundant or unrelated to SP. Thus, fea-

ture selection was used to choose a subset of characteristics with the highest predictive potential to improve performance and minimize overfitting. Four distinct regression-based ML models from the Sklearn Python Package were used to pick the features. First of all, the relative relevance of features was determined by fitting each of these regressors independently to the 1112 retrieved features. Then, a subset of the optimum number of excellent features was picked for each regressor by setting different thresholds for each regressor.

The feature selection was performed by using five ML-based algorithms. Out of 1112 features, Least Absolute Shrinkage and Selection Operator (LASSO) regressor selected 51 significant features using grid search with CV, K-nearest neighbors (KNN) regressor selected 26 features using permutation importance, Extreme Gradient Boosting (XGB) regressor selected 230 features, RF selected 248 features, and DT selected 209 features.

In all, 186 features were selected based on the criteria that a feature must be selected by at least two ML feature selection algorithms as mentioned above. The features selected by a particular feature selection algorithm were highlighted in green color and those not selected by yellow color in Table 3.12 (Appendix). The selected features were derived from both original and derived images (original: 18, square: 11, squareroot: 4, wavelet: 57, log-sigma: 54, logarithm: 14, exponential: 23, diagnostics: 3, age and tumor volume). It was found that most of the features selected originated from wavelet and log-sigma image decompositions.

3.2.6 Machine Learning Algorithms for Survival Prediction

Different ML regression models from the PyCaret library were trained for the purpose of this research to see which algorithms performed well for the selected feature set. These included AdaBoost (ADA) Regressor, Bayesian Ridge (BR), DT Regressor, Elastic Net (EN), Extra Trees (ET) Regressor, Gradient Boosting Regressor (GBR), Huber Regressor, KNN Regressor, Lasso Least Angle Regression (LLAR), LASSO Regression, Least Angle Regression (LAR), Light Gradient Boosting Machine (LightGBM), Linear Regression (LR), Orthogonal Matching Pursuit (OMP), Passive Aggressive Regressor (PAR), RF Regressor, and Ridge Regression. The Mean Absolute Error (MAE), MSE, and Root Mean Square Error (RMSE) loss metrics were used to check the performance of all the regression models [171]. The description of regression-based ML algorithms is given in Table 3.2.

Table 3.2: Description of Machine Learning Algorithms Trained for Survival Prediction

Model	**Description**	**Advantages**	**Disadvantages**
AdaBoost Regressor	Adaptive Boosting algorithm. The final output of the boosted classifier is a weighted sum of the various learning algorithms ('weak learners'). AdaBoost adapts weak learners to misclassified examples from earlier classifiers.	Less likely to overfit. Captures several non-linear correlations.	Requires a high-quality dataset. Before using the AdaBoost algorithm, avoid noisy data and outliers.
Bayesian Ridge	Calculates the values of the regression coefficients using Bayes' theorem. Rather than computing the least squares, this method estimates the posterior distribution of the features. Very helpful when we do not have enough data or when the data is not spread out well. The outcome of a Bayesian Regression model is derived from a probability distribution, in contrast to conventional regression approaches in which the output is derived from a single value for each feature.	Extremely effective when the dataset size is small. Ideal for real-time learning. No prior knowledge of the dataset is required. In terms of stability, it outperforms standard linear regression.	Model inference may take a long time. Unsuitable for big amounts of data.

Table 3.2 continued from previous page

Model	Description	Advantages	Disadvantages
Decision Tree Regressor	A decision-making tool employs a tree-like flowchart structure or is a model of choices and all potential outputs, including outcome, input costs, and utility. Decision tree regression evaluates an object's properties and trains a model with a tree structure to forecast data in the future to create meaningful continuous output. When making a choice, it usually mimics human reasoning abilities. Because logic has a tree-like structure, it is simple to grasp.	Easy to grasp and interpret. Little data preparation is required. The tree's cost is proportional to the number of data points utilized to train it. Capable of working with both numerical and category data. Capable of dealing with multi-output difficulties. Makes use of a white-box model. Statistical tests may be used to validate a model.	Does not function well with smooth borders. Does not function when the variables are unrelated. It has a large variance due to the greedy technique. More time is required to train the model. It can grow intricate, resulting in overfitting. Unstable because small changes in the data might result in the generation of an entirely new tree. Concepts like XOR, parity, and multiplexer difficulties are difficult to grasp since decision trees do not clearly describe them. Make biased trees if certain classes predominate.
Elastic Net	Combining ridge and lasso regression methods is especially beneficial when dealing with highly correlated data. Uses the penalties with the ridge and lasso regression methods to make regression models more accurate. Encourages group effect in the case of strongly linked variables There are no restrictions on the number of variables that may be chosen. Double shrinking is possible.	Does not have the issue of picking more than n factors when n<<p, but LASSO does. When we have a large dataset, elastic regression works nicely. Elastic net is essentially a hybrid of L1 and L2 regularization.	A more intensive computational process than either LASSO or Ridge.

Table 3.2 continued from previous page

Model	Description	Advantages	Disadvantages
Extra Trees Regressor (Extremely Randomized Trees)	Algorithm for ensemble learning Builds the collection of decision trees. The decision rule is chosen at random during tree building. Comparable to Random Forest, except that split values are selected at random. Uses the full original sample, *i.e.*, no bootstrapping, unlike random forest.	Lower computational cost. Extra trees tend to perform better with noisy characteristics. Three times quicker than the random forest technique.	Selects the split point at random rather than calculating the ideal one.
Gradient Boosting Regressor	The most widely used method for tabular datasets. Capable of detecting any non-linear connection between the model target and features and dealing with missing values, outliers, and large related categorical values on data features without additional treatment.	Increased precision. Requires less pre-processing. Greater adaptability. Deals with missing data.	Can overemphasize outliers, leading to overfitting. Expensive to compute. During tuning, a huge grid search is required. It is less interpretive.
Huber Regressor	Robust regression is aware of outliers in a dataset and gives them less weight than other instances in the dataset. Huber loss, square loss (centered on the mean), and absolute value loss (centered around the median).	Widely used in instances where outliers might affect the effectiveness and precision of least-squared-loss error-based regression.	Resistant to outliers but will not disregard them.

Table 3.2 continued from previous page

Model	Description	Advantages	Disadvantages
K-Nearest Neighbors Regressor	Lazy learner algorithm because it does not instantly learn from the training set. Instead, it saves the dataset and takes action on it at the time of prediction. A non-parametric algorithm implies it makes no assumptions about the underlying data.	There is no training period. Simple to implement. New data may be smoothly added without affecting the algorithm's accuracy.	Does not perform well with large sizes. Susceptible to noise, missing values, and outliers. Does not perform well with big data sets because the cost of computing distance is very expensive. Requires feature scaling.
LASSO Regression	If a set of predictors is strongly connected, the lasso selects just one and reduces the rest to zero. Penalizes regression coefficient magnitude. Reduces coefficients to zero (exactly zero), which aids in feature selection. L1 regularization is carried out. Lasso regression assumes the same as that of least squared regression, except normality.	Performs feature selection automatically. More precise forecast. Promotes basic, sparse models (*i.e.*, models with fewer parameters). Suitable for models with high degrees of multicollinearity.	Selected characteristics will be heavily skewed. LASSO chooses mostly n features for n<<p (n-number of data points, p-number of features). LASSO will choose just one characteristic from a set of associated features; the choice is random. The feature used for each bootstrapped data might be significantly different. Prediction accuracy is lower than that of Ridge regression.

Table 3.2 continued from previous page

Model	Description	Advantages	Disadvantages
Lasso Least Angle Regression	LLAR is a regression technique for high-dimensional data. Comparable to forward stepwise regression. Because it is used with data having a number of attributes, LLAR determines the attribute most significantly associated with the goal value at each stage. There might be several attributes with the same association. In this case, LLAR averages the attributes and advances in the same direction as the attributes. That is why the technique is known as Least Angle Regression. In essence, LLAR jumps in the ideally determined direction while avoiding overfitting the model. Normalizes all data such that they have a zero mean and unit variance.	Computationally as quick as forwarding selection, but may be more accurate sometimes. When the number of features is substantially more than the number of data instances, it is numerically highly efficient. It is readily adaptable to generate answers for various estimators.	LLAR is quite sensitive to noise and might provide unexpected results at times.
Light Gradient Boosting Machine	The gradient boosting approach based on decision trees improves model efficiency and minimizes memory use. LightGBM enhances the gradient boosting technique by including a form of autonomous feature selection and concentrating on boosting cases with greater gradients. Leaf-wise approach reduces more loss than the level-wise technique, resulting in superior accuracy to current boosting algorithms.	Increased training pace and efficiency. Reduce memory consumption. Greater precision than any other boosting technique. Support for Large Datasets. Support for concurrent learning.	LightGBM splits the tree leaf-by-leaf, which might result in overfitting since it creates more complex trees. Because LightGBM is susceptible to overfitting, it may easily overfit tiny data sets.

Table 3.2 continued from previous page

Model	Description	Advantages	Disadvantages
Linear Regression	Independent and dependent variables must possess a linear connection. Multicollinearity may increase coefficient variance and sensitivity to model modifications. In the case of several independent variables, the model may use forward selection, backward elimination, and stepwise method.	Easy approach. Excellent interpretation. Simple to implement.	Assumes linear connection, which is usually inaccurate. Susceptible to outliers. Overfitting occurs when the number of observations is small. Has multicollinearity, autocorrelation, and heteroskedasticity.
Orthogonal Matching Pursuit	A sparse approximation algorithm finds multi-dimensional data's 'best matching' projection onto an over-complete dictionary. The most common greedy approach is discovering a sparse linear system solution vector. OMP uses projection to locate sparse solution vector support indices.	Recovers a high-dimensional sparse signal from noisy linear measurements.	Time-consuming if data is large.
Passive Aggressive Regressor	Progressively trained by feeding examples sequentially, singly, or in groups called mini-batches. Applied to large-scale learning. Algorithm for online learning.	Useful when training on the complete dataset is computationally impossible.	Works well for real-time data only.

Table 3.2 continued from previous page

Model	Description	Advantages	Disadvantages
Random Forest Regressor	The ensemble approach, which employs numerous decision trees and techniques such as Bootstrap and Aggregation, is frequently called bagging. Instead of depending on individual decision trees, combine many decision trees to determine the outcome. Typically outperforms on a wide range of problems, particularly those with nonlinear connections.	Performs well on huge datasets. Can handle hundreds of input variables without deleting any of them. Estimates which variables are significant. Estimates generalization error internally as the forest grows. An efficient approach for predicting missing data that retains accuracy even when a considerable part of the data is missing. High performance and precision. Estimates the significance of features. There is no need for feature scaling.	Some data sets are overfitted. Favors category qualities with higher levels, rendering variable significance ratings unreliable. A black box approach with less interpretability. More computing resources are required. The prediction time is short. We must determine the number of trees included in the model.
Ridge Regression	Used when the independent variables have a strong correlation. While least square estimates are impartial in the presence of multicollinearity, their variances are large enough to cause the observed value to deviate from the true value. Ridge regression minimizes standard errors by skewing the estimates of regression. Model tuning method. L2 regularization is carried out.	Parameter reduction to avoid multicollinearity.Coefficient reduction and model simplification. Works effectively with big multivariate data sets when the number of predictors (p) exceeds the number of observations (n).Prevents the overfitting of models. No need for unbiased estimators.	Incorporates all predictors into the final model. Unable to conduct feature selection. Reduces coefficients to zero. Exchanges variance for bias.

Table 3.3: Description of Error Metrics Used to Compare Model Performance on the Training Dataset

Metric	**Description**	**Formula**	**Advantages**	**Disadvantages**
Mean Absolute Error (MAE)	The average of the absolute discrepancies between the actual and anticipated values from the model. A small MAE indicates that the model is excellent at prediction, while a big MAE indicates that the model may struggle in specific areas.	$MAE = \frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - \hat{Y}_i\right\|$ (3.7) where, Y_i= actual label value, $\hat{Y}_i$= predicted label value.	MAE is in the same unit as the output variable. Most resistant to outliers.	Because MAE is not differentiable, we must use differentiable optimizers such as Gradient descent.
Mean Squared Error (MSE)	MSE is the mean of the discrepancies between anticipated and expected target values.	$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$ (3.8) where, Y_i= actual label value, $\hat{Y}_i$= predicted label value.	A squared difference removes the sign, so the error value is now positive. Squares also have the effect of making big mistakes even bigger.	If the dataset contains outliers, it penalizes them the most, and the estimated MSE is larger. As a result, it is less resistant to outliers, which is a benefit in MAE. Squaring a little positive error overstates the model's poor performance, whereas squaring a little negative error understates it.
Root Mean Squared Error (RMSE)	The average root-squared difference between the actual and anticipated values. The lower the RMSE score, the better the model's predictions. A higher RMSE implies significant differences between the expected and actual values.	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)}$ (3.9) where, Y_i= actual label value, $\hat{Y}_i$= predicted label value.	The output value is in the same unit as the desired output variable, simplifying loss interpretation.	When compared to MAE, it is less resistant to outliers.

The description of the metrics used for comparison of ML models is given in Table 3.3. The ridge regression model based on the least MAE was selected, as reflected in Table 3.4. The available regression models were cross-validated using ten-fold CV. The MAE metric results for CV benchmark criteria of all the trained regression models are presented in Tables 3.13 and 3.14 (Appendix).

Table 3.4: Performance Of Different ML Models on the Training Dataset Before Hyperparameter Tuning (n=119 GBM Cases)

Index	**Model**	**MAE**	**MSE**	**RMSE**
Ridge	Ridge Regression	256.9772	125770.1	334.9074
ADA	AdaBoost Regressor	269.2462	143733	363.1874
BR	Bayesian Ridge	273.1253	141722.1	350.0678
LLAR	Lasso Least Angle Regression	281.0763	156827.9	383.1525
RF	Random Forest Regressor	281.7306	145305.3	368.9452
ET	Extra Trees Regressor	283.7394	153340	379.9375
KNN	K-Nearest Neighbors Regressor	284.002	157157	378.8889
LightGBM	Light Gradient Boosting Machine	292.1212	154101.3	377.5567
GBR	Gradient Boosting Regressor	293.0963	159051.9	387.6862
OMP	Orthogonal Matching Pursuit	322.6399	179952.3	408.5865
Huber	Huber Regressor	354.6148	229809.6	457.0904
DT	Decision Tree Regressor	357.4295	244305.2	475.409
LR	Linear Regression	397.3329	339780.8	523.0583
EN	Elastic Net	625.0217	1496322	979.4278
LASSO	Lasso Regression	679.9512	1543719	1067.025
PAR	Passive Aggressive Regressor	3238.936	37775493	4935.378

Each ML algorithm needed hyperparameter optimization to enable the best results. The ridge regression model was tuned to the hyperparameters shown in Table 3.5 using the PyCaret Package.

Table 3.5: Hyperparameter Tuning on the Training Dataset

Parameter	**Value**
Regularization strength	alpha=8.95
Number of iterations	10
Normalize	True
Metric optimized	R2 score

Table 3.5 continued from previous page

Parameter	Value
Fit intercept	True
Copy_X	True
Tol precision	0.001
Solver	auto

3.3 Experiments

Sections 3.3.1 and 3.3.2 explain the experimental set-up and evaluation parameters.

3.3.1 Experimental Set-up

This work aims to predict the survival days of GBM patients using the proposed methodology based on the combination of DL, radiomics, and ML. Ranger [172] optimizer trains the deep–supervised 3D UNet model to segment tumor subregions with a learning rate of $1e-4$. For the first five epochs, the linear warm-up was also employed. With a batch size of 1, the model was trained on 369 patients of the BraTS 2020 training dataset for 200 epochs using SWA for every five epochs. Further, the trained 3D deep-supervised UNet model generated segmentation maps for the BraTS 2020 validation set of 125 patients. The segmented and raw FLAIR MRI scans were used for radiomics-based feature extraction for SP. The segmentation architecture was implemented in PyTorch. The radiomic features were extracted using a Python-based PyRadiomics Package and further combined with clinical features of selected 119 patients (patients with GTR surgery status only) to make a complete set of 1112 features. The final features were selected using ML-based algorithms written in the Python-based scikit-learn library. Lastly, the selected set of 186 features was used to train regression-based ML models using the PyCaret Python Package. The implementation was completed using a 16 GB NVIDIA Tesla P100-PCIE graphics card available with the Google Colaboratory Pro facility. Training is done without weight decay; however, we reduce the learning rate using a cosine decay schedule as given in *Eq.* 3.10:

$$\eta_t = \eta_{\min} + 0.5\left(\eta_{\max} - \eta_{\min}\right)\left(1 + \cos\left(\pi \frac{T_{cur}}{T}\right)\right) \tag{3.10}$$

where, ' η_{max}' represents our starting learning rate (set to $1e-4$), ' η_{min}' represents our final learning rate (set to $1e-7$), ' T_{cur}' represents the current iteration counter, and ' T' is the total number of iterations to train for (set to 200 epochs). The length of one epoch of training is

30 minutes due to the massive input and intermediate tensors, as well as the switching time between GPU and CPU.

3.3.2 Evaluation Parameters

Figure 3.5 depicts the vocabulary for understanding the formulae for performance metrics.

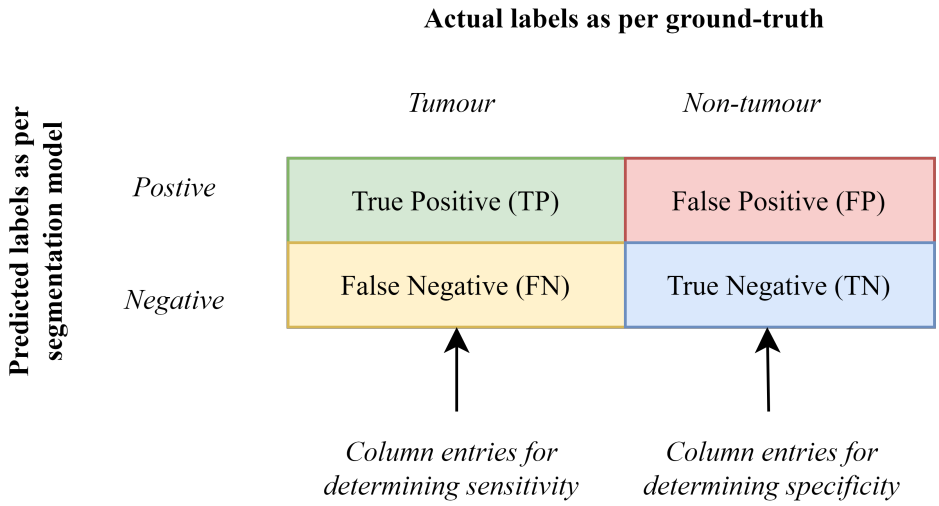


Figure 3.5: Terminology Used for Determining Sensitivity and Specificity

For an online validation of the segmentation results, the metrics evaluated include Dice Similarity Coefficient (DSC), Sensitivity, Specificity, and 95% Hausdorff Distance (HD).

(i) DSC: The DSC determines the overlapping area between the automated model and MRI ground-truth segmentation maps, as stated in *Eq.* 3.11.

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{3.11}$$

where: TP = True Positive,
FP = False Positive,
FN = False Negative.

(ii) Sensitivity: A model's sensitivity is its ability to reliably identify an MRI voxel as a tumor, as given in *Eq.* 3.12. The other terms used for sensitivity include True Positive Rate (TPR) or recall.

$$\text{Sensitivity or TPR or Recall} = \frac{TP}{TP + FN} \tag{3.12}$$

where: TP = True Positive,
FN = False Negative.

(iii) Specificity: A model's specificity relates to its capacity to correctly identify an MRI voxel as tumor-free, as shown in *Eq.* 3.13. True Negative Rate (TNR) is another phrase for specificity.

$$\text{Specificity or } TNR = \frac{TN}{TN + FP} \tag{3.13}$$

where: TN = True Negative,
FP = False Positive.

(iv) HD95: The average HD95 between two X and Y finite sets is shown in *Eq.* 3.14.

$$d_{AHD}(X,Y) = \left(\frac{1}{X} \sum_{x \in X} \min_{y \in Y} d(x,y) + \frac{1}{Y} \sum_{y \in Y} \min_{x \in X} d(x,y) \right) / 2 \tag{3.14}$$

where: $\frac{1}{X} \sum_{x \in X} \min_{y \in Y} d(x,y)$ = average directed HD from X to Y,
$\frac{1}{Y} \sum_{y \in Y} \min_{x \in X} d(x,y)$ = average directed HD from Y to X.

The average directed HD from point set X to Y is calculated by dividing the total of the minimum distances from all points in X point set to Y by the number of points in X.

For an online validation of the SP results, the metrics evaluated include accuracy, MSE, medianSE, stdSE, and Spearman. For classification, OS time was quantified into three survival categories: short survival (< 10 $months$), intermediate survival (10–15 $months$) and long survival (> 15 $months$). The three-class accuracy metric was evaluated. Evaluation metrics for regression include the MSE, mSE and Spearman correlation coefficient (ρ).

(v) Accuracy: The percentage of correct predictions made by the proposed model is its accuracy, as reflected in *Eq.* 3.15.

$$\text{Accuracy } = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.15}$$

where: TP = True Positive,
TP = True Negative,
FP = False Positive,
FN = False Negative.

(vi) MSE: MSE is the mean squared deviation between estimated and actual values. Mean Squared Error, or MSE is another name for Mean Squared Deviation (MSD). It assesses the performance of the regression model as shown in *Eq.* 3.16.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 \tag{3.16}$$

where: n = number of data points,
Y_i = observed values,
$\hat{Y}_i$ = predicted values.

(vii) Spearman Correlation Coefficient (ρ): It is a rank correlation measure that is non-parametric. It evaluates a monotonic function's ability to describe two variables. It is used to assess the performance of the regression model, as shown in *Eq.* 3.17.

$$\rho = 1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)} \tag{3.17}$$

where: n = number of data points,
d_i = difference between ranks.

3.4 Results and Discussion

A discipline of artificial intelligence known as ML is one in which computers do not require to be explicitly programmed, but can assess and execute tasks based on various data types without programming. Radiomics, on the other hand, is a method that takes high-volume quantitative details from radiographic images and uses these data to make automated prediction models that predict how a patient will do based on the image attributes. This research aims to establish a relationship between the biological factors underpinning pre-operative MRI and the OS of GBM patients through the use of radiomics.

The chapter discusses the technique for automatically predicting survival in glioma patients using mpMRI images and clinical data. Radiomic features were extracted from MRI scans after segmenting the tumor region using a DCNN. These radiomic characteristics were combined with clinical information to predict OS. The 3D UNet with deep supervision and SWA was used to achieve high-performance tumor segmentation. After training the model on

369 patients, predictions were made on the validation set of 125 glioma patients. Figure 3.6 highlights segmentation results as compared to the ground truth.

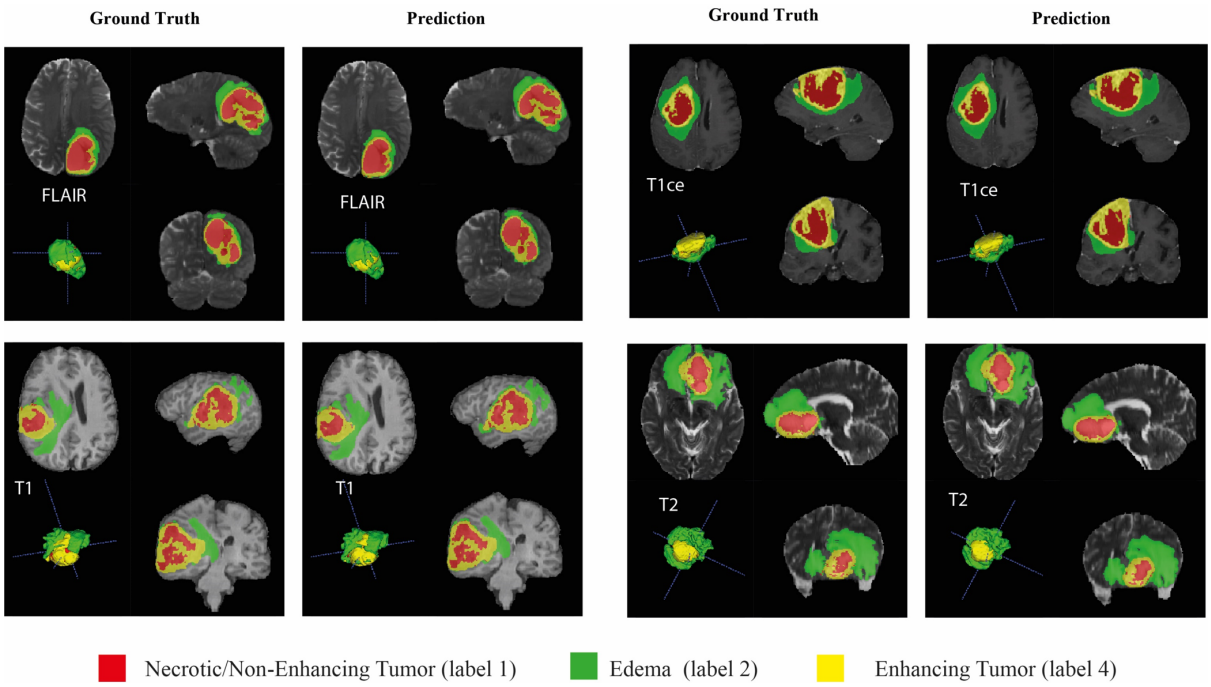


Figure 3.6: Illustration of Segmentations Generated in Comparison to the Ground-truth for Four Different Patients using Different MRI Sequences (FLAIR, T1CE, T1, and T2)

The generated segmentation labels were evaluated using the IPP. The three tumor subregions, *i.e.*, ET, WT, and TC were defined to validate the segmentation model's performance. All the statistical metrics were calculated using the IPP organized by the University of Pennsylvania. The DSC, HD95, Sensitivity, and Specificity of ET, WT, and TC were determined to assess segmentation performance. According to the evaluation results, as reflected in Table 3.7, the segmentation model performed better with respect to WT and TC than the ET region in DSC and HD95. It has been found that the DSC for segmentations was excellent; however, we discovered that the model's specificity was much greater than its sensitivity which suggested that the predictions were under-segmented.

The MRI segmentation results were used to train the SP model on characteristics collected from brain tumor subregions of 119 patients for whom clinical data such as their age and GTR status was available. The model was then utilized to forecast patient's OS on the validation set of 29 GBM patients. The performance of OS prediction was examined through two different evaluation methods: classification and regression on IPP. The OS predicted in days was submitted on IPP, which automatically divided the patients into three groups: short-survivors (those who survive < 10 months), mid-survivors (those who survive between 10 and 15 months), and long-survivors (those who survive > 15 months). Initially, this catego-

rization of patients was examined to see the model's accuracy. Additional pair-wise error analysis of the regression model was performed between the anticipated and actual OS (in days) using measures such as MSE, Spearman, medianSE, and stdSE.

The evaluation results presented in Table 3.8 indicate that the ridge regressor trained on the provided feature set achieved the best accuracy of 69% with a 0.635 Spearman score. Figure 3.7 shows the plot of important features for SP as per the ridge regressor.

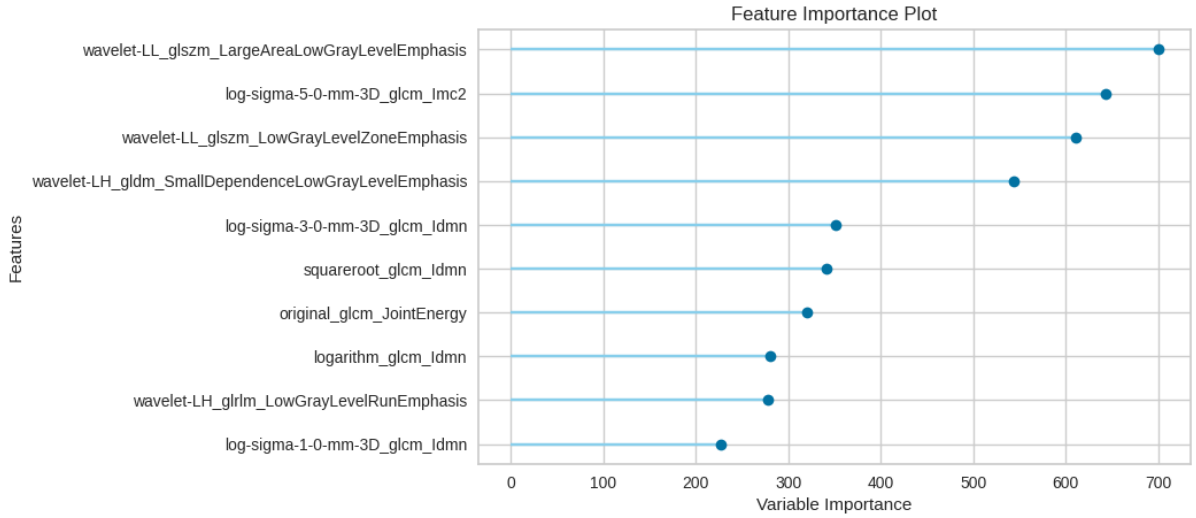


Figure 3.7: Feature Importance Plot of Ridge Regressor Model

Table 3.6 lists the radiomic features directly and inversely related to the prediction of OS time of GBM patients.

Table 3.6: List of Positively and Negatively Related Features with OS time of GBM Patients

Twenty Most Positively Related Features	**Twenty Most Negatively Related Features**
wavelet-LL_glszm_LargeAreaLowGrayLevelEmphasis	original_glcm_JointEnergy
log-sigma-5-0-mm-3D_glcm_Imc2	wavelet-HL_glrlm_LongRunLowGrayLevelEmphasis
wavelet-LL_glszm_LowGrayLevelZoneEmphasis	square_glrlm_GrayLevelNonUniformityNormalized
wavelet-LH_gldm_SmallDependenceLowGrayLevelEmphasis	log-sigma-3-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
log-sigma-3-0-mm-3D_glcm_Idmn	logarithm_gldm_SmallDependenceLowGrayLevelEmphasis
squareroot_glcm_Idmn	log-sigma-5-0-mm-3D_glcm_Idn
logarithm_glcm_Idmn	log-sigma-3-0-mm-3D_glcm_MaximumProbability
wavelet-LH_glrlm_LowGrayLevelRunEmphasis	exponential_glcm_Idmn
log-sigma-1-0-mm-3D_glcm_Idmn	wavelet-LH_glcm_MaximumProbability
log-sigma-3-0-mm-3D_glrlm_LongRunLowGrayLevelEmphasis	log-sigma-5-0-mm-3D_glszm_GrayLevelNonUniformityNormalized
exponential_gldm_SmallDependenceLowGrayLevelEmphasis	square_glrlm_LongRunLowGrayLevelEmphasis
log-sigma-3-0-mm-3D_glrlm_ShortRunLowGrayLevelEmphasis	square_glrlm_RunVariance
wavelet-HH_gldm_SmallDependenceLowGrayLevelEmphasis	log-sigma-3-0-mm-3D_glcm_Idn
original_glcm_Idn	exponential_glcm_Imc1
log-sigma-1-0-mm-3D_glcm_Idn	wavelet-HH_glrlm_LongRunLowGrayLevelEmphasis
log-sigma-3-0-mm-3D_gldm_LargeDependenceLowGrayLevelEmphasis	wavelet-HL_glcm_Idmn
log-sigma-3-0-mm-3D_glcm_Imc2	original_shape_Flatness
wavelet-HH_glcm_Idn	logarithm_glcm_Idn
original_glrlm_LongRunLowGrayLevelEmphasis	log-sigma-1-0-mm-3D_glrlm_LongRunLowGrayLevelEmphasis
exponential_glcm_Imc2	exponential_gldm_DependenceNonUniformityNormalized

A descriptive analysis of continuous variable age as mean is given in Table 3.1. No categorical variables for analysis have been taken up. The Shapiro 1D rank-based statistical analysis (shown in Appendix from Figure 3.9 to Figure 3.16) performed using the Yellowbrick Python Package plotted via matplotlib library shows that none of the selected features is irrelevant to OS.

3.4.1 Comparison with State-of-the-Art Techniques

For the purpose of this work, a CNN-based technique was used to perform automatic segmentation of glioma MRI scans to account for the tumor's heterogeneous imaging findings. A different validation set was used to properly verify the segmentation model. The performance of the proposed segmentation model has been compared with that of other existing techniques, as shown in Table 3.7.

Table 3.7: Performance of Segmentation Model on the Validation Dataset (N=125 Cases)

Reference	**Dice score**			**Hausdorff95 distance**		
	WT	**TC**	**ET**	**WT**	**TC**	**ET**
González *et al.* [71]	0.90	0.81	0.77	6.16	7.55	21.80
Agravat and Raval [75]	0.87	0.75	0.76	7.03	10.87	27.70
Miron *et al.* [76]	0.90	0.84	0.78	-	-	-
Anand *et al.* [74]	0.88	0.74	0.71	6.88	32.00	38.31
Parmar and Parikh [72]	0.90	0.84	0.74	5.08	8.69	36.04
Soltaninejad *et al.* [82]	0.87	0.80	0.66	6.91	7.80	47.33
Pang *et al.* [173]	0.88	0.76	0.75	18.09	29.05	34.23
Ali *et al.* [174]	0.87	0.74	0.74	9.42	10.09	3.92
Russo *et al.* [175]	0.89	0.82	0.76	6.07	8.12	27.37
Carmo *et al.* [73]	0.89	0.82	0.77	-	-	-
Han [87]	0.77	0.66	0.53	8.71	11.67	7.18
Pei *et al.* [81]	0.89	0.81	0.76	5.28	7.74	33.26
Zhao *et al.* [80]	0.90	0.85	0.79	5.19	6.18	23.56
Patel *et al.* [77]	0.90	0.85	0.79	4.88	5.85	15.70
Dai *et al.* [78]	0.90	0.84	0.79	4.19	5.93	26.55
Proposed	**0.89**	**0.84**	**0.76**	**6.28**	**6.73**	**26.28**

The results showed that the 3D UNet trained with deep supervision and SWA enhanced the segmentation model's generalizability. As CNN-based tumor segmentation is fast and repeatable, it has greater practical applications than which require manual tumor segmentation. Automating time-consuming, user-dependent operations can reduce user bias. Further, the data was obtained using different MRI scanners with varied acquisition settings. Despite this variability, the CNN method successfully identified and segmented the glioblastoma subregions; we hypothesized that the image pre-processing steps (*i.e.*, min-max scal-

ing) reduced the heterogeneities caused by inter-scanner variance and acquisition dimensional space, resulting in a more consistent outcome. External validation demonstrated that the segmentation model's prediction abilities were robust. Several other ML models were also trained to predict GBM patient survival. As shown in Table 3.8, the proposed methodology achieved the highest accuracy but second-highest MSE.

Table 3.8: Performance of Survival Model on the Validation Dataset (N=29 GBM Cases)

Reference	**Accuracy**	**MSE**	**medianSE**	**stdSE**	**(ρ)**
González *et al.* [71]	0.552	87581.00	51529.00	141983.00	0.442
Agravat and Raval [75]	0.517	116083.47	43974.09	168176.16	0.217
Miron *et al.* [76]	0.414	87744.14	37636.00	118896.90	0.321
Anand *et al.* [74]	0.448	110677.44	22874.18	142423.69	0.169
Parmar and Parikh [72]	0.414	101278.00	49248.00	132237.66	0.228
Soltaninejad *et al.* [82]	0.450	109564.00	39601.00	192755.00	0.280
Pang *et al.* [173]	0.517	151013.39	91990.89	185785.14	0.155
Ali *et al.* [174]	0.483	105079.40	37004.93	146376.00	0.134
Russo *et al.* [175]	0.586	99776.58	24539.13	168419.90	0.338
Carmo *et al.* [73]	0.552	106659.89	-	-	-
Han [87]	0.552	147898.50	52900.00	162551.70	0.333
Pei *et al.* [81]	0.552	101697.00	56169.00	116680.00	0.329
Zhao *et al.* [80]	0.655	71529.95	25869.75	125672.00	0.584
Patel *et al.* [77]	0.655	152467.00	39601.00	196569.00	0.479
Dai *et al.* [78]	0.450	87008.00	32400.00	-	0.430
Proposed	**0.690**	**83666.22**	**16620.98**	**161480.20**	**0.635**

After examining all the statistical variables obtained through IPP, the ridge regressor model showed the greatest performance on previously unknown data. The proposed improved technique excelled all other techniques in terms of accuracy and Spearman score, followed by MSE [71, 77, 80, 81, 87, 175]. The study attains utmost significance for being the first to register 69% classification accuracy for three survivor groups using the benchmark BraTS dataset to assess the various research workflows. A comparison of the SP model with other existing techniques has been presented in Table 3.8. The proposed technique performed even better than the cox proportional model [77] state-of-the-art in the field of SP.

The investigation undertaken in this work demonstrated that several feature selection strategies increased the Spearman correlation and classification accuracy despite the availability of minimal data for model training. The proposed SP technique, when compared with the existing methodologies, revealed that ML-based models built on selected radiomic features using multiple ML-based feature selection techniques can even outperform fully trained DL-based models for the survival prognosis of GBM patients [167, 176]. Also, we extracted filtered radiomic images, which improved the overall performance of the proposed method-

ology in comparison to existing techniques in which there is a lack of filter-based features such as wavelet decomposition, LoG, square, squareroot, logarithm, exponential, *etc.* However, further research can help to assess the robustness and repeatability of the feature selection method. The proposed technique has the significant benefit of being transferable to new data collected from various scanners and institutions.

In the current study, the accuracy is considered a better metric than MSE because the BraTS dataset includes data of some patients, which work as outliers during model training. Thus, MSE penalizes the wrong predictions more by squaring the error gap between the actual and predicted OS values, and the estimated MSE is larger. As a result, it is less resistant to outliers as compared to accuracy, which measures the percentage of correct predictions only.

3.4.2 Strengths and Limitations of the Proposed Methodology

The current study discovered that mpMRI-based radiomics enhanced the predictive power of the radiomic and clinical prognostic factors for GBM patients. As a result, this study lends sufficient support for using radiomics in risk stratification, personalized treatment, and care of GBM patients. For the purposed of quantitative analysis, four standard MRI sequences were used with multiple subregions, allowing for a comprehensive 3D description of gliomas [177]. Also, radiomic attributes were optimized for consistency and reproducibility through rigorous feature selection in a separate training set, by applying standardization of mpMRI scans obtained from different manufacturers, and validating prognostic models in a separate validation set.

As suggested by Choi *et al.* [178], radiomic analysis based on a mix of standard sequences would not prefer one sequence over another; therefore, we selected the FLAIR MRI sequence to extract radiomic features combined with segmented tumor regions. Furthermore, the chosen features were mostly texture-based; and very few were first-order or shape-based. Thus, it can be anticipated that assessing more complex imaging aspects through the use of radiomics may be more effective in predicting prognosis than the traditional shape-based image findings.

By incorporating many variables and analyzing complex non-linear relationships, ML models enable more precise and subject-specific predictions [179] instead of fitting the subjects' features into predetermined models with selected and weighted variables according to statistical significance [180, 181]. ML models may also grow and update themselves automatically as new data is added, resulting in a dynamic model that can improve its accuracy over time. The basic difference between ML models and conventional statistical techniques is that ML models may learn independently from data rather than being coded to do a job.

Overfitting is another problem that arises while developing models in ML. Overfitting is common whenever an overly complex and accurate model is trained on irrelevant data or unpredictability in the training dataset. As a result, estimates made through novel datasets will be imprecise. However, this issue was avoided by using CV and feature elimination. Besides this, the train-test split may indicate overfitting, *i.e.*, when a model performs better on the training sample than the test data, overfitting is sure to occur. In this regard, no variations in train-test splits were observed in the proposed ridge regressor model.

The present research work is subject to certain limitations. Firstly, its retrospective approach may have introduced selection bias. Pre-selection of patients is both a strength and a weakness. All patients had surgical excision with a minimal residual tumor volume and concurrent treatment. As a result, the findings cannot be generalized to all individuals with glioblastoma. However, an attempt was made to mitigate this risk by strictly following the criteria for inclusion. Moreover, the prognostic models were verified using only BraTS data; external validation at a different location would have increased the model's generalizability. Secondly, only two clinical variables were available in this data set (patient age and resection status); other predictive molecular markers [182] and clinical factors like gender [183], mutational status [184], functionals status [185] such as Karnofsky's performance status may have had a greater impact on the prognostic model [186]. Thirdly, the 69% accuracy attained for the three-class classification still has the potential for improvement. As the regression-based survival model was trained on a limited sample size of 119 patients only. Imaging characteristics and limited clinical data can explain only a portion of patient survival.

3.5 Inference and Findings

This work examined the effectiveness of using radiomic characteristics and an ML-based regressor to estimate the OS of patients who had been diagnosed with GBM. Using pre-operative and 3D deep-supervised UNet model-based segmented MRI scans, multi-modal radiomic characteristics were extracted from the FLAIR channel. Ridge regressor was selected based on minimum MSE to make the OS prediction for the GBM patients. The classification accuracy demonstrated that the Ridge regressor trained on specific characteristics outperformed the conventional approaches. Multiple feature selection strategies helped in identifying more relevant characteristics based on their relationship with the goal survival variable and improved the capacity of ML algorithms to classify GBM patients into three distinct survival groups, *viz.*, short, mid, and long. On the BraTS 2020 validation set, the suggested technique recorded the highest accuracy and Spearman correlation. Further, radiomic characteristics that contributed the most to the NN's predicting abilities were discovered.

Chapter 4

Deep Learning Features-based Survival Prognosis of Glioblastoma Multiforme Patients

This chapter proposes DL-based 3D U-shaped CNN inspired encoder-decoder architecture to segment the brain tumor. Further, feature extraction was performed on these segmented and raw MRI scans using a pre-trained 2D RNN. The dimension-reduced principal components were integrated with clinical data and the handcrafted features of tumor subregions to compare the performance of regression-based automated ML techniques. The proposed methodology was able to achieve the MSE of 87067.328, medianSE of 30915.66, and a Spearman rank correlation of 0.326 for SP with the validation set of BraTS 2020 dataset. These results made the MSE far better than the existing automated techniques for the same patients. Automated SP of GBM patients is a crucial topic with its relevance in clinical use. The results proved that DL-based feature extraction using 2D pre-trained networks is better than many heavily trained 3D and 2D prediction models from scratch. The ensembled approach has produced better results than single models. The most crucial feature affecting GBM patients' survival is the patient's age, as per the feature importance plots presented in this work. The most critical MRI modality for SP of GBM patients is the T2-FLAIR, as evident from the feature importance plots.

4.1 Contribution of the Proposed Technique

The development of automated SP procedures for patients with HGG tumors has been a latest area of research focus. Using the Multimodal BraTS 2020 validation set, the efficacy of proposed DL, and ML-based SP approach was tested. Using a pre-trained 2D model and an automated ML-based SP approach, this work has successfully brought an improvement in SP job performance. SP is, broadly, a two-step procedure in which the BTS output is used as an input to the survival prognosis step with feature extraction/selection and model selection as the in-between steps.

The proposed technique has contributed significantly as here under:

– The tumor and its subregions were segmented using 3D UNet-based encoder-decoder

architecture.

- The segmented and raw MRI scans were used for DL-based feature extraction using pre-trained 2D RNN.
- The feature space was reduced using PCA feature reduction algorithm.
- The reduced principal components were combined with handcrafted features to train automated ML regression-based models.
- The best-performing six regression models were ensembled together to predict patients' survival time in days.
- The results obtained through the proposed technique were compared with those of other state-of-the-art techniques to validate its performance.

The chapter has been organized into various sections. Section 4.1 describes the contribution, and organization of this chapter. Section 4.2 deals with the dataset and explains the proposed methodology. The experimental set-up and evaluation parameters have been explained in Section 4.3. The results and discussion are presented in Section 4.4. Section 4.5 appears with the findings of the proposed technique.

4.2 Materials and Methods

4.2.1 Dataset and Proposed Methodology

The BraTS 2020 dataset has been used for the implementation of proposed methodology; and a detailed explanation of the dataset is given in Section 3.2.1 of Chapter 3. As already mentioned in Table 3.1, there are 236 patient data items for survival statistics, including ID, age, OS (in days), and surgical status on 119 of the 236 data patients, GTR surgery was conducted, whereas the remaining 117 patients had either STR surgery or no surgery, and were listed as NA in the patient's data record. Out of 119 patients who underent GTR surgery, only a single patient was surviving; thus, 118 individuals were available for survival training. The researchers need to develop models to estimate the patients' OS based on the reduced amount of data. Only patients having the resection status 'GTR' are eligible for examination. Validation dataset of 125 patients for segmentation and 29 patients for survival prognosis may be used to evaluate the performance of proposed models.

For BraTS 2020 data, the SP task should be considered a regression problem instead of the classification problem. The reasons for the same are explained as below:

- The training data for the SP task included only 236 patients only, out of which only

119 underwent GTR surgery. Out of total 236, 89 were short-survivors (with <300 survival days), 60 mid-survivors (between 300-450 survival days), and 87 long-survivors (>450 survival days). It is evident from the data that the survivors do not represent this classification in equal numbers. Thus, more number of patients representing the short and long categories will deviate the survival prognosis towards these two classes instead of mid survivors.

– The existing literature reveals that most of the researchers have made predictions in survival number of days, but the online evaluation system also calculates accuracy as a classification metric by converting the predicted survival days to one of the survivor categories. For instance, if a patient has an actual survival of 440 days and the proposed model predicts it as 458 days, so as per classification, it will be considered under the long survivor category instead of the actual medium survivor category. However, if SP is considered a regression problem only, then the prediction of 458 days stands quite close to the actual value of 440 days. Thus, according to the current study, SP of GBM patients is a regression task.

Pre-processing, segmentation, feature extraction, feature selection, automated model selection and SP are the essential steps of the proposed methodology as shown in Figure 4.1. All these steps have been explained in detail in the following sections.

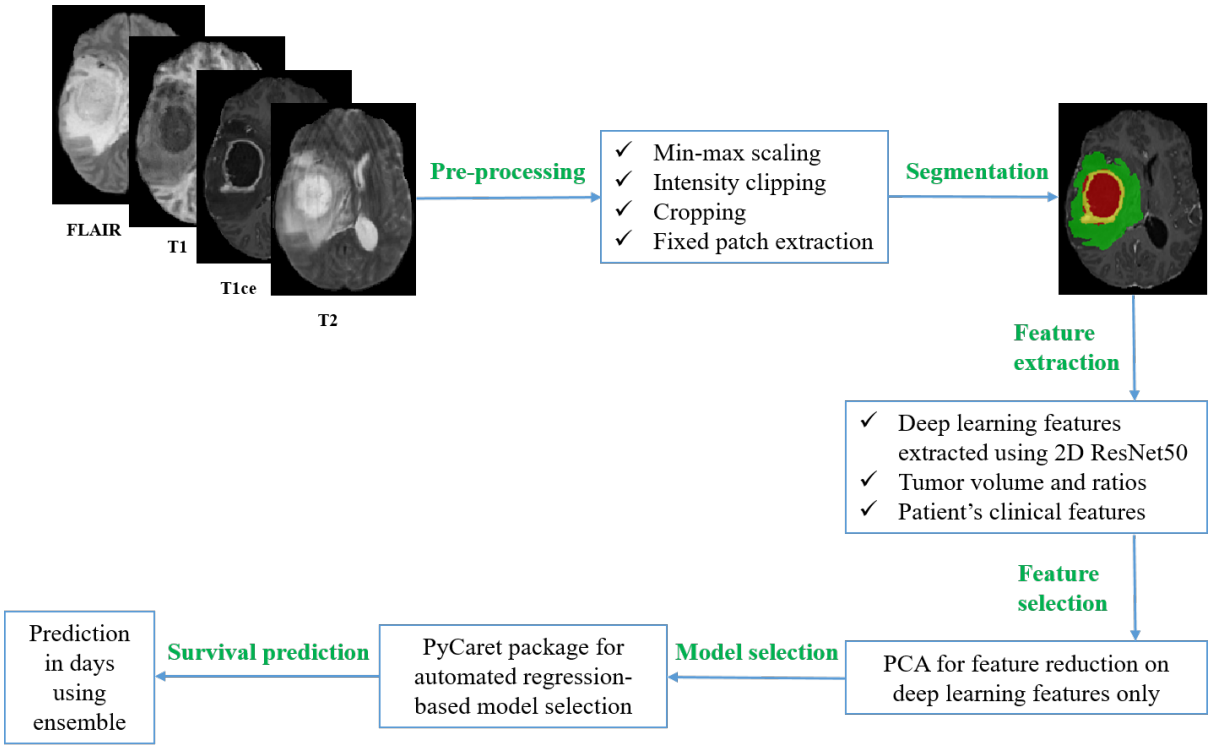


Figure 4.1: The Proposed Methodology for Early SP of Glioblastoma Patients

4.2.2 Pre-processing

As the data taken up for investigation has been collected from various institutions, and it is generated by different MRI acquisition machines, thus, there is a need for intensity standardization of the raw MRI volumes for better outcomes. Pre-processing has been performed in the same manner as mentioned in Chapter 3, Section 3.2.2.

4.2.3 Segmentation Architecture

The proposed methodology has performed semantic segmentation using U-shaped CNN based on 3D encoder-decoder architecture to segment tumor subregions, *i.e.*, WT, TC and ET. The 3D architecture captures global spatial relationships between the voxels from different portions of the human brain. Our BTS architecture is an end-to-end 3D modified version inspired from the original 2D UNet [62] explicitly proposed for performing medical image segmentation. The BTS architecture is shown in Figure 4.2. It includes a contracting path (encoder), midway (bottleneck), and an expansive path (decoder).

Encoder: The 3D UNet encoder takes four different 3D MRI modalities with $128 \times 128 \times 128 \times 4$ as input dimensions. The contracting path follows the pattern of a typical CNN. It consists of the repeated application of two 3D convolutions (with the kernel as '3', stride as '1' and padding as '1'), each followed by GN [119] and ReLU [168] non-linear activation function. The use of GN is preferred because of its smaller batch size. Further, a 3D max-pooling operation (with kernel '2' and stride '2') is used for downsampling. The number of feature channels are doubled at each stage in the encoder as 48, 96, 192 and 384.

Bottleneck: It includes the repeated application of two dilated 3D convolutions (with the kernel as '3', stride as '1', padding as '2' and dilation as '2'), each followed by GN and ReLU to generate 384 feature maps same as the input. Dilated convolutions increase network depth without further increasing the feature maps and consumption of memory resources. Lastly, one 3D convolution operation is applied (with the kernel as '3', stride as '1' and padding as '1') to convert 768 feature maps to 384 feature maps, where the first input of 384 feature maps is taken from the output of dilated convolutions and the second from encoder stage 4 output. These two 384 feature maps are concatenated before convolution operation, followed by GN and ReLU.

Decoder: Each step on the expansive path includes a spatial upsampling operation of the feature map with trilinear interpolation mode and a scale factor of 2. Each upsampling layer is followed by two convolutions (known as "up-convolution" with the kernel as '3', stride as '1' and padding as '1') which reduce the number of feature channels to half, concatenated with the parallel feature map obtained from the contracting path, each followed by GN and

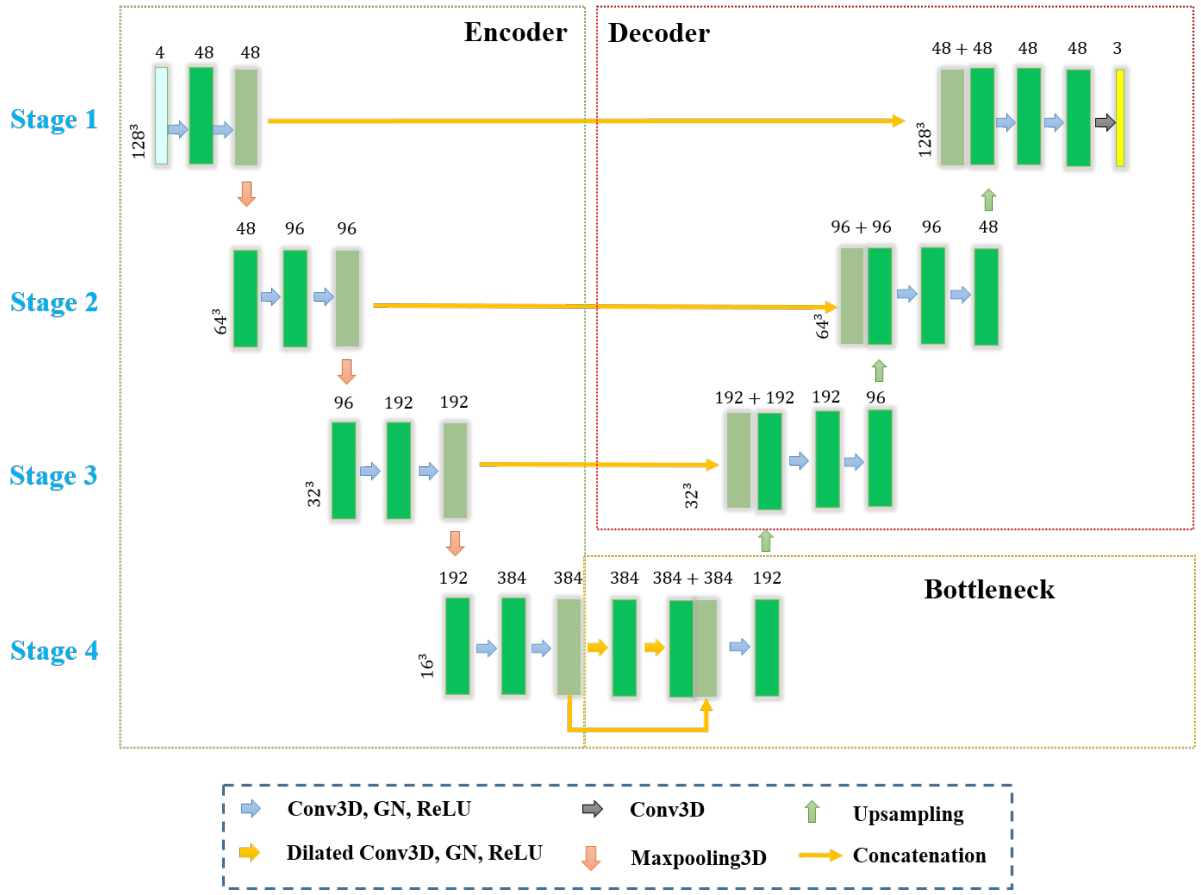


Figure 4.2: Illustration of 3D UNet Encoder-Decoder based Segmentation Architecture which takes four Modalities as Input and Generates a Probability-based Segmentation Map

ReLU operations. At the final layer, a 3D convolution (with the kernel as '1' and stride as '1') maps each 48-component feature vector to the desired number of class labels (labels 1, 2 and 4). The output layer denotes the probabilities for three different tumor subregions. Overall, the network consists of 18 convolutional layers. Loss function is applied in the same manner as mentioned in Chapter 3 Section 3.2.3.

4.2.4 Deep Learning and Handcrafted Feature Extraction

Several CNN models are available now with increased performance and deeper design. However, the task of training deeper networks is challenging since they demand an enormous quantity of data and millions of parameters. For more accurate and useful models, it is quite important to have a large number of precisely labeled datasets. However, there is no large-scale annotated dataset accessible for medical imaging problems. Also, training a deep CNN with small or unbalanced training samples is difficult since the dimensionality of the data is substantially more significant than the number of samples, which may lead to overfitting. However this issue can be addressed through the transfer learning approach, in which the

model is first pre-trained on huge datasets, such as ImageNet, and then fine-tuned for the situation at hand. For feature extraction, a ResNet50 network pre-trained on ImageNet was employed in this study. Efficient feature extraction is critical in developing relevant and detailed representations using raw input data to get accurate and robust outcomes. Through the suggested technique, we employed a pre-trained, well-established 2D ResNet50 [100] trained on the ImageNet dataset as a visual feature extractor to encode the input MRI scans into a feature vector of sparse descriptors with a low dimension. ResNet50's better performance in image classification enables it to extract high-quality features from ImageNet pictures. It was hypothesized that the ResNet50 feature extraction layers would also perform well on GBM patients' SP. The segmented and raw MRI scans are used for DL-based feature extraction for SP via Keras 2D ResNet50 model pre-trained on the ImageNet dataset. Figure 4.3 illustrates the architecture of the 2D ResNet50 pre-trained model.

The extracted features are then utilized as an input signal for PyCaret regression-based models. ResNet50 uses many stacked various size kernels and filters to enhance the network depth, allowing the network to extract more complex features at a lesser cost. The input layer in ResNet50 expects a $224 \times 224 \times 3$ image, however, the 3D MRI scans in the BraTS dataset are $240 \times 240 \times 155$. The 3D $240 \times 240 \times 155$ dimensions of MRI scans are cropped into $224 \times 224 \times 3$ non-zero array dimensions suitable as the input required for the 2D ResNet50 pre-trained model. The labeled segmentation map is combined with one of the input MRI modalities for cropping the ROI for feature extraction.

The 3D scans' height and breadth were cropped to 224×224, and the unnecessary black region surrounding the brain was removed. Further, the depth of 3D scan was decreased to three slices with the largest brain region, *i.e.*, non-zero 3D voxels detected using the numpy python library's argwhere function. The last fully connected dense layer of 2D ResNet50 provides 1000 features as an output for all the four MRI modalities with a total of 4×1000 features. The extracted features are numeric values that describe specifications of MRI scans. To complement the deep features extracted through a 2D pre-trained model, we extracted the handcrafted features such as the amount of necrotic/edema/enhancing, tumor to brain ratio, and extent of tumor from raw and segmented 3D MRI scans. Also, the patient's age was considered a crucial clinical factor in predicting the survival time of GBM patients.

4.2.5 Feature Selection

It is often beneficial to minimize their dimensionality when analyzing big multivariate datasets. PCA [187] is one of the significant approaches to select informative principal components obtained from DL features. It substitutes a smaller number, 'q' of derived variables, called principal components, linear combinations of the original variables. Often, with 'q' signif-

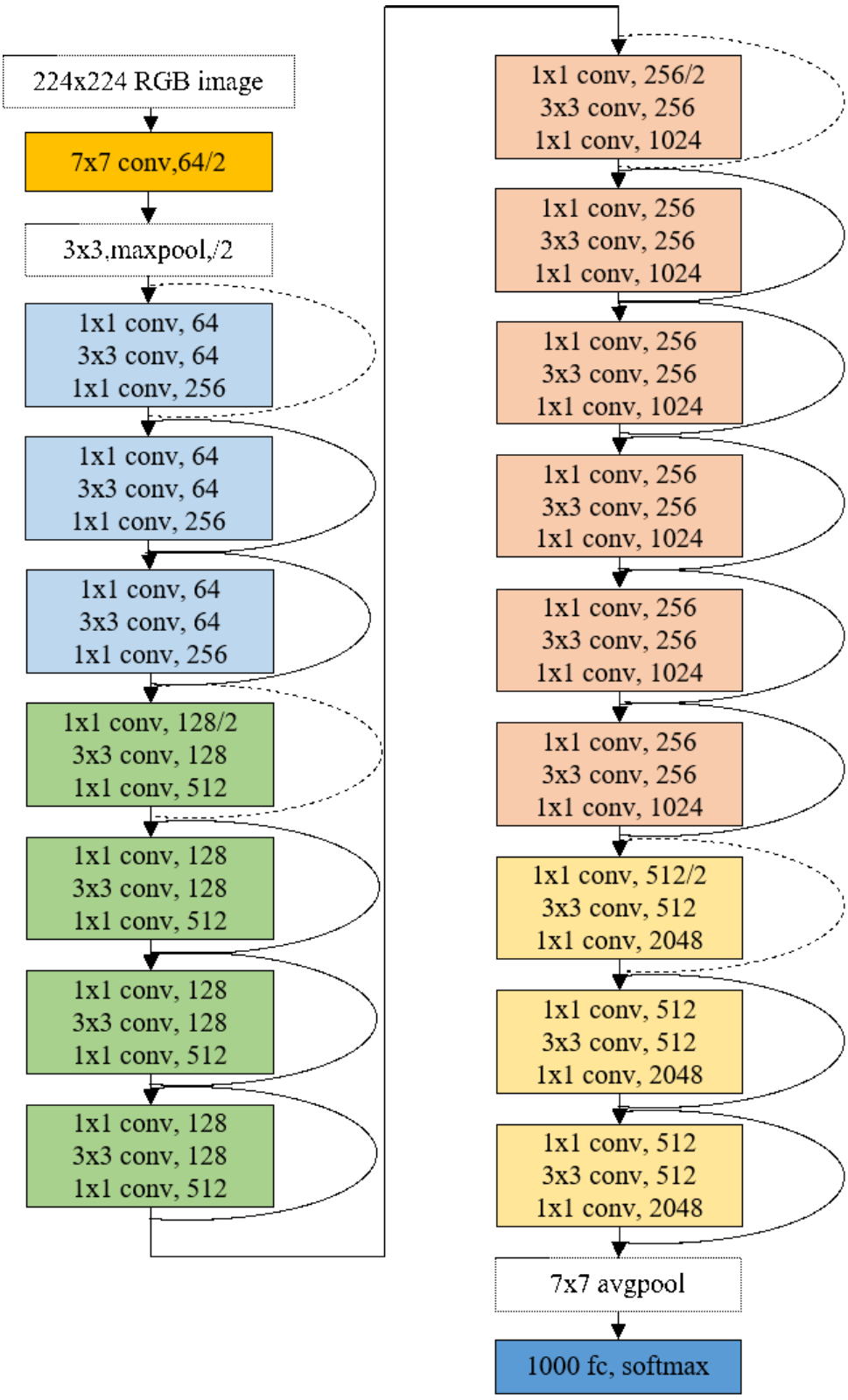


Figure 4.3: Illustration of Pre-Trained 2D Resnet50 Architecture used for Extraction of Deep Features from Raw and Segmented 3D MRI Scans

icantly less than 'p', it is feasible to keep the majority of the variability in the original variables. For the purpose of this work, PCA was used to reduce the dimensions of feature vectors extracted using a 2D ResNet50 feature extractor, as shown in Figure 4.4 (a). PCA stores information in a set of data with fewer dimensions. This is done by projecting the data set integrally into a subspace made by an orthogonal axis system. The reduced dimension computational content is chosen so that the most important data properties can be found with little loss of information. This kind of reduction is helpful in many areas, such as image compression, data representation, *etc.* So, PCA can be used in many different ways in the field of Medical Image Processing. Image feature extraction, fusion, composition, segmentation, registration, and denoising are possible [188].

The basic principle of PCA is to minimize the number of variables in a data collection while retaining as much data as feasible. Depending on the need, the researcher may decide to preserve every component or drop out the less important ones. The number of principal components selected for each modality, in this study, are shown in Figure 4.4 (a, b, c, d). Since the values in the remaining dimensions are often small, there will be little information lost, if they are removed. The amount of information lost due to the selection of the fixed initial number of important principal components from each MRI modality is also provided in Figure 4.4 (a, b, c, d). But the remaining principal components will not make much difference in the final analysis because as shown in Figure 4.4 (a, b, c, d), the cumulative variance achieved with the considered number of prinicpal components for each modality is more than 90%. To achieve this level of cumulative variance, maximum number of principal components are selected from FLAIR modality rather than from T1, T1CE and T2 modalities. Figure 4.4 (e) illustrates the pair-wise bivariate relationships between the different clinical and handcrafted features helpful in predicting the survival days for GBM patients.

4.2.6 Automated Model Selection

This study used an automated ML regression module in the PyCaret Package to predict the survival time of GBM patients. PyCaret is an open-source, low-code ML library written in Python that automates ML workflows. The PyCaret Package aims to automate the most time-consuming steps involved in evaluating and comparing ML algorithms for classification and regression. It is an end-to-end ML and model management solution that exponentially speeds up the trial cycle and increases productivity. The primary advantage of the library is that it enables a great deal to be accomplished with very few lines of code and minimal manual configuration. The Regression Module in PyCaret Package is a supervised ML module used to estimate the relationships between a target variable and more feature variables. For the SP task, the survival days is the 'target' variable and 35 'feature' variables are

as below:

– *DL features:* Fifteen principal components from the FLAIR modality, five principal components from the T1 modality, five principal components from the T2 modality, four principal components from the T1CE modality

– *Handcrafted features:* Amount of NCR, amount of ED, amount of ET, tumor to brain ratio, extent of tumor

– *Clinical feature:* Patient's age

The workflow of PyCaret for SP of GBM patients is given below:

- *Set-up:* This model function from the PyCaret regression module is used to set data and target variables.

- *Compare Models:* This model function is used to compare the performance of all the regression models available in the PyCaret package based on the 'MSE' metric.

- *Blend Models:* This model function is used to ensemble the results of the first six best-performing regression models sorted, based on the 'MSE' metric.

- *Tune Model:* This model function is used to perform automatic tuning of the hyper-parameters required for a blended regression model obtained by performing 10-fold CV.

- *Evaluate Model:* This model function is used to visualise different plots like residuals, prediction error, feature importance, and learning curve.

- *Finalise Model:* This model function is used to select the best performing model.

- *Predict and Save Model:* The predict model function is used to make predictions on unseen data, whereas the save model function is used to save the model for future usage.

The MAE, and MSE loss metrics were used to check the performance of all the regression models. A comparison of the first six best-performing regression models sorted on the basis of 'MAE' metric is shown in Table 4.1. The available regression models were cross-validated using ten-fold CV. Each ML algorithm needed hyperparameter optimization to enable the best results. The six regression models which attained minimum MAE were tuned to the best hyperparameters as shown in Table 4.1 using the PyCaret Package.

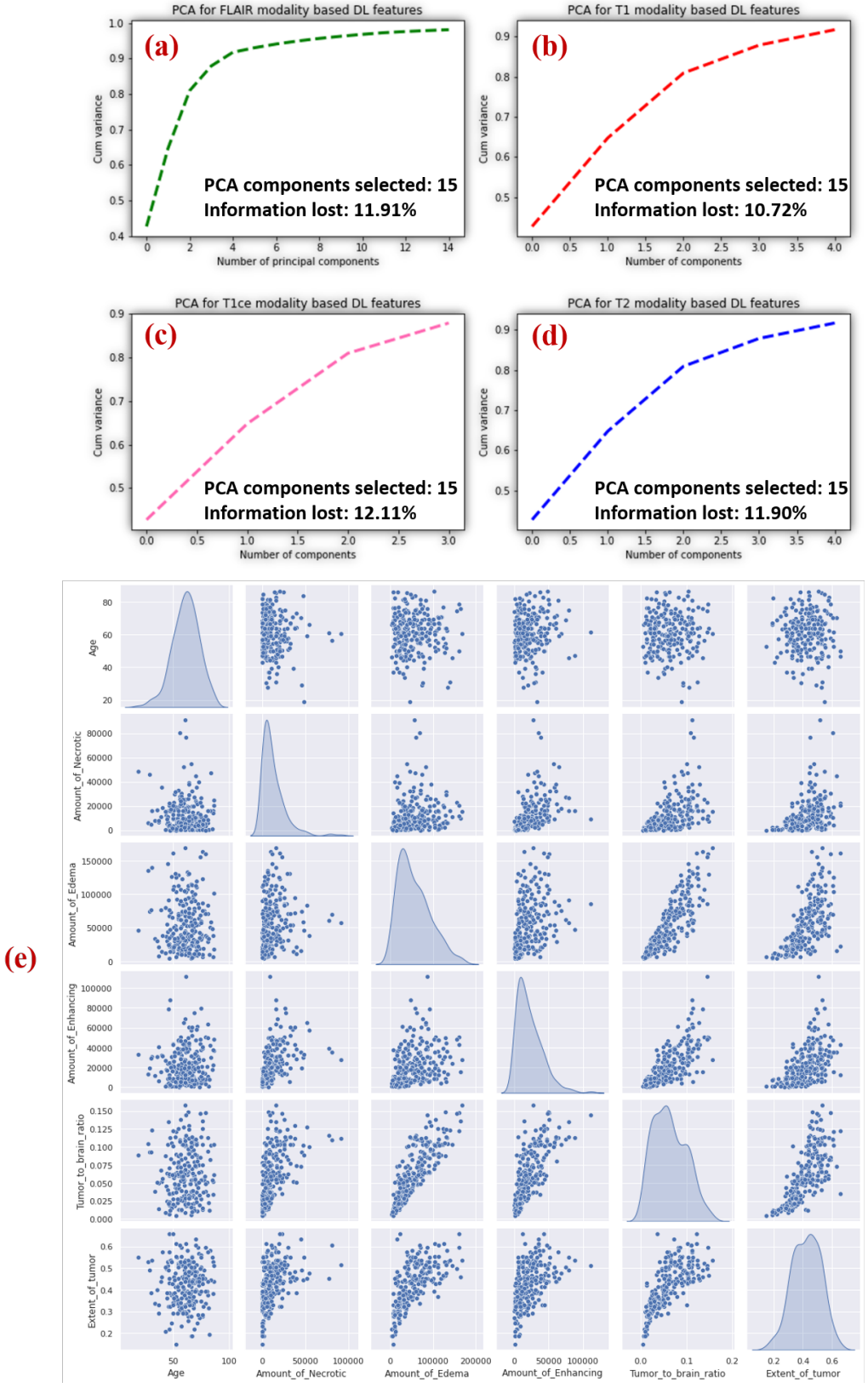


Figure 4.4: (a, b, c, d) Plots for the Cumulative Proportion of Variance for Principal Components Generated using Matplotlib And Sklearn Libraries, (e) Pairplot for Different Handcrafted and Clinical Features Generated Using Seaborn Library

Table 4.1: Summary of Hyperparameters, Weights, MAE and MSE Metrics for Six Best Performing Models on Training Data generated using Automated ML Pycaret Package

Model	**Description**	**Hyperparameters**	**Weights**	**MAE**	**MSE**
BR	Bayesian Ridge	alpha_1=$1e-06$, alpha_2=$1e-06$, alpha_init=None, compute_score=False, copy_X=True, fit_intercept=True, lambda_1=$1e-06$, lambda_2=$1e-06$, lambda_init=None, n_iter=300, normalize=False, tol=0.001, verbose=False	0.610	246.5433	110822.4619
EN	Elastic Net	alpha=1.0, copy_X=True, fit_intercept=True, l1_ratio=0.5, max_iter=1000, normalize=False, positive=False, precompute=False, random_state=399, selection='cyclic', tol=0.0001, warm_start=False	0.099	247.4187	111327.1768
OMP	Orthogonal Matching Pursuit	fit_intercept=True, n_nonzero_coefs=None, normalize=True, precompute='auto', tol=None	0.069	246.7298	111344.5015
LLAR	Lasso Least Angle Regression	alpha=1.0, copy_X=True, eps=2.2204e-16, fit_intercept=True, fit_path=True, jitter=None, max_iter=500, normalize=True	0.270	246.2395	112704.9955
LASSO	Lasso Regression	alpha=1.0, copy_X=True, fit_intercept=True, max_iter=1000, normalize=False, positive=False, precompute=False, random_state=915, selection='cyclic', tol=0.0001, warm_start=False	0.030	254.9230	118759.5081
Ridge	Ridge Regression	alpha=7.28, copy_X=True, fit_intercept=True, max_iter=None, normalize=True, random_state=7103, solver='auto', tol=0.001	0.110	257.4644	119951.9934

The top six performing regression models are: BR, EN, OMP, LLAR, LASSO, and Ridge Regression. The process of determining the characteristics in a dataset that contribute the most to forecasting the target variable is called feature importance. Focusing on a subset of characteristics rather than all of them reduces the risk of overfitting. Feature reduction helps to improve the results and the training time also decreases. Feature selection is accomplished in PyCaret using the feature selection parameter. It selects the subset of characteristics that are most important for modelling by combining different supervised feature selection techniques. The subset's size can be adjusted via the set-up function's feature selection threshold argument. The feature importance plots shown in Figure 4.5 indicate which features have the greatest impact on predicting the survival time of GBM patients. It is evident from the feature importance plots that a patient's 'age' plays a crucial role in affecting the OS time prediction of GBM patients. Out of four MRI modalities, 'FLAIR' is the most critical modality to extract deep features for SP. Out of five handcrafted features, 'Extent_of_tumor' is the most important feature, followed by 'Tumor_to_brain_ratio'.

Table 4.2 demonstrates the range of different feature vectors selected for training PyCaret's regression-based models. It is evident from the analysis of Figure 4.5 in combination with Table 4.2 that the features having a wide range are more important in comparison to the rest of the feature set. Thus, it can be considered that any change in the range of feature vectors will impact their importance for different regression models and the overall final prediction results.

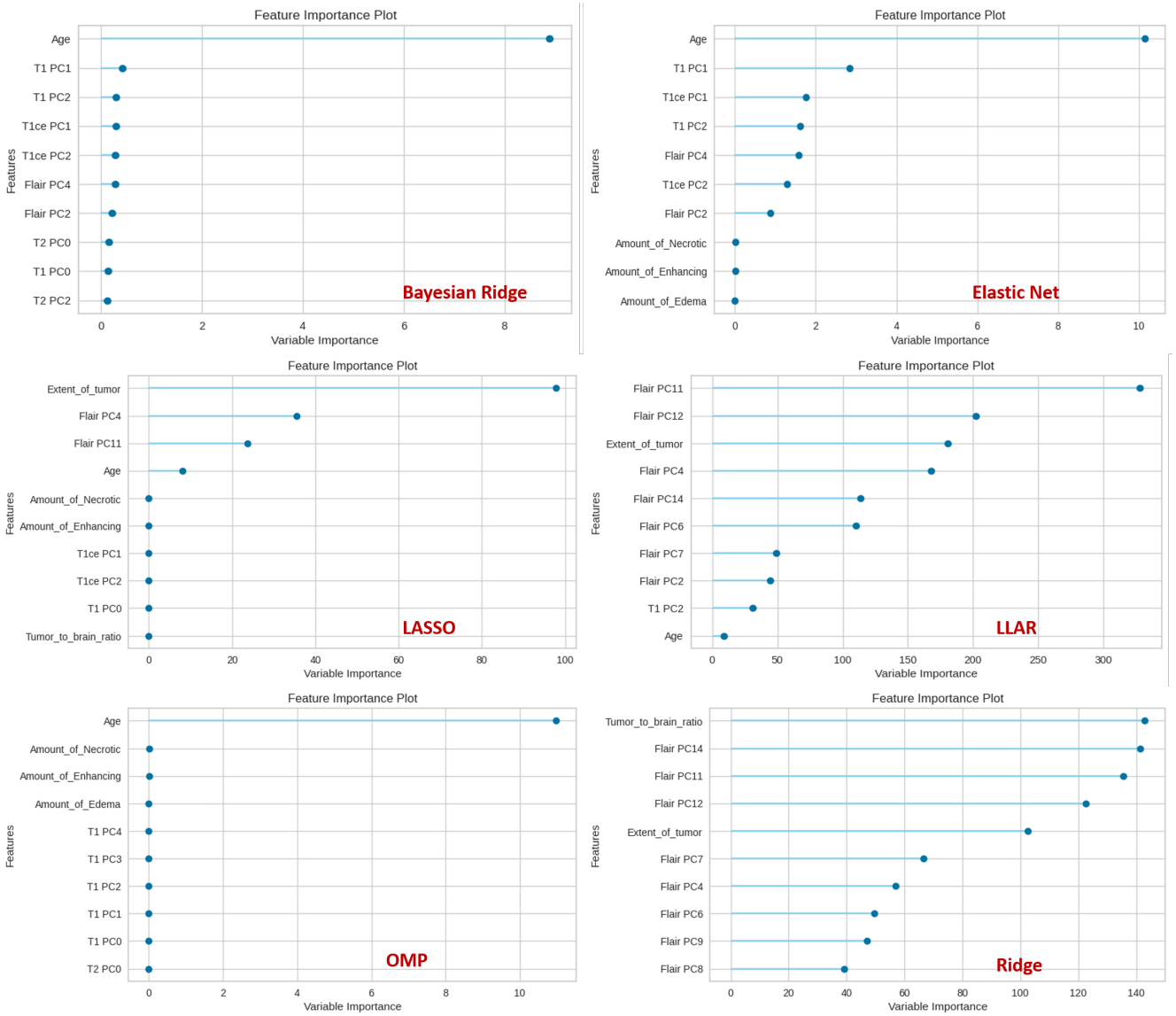


Figure 4.5: Feature Importance Plots for Top Six Regression Models

Table 4.2: Range of Feature Vectors Used for Training and Validating Different Regression-based Models

Feature Vector	**Training Feature Vectors Range**				**Validation Feature Vectors Range**			
	Min.	**Max.**	**Mean**	**Median**	**Min.**	**Max.**	**Mean**	**Median**
Age	18.9750	86.6520	61.2232	61.4710	21.7260	85.6770	57.2690	58.0000
Amount_of_Necrotic	0.0000	91299.0000	12329.1780	8542.5000	464.0000	33696.0000	7906.7930	4793.0000
Amount_of_Edema	3676.0000	168794.0000	56313.3814	47620.0000	8937.0000	132267.0000	65327.3800	63096.0000
Amount_of_Enhancing	106.0000	111250.0000	22501.4703	17783.5000	3575.0000	63705.0000	20090.9300	15603.0000
Amount_of_Tumor	0.0045	0.1579	0.0632	0.0596	0.0113	0.1314	0.0617	0.0598
Extent_of_tumor	0.1484	0.6581	0.4243	0.4323	0.2710	0.7355	0.4418	0.4387
Flair PC0	-0.5890	0.6855	-0.0233	-0.1399	-0.3918	0.8459	0.0899	0.0159
Flair PC1	-0.4333	0.6682	-0.0124	-0.0848	-0.2323	0.4236	0.0004	-0.1732
Flair PC2	-0.3573	0.7308	-0.0472	-0.1171	-0.5989	0.6741	-0.0014	-0.0232
Flair PC3	-0.3259	0.8083	-0.0234	-0.1026	-0.2058	0.0859	-0.0339	-0.0339
Flair PC4	-0.1410	0.8412	-0.0037	-0.0397	-0.1496	0.5468	-0.0126	-0.0435
Flair PC5	-0.2270	0.6174	-0.0034	-0.0078	-0.327	0.1654	-0.0161	-0.0176
Flair PC6	-0.3334	0.4472	-0.0009	-0.0072	-0.1734	0.7045	-0.0017	-0.0158
Flair PC7	-0.3150	0.7021	0.0022	-0.0017	-0.1446	0.6539	0.0138	-0.0005
Flair PC8	-0.2920	0.7397	-0.0011	-0.0003	-0.4553	0.0261	-0.0192	-0.0004
Flair PC9	-0.3439	0.5255	0.0061	0.0013	-0.2953	0.2442	-0.0081	0.0009
Flair PC10	-0.4608	0.5965	0.0006	-0.0007	-0.2415	0.4055	0.0062	0.0003
Flair PC11	-0.2121	0.4155	-0.0018	-0.0036	-0.1935	0.5289	0.0113	0.0006
Flair PC12	-0.2432	0.6216	0.0004	-0.0002	-0.05	0.5228	0.0263	-0.0019
Flair PC13	-0.4080	0.6865	0.0000	0.0000	-0.2195	0.2741	-0.0034	0.0001
Flair PC14	-0.2116	0.4340	-0.0005	-0.0012	-0.0693	0.234	0.0031	-0.0015

Table 4.2 continued from previous page

Feature Vector	Training Feature Vectors Range				Validation Feature Vectors Range			
	Min.	**Max.**	**Mean**	**Median**	**Min.**	**Max.**	**Mean**	**Median**
T1 PC0	-0.3812	0.8624	-0.0650	-0.1009	-0.3851	0.7864	-0.1135	-0.2493
T1 PC1	-0.3848	0.8661	0.0392	-0.0868	-0.4216	0.7968	0.0238	-0.0446
T1 PC2	-0.6095	0.7734	-0.0425	-0.0270	-0.4995	0.6719	0.0261	0.0035
T1 PC3	-0.3045	0.4527	0.0042	-0.0862	-0.2396	0.4551	0.0891	0.0659
T1 PC4	-0.2661	0.7119	0.0133	0.0022	-0.0724	0.2653	0.0078	-0.0055
T2 PC0	-0.5132	0.8374	-0.0453	-0.0499	-0.4884	0.8009	-0.0945	-0.305
T2 PC1	-0.5836	0.7045	0.0441	0.1673	-0.6277	0.6531	-0.0514	-0.0491
T2 PC2	-0.4746	0.7541	0.0351	-0.0173	-0.3104	0.4785	0.0406	-0.0665
T2 PC3	-0.5939	0.6388	0.0216	-0.0225	-0.0896	0.2232	-0.007	-0.0193
T2 PC4	-0.4784	0.6252	0.0047	0.0127	-0.0666	0.7039	0.0663	0.0042
T1CE PC0	-0.5143	0.8425	-0.1085	-0.1869	-0.4427	0.7824	-0.1248	-0.2138
T1CE PC1	-0.5104	0.7147	-0.0072	-0.1036	-0.4063	0.7059	0.1147	0.0884
T1CE PC2	-0.5480	0.8001	-0.0528	-0.0718	-0.3452	0.7714	-0.0773	-0.1304
T1CE PC3	-0.3806	0.5441	-0.0093	-0.0508	-0.1433	0.7463	0.0388	0.0174

4.2.7 Survival Prediction

In this research work, the blended model for SP has been used by considering the six best-performing regression models, *viz.*, BR, EN, OMP, LLAR, LASSO, and Ridge from the PyCaret automated ML package. SP is made using DL features extracted through 2D ResNet50 pre-trained model, handcrafted features calculated from segmented and raw MRI modalities, and clinical features like patient's age. The trained automated ML-based blended regression model is used to make predictions for the BraTS 2020 validation unseen patient's data as shown in Figure 4.6.

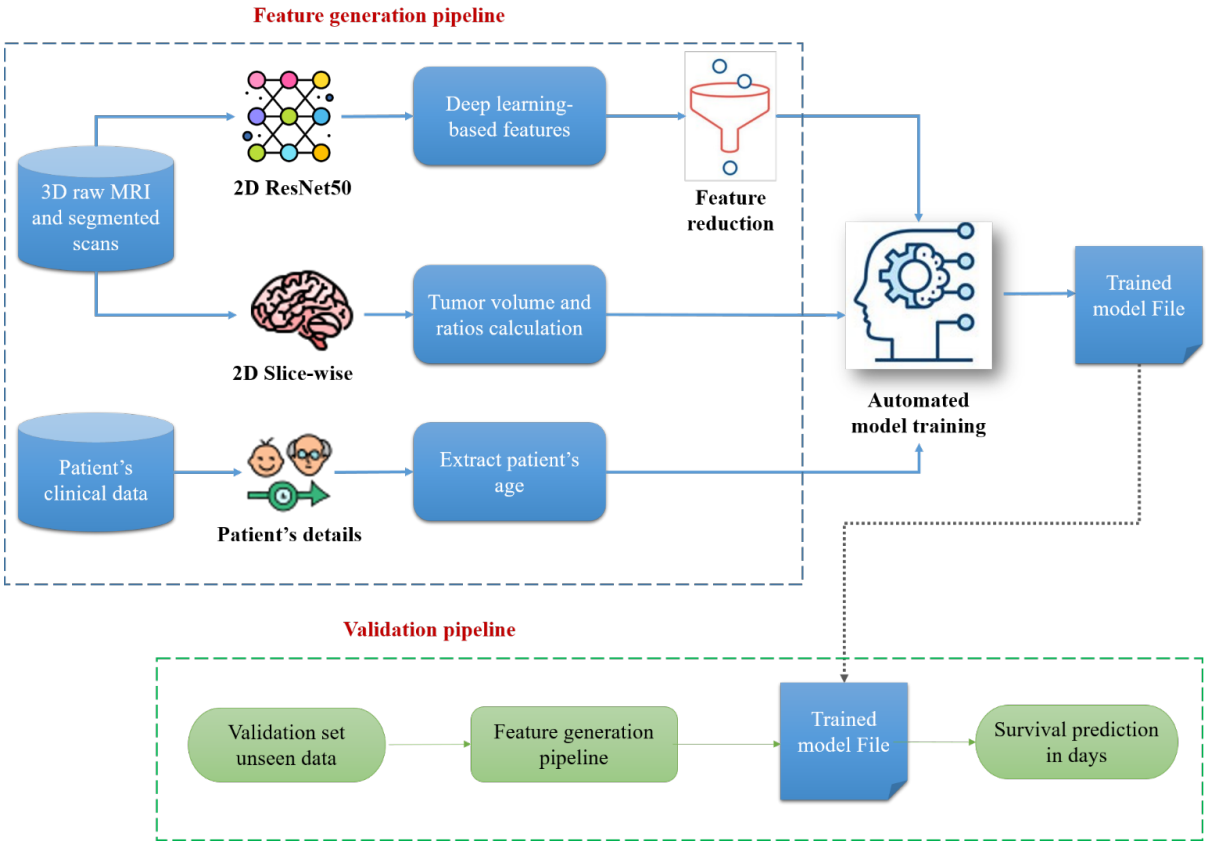


Figure 4.6: Overview of Complete Survival Prediction Pipeline

4.3 Experiments

Sections 4.3.1 and 4.3.2 explain the experimental set-up and evaluation parameters.

4.3.1 Experimental Set-up

This work aims to predict the OS of GBM patients using the proposed methodology based on the combination of DL and ML. Ranger [172] optimizer trains the 3D UNet model to

segment tumor subregions with $1e-4$ learning rate. For the first five epochs, the linear warm-up was also employed. The 3D UNet model was trained on 295 patients out of 369 and cross-validated on 74 patients of the BraTS training patient's data for 200 epochs using a batch size of 1. Further, the trained 3D UNet model generates segmentation maps for the validation set of 125 patients. The segmentation architecture was implemented in PyTorch. The training was completed using a 16 GB NVIDIA Tesla P100-PCIE graphics card available with the Google Colaboratory Pro facility. Training is done without weight decay; however, we reduce the learning rate using a cosine decay schedule as given in *Eq.* 4.1:

$$\eta_t = \eta_{\min} + 0.5\left(\eta_{\max} - \eta_{\min}\right)\left(1 + \cos\left(\pi \frac{T_{cur}}{T}\right)\right) \tag{4.1}$$

where η_{max} denotes our starting learning rate (set to $1e-4$); η_{min} represents our final learning rate (*i.e.*, $1e-7$); T_{cur} reflects the current iteration counter; and T is the total number of iterations to train (set to 200 epochs). The length of one epoch of training is 30 minutes due to the massive input and intermediate tensors, as well as the switching time between GPU and CPU. Further, the segmented and raw MRI scans were used for extracting the handcrafted and DL-based features for SP. The DL features were extracted using a Keras 2D ResNet50 model pre-trained on the Image dataset. These DL-based features were further combined with clinical features of selected 236 patients (patients whose survival data was available) like patient's age and handcrafted features such as the amount of necrotic/edema/enhancing, tumor to brain ratio and extent of tumor. The final features were selected using ML-based feature selection algorithm, *i.e.*, PCA written in the Python-based Scikit-learn library. Lastly, the selected set of 35 feature vectors (29 DL-based features, 5 handcrafted features and one clinical feature) were used to train regression-based ML models using the PyCaret Python Package.

4.3.2 Evaluation Parameters

The parameters used for the evaluation of the proposed segmentation model (such as DSC, Sensitivity, Specificity, 95%HD) and survival model (such as MSE, medianSE, stdSE, and Spearman) are same as explained in Chapter 3, Section 3.3.2. For classification, OS time was quantified into three survival categories: short survival ($< 10\ months$), intermediate survival (10–15 $months$) and long survival ($> 15\ months$). The three-class accuracy metric was evaluated. Evaluation metrics for regression include the MSE, mSE and Spearman correlation coefficient (ρ).

4.4 Results and Discussion

The chapter highlights the technique for automatically predicting survival in glioma patients using mpMRI images and clinical data. Figure 4.7 depicts the segmentation maps of tumor subregions created by the segmentation module which have been utilized to forecast the survival time of patients. The segmentation maps produced by the 3D UNet architecture are of high quality. However, a few slices from the validation set were selected to demonstrate the segmentation model's performance graphically. Figure 4.7 exhibits 2D and 3D visualizations created through the use of ITK-SNAP Tool.

The generated segmentation labels were evaluated using the IPP. The three tumor subregions, *i.e.*, ET, WT, and TC were defined to validate the segmentation model's performance. All the evaluation metrics were calculated using the IPP organized by the University of Pennsylvania. The DSC, HD95, Sensitivity, and Specificity of ET, WT, and TC were determined to assess the segmentation performance. According to the evaluation results, as reflected in Table 4.3, the segmentation model performed better with respect to WT than the TC and ET region in DSC and HD95. It was found that the DSC (ET:0.768, WT: 0.877, TC: 0.794) and HD95 (ET:32.689, WT: 7.799, TC: 13.859) distances for segmentations were good. However, the model's specificities (ET: 0.999, WT: 0.998, TC: 0.999) were far greater than its sensitivities (ET: 0.796, WT:0.931, TC:0.799) for different tumor subregions which suggested that the predictions were under-segmented. The MRI segmentation results were used to train the SP model on the characteristics collected from brain tumor subregions of 236 patients for whom clinical data such as their age was available. The blended model was then utilized to forecast patients' OS on the validation set of 29 GBM patients. The performance of OS prediction was examined through regression evaluation method on IPP. The OS predicted in days was submitted on IPP. Pair-wise error analysis of the regression model was performed between the anticipated and actual OS (in days) using measures such as MSE, Spearman, medianSE, and stdSE. The evaluation results presented in Table 4.3 indicate that the proposed blended model trained on the selected feature set attained the highest ever recorded MSE of 87067.33, medianSE of 30915.66, stdSE of 124721.48, and an excellent Spearman coefficient of 0.326 comparable to the existing procedures verified on the same group of patients. The proposed technique performed well with the DL features extracted through the use of a 2D pre-trained ResNet50 model instead of training a 3D SP model from the scratch. Spearman Correlation Coefficient with a value of 0.326 demonstrates excellent correlation between predicted and actual survival values. The BraTS web portal was used to evaluate the validation data regression metrics.

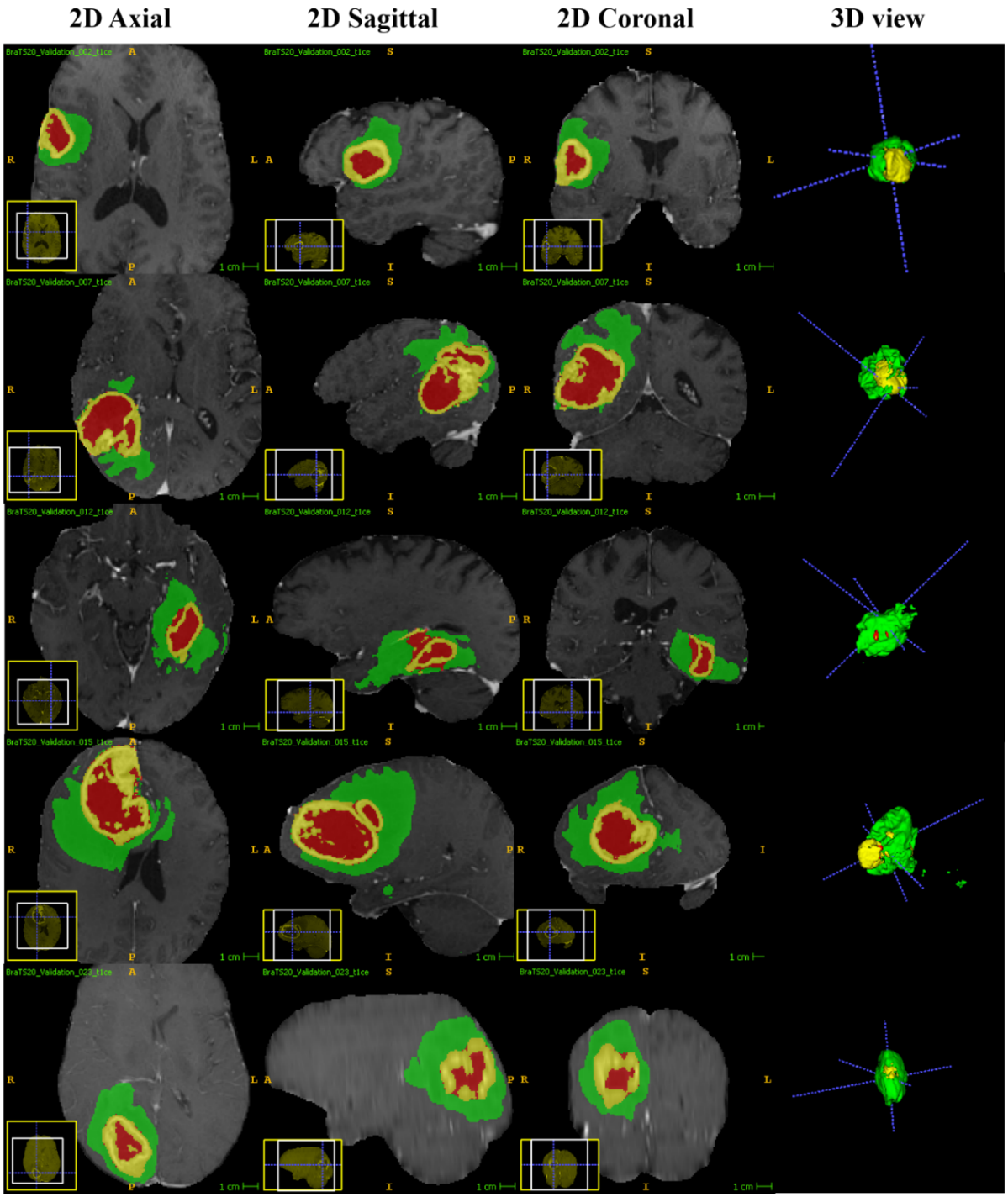


Figure 4.7: Zoomed Segmentation Maps Generated Using 3D UNet from the Brats Validation Data for Subject Ids: 2, 7, 12, 15, 23. The Generated Segmentation Maps were Overlayed On The T1CE Modality Of Individual Patients With Labels 1 For NCR And NET (Red Colour), 2 For ED (Green Colour), 4 For ET (Yellow Colour)

4.4.1 Comparison with State-of-the-Art Techniques

For the proposed methodology, a CNN-based UNet model was used to perform automatic segmentation of glioma MRI scans which accounted for the tumor's heterogeneous image findings. A different validation set was used to properly verify the segmentation model. The results showed that the 3D UNet trained with 3D MRI scans and ranger optimizer enhanced the segmentation model's generalizability. As CNN-based tumor segmentation is fast and repeatable, it has greater practical applications than manual tumor segmentation. Automating time-consuming, user-dependent operations can reduce user bias. Further, the data was obtained using different MRI scanners with varied acquisition settings. Despite this variability, the CNN method successfully identified and segmented the glioblastoma subregions. It was hypothesized that the image pre-processing steps (*i.e.*, min-max scaling) reduced the heterogeneities caused by inter-scanner variance and acquisition dimensional space, resulting in a more consistent outcome. External validation demonstrated that the segmentation model's prediction abilities were robust. The performance of the proposed segmentation model has been compared with that of other existing techniques as shown in Table 4.3.

Table 4.3: Performance of Proposed Segmentation Model on the Validation Dataset (N=125 Cases)

Reference	**Dice score**			**Hausdorff95 distance**		
	WT	**TC**	**ET**	**WT**	**TC**	**ET**
Patel *et al.*, 2021 [77]	0.90	0.85	0.79	4.88	5.85	15.70
Pang *et al.*, 2021 [173]	0.88	0.76	0.75	18.09	29.05	34.23
Han, 2021 [87]	0.77	0.66	0.53	8.71	11.67	7.18
Agravat and Raval, 2021 [75]	0.87	0.75	0.76	7.03	10.87	27.70
Anand *et al.*, 2021 [74]	0.88	0.74	0.71	6.88	32.00	38.31
Soltaninejad *et al.*, 2021 [82]	0.87	0.80	0.66	6.91	7.80	47.33
Ali *et al.*, 2021 [174]	0.87	0.74	0.74	9.42	10.09	3.92
Pei *et al.*, 2021 [81]	0.89	0.81	0.76	5.28	7.74	33.26
Parmar and Parikh [72]	0.90	0.84	0.74	5.08	8.69	36.04
Russo *et al.* [175]	0.89	0.82	0.76	6.07	8.12	27.37
Miron *et al.* [76]	0.90	0.84	0.78	-	-	-
González *et al.* [71]	0.90	0.81	0.77	6.16	7.55	21.80
Proposed	**0.88**	**0.79**	**0.77**	**7.79**	**13.86**	**32.69**

Table 4.4 compares the findings of the proposed methodology with those attained through the recent existing techniques. As shown in the Table 4.4, the proposed methodology achieved the lowest MSE with lowest medianSE. It is pertinent to mention here that we only employed principal components based on DL features, tumor region-based volume/ratio related handcrafted features, and the patient's age as predictors. The methodology suggested

in this work has performed better than many existing techniques and delivered superior predictions for SP. Using a pre-trained model saves the training time along with minimizing the resource utilization and avoid model overfitting on a small patient's data available for training SP models. The performance of the proposed methodology makes it evident that information collected through 2D ResNet50 architecture from segmented and raw MRI images, tumor-based handcrafted features, and the patients' age had a critical influence on GBM patients' survival prognosis.

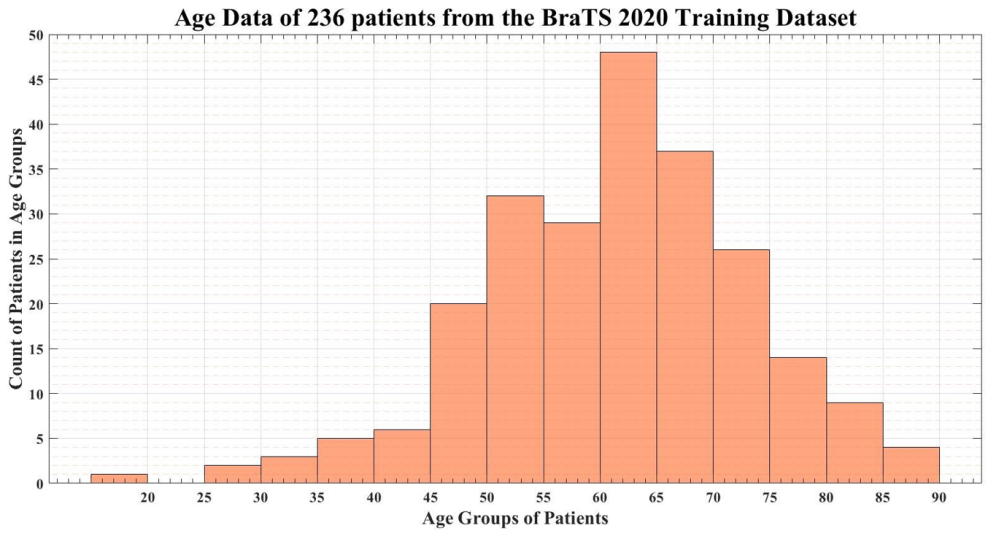
Table 4.4: Performance of Proposed Survival Model on the Validation Dataset (N=29 GBM Cases)

Reference	**MSE**	**medianSE**	**stdSE**	**Spearman**
Patel *et al.* [77]	152467	39601	196569	0.479
Pang *et al.* [173]	151013.4	91990.89	185785.1	0.155
Han [87]	147898.5	52900	162551.7	0.333
Agravat and Raval [75]	116083.5	43974.09	168176.2	0.217
Anand *et al.* [74]	110677.4	22874.18	142423.7	0.169
Soltaninejad *et al.* [82]	109564	39601	192755	0.28
Ali *et al.* [174]	105079.4	37004.93	146376	0.134
Pei *et al.* [81]	101697	56169	116680	0.329
Parmar and Parikh [72]	101278	49248	132237.7	0.228
Russo *et al.* [175]	99776.58	24539.13	168419.9	0.338
Miron *et al.* [76]	87744.14	37636	118896.9	0.321
González *et al.* [71]	87581	51529	141983	0.442
Proposed (With patient' sage)	**87067.33**	**30915.66**	**124721.48**	**0.326**
Proposed (Without patient's age)	**102223.558**	**26404.996**	**207142.707**	**0.222**

The data relating to the age of the patients drawn from the BraTS 2020 training dataset is displayed in Figure 4.8 (a). The age of the patients ranges from 18 years to 86 years. However, most of the patients belong to the 60-65 years age group. The categorization of the patients made on the basis of their age in the validation set is reflected in Figure 4.8 (b). The age of the patients ranges between 21 years and 85 years. However, most of the patients belong to the 60-80 years.

In some of the previous studies also [43, 150, 156, 157, 189, 190], age has been reported as the most significant factor in the case of patients with GTR status. Thus, the patient's age acts as a sole predictor of patient's survival. Table 4.4 also presents the survival results both with and without considering the age of the patients as a predictor for the proposed methodology. It is evident from the results that there is an increase in the error and a decrease in the correlation without considering age as a predictor in the SP of GBM patients. A summary of the features extracted by previous researchers along with the MSE results achieved using the same validation set of GBM patients is provided in Table 4.5.

(a)



(b)

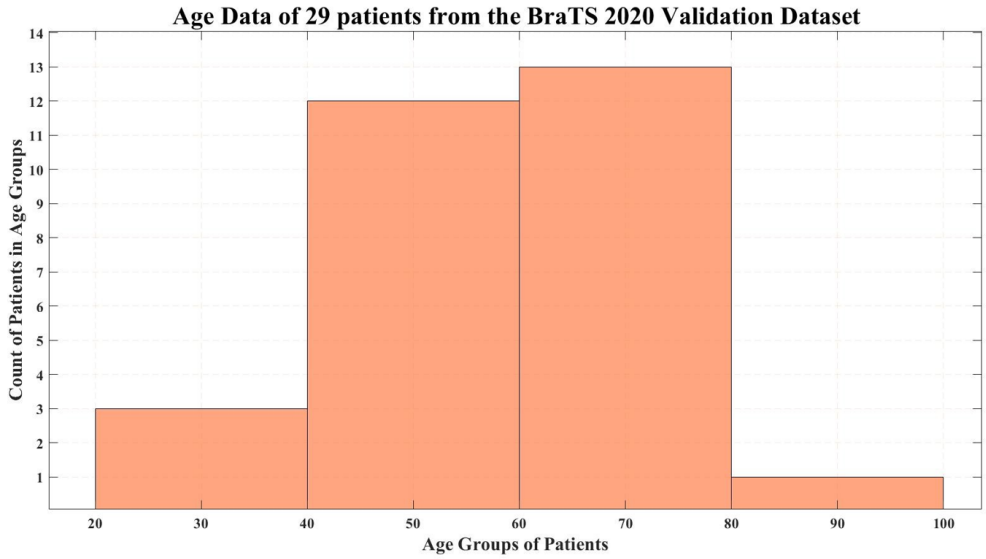


Figure 4.8: (a) Age Groups of 236 Patients from the BraTS 2020 Training Set, (b) Age Groups of 29 Patients from the BraTS 2020 Validation Set

Table 4.5: Summary of Features Extracted by Previous Studies and the Current Study To Predict The Survival Time Of GBM Patients using the Same Set Of Patients

Reference	**Feature Category**	**Imaging-based Features Extracted**	**MSE**
Pang *et al.* [173]	Busyness extracted from segmentation results in the enhanced area as the mask, patient's age	Minor axis length, sphericity extracted from segmentation result with ED area as the mask	151013.4
Han [87]	Patient's age, complicated 16 parameters	-	147898.5
Agravat and Raval [75]	Patient's age, statistical, and necrosis shape features	Amount of ED/NCR/ET, the extent of tumor and proportion of tumor, elongation, flatness, minor axis length, primary axis length, 2D diameter row, 2D diameter column, sphericity, surface area, 2D diameter slice, 3D diameter	116083.5
Pei *et al.* [81]	Shape features, number of NC, ED, ET, shape elongation, flatness, least axis length, surface area, and non-radiomics features, such as patient's age	-	101697
Parmar and Parikh [72]	Volumetric and shape features, patient's age.	Major axis dimension, minor axis dimension, elongation, maximum 2D diameter, maximum 3D diameter mess volume, spherical index, flatness for the whole brain, and tumor substructure	101278
Russo *et al.* [175]	Deep-learning features (LesionEncoder), latent variables of the input MRI scans, resection status, and patient's age	-	99776.58
Proposed	**DL features: Principal components from the FLAIR, T1, T2, T1CE modality, Clinical feature: Patient's age**	**Handcrafted features: Amount of NCR, ED, ET, Tumor to brain ratio, Extent of tumor**	**87067.33**

4.4.2 Strengths and Limitations of the Proposed Methodology

Although the proposed model's Spearman has not the highest of all the current methodologies, yet it has successfully generalized the results during training which are no less significant in any way as compared to the top-performing models. The Spearman correlation coefficient is a non-parametric test based on the ranked values for each variable rather than the raw data. The Spearman rank correlation calculates the P-value in the same way as by linear regression and correlation, except that it is calculated on ranks, not measurements. It finds whether a correlation exists between two variables. When the calculated value stands close to 1, there is a positive correlation; when it is close to -1, there exists a negative correlation; and when it appears close to 0, there is a limited correlation. This study with Spearman of 0.32 represents a moderate association between predicted and true values, but this value is clinically quite useful. Further, Spearman method is less significant as compared to other metrics used in this study to validate the proposed methodology because it cannot be used for calculating correlation in grouped frequency distributions. It can be applied to individual observations only. It also includes the commonly used metric of MSE to evaluate the regression-based problem. The SP should be formulated as a regression problem with MSE as the optimization target. The MSE is considered significant as it ensures that the trained model has no outlier predictions with huge errors. The MSE puts larger weight on theses errors due to the squaring part of the function. It is a measure used to find how close a fitted line is to data points. Smaller the Mean Squared Error, closer would be the fit to the data. Thus, the lower MSE achieved by the proposed methodology proves its effectiveness. The suggested methodology is able to extract more abundant features from 3D MRI scans using the 2D ResNet50 model trained on the ImageNet dataset. Further, it significantly minimizes the network parameters and computing resources used for the purpose of this study. The proposed methodology projects the survival time of patients in terms of days. SP can also be considered a classification problem with accuracy as the primary metric, if the training data contains more samples for all the three categories, *viz.*, short, mid and long survivors. The patient's balanced data leads to develop a more generalized model. The number of patients tested for OS is quite less than the segmentation task. Without extensive training in patient's data, the techniques cannot extract accurate information. There lies the scope for further research in the domain of survival prognosis which can consider tumor grade as an input feature. In this work, the segmentation model was trained on both LGG and HGG types of tumors because the labeled ground truth was available for such patients. But only the survival time of HGG patients who had grade IV GBM tumor was predicted due to the unavailability of survival data for patients of different tumor grades. Thus, all the features used for training the SP model were extracted from the MRI scans of GBM HGG patients only. The BraTS dataset used in this study contained the survival data for GBM HGG patients only

which did not allow to consider tumor grade as an input feature vector. But whether features extracted from MRI scans of patients with different tumor grades can affect the patient survival or not, this can be the focus of future work. A greater effort is required for the collection of data related to LGG patients.

4.5 Inference and Findings

Early prediction of survival time of GBM is vital to plan the treatment and patient care properly. Diagnostic modalities such as T1, T1CE, T2 and FLAIR MRI scans play a crucial role in monitoring the progression of tumorous regions and the severity of the disease in GBM patients. This chapter presents a pre-trained 2D ResNet feature extractor to extract the features from raw and segmented MRI scans. The scans have been segmented using 3D UNet architecture. Along with deep features, handcrafted features have also been calculated from raw and segmented tumor subregions. PCA has been used as a feature reduction technique to curtail the dimensions of deep features extracted through a pre-trained 2D ResNet50 model. Automated regression-based ML has been used to make predictions for survival time of GBM patients using reduced DL feature-based principal components, handcrafted tumor subregion-based volumes/ratios, and clinical data like patient's age. The top six best performing regression models obtained through the PyCaret Package have been blended to make final predictions on unseen data. The experimental results have demonstrated that the proposed methodology has performed better than many other existing 2D and 3D techniques proposed for SP of GBM patients.

Chapter 5

Automated Neural Network-based Survival Prognosis of Glioblastoma Multiforme Patients

This chapter proposes a lightweight 2D methodology to predict the survival time of GBM patients based on the pre-operative 3D MRI scans and clinical data provided in the publicly available BraTS 2020 dataset. Firstly, the 2D Residual UNet for Segmentation (ResUNet-SEG) model was trained to perform semantic segmentation on brain tumor subregions. Then, the raw and segmented MRI volumes along with clinical data were used to train the 2D CNN for SP (CNN-SP) model to predict the survival time in days. The experiments showed that the proposed methodology achieved an accuracy of 0.517 with MSE of 136783.42, MSE, medianSE of 106608.6, stdSE of 139210.8, and Spearman regression correlation score of 0.299 on the multimodal BraTS 2020 validation set. The results of the proposed methodology are better than those of other state-of-the-art automated techniques used for survival prognosis of GBM patients validated on the same set of patients. Automated prediction of the OS time of GBM patients is a vital topic due to its clinical applications. The results have proved that DL-based feature learning is better than existing ML-based techniques with handcrafted radiomics-based feature extraction. It eliminates the need for feature selection as well. However, the applicability results is limited due to unavailability of vast clinical data required to train CNN-based deep architectures.

5.1 Contribution of the Proposed Technique

It has been observed through the study of literature that UNet design delivers excellent semantic segmentation results. As a result, a 2D UNet-based segmentation module with ResNet modifications was adopted in the proposed methodology. The raw and segmented MRI images were used to extract the deep features. Age and extracted DL features were utilized for training CNN-based architecture for OS prediction.

The proposed technique has contributed significantly as follows:

– The tumor and its subregions were segmented using 2D residual UNet-based encoder-decoder architecture.

– The clinical data, segmented scans and raw MRI scans together were used for deep feature extraction and selection.

– The selected input set was used to train 2D CNN DL-based model and predict patients' survival time in days.

– The results of the proposed technique were compared with those of other state-of-the-art techniques to validate its performance.

This chapter has been organized into various sections. Section 5.1 explains the contribution, and organization of the chapter. Section 5.2 deals with the dataset and describes the proposed methodology. The experimental set-up and evaluation parameters have been discussed in Section 5.3. The results and discussion are presented in Section 5.4. Section 5.5 highlights the findings of the proposed technique.

5.2 Materials and Methods

5.2.1 Dataset and Proposed Methodology

The BraTS 2020 dataset has been used for implementation of the proposed methodology and the detailed explanation of the dataset is mentioned in Chapter 3, Section 3.2.1. Our proposed methodology consists of steps *viz.*, MRI pre-processing followed by data augmentation in section, segmentation module, post-processing and lastly, SP module. All the steps are explained with complete details in the following sub-sections.

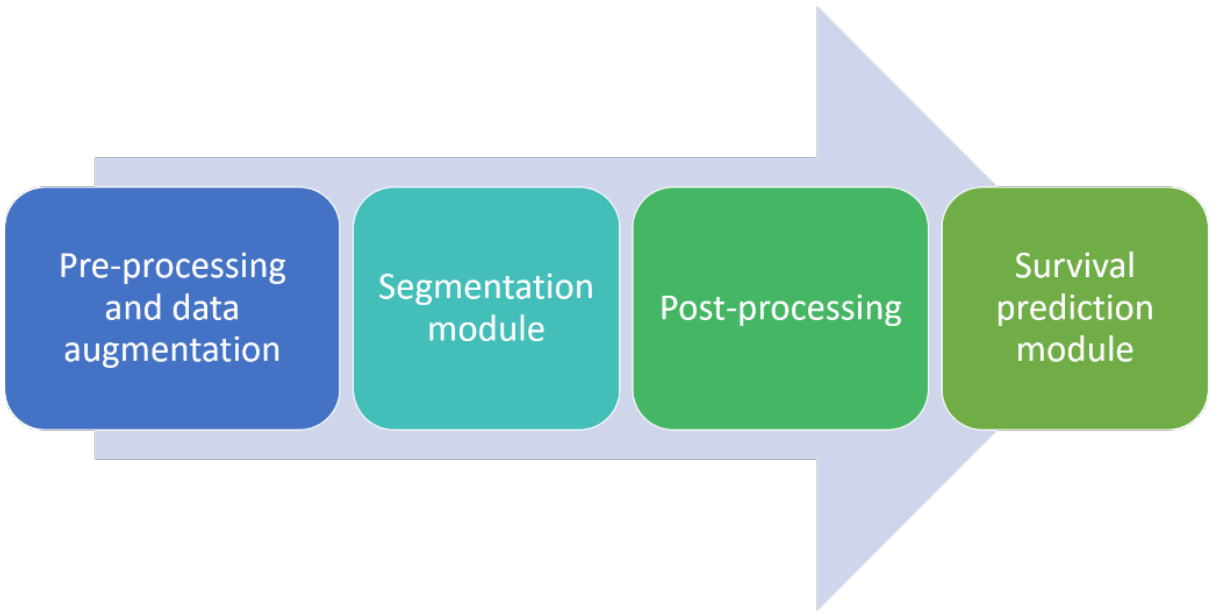


Figure 5.1: Top-level Schematic Diagram of the Proposed Methodology

As shown in Figure 5.1, the raw input 3D MRI NIfTI scans are given to the DL segmentation model in the processed form to generate labels for tumor subregions. Further, generated segmentation maps are post-processed into 3D MRI NIfTI scans followed by a regression-based DL SP model, which predicts the OS of patients in days.

5.2.2 Pre-processing and Data Augmentation

The 3D brain MRI NifTI scans are converted into Portable Network Graphics (PNG) format in this pre-processing stage. Firstly, label 4 for ET (yellow colour) is modified to label 3 because label 3 is missing in segmentation labels. All the modalities are read using a nibabel Python package to process neuroimaging file format like NifTI files. These MRI scans are normalized as per the *Eqs.* 5.1, 5.2 and 5.3 given below:

$$i_{Mean} = mean(img[img > 0]) \tag{5.1}$$

$$i_{SD} = std((img[img > 0]) * 3)/128 \tag{5.2}$$

$$img = \begin{cases} img[img > 0] = \left(\frac{img[img>0] - i_{Mean}}{i_{SD}}\right) + 127 \\ img[img < 0] = 0 \\ img[img > 255] = 255 \end{cases} \tag{5.3}$$

where: img = Brain MRI scan image with background,
i_{Mean} = Mean of the brain MRI scan without background,
i_{SD} = Standard deviation of the brain MRI scan without background.

These MRI scans are normalized before performing segmentation. The height and width of scans are kept the same, *i.e.*, 240×240 but the depth is changed from 155 slices to one slice. Raw modalities are saved as PNG files with Red-Green-Blue-Alpha (RGBA) format. The FLAIR, T1CE, T1 and T2 modalities are saved in Red, Green, Blue and Alpha channels, respectively, with a single slice at a time. Thus for each patient, 155 PNG files are generated with all four modalities are four channels in RGBA format, as shown in Figure 5.2. Similarly, segmentation labels are also saved as PNG files with the same labels in all channels of RGB format.

In data analysis, data augmentation refers to increasing the quantity of data by adding slightly changed copies of previously collected data or by creating new synthetic data from previ-

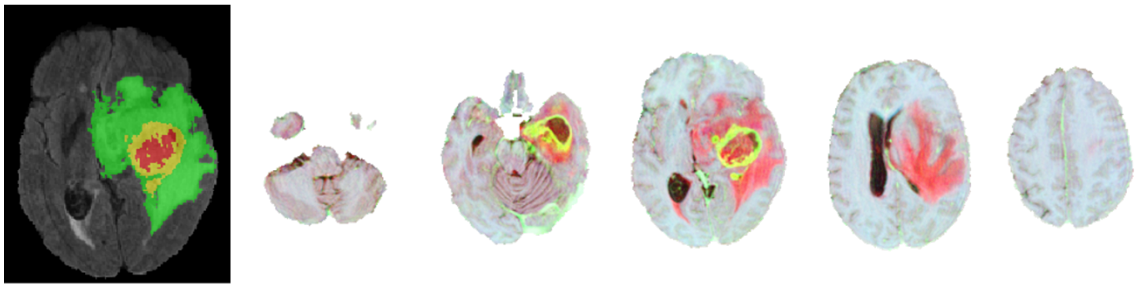


Figure 5.2: Different PNG Slices Generated from NIfTI After Pre-processing

ously collected data. When training a machine learning model, it functions as a regularizer and aids in reducing overfitting. As shown in Figure 5.3, horizontal and vertical flipping is applied to increase the count of PNG slices and reduce overfitting.

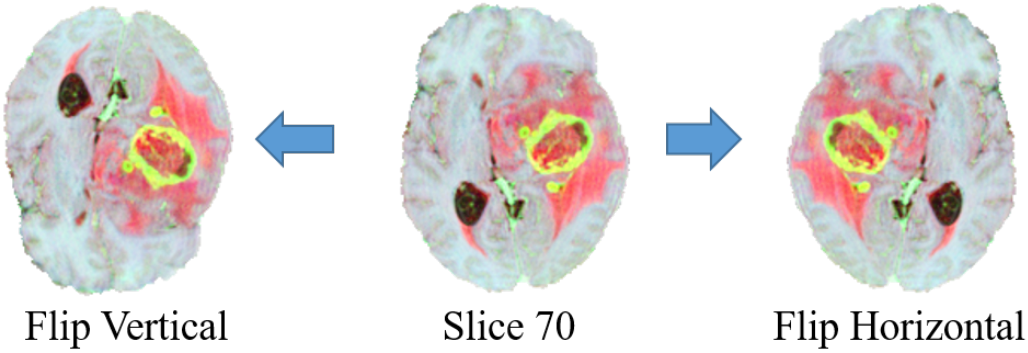


Figure 5.3: Different PNG Slices Generated After Data Augmentation

5.2.3 Segmentation Module

Using a 2D version of the UNet [62] as the backbone of the segmentation module for our proposed technique, we train it to segment subregions like WT, TC, and ET concurrently using a single network. It accepts four modalities of brain tumor MRI images and ground truth segmentation as 2D PNG input files.

UNet design is a significant advancement in computer vision, revolutionizing segmentation in medical imaging and other domains. The UNet's defining characteristic is the extended skip link between each level of contracting and growing paths. It is as though FCN is being dragged upward from both ends resulting in the U-shaped architecture. ResNet was another game-changing development in computer vision. Instead of learning unreferenced functions, Residual Networks (ResNets) learn residual functions from the level inputs. Residual

nets enable these layers to suit a more generic mapping instead of depending on a particular mapping for each layer. ResNet-50, for example, has fifty layers and is built utilizing residual blocks sequentially. The residual blocks and the identity mapping connections of ResNet aided in constructing a deeper CNN, which delivered record-breaking classification results on the ImageNet dataset. By substituting ResBlock for convolutions on each level of UNet, we can get more outstanding performance than the original UNet. Figure 5.4, given below, illustrates the complete segmentation model architecture.

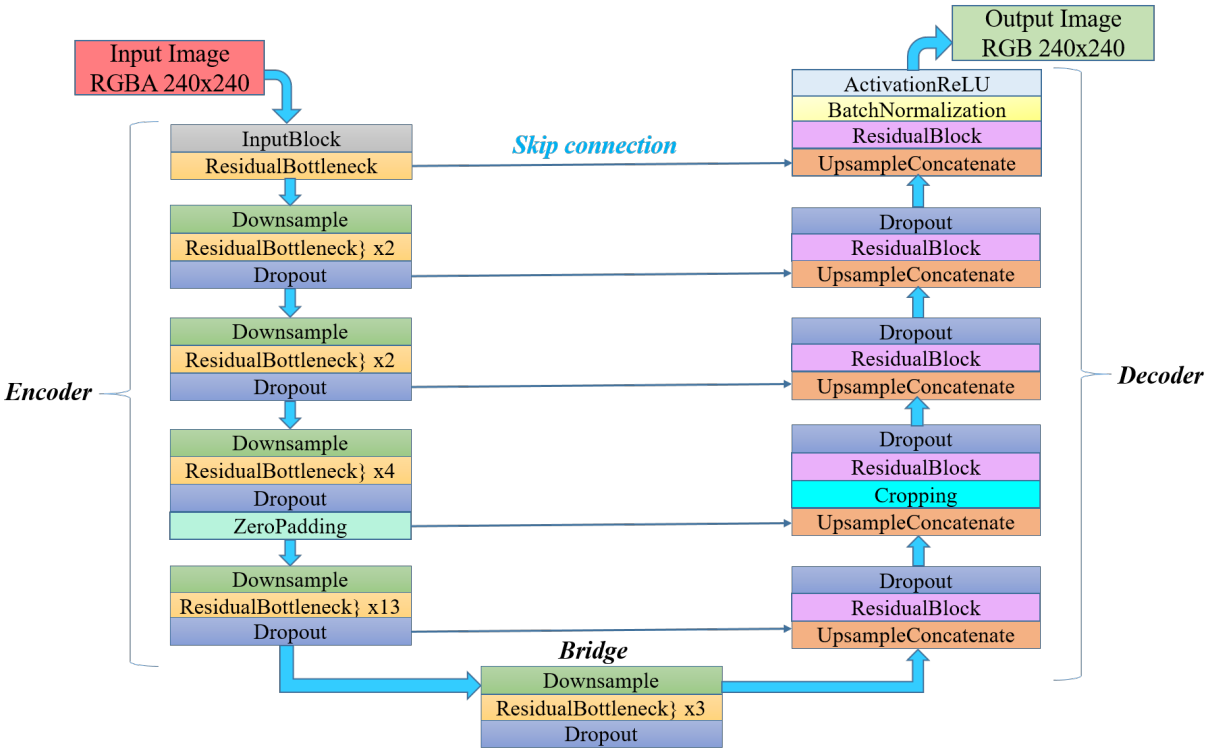


Figure 5.4: The Architecture of the 2D ResUNet-SEG Model

The proposed network is a 2D UNet variation, as previously stated. Although we kept UNet with skip connections' encoder-decoder backbone intact, we updated its design as follows:

- 'Residual bottleneck' and 'downsample' blocks in every encoder stage are based on the ResNet.
- 'Residual' blocks are also based on the ResNet architecture in every decoder stage.

Thus, the architecture is named the 2D ResUNet-SEG model. Dropout is also included to improve model regularization. As it is shown in Figure 5.4, the segmentation network encompasses five stages in encoding and decoding pathways with a bridge in between both the pathways. The complete details of each block used in segmentation architecture are illustrated in Figure 5.5.

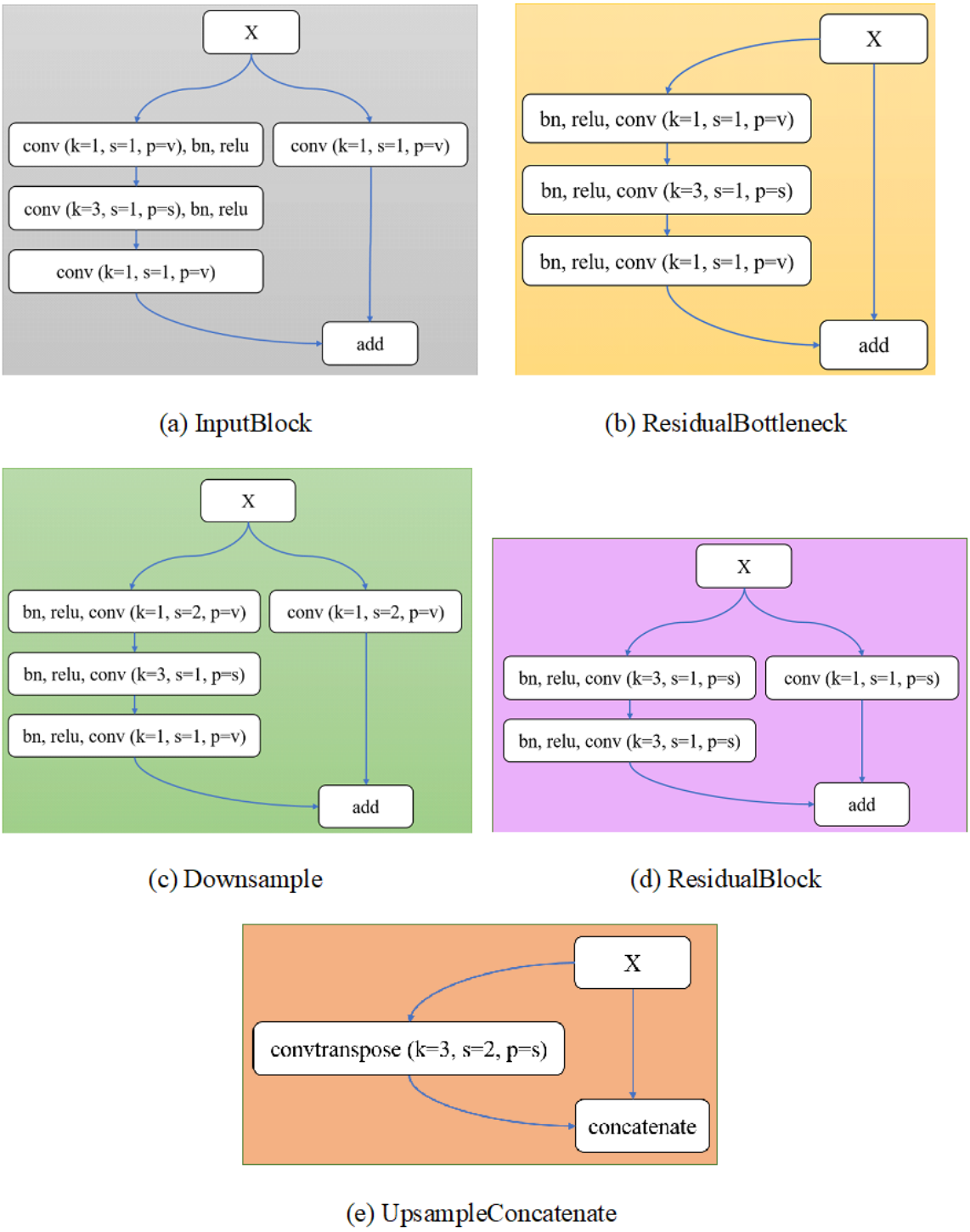


Figure 5.5: Different Blocks Used in Encoding and Decoding Pathways of Segmentation Architecture

Some abbreviations used in Figure 5.5 representing different blocks used in segmentation architecture.

- *Input:* 'X' denotes the input to every block.
- *Operations:*
  - 'conv' means 2D Convolution operation. Matrix operations like convolution use kernels to glide across input data and conduct element-by-element multiplication with the section of the input they are on, then the summation of the results. Weight sharing and picture translation may be achieved by using convolutions, which reduce the number of effective parameters. In every 'conv' operation, the 'k' denotes kernel size, 's' denotes stride and 'p' denotes padding with 'v' as valid and 's' as same.
  - 'bn' means Batch Normalization. To speed up the training of DNNs, Batch Normalization tries to decrease the internal covariate shifts. A normalization procedure fixes layer inputs means and variances. Gradient flow across the network is further facilitated by Batch Normalization, which reduces the dependency of gradients on the parameters or their starting values. It eliminates the danger of divergence by allowing considerably larger learning rates.
  - 'relu' are activation functions with linear positive dimensions but zero negative dimensions, referred to as "rectified linear units". A kink causes the non-linearity in the function. However, linearity is appealing because it precludes gradient non-saturation (contrast to sigmoid activations), even if the gradient is zero for half of the main diagonal.
  - 'add' denotes the add operation, which adds two given inputs.
  - 'concatenate' denotes concatenate operation, which appends two given inputs.
  - 'convtranspose' denotes the Transposed convolution. ConvTranspose is a convolutional procedure whose kernel is learned during model training (like a regular Conv operation).

Table 5.1 provides a detailed breakdown of the architectural description, including the number of feature maps in each layer. An encoder, a decoder, and a bridge form the standard U-shape network used to perform semantic segmentation on tumor subregions. The output will be a segmentation map with four tumor classes.

Table 5.1: The Details of Feature Maps Generated at Each Stage of Segmentation Architecture

Stage	**Image dimensions**	**Encoder Block**	**Encoder feature maps**	**Decoder Block**	**Decoder feature maps**
Level 1	240 × 240	InputBlock	16/16/64 × 1	ResidualBlock	Output: 4 classes 32/32 × 1
		ResidualBottleneck	16/16/64 × 1	UpsampleConcatenate	128 × 1
Level 2	120 × 120	Downsample	32/32/128 × 1	ResidualBlock	64/64 × 1
		ResidualBottleneck	32/32/128 × 1	UpsampleConcatenate	256 × 1
Level 3	60 × 60	Downsample	64/64/256 × 1	ResidualBlock	128/128 × 1
		ResidualBottleneck	64/64/256 × 2	UpsampleConcatenate	512 × 1
Level 4	30 × 30	Downsample	128/128/512 × 1	ResidualBlock	256/256 × 1
		ResidualBottleneck	128/128/512 × 4	UpsampleConcatenate	1024 × 1
Level 5	16 × 16	Downsample	256/256/1024 × 1	ResidualBlock	512/512 × 1
		ResidualBottleneck	256/256/1024 × 13	UpsampleConcatenate	2048 × 1
Bridge	8 × 8	Downsample	512/512/2048 × 1		
		ResidualBottleneck	512/512/2048 × 3		

The encoder path compresses the visual features, then is recovered in the decoder pathway for pixel-by-pixel classification. Figures 5.5 (b) and (d) depict the layout of the residual blocks within encoder and decoder routes. The encoder route and bridge connection use the bottleneck design, which consists of a 1×1 convolution layer for reducing the depth of the feature channel, a 3×3 convolution layer for restoring dimension, and lastly, a 1×1 convolution layer for restoring dimension. The typical residual block consists of two stacked 3×3 2D convolutions in the decoder path. We use BN and ReLU activation in all residual blocks (Figure 5.5).

Considering the several levels of a deep architecture, the issues of overfitting and deactivating dead neurons must be handled during network training. Following the work [191], a regularisation technique is adopted that employs a low-rate dropout (e.g., 0.2) after each level. To aid in feature mapping, we concatenate the levels of each decoder (after upsampling) through a connection from the encoder's output. Downsampling is performed using a two-step stride in each level's initial convolution (except for level 1). Two-dimensional transpose convolution with a kernel size of 2×2 and a stride size of 2 is used for upsampling. Convoluted using a 1×1 kernel, the final layer gives pixel-by-pixel classification scores to depict the three tumor subclasses and the background class; the channel decides the pixel's class with the maximum output (argmax); softmax is used during training. The input slice is a 2D image representing each MRI modality with four channels.

Loss of Segmentation: Categorical Cross-Entropy (CCE) was employed as the loss function, derived using below given *Eq.* 5.4. The CCE helps solve multi-class classification problems. It was selected for simplicity; however, alternative loss functions, such as soft dice loss, may have yielded superior performance [94].

$$CCE = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} 1_{y_i \in C_c} \log p_{\text{model}}\left[y_i \in C_c\right] \tag{5.4}$$

The double summation is computed over the instances 'i', which have a count of 'N' , and the categories 'c', which have a count of 'C'. The term $1_{y_i \in C_c}$ denotes the indicator function for the i$^{\text{th}}$ instance that falls within the c$^{\text{th}}$ category. The $p_{\text{model}}\left[y_i \in C_c\right]$ is the probability predicted that the i$^{\text{th}}$ instance belongs to the c$^{\text{th}}$ category as predicted by the model. When there are more than two categories, the NN outputs a vector of 'C' probabilities, each indicating the likelihood that the network input should be categorized as belonging to the specified category.

5.2.4 Post-processing

After segmentation, post-processing is performed to convert generated 2D 240 × 240 PNG masks into the 240 × 240 × 155 NIfTI formats. The generated NIfTI files are used further to predict the OS of GBM patients.

5.2.5 Survival Prediction Module

With the help of segmentation architecture, a tumor structure map is constructed for each patient with four discrete values, *i.e.*, '1' for NCR and NET (red colour), '2' for ED (green colour), '4' for ET (yellow colour), and '0' for background and normal brain tissues. The spatial distribution of tumor subregions provides information on tumor heterogeneity and tumor invasiveness. All the 3D raw MRI scans are standardized slice-wise using the following *Eqs.* 5.5, 5.6, 5.7 and 5.8:

$$image_{slice} = \sum_{z=1}^{155} raw_{mri}[:,:,z] \tag{5.5}$$

$$centered = image_{slice} - mean\left(image_{slice}\right) \tag{5.6}$$

$$centered = \frac{centered}{std(centered)}, if \quad std(centered) \neq 0 \tag{5.7}$$

$$standardized_{mri}[:,:,z] = centered \tag{5.8}$$

where: $image_{slice}$ = Sum of intensities of brain MRI scan slice-wise with background,

$centered$ = Scan after subtracting mean of whole scan from each slice divided by standard deviation of whole scan,

$standardized_{mri}$ = Brain MRI scan after standardization.

To predict OS, we trained a CNN architecture named 2D CNN-SP model to extract deep features from raw MRI scans and segmented tumor structure maps. Input images have dimensions of 240×240×5, which contain the 75^{th} slice for four standardized MRI modalities, *i.e.*, T1, T1CE, T2, FLAIR and one slice for segmented tumor structure map. More specifically, we pass an image data with five channels containing the centre slice of all the four modalities, *i.e.*, T1, T1CE, T2, FLAIR and segmented structure map into our 2D CNN-SP model and

extracted deep features from the end of the flatten layer. Previous works have shown that a patient's age alone can predict survival relatively well and outperform approaches integrating more complicated radiomic features [156, 192]. Considering the importance of the age factor in SP, the extracted feature map was multiplied by the patient's age, followed by other layers, as shown in Figure 5.6. The proposed 2D CNN-SP model was trained on a set of 236 patients, for which the BraTS organizers shared the survival data publicly.

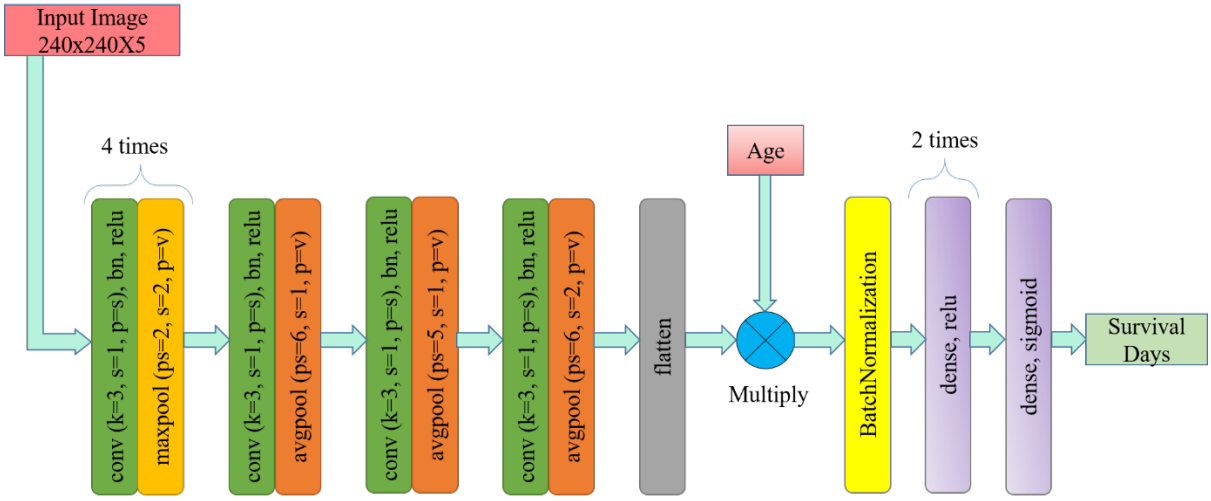


Figure 5.6: The Architecture of the 2D CNN-SP Model

2D CNN-SP architecture with seven-layered convolutions, as shown in Figure 5.6, was trained to predict the survival time of GBM patients. All convolutional layers have a kernel size of 3×3 with stride 2×2 and 'same' padding, followed by the BN and ReLU layers. The first four convolutional layers are followed by the Max Pooling layer having a pool size of 2×2 with stride 2×2 and 'valid' padding. The later three convolutional layers are followed by the Average Pooling layer having a pool size of 6×6, 5×5 and 6×6 with stride 1×1 and ' valid' padding, respectively. The trainable parameters are significantly reduced using Max Pooling and Average Pooling layers. Further, passing through the flatten layer, inputs are converted into 32 features multiplied by the patient's age, followed by BN and two dense layers with ReLU activation function. Finally, survival time is obtained in days as the output from the last dense layer with sigmoid activation. The output shape, *i.e.*, the shape of the feature map generated after every stage, is mentioned in Table 5.2.

Table 5.2: The Details of Feature Maps Generated at Each Stage of Survival Prediction Architecture

Stage	**Output Shape**
InputLayer1 (Image)	240, 240, 5
Conv2D, BatchNormalization, Activation	240, 240, 16

Table 5.2 continued from previous page

Stage	Output Shape
MaxPooling2D	120, 120, 16
Conv2D, BatchNormalization, Activation	120, 120, 24
MaxPooling2D	60, 60, 24
Conv2D, BatchNormalization, Activation	60, 60, 32
MaxPooling2D	30, 30, 32
Conv2D, BatchNormalization, Activation	30, 30, 32
MaxPooling2D	15, 15, 32
Conv2D, BatchNormalization, Activation	15, 15, 32
AveragePooling2D	10, 10, 32
Conv2D, BatchNormalization, Activation	10, 10, 32
AveragePooling2D	6, 6, 32
Conv2D, BatchNormalization, Activation	6, 6, 32
AveragePooling2D	1, 1, 32
Flatten	32
InputLayer (Age)	1
Multiply	32
BatchNormalization	32
Dense	32
Dense	16
Dense	1

Loss of Survival Prediction: A smaller MSE means a more accurate SP result. The MSE between a prediction P_i and it is the ground truth G_i can be calculated as given in *Eq.* 5.9.

$$L_{MSE}(P,G) = \frac{\sum_{i=1}^{N}(P_i - G_i)^2}{N} \tag{5.9}$$

5.3 Experiments

The experimental setup and the evaluation parameters are explained in below given subsections.

5.3.1 Experimental Set-up

We aim to predict survival days of GBM patients based on 2D ResUNet-SEG plus 2D CNN-SP methodology. Our proposed methodology is implemented in Keras with Tensorflow backend and trained on 16 GB NVIDIA Tesla P100-PCIE graphics card available with Google Colaboratory Pro facility. ADAM optimizer trains the 2D ResUNet-SEG model to segment tumor subregions with a learning rate of $1e-4$ and He initialization. We trained the ResUNet model on 369 patients of the BraTS 2020 training dataset for 50 epochs using batch sizes of 10. The scans of all patients are converted to single PNG images with four channels for each MRI modality, as indicated in the pre-processing section, and then randomly inserted into the model. The 2D CNN-SP model is also trained with ADAM optimizer to predict the survival time of GBM patients with a learning rate of $1e-3$. We trained the CNN-SP model for 300 epochs with batch sizes of 64 on 236 patients with clinical data from the BraTS 2020 training dataset. The raw and segmented scans of all patients are given as five-channel input to the model and patient's age to predict the survival days as the output.

5.3.2 Evaluation Parameters

The parameters used for the evaluation of the proposed segmentation model (such as DSC, Sensitivity, Specificity, 95%HD) and survival model (such as accuracy, MSE, medianSE, stdSE, and Spearman) are same as explained in Chapter 3, Section 3.3.2. For classification, OS time was quantified into three survival categories: short survival ($< 10\ months$), intermediate survival (10–15 $months$) and long survival ($> 15\ months$). The three-class accuracy metric was evaluated. Evaluation metrics for regression include the MSE, mSE and Spearman correlation coefficient (ρ).

5.4 Results and Discussion

Figure 5.7 shows the segmentation maps of tumor subregions generated using the segmentation module, further used to predict patients' survival time. The segmentation maps obtained using ResUNet-SEG architecture are quite good. In order to visually show the performance of the segmentation model, we selected a few slices from the validation set. The visualizations shown in Figure 5.7 are produced with ITK-Snap.

The available age groups of the training set can be seen in Figure 5.8 (a), with a range of 18-86 years. Maximum patients belong to the 60-65 years age group. The available survival day groups of the training set can be seen in Figure 5.8 (b), with range of 5-1767 days. Maximum patients have survival days within the range of 200-400. The available age groups of the

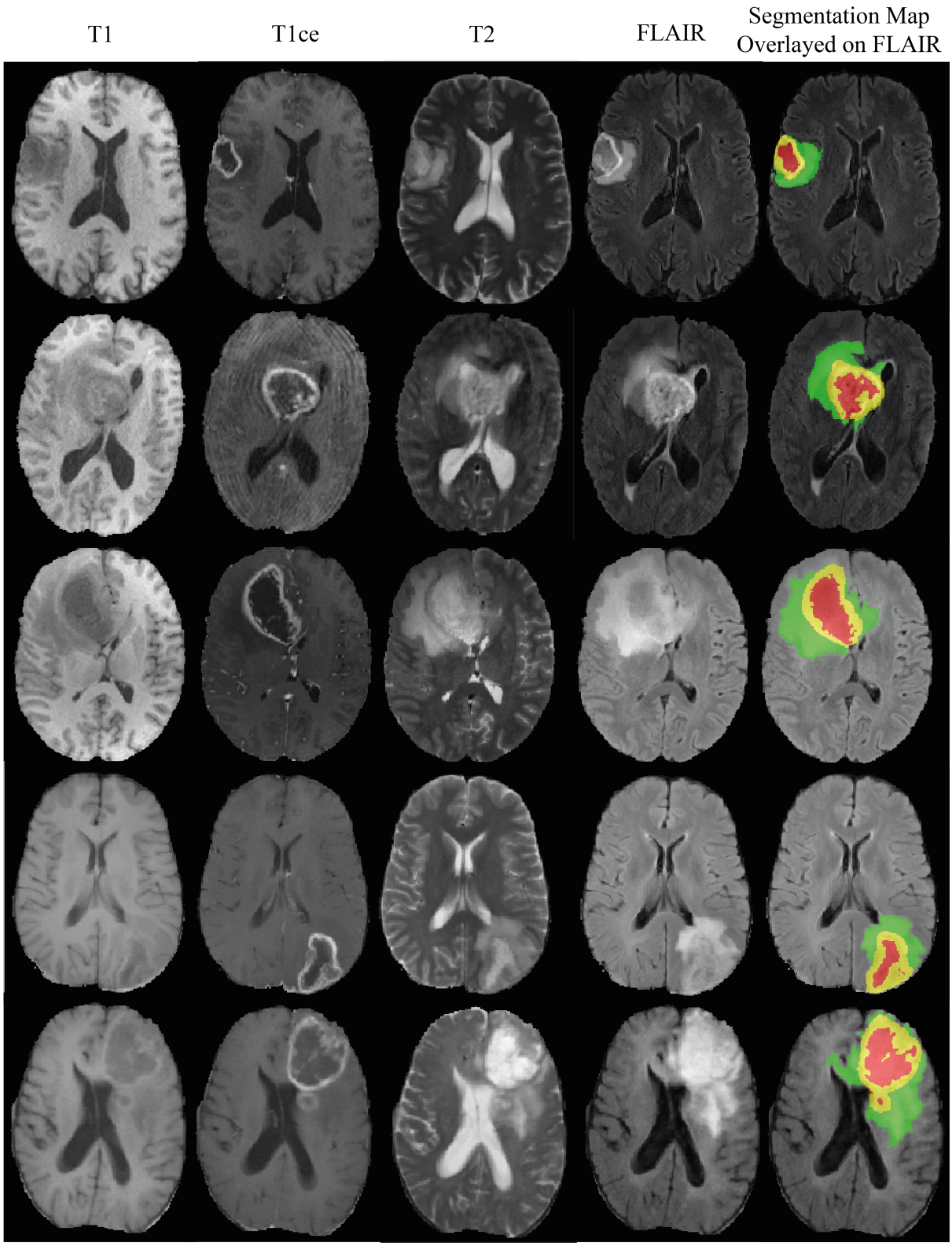


Figure 5.7: Segmentation Maps Generated Using 2D ResUNet-SEG from the BraTS Validation Dataset for Patient IDs: 2, 8, 15, 22, 34. The Generated Segmentation Maps are Overlayed on the FLAIR Modality of Individual Patients with Labels 1 for NCR and NET (Red colour), 2 for ED (Green colour), 4 for ET (Yellow colour).

validation set can be seen in Figure 5.8 (c), with a minimum age of 21 years and maximum age of 85 years approximately. Maximum patients belong to the 60-80 years age group. The predicted survival day groups of the validation set can be seen in Figure 5.8 (d), with range of 145-1286 days. Maximum patients have survival days within the range of 600-900, *i.e.*, long survivors category, followed by 300-600 range, *i.e.*, mid survivors category.

We verified our proposed methodology using the BraTS 2020 validation dataset with 29 patients, and the results are displayed in Table 5.3. Our proposed methodology achieved 51.7% accuracy, which is competitive to the state-of-the-art techniques validated on the same set of patients. The Spearman Correlation Coefficient value is the maximum for our proposed methodology among the three models with the same accuracy, proving the best correlation between predicted and actual survival values. For survival validation, CNN-SP achieved 136783 days2 MSE based on the regression phenomenon.

5.4.1 Comparison with State-of-the-Art Techniques

The proposed methodology results are compared with the existing techniques in Table 5.3. Our CNN-SP model achieved competitive results for SP among existing techniques. It is noted that we used only deep features and the patient's age as predictors. Although more complex features were integrated into existing ML and hybrid techniques, our proposed methodology outperforms all such techniques and provides better predictions. It means that the information extracted using CNN-SP architecture from segmented and raw MRI scans and patients' age played a vital role in the survival prognosis of GBM patients.

Table 5.3: Comparison of the Proposed Methodology with State-of-the-art Techniques Evaluated Using CBICA Online Portal for the Validation Set (n=29 Patients)

Reference	**Accuracy**	**MSE**	**medianSE**	**stdSE**	**Spearman**
Akbar *et al.* [193]	0.345	149764.87	60314.09	190758.94	0.123
Zhang *et al.* [133]	0.379	-	-	-	-
Miron *et al.* [76]	0.414	87744.14	37636	118896.9	0.321
Parmar *et al.* [72]	0.414	101278	49248	132237.66	0.228
Anand *et al.* [74]	0.448	110677.44	22874.18	142423.69	0.169
Soltaninejad *et al.* [82]	0.45	109564	39601	192755	0.28
Dai *et al.* [78]	0.45	87008	32400	-	0.43
Ali *et al.* [174]	0.483	105079.4	37004.93	146376	0.134
Agravat *et al.* [75]	0.517	116083.48	43974.09	168176.16	0.217
Pang *et al.* [173]	0.517	151013.39	91990.89	185785.15	0.155
Proposed	**0.517**	**136783.42**	**106608.6**	**139210.8**	**0.299**

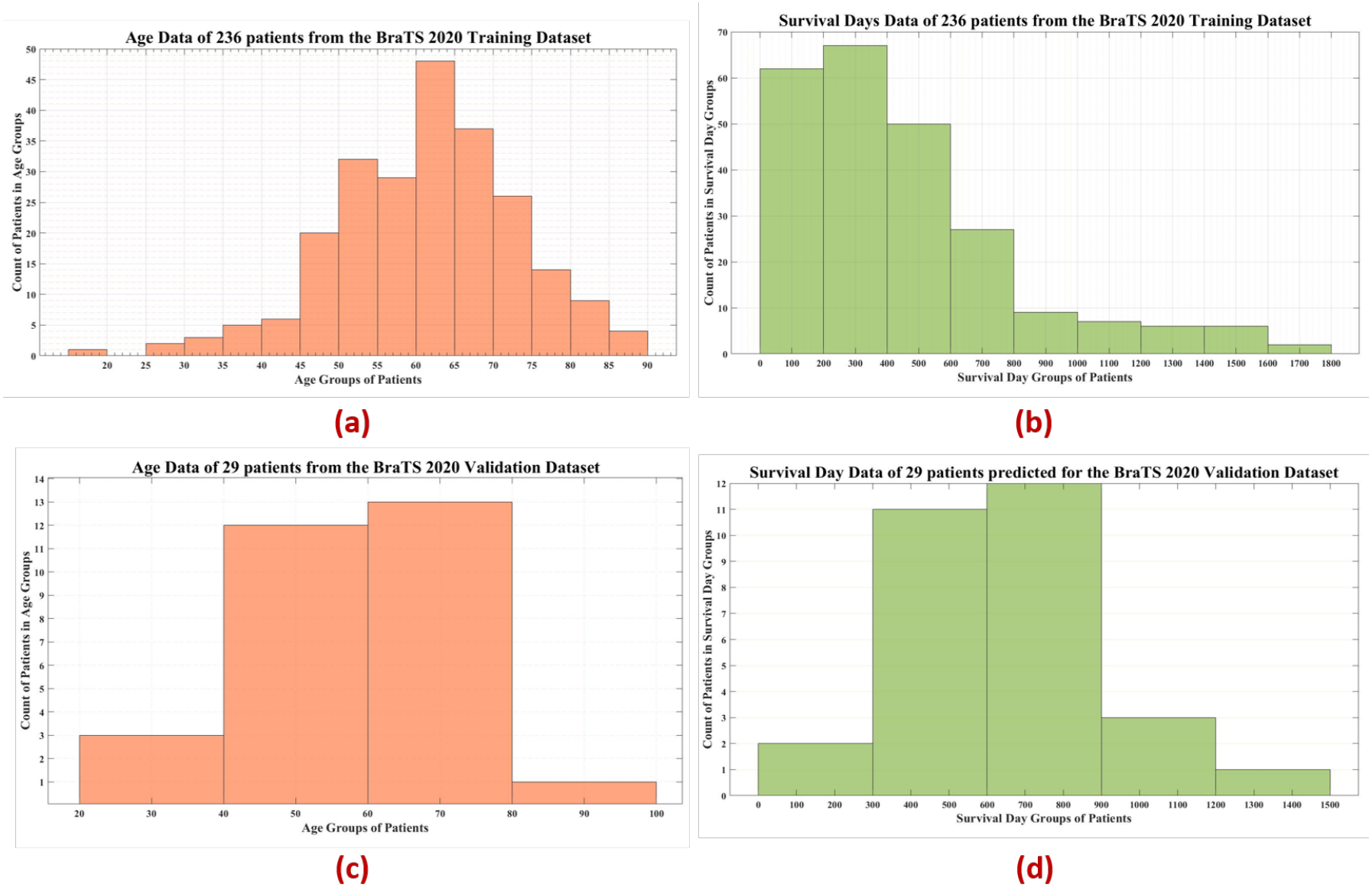


Figure 5.8: (a) Age Groups of 236 Patients from the BraTS 2020 Training Set, (b) Survival Day Groups of 236 Patients from the BraTS Training Set, (c) Age Groups of 29 Patients from the BraTS 2020 Validation Set, (d) Survival Day Groups of 29 Patients from the BraTS Validation Set

5.4.2 Strengths and Limitations of the Proposed Methodology

Our proposed methodology's detailed strengths and limitations are elaborated in this subsection. Although the accuracy of our model is not completely the best among existing techniques, it is evident from the results that our model generalized well during training and provided the best stdSE and Spearman coefficient with an accuracy of 0.517. It is expected to formulate the SP as a regression problem, and the accuracy is not the optimization goal. However, our model got significantly lower performance in terms of MSE and mSE. The most important feature of our proposed methodology is that it can extract more abundant features. At the same time, it greatly reduces the number of network parameters and the consumption of computing resources.

A potential limitation of this work is that it is not the best technique among 3D architectures. Most models predicted only between two classes: long or mid survivors. Only a few models could predict short survivors, which means that distinguishing between three classes is more difficult than separating between two groups. More patients in the group of short survivors are required. Using loss functions to favour short survivors' class predictions might help to alleviate this problem. We realize from the evaluated SP that the estimated values are in a narrow range of 145–1286, whereas, training dataset has a wide range of 5–1767. The numbers of evaluated subjects for OS are fairly lower than that of the segmentation task. The strategies cannot extract proper information without a large training dataset. Even though fewer parameters have been considered, the network could not generalize that well.

5.5 Inference and Findings

In this chapter, we proposed a 2D ResUNet-SEG plus 2D CNN-SP architectures based methodology to predict the survival time of GBM patients using 3D pre-operative raw brain MRI scans and other clinical data available in the BraTS 2020 dataset. The results showed that the performance of the survival results on the BraTS 2020 validation dataset could be improved by extracting deep features using 2D CNN-SP from raw and scans segmented using 2D ResUNet-SEG architecture and patient's age. Based on the results achieved, the proposed methodology is adequate among 2D architectures, although it still needs to be improved to generate higher performance metrics.

Chapter 6

Conclusions and Scope for further research

6.1 Conclusion

The availability of the benchmark BraTS dataset has significantly contributed in the field of computer-assisted medical image analysis for BTS and SP. This research work conducts a detailed literature survey with a focus on BTS and SP techniques. Early prediction of survival time of GBM is vital to plan proper treatment and patient care. Diagnostic modalities such as T1, T1CE, T2 and T2-FLAIR MRI scans play a crucial role in monitoring the progression of tumorous regions and the severity of the disease in GBM patients.

The *first technique* examined the effectiveness of using radiomic characteristics and an ML-based regressor to estimate the OS of GBM patients. Using pre-operative and 3D deep-supervised UNet model-based segmented MRI scans, multi-modal radiomic characteristics were extracted from the FLAIR channel. Ridge regressor was selected based on minimum MSE to make the OS prediction for the GBM patients. The classification accuracy demonstrated that the Ridge regressor trained on specific selected characteristics outperformed the conventional approaches. Multiple feature selection strategies helped in identifying more relevant characteristics based on their relationship with the goal survival variable and improved the capability of ML algorithms to classify GBM patients into three distinct survival groups, *viz.* short, mid, and long. On the BraTS validation set, the suggested technique recorded the highest accuracy and Spearman regression correlation. Further, radiomic characteristics that contributed the most to the NN's predicting abilities were discovered. A strong correlation existed between age and survival prognosis on the validation set.

Under the *second technique*, a pre-trained 2D ResNet feature extractor was used to extract the features from raw MRI scans; and MRI scans were segmented using 3D UNet model. The scans have been segmented using 3D UNet architecture. Along with deep features, handcrafted features have also been calculated from raw and segmented tumor subregions. PCA has been used as a feature reduction technique to curtail the dimensions of deep features extracted through a pre-trained 2D ResNet50 model. Automated regression-based ML has been used to make predictions for survival time of GBM patients using reduced DL feature-

based principal components, handcrafted tumor subregion-based volumes/ratios, and clinical data like patient's age. The top six best performing regression models obtained through the PyCaret Package have been blended to make final predictions on unseen data. The experimental results have demonstrated that the proposed methodology has performed better than many other existing 2D and 3D techniques proposed for SP of GBM patients. Features-based and blended decision-level fusion improved regressor's performance. Blended model is used for decision-level fusion because this model combines output prediction probabilities from many models to produce a single output prediction probability, which means that local and global knowledge is combined beneficially.

In *third technique*, the 2D ResUNet-SEG plus 2D CNN-SP architectures based methodology was proposed to predict the survival time of GBM patients using 3D pre-operative raw brain MRI scans and other clinical data available in the BraTS dataset. The results showed that the performance of the survival results on the BraTS validation dataset could be improved by extracting deep features using 2D CNN-SP from raw and scans segmented using 2D ResUNet-SEG architecture and patient's age. Based on the results achieved, the proposed methodology is adequate among 2D architectures, although it still needs to be improved to generate higher performance metrics. The CNN technique for SP get overfitted within a few epochs, regardless of the number of layers in the model; thus, it confirms the need for additional training data for the model to learn. A deep-learning method, on the other hand, yielded a higher Spearman coefficient and reduced the error in comparison to many existing 2D techniques. CNN models were compared to feature-based models that were handcrafted. Several CNN architectures with many model parameters outperformed biophysical characteristics and age-based models. Given the present state of research, it raises the concern that generic DL interpretation models for OS prediction have limited performance.

The effectiveness of 2D models is inferior to that of 3D models; however, combining both the designs can enhance the ensemble's overall performance. The 3D models provide depth information, while the 2D models correctly hyper-tune model parameters and improve the overall performance. Though 2D network techniques require less computer resources and computation time, the apparent benefits of 3D NN's and ensemble designs demonstrate a performance difference in tumour segmentation between both the 2D and the 3D network-based strategies.

6.2 Scope for further research

Various methods can be suggested to enhance the OS time prediction system. Data denoising methods for MRI pre-processing, raising the complexity of the DL model to learn more

semantic features for glioma segmentation, and employing multi-loss functions to handle class imbalance may be used to improve the segmentation performance. ML models trained on texture characteristics derived through a co-occurrence matrix may improve the SP classification results.

–

–

–

Future work can also include surface and volume attributes as a part of the embedded statistics of SP branches. Further investigation is required to bridge the difference between existing GBM radiomics research outside the BraTS, which promotes intensity-based features concerning normalisation methods to derive benefit from the additional intensity information in a multi-centre environment and the relevant details in pre-treatment MRI about the OS. Because the therapy procedure and some other critical factors, such as methylation status, are unknown in the BraTS dataset, it anticipates achieving a strict performance based on the limited available data. A subset of the BraTS dataset was collected before the latest imaging guidelines, and treatment practices, so studying how this affects the relevance of models built from BraTS data to actual clinical data will be quite beneficial. The researchers aim to fully evaluate the usage of spatial dropout layers, determining where such dropout layers should be put inside the network and with what proportion these should be used. They plan to evaluate the effectiveness of more complex techniques of test-time augmentation. They also want to investigate Generative Adversarial Networks (GANs) to augment data during training. In future, by integrating atlas-based techniques with DL to get a more precise brain parcellation, the accuracy of BTS and SP techniques can be further improved. The future research directions are also presented in Figure 6.1.

Furthermore, the study may also entail certain other prognostic factors in the prediction model, such as pathology evaluation information, molecular abnormality data, comorbidities, and Karnofsky's performance status. New network topologies and training methodologies can also be examined to improve feature extraction and selection to get better results. Further research can also be undertaken by exploring the performance of pre-trained models in the BTS task as well. The performance of a 3D feature extractor in place of a 2D pre-trained model can also be examined.

In the future, researchers can integrate atlas-based approaches with deep learning to get a more precise brain parcellation, which might help to further improve the accuracy of BTS and SP. The proposed methodology can be extended by embedding the extracted deep features extracted using 3D CNN architectures. It will also focus on identifying architectural

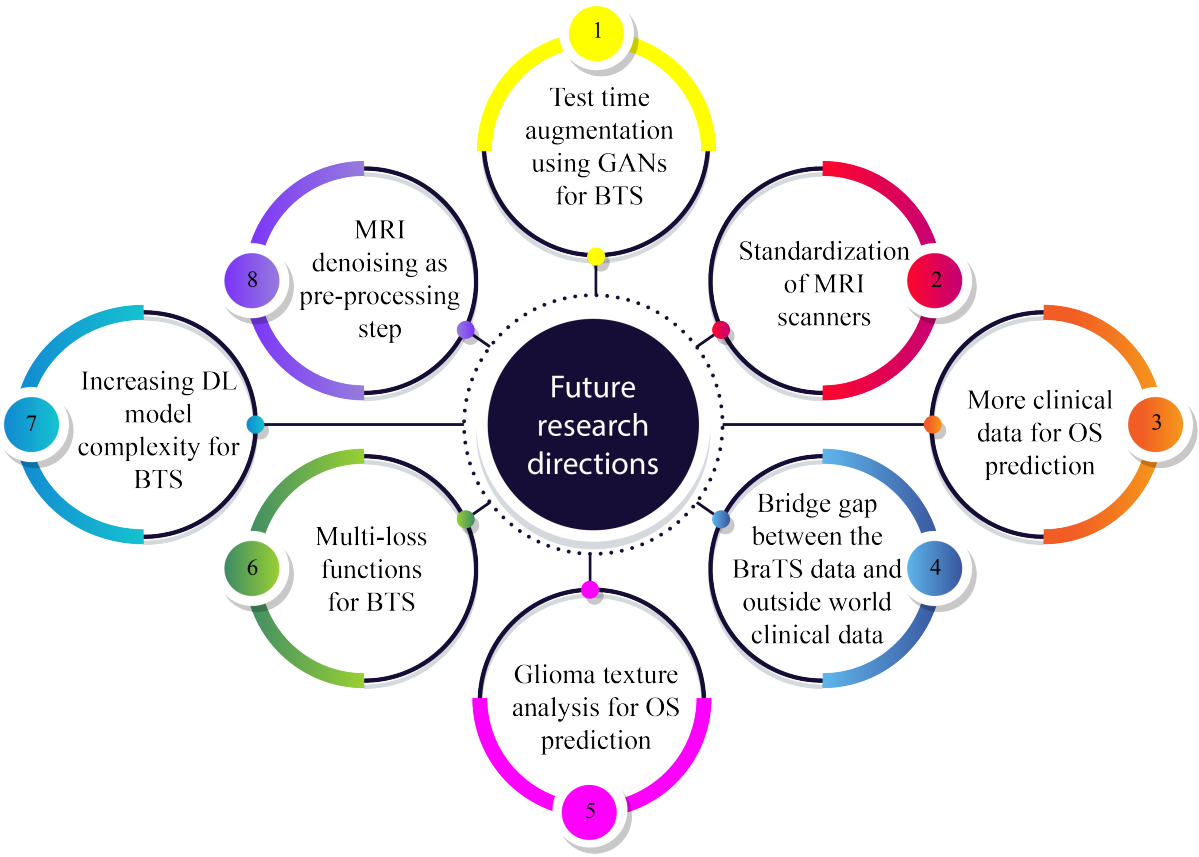


Figure 6.1: Future Research Directions for BTS and OS Prediction

changes that can improve the generalizability of the proposed model. GANs can be explored as a means of data augmentation during training and test-time augmentation methods such as scaling and rotating.

Indeed in its recent classification, WHO included molecular gliomas as well as the evaluation of Isocitrate dehydrogenase (IDH) status for diffusing astrocytomas and GBM, as well as 1p/19q co-deletion for oligodendroglioma. These molecular features are required for tumour diagnosis and survival prognosis, but they are not included in the BraTS dataset. It is possible to enhance OS correctness by including the 'gender' feature in the current list. A combination of classification and regression can help to solve the prediction issue. If the classification model predicted a different class with greater probability than the regression model, the projected survival time would be the output of that model in order to improve resilience. It is apparent that post-processing can substantially improve the model's efficiency on the BraTS dataset. The researchers also believe that using transfer learning to match volumetric data from 3D models with 2D slice-wise predictions is the best way forward. The results can also be improved further by combining 3D and 2D techniques.

References

[1] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[2] S. Rathore, H. Akbari, J. Doshi, G. Shukla, M. Rozycki, M. Bilello, R. A. Lustig, and C. A. Davatzikos, "Radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma: implications for personalized radiotherapy planning," *Journal of Medical Imaging*, vol. 5, no. 2, p. 021219, 2018.

[3] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, "Data augmentation for brain-tumor segmentation: a review," *Frontiers in computational neuroscience*, vol. 13, p. 83, 2019.

[4] L. Wang, S. Wang, R. Chen, X. Qu, Y. Chen, S. Huang, and C. Liu, "Nested dilation networks for brain tumor segmentation based on magnetic resonance imaging," *Frontiers in Neuroscience*, vol. 13, p. 285, 2019.

[5] S. Rajput and M. S. Raval, "A review on end-to-end methods for brain tumor segmentation and overall survival prediction," *arXiv preprint arXiv:2006.01632*, 2020.

[6] L. Guo, L. Zhao, Y. Wu, Y. Li, G. Xu, and Q. Yan, "Tumor detection in mr images using one-class immune feature weighted svms," *IEEE Transactions on Magnetics*, vol. 47, no. 10, pp. 3849–3852, 2011.

[7] L. M. DeAngelis, "Brain tumors," *New England journal of medicine*, vol. 344, no. 2, pp. 114–123, 2001.

[8] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, "Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2013–2017," *Neuro-oncology*, vol. 22, no. Supplement_1, pp. iv1–iv96, 2020.

[9] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on mri brain tumor segmentation," *Magnetic resonance imaging*, vol. 31, no. 8, pp. 1426–1438, 2013.

[10] G. L. Qurat-Ul-Ain, S. B. Kazmi, M. A. Jaffar, and A. M. Mirza, "Classification and segmentation of brain tumor using texture analysis," *Recent advances in artificial intelligence, knowledge engineering and data bases*, pp. 147–155, 2010.

[11] F. Hanif, K. Muzaffar, K. Perveen, S. M. Malhi, and S. U. Simjee, "Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment," *Asian Pacific journal of cancer prevention: APJCP*, vol. 18, no. 1, p. 3, 2017.

[12] G. Mohan and M. M. Subashini, "Mri based medical image analysis: Survey on brain tumor grade classification," *Biomedical Signal Processing and Control*, vol. 39, pp. 139–161, 2018.

[13] S. Aswathy, G. G. Devadhas, and S. Kumar, "Mri brain tumor segmentation using genetic algorithm with svm classifier," *Journal of Electronics and Communication Engineering, e-ISSN*, pp. 2278–2834, 2017.

[14] K. Sridhar, S. Baskar, P. M. Shakeel, and V. Dhulipala, "Developing brain abnormality recognize system using multi-objective pattern producing neural network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 8, pp. 3287–3295, 2019.

[15] J. Amin, M. Sharif, M. Yasmin, H. Ali, and S. L. Fernandes, "A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions," *Journal of Computational Science*, vol. 19, pp. 153–164, 2017.

[16] G. Mohan and M. M. Subashini, "Medical imaging with intelligent systems: a review," *Deep learning and parallel computing environment for bioengineering systems*, pp. 53–73, 2019.

[17] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of mri-based medical image analysis for brain tumor studies," *Physics in Medicine & Biology*, vol. 58, no. 13, p. R97, 2013.

[18] Q. T. Ostrom, H. Gittleman, P. Liao, T. Vecchione-Koval, Y. Wolinsky, C. Kruchko, and J. S. Barnholtz-Sloan, "Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2010–2014," *Neuro-oncology*, vol. 19, no. suppl_5, pp. v1–v88, 2017.

[19] J. Liu, M. Li, J. Wang, F. Wu, T. Liu, and Y. Pan, "A survey of mri-based brain tumor segmentation methods," *Tsinghua science and technology*, vol. 19, no. 6, pp. 578–595, 2014.

[20] C. Dupont, N. Betrouni, N. Reyns, and M. Vermandel, "On image segmentation methods applied to glioblastoma: state of art and new trends," *Innovation and Research in BioMedical engineering*, vol. 37, no. 3, pp. 131–143, 2016.

[21] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[22] R. L. Lym, Q. T. Ostrom, C. Kruchko, M. Couce, D. J. Brat, D. N. Louis, and J. S. Barnholtz-Sloan, "Completeness and concordancy of who grade assignment for brain and central nervous system tumors in the united states, 2004–2011," *Journal of neuro-oncology*, vol. 123, no. 1, pp. 43–51, 2015.

[23] Z. A. Shboul, M. Alam, L. Vidyaratne, L. Pei, M. I. Elbakary, and K. M. Iftekharuddin, "Feature-guided deep radiomics for glioblastoma patient survival prediction," *Fron-*

tiers in neuroscience, vol. 13, p. 966, 2019.

[24] Q. T. Ostrom, H. Gittleman, J. Xu, C. Kromer, Y. Wolinsky, C. Kruchko, and J. S. Barnholtz-Sloan, "Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2009–2013," *Neuro-oncology*, vol. 18, no. suppl_5, pp. v1–v75, 2016.

[25] J. Chen, R. M. McKay, and L. F. Parada, "Malignant glioma: lessons from genomics, mouse models, and stem cells," *Cell*, vol. 149, no. 1, pp. 36–47, 2012.

[26] Y. Wang, Y. Zhang, F. Hou, Y. Liu, J. Tian, C. Zhong, Y. Zhang, and Z. He, "Modality-pairing learning for brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, pp. 230–240, Springer, 2020.

[27] C. Santiago, J. C. Nascimento, and J. S. Marques, "Fast segmentation of the left ventricle in cardiac mri using dynamic programming," *Computer methods and programs in biomedicine*, vol. 154, pp. 9–23, 2018.

[28] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.

[29] P. M. a. Rosenschöld, S. Engelholm, L. Ohlhues, I. Law, I. Vogelius, and S. A. Engelholm, "Photon and proton therapy planning comparison for malignant glioma based on ct, fdg-pet, dti-mri and fiber tracking," *Acta Oncologica*, vol. 50, no. 6, pp. 777–783, 2011.

[30] R. Kumari, "Svm classification an approach on detecting abnormality in brain mri images," *International Journal of Engineering Research and Applications*, vol. 3, no. 4, pp. 1686–1690, 2013.

[31] N. Porz, S. Bauer, A. Pica, P. Schucht, J. Beck, R. K. Verma, J. Slotboom, M. Reyes, and R. Wiest, "Multi-modal glioblastoma segmentation: man versus machine," *PloS one*, vol. 9, no. 5, p. e96873, 2014.

[32] N. Upadhyay and A. Waldman, "Conventional mri evaluation of gliomas," *The British journal of radiology*, vol. 84, no. special_issue_2, pp. S107–S111, 2011.

[33] T. Magadza and S. Viriri, "Deep learning for brain tumor segmentation: a survey of state-of-the-art," *Journal of Imaging*, vol. 7, no. 2, p. 19, 2021.

[34] B. H. Menze, K. v. Leemput, D. Lashkari, M.-A. Weber, N. Ayache, and P. Golland, "A generative model for brain tumor segmentation in multi-modal images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 151–159, Springer, 2010.

[35] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig, "A brain tumor segmentation framework based on outlier detection," *Medical image analysis*, vol. 8, no. 3, pp. 275–283, 2004.

[36] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. M. Thomas, T. Das, R. Jena, and S. J. Price, "Decision forests for tissue-specific segmentation of

high-grade gliomas in multi-channel mr," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 369–376, Springer, 2012.

[37] Y. Wu, W. Yang, J. Jiang, S. Li, Q. Feng, and W. Chen, "Semi-automatic segmentation of brain tumors using population and individual information," *Journal of digital imaging*, vol. 26, no. 4, pp. 786–796, 2013.

[38] T. Yang, J. Song, and L. Li, "A deep learning model integrating sk-tpcnn and random forests for brain tumor segmentation in mri," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 3, pp. 613–623, 2019.

[39] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4itk: improved n3 bias correction," *IEEE transactions on medical imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[40] M. Goetz, C. Weber, F. Binczyk, J. Polanska, R. Tarnawski, B. Bobek-Billewicz, U. Koethe, J. Kleesiek, B. Stieltjes, and K. H. Maier-Hein, "Dalsa: domain adaptation for supervised learning from sparsely annotated mr images," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 184–196, 2015.

[41] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, and X. Ye, "Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in flair mri," *International journal of computer assisted radiology and surgery*, vol. 12, no. 2, pp. 183–203, 2017.

[42] Y. Xue, S. Chen, J. Qin, Y. Liu, B. Huang, and H. Chen, "Application of deep learning in automated analysis of molecular images in cancer: a survey," *Contrast media & molecular imaging*, vol. 2017, 2017.

[43] L. Weninger, C. Haarburger, and D. Merhof, "Robustness of radiomics for survival prediction of brain tumor patients depending on resection status," *Frontiers in computational neuroscience*, vol. 13, p. 73, 2019.

[44] O. Maier, M. Wilms, and H. Handels, "Image features for brain lesion segmentation using random forests," in *BrainLes 2015*, pp. 119–130, Springer, 2015.

[45] T. Singh, N. M. Garg, and S. R. Iyengar, "Nondestructive identification of barley seeds variety using near-infrared hyperspectral imaging coupled with convolutional neural network," *Journal of Food Process Engineering*, vol. 44, no. 10, p. e13821, 2021.

[46] S. C. Raikwar, S. Tapaswi, and S. Chakraborty, "Bounding function for fast computation of transmission in single image dehazing," *Multimedia Tools and Applications*, vol. 81, no. 4, pp. 5349–5372, 2022.

[47] N. Rani and D. K. Gupta, "Performance based analysis of classification techniques in breast cancer screening–a review," *International Journal of Research in Engineering, Science and Management*, vol. 5, no. 2, pp. 130–135, 2022.

[48] B. Singh, P. K. Sharma, S. A. Huddedar, A. Sur, and P. Mitra, "Steggan: hiding image

within image using conditional generative adversarial networks," *Multimedia Tools and Applications*, pp. 1–23, 2022.

[49] K. Wang, Z.-J. Xu, Y. Gong, and K.-L. Du, "Mechanical fault prognosis through spectral analysis of vibration signals," *Algorithms*, vol. 15, no. 3, p. 94, 2022.

[50] M. M. Richter, S. Paul, V. Këpuska, and M. Silaghi, "Feature extraction," in *Signal Processing and Machine Learning with Applications*, pp. 221–250, Springer, 2022.

[51] R. Cantini, F. Marozzo, A. Orsino, D. Talia, P. Trunfio, R. M. Badia, J. Ejarque, and F. Vazquez, "Block size estimation for data partitioning in hpc applications using machine learning techniques," *arXiv preprint arXiv:2211.10819*, 2022.

[52] K. Ma, Z. Wang, W. Tian, S. Chen, and L. Hanzo, "Deep learning for mmwave beam-management: State-of-the-art, opportunities and challenges," *IEEE Wireless Communications*, 2022.

[53] H. Zhang and L. Hanzo, "Federated learning assisted multi-uav networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14104–14109, 2020.

[54] L. Belcastro, F. Marozzo, D. Talia, P. Trunfio, F. Branda, T. Palpanas, and M. Imran, "Using social media for sub-event detection during disasters," *Journal of big data*, vol. 8, no. 1, pp. 1–22, 2021.

[55] M. M. Richter, S. Paul, V. Këpuska, and M. Silaghi, "Signal processing and machine learning with applications," 2022.

[56] K.-L. Du, C.-S. Leung, W.-H. Mow, and M. Swamy, "Perceptron: Learning, generalization, model selection, fault tolerance, and role in the deep learning era," *Mathematics*, vol. 10, no. 24, p. 4730, 2022.

[57] B. Singh, A. Sur, and P. Mitra, "Steganalysis of digital images using deep fractal network," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 3, pp. 599–606, 2021.

[58] A. Saxena, D. K. Gupta, and S. Singh, "An animal detection and collision avoidance system using deep learning," in *Advances in Communication and Computational Technology*, pp. 1069–1084, Springer, 2021.

[59] S. C. Raikwar and S. Tapaswi, "Lower bound on transmission using non-linear bounding function in single image dehazing," *IEEE Transactions on Image Processing*, vol. 29, pp. 4832–4847, 2020.

[60] T. Singh, N. M. Garg, and S. Iyengar, "Identification of harvesting year of barley seeds using near-infrared hyperspectral imaging combined with convolutional neural network," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 3–8, Springer, 2021.

[61] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *Neu-*

roImage, vol. 108, pp. 214–224, 2015.

[62] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[63] M. Balwant, "A review on convolutional neural networks for brain tumor segmentation: Methods, datasets, libraries, and future directions," *Innovation and Research in BioMedical engineering*, 2022.

[64] G. Kaur, P. S. Rana, and V. Arora, "State-of-the-art techniques using pre-operative brain mri scans for survival prediction of glioblastoma multiforme patients and future research directions," *Clinical and Translational Imaging*, pp. 1–35, 2022.

[65] R. R. Agravat and M. S. Raval, "A survey and analysis on automated glioma brain tumor segmentation and overall patient survival prediction," *Archives of Computational Methods in Engineering*, vol. 28, no. 5, pp. 4117–4152, 2021.

[66] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, no. 1, pp. 1–13, 2017.

[67] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, *et al.*, "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*, 2018.

[68] S. B. H. Akbari et al., "Segmentation Labels for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive." `https://wiki.cancerimagingarchive.net/x/KoZyAQ`, 2017.

[69] S. B. H. Akbari et al., "Segmentation Labels for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive." `https://wiki.cancerimagingarchive.net/x/LIZyAQ`, 2017.

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

[71] S. R. González, I. Zemmoura, and C. Tauber, "3d brain tumor segmentation and survival prediction using ensembles of convolutional neural networks," in *International MICCAI Brainlesion Workshop*, pp. 241–254, Springer, 2020.

[72] B. Parmar and M. Parikh, "brain tumor segmentation and survival prediction using patch based modified 3d u-net," in *International MICCAI Brainlesion Workshop*, pp. 398–409, Springer, 2020.

[73] D. Carmo, L. Rittner, and R. Lotufo, "Multiattunet: brain tumor segmentation and

survival multitasking," in *International MICCAI Brainlesion Workshop*, pp. 424–434, Springer, 2020.

[74] V. K. Anand, S. Grampurohit, P. Aurangabadkar, A. Kori, M. Khened, R. S. Bhat, and G. Krishnamurthi, "Brain tumor segmentation and survival prediction using automatic hard mining in 3d cnn architecture," in *International MICCAI Brainlesion Workshop*, pp. 310–319, Springer, 2020.

[75] R. R. Agravat and M. S. Raval, "3d semantic segmentation of brain tumor for overall survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 215–227, Springer, 2020.

[76] R. Miron, R. Albert, and M. Breaban, "A two-stage atrous convolution neural network for brain tumor segmentation and survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 290–299, Springer, 2020.

[77] J. Patel, K. Chang, K. Hoebel, M. Gidwani, N. Arun, S. Gupta, M. Aggarwal, P. Singh, B. R. Rosen, E. R. Gerstner, *et al.*, "Segmentation, survival prediction, and uncertainty estimation of gliomas from multimodal 3d mri using selective kernel networks," in *International MICCAI Brainlesion Workshop*, pp. 228–240, Springer, 2020.

[78] C. Dai, S. Wang, H. Raynaud, Y. Mo, E. Angelini, Y. Guo, and W. Bai, "Self-training for brain tumour segmentation with uncertainty estimation and biophysics-guided survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 514–523, Springer, 2020.

[79] R. McKinley, M. Rebsamen, K. Daetwyler, R. Meier, P. Radojewski, and R. Wiest, "Uncertainty-driven refinement of tumor-core segmentation using 3d-to-2d networks with label uncertainty," in *International MICCAI Brainlesion Workshop*, pp. 401–411, Springer, 2020.

[80] G. Zhao, B. Jiang, J. Zhang, and Y. Xia, "Segmentation then prediction: a multi-task solution to brain tumor segmentation and survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 492–502, Springer, 2020.

[81] L. Pei, A. Murat, and R. Colen, "Multimodal brain tumor segmentation and survival prediction using a 3d self-ensemble resunet," in *International MICCAI Brainlesion Workshop*, pp. 367–375, Springer, 2020.

[82] M. Soltaninejad, T. Pridmore, and M. Pound, "Efficient mri brain tumor segmentation using multi-resolution encoder-decoder networks," in *International MICCAI Brainlesion Workshop*, pp. 30–39, Springer, 2020.

[83] M. J. Ali, M. T. Akram, H. Saleem, B. Raza, and A. R. Shahid, "Glioma segmentation using ensemble of 2d/3d u-nets and survival prediction using multiple features fusion," in *International MICCAI Brainlesion Workshop*, pp. 189–199, Springer, 2020.

[84] A. S. Akbar, C. Fatichah, and N. Suciati, "Modified mobilenet for patient survival pre-

diction," in *International MICCAI Brainlesion Workshop*, pp. 374–387, Springer, 2020.

[85] E. Pang, W. Shi, X. Li, and Q. Wu, "Glioma segmentation using encoder-decoder network and survival prediction based on cox analysis," in *International MICCAI Brainlesion Workshop*, pp. 318–326, Springer, 2020.

[86] C. Russo, S. Liu, and A. D. Ieva, "Impact of spherical coordinates transformation pre-processing in deep convolution neural networks for brain tumor segmentation and survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 295–306, Springer, 2020.

[87] I. S. Han, "Multimodal brain image analysis and survival prediction using neuromorphic attention-based neural networks," in *International MICCAI Brainlesion Workshop*, pp. 194–206, Springer, 2020.

[88] W.-S. Han and I. S. Han, "Multimodal brain image segmentation and analysis with neuromorphic attention-based learning," in *International MICCAI Brainlesion Workshop*, pp. 14–26, Springer, 2019.

[89] J. Marti Asenjo and A. Martinez-Larraz Solís, "Mri brain tumor segmentation using a 2d-3d u-net ensemble," in *International MICCAI Brainlesion Workshop*, pp. 354–366, Springer, 2020.

[90] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[91] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "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*, pp. 240–248, Springer, 2017.

[92] C. Chen, X. Liu, M. Ding, J. Zheng, and J. Li, "3d dilated multi-fiber network for real-time brain tumor segmentation in mri," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 184–192, Springer, 2019.

[93] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge," in *International MICCAI Brainlesion Workshop*, pp. 287–297, Springer, 2017.

[94] Z. Jiang, C. Ding, M. Liu, and D. Tao, "Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task," in *International MICCAI brainlesion workshop*, pp. 231–241, Springer, 2019.

[95] A. Myronenko, "3d mri brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*, pp. 311–320, Springer, 2018.

[96] C. Zhou, C. Ding, Z. Lu, X. Wang, and D. Tao, "One-pass multi-task convolutional neural networks for efficient brain tumor segmentation," in *International Conference on*

Medical Image Computing and Computer-Assisted Intervention, pp. 637–645, Springer, 2018.

[97] R. R. Agravat and M. S. Raval, "Deep learning for automated brain tumor segmentation in mri images," in *Soft computing based medical image analysis*, pp. 183–201, Elsevier, 2018.

[98] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, pp. 565–571, IEEE, 2016.

[99] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[100] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[101] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "Densenet: Implementing efficient convnet descriptor pyramids," *arXiv preprint arXiv:1404.1869*, 2014.

[102] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "No new-net," in *International MICCAI Brainlesion Workshop*, pp. 234–244, Springer, 2018.

[103] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, *et al.*, "Ensembles of multiple models and architectures for robust brain tumour segmentation," in *International MICCAI brainlesion workshop*, pp. 450–462, Springer, 2017.

[104] R. McKinley, M. Rebsamen, R. Meier, and R. Wiest, "Triplanar ensemble of 3d-to-2d cnns with label-uncertainty for brain tumor segmentation," in *International MICCAI brainlesion workshop*, pp. 379–387, Springer, 2019.

[105] Y.-X. Zhao, Y.-M. Zhang, and C.-L. Liu, "Bag of tricks for 3d mri brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, pp. 210–220, Springer, 2019.

[106] A. Myronenko and A. Hatamizadeh, "Robust semantic segmentation of brain tumor regions from 3d mris," in *International MICCAI Brainlesion Workshop*, pp. 82–89, Springer, 2019.

[107] L. Chato, P. Kachroo, and S. Latifi, "An automatic overall survival time prediction system for glioma brain tumor patients based on volumetric and shape features," in *International MICCAI Brainlesion Workshop*, pp. 352–365, Springer, 2020.

[108] Z. Dai and R. Heckel, "Channel normalization in convolutional neural network avoids vanishing gradients," *arXiv preprint arXiv:1907.09539*, 2019.

[109] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by

reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, PMLR, 2015.

[110] M. Khened, A. Kori, H. Rajkumar, G. Krishnamurthi, and B. Srinivasan, "A generalized deep learning framework for whole-slide image segmentation and analysis," *Scientific reports*, vol. 11, no. 1, pp. 1–14, 2021.

[111] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[112] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

[113] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[114] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, 2017.

[115] X. Kong, G. Sun, Q. Wu, J. Liu, and F. Lin, "Hybrid pyramid u-net model for brain tumor segmentation," in *International conference on intelligent information processing*, pp. 346–355, Springer, 2018.

[116] Y. Suter, U. Knecht, R. Wiest, and M. Reyes, "Overall survival prediction for glioblastoma on pre-treatment mri using robust radiomics and priors," in *International MICCAI Brainlesion Workshop*, pp. 307–317, Springer, 2020.

[117] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

[118] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, *et al.*, "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *arXiv preprint arXiv:1809.10486*, 2018.

[119] Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.

[120] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 510–519, 2019.

[121] B. Kayalibay, G. Jensen, and P. van der Smagt, "Cnn-based segmentation of medical imaging data," *arXiv preprint arXiv:1701.03056*, 2017.

[122] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.

[123] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[124] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018.

[125] Y.-Z. Feng, S. Liu, Z.-Y. Cheng, J. C. Quiroz, D. Rezazadegan, P.-K. Chen, Q.-T. Lin, L. Qian, X.-F. Liu, S. Berkovsky, *et al.*, "Severity assessment and progression prediction of covid-19 patients based on the lesionencoder framework and chest ct," *Information*, vol. 12, no. 11, p. 471, 2021.

[126] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

[127] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941*, 2017.

[128] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.

[129] D. Carmo, B. Silva, C. Yasuda, L. Rittner, R. Lotufo, A. D. N. Initiative, *et al.*, "Hippocampus segmentation on epilepsy and alzheimer's disease studies with multiple convolutional neural networks," *Heliyon*, vol. 7, no. 2, p. e06226, 2021.

[130] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European conference on computer vision*, pp. 483–499, Springer, 2016.

[131] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[132] A. Rafi, J. Ali, T. Akram, K. Fiaz, A. Raza Shahid, B. Raza, and T. Mustafa Madni, "U-net based glioblastoma segmentation with patient's overall survival prediction," in *International Symposium on Intelligent Computing Systems*, pp. 22–32, Springer, 2020.

[133] Y. Zhang, J. Wu, W. Huang, Y. Chen, E. X. Wu, and X. Tang, "Utility of brain parcellation in enhancing brain tumor segmentation and survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 391–400, Springer, 2020.

[134] Z. Liu, L. Tong, L. Chen, Z. Jiang, F. Zhou, Q. Zhang, X. Zhang, Y. Jin, and H. Zhou, "Deep learning based brain tumor segmentation: a survey," *Complex & Intelligent Systems*, pp. 1–26, 2022.

[135] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and*

pattern recognition, pp. 4700–4708, 2017.

[136] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, *et al.*, "Api design for machine learning software: experiences from the scikit-learn project," *arXiv preprint arXiv:1309.0238*, 2013.

[137] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[138] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[139] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[140] K. Swanson, R. Rostomily, and E. Alvord, "A mathematical modelling tool for predicting survival of individual patients following resection of glioblastoma: a proof of principle," *British journal of cancer*, vol. 98, no. 1, pp. 113–119, 2008.

[141] K. R. Swanson, C. Bridge, J. Murray, and E. C. Alvord Jr, "Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion," *Journal of the neurological sciences*, vol. 216, no. 1, pp. 1–10, 2003.

[142] W.-S. Han and I.-S. Han, "All weather human detection using neuromorphic visual processing," in *Intelligent Systems for Science and Information*, pp. 25–44, Springer, 2014.

[143] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.

[144] M. Reuter, H. D. Rosas, and B. Fischl, "Highly accurate inverse consistent registration: a robust approach," *Neuroimage*, vol. 53, no. 4, pp. 1181–1196, 2010.

[145] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical image analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[146] G. N. Ramien, P. F. Jaeger, S. A. Kohl, and K. H. Maier-Hein, "Reg r-cnn: lesion detection and grading under noisy labels," in *Uncertainty for safe utilization of machine learning in medical imaging and clinical image-based procedures*, pp. 33–41, Springer, 2019.

[147] C. Li, S. Wang, P. Liu, T. Torheim, N. R. Boonzaier, B. R. van Dijken, C.-B. Schönlieb, F. Markowetz, and S. J. Price, "Decoding the interdependence of multiparametric magnetic resonance imaging to reveal patient subgroups correlated with survivals," *Neo-*

plasia, vol. 21, no. 5, pp. 442–449, 2019.

[148] A. L. Baldock, S. Ahn, R. Rockne, S. Johnston, M. Neal, D. Corwin, K. Clark-Swanson, G. Sterin, A. D. Trister, H. Malone, *et al.*, "Patient-specific metrics of invasiveness reveal significant prognostic benefit of resection in a predictable subset of gliomas," *PLoS One*, vol. 9, no. 10, p. e99057, 2014.

[149] C. Li, S. Wang, J.-L. Yan, T. Torheim, N. R. Boonzaier, R. Sinha, T. Matys, F. Markowetz, and S. J. Price, "Characterizing tumor invasiveness of glioblastoma using multiparametric magnetic resonance imaging," *Journal of neurosurgery*, vol. 132, no. 5, pp. 1465–1472, 2019.

[150] S. Wang, C. Dai, Y. Mo, E. Angelini, Y. Guo, and W. Bai, "Automatic brain tumour segmentation and biophysics-guided survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 61–72, Springer, 2019.

[151] X. Feng, Q. Dou, N. Tustison, and C. Meyer, "Brain tumor segmentation with uncertainty estimation and overall survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 304–314, Springer, 2019.

[152] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, no. 1, pp. 3–42, 2006.

[153] C. E. McCulloch, "Generalized linear models," *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1320–1324, 2000.

[154] M. Górriz, J. Antony, K. McGuinness, X. Giró-i Nieto, and N. E. O'Connor, "Assessing knee oa severity with cnn attention-based end-to-end architectures," in *International conference on medical imaging with deep learning*, pp. 197–214, PMLR, 2019.

[155] U. Baid, S. Talbar, S. Rane, S. Gupta, M. H. Thakur, A. Moiyadi, S. Thakur, and A. Mahajan, "Deep learning radiomics algorithm for gliomas (drag) model: a novel approach using 3d unet based deep convolutional neural network for predicting survival in gliomas," in *International MICCAI Brainlesion Workshop*, pp. 369–379, Springer, 2018.

[156] L. Weninger, O. Rippel, S. Koppers, and D. Merhof, "Segmentation of brain tumors and patient survival prediction: Methods for the brats 2018 challenge," in *International MICCAI brainlesion workshop*, pp. 3–12, Springer, 2018.

[157] Y. Suter, A. Jungo, M. Rebsamen, U. Knecht, E. Herrmann, R. Wiest, and M. Reyes, "Deep learning versus classical regression for brain tumor patient survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 429–440, Springer, 2018.

[158] X. Guo, C. Yang, P. L. Lam, P. Y. Woo, and Y. Yuan, "Domain knowledge based brain tumor segmentation and overall survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 285–295, Springer, 2019.

[159] F. Wang, R. Jiang, L. Zheng, C. Meng, and B. Biswal, "3d u-net based brain tumor segmentation and survival days prediction," in *International MICCAI Brainlesion Work-*

shop, pp. 131–141, Springer, 2019.

[160] M. Islam, V. Jose, and H. Ren, "Glioma prognosis: segmentation of the tumor and survival prediction using shape, geometric and clinical information," in *International MICCAI Brainlesion Workshop*, pp. 142–153, Springer, 2018.

[161] R. R. Agravat and M. S. Raval, "Brain tumor segmentation and survival prediction," in *International MICCAI Brainlesion Workshop*, pp. 338–348, Springer, 2019.

[162] L. Sun, S. Zhang, H. Chen, and L. Luo, "Brain tumor segmentation and survival prediction using multimodal mri scans with deep learning," *Frontiers in neuroscience*, vol. 13, p. 810, 2019.

[163] M. Rebsamen, U. Knecht, M. Reyes, R. Wiest, R. Meier, and R. McKinley, "Divide and conquer: stratifying training data by tumor grade improves deep learning-based brain tumor segmentation," *Frontiers in neuroscience*, vol. 13, p. 1182, 2019.

[164] S. Banerjee, S. Mitra, and B. U. Shankar, "Multi-planar spatial-convnet for segmentation and survival prediction in brain cancer," in *International MICCAI Brainlesion Workshop*, pp. 94–104, Springer, 2018.

[165] K. Aftab, F. B. Aamir, S. Mallick, F. Mubarak, W. B. Pope, T. Mikkelsen, J. P. Rock, and S. A. Enam, "Radiomics for precision medicine in glioblastoma," *Journal of neuro-oncology*, pp. 1–15, 2022.

[166] J. Fu, K. Singhrao, X. Zhong, Y. Gao, S. X. Qi, Y. Yang, D. Ruan, and J. H. Lewis, "An automatic deep learning–based workflow for glioblastoma survival prediction using preoperative multimodal mr images: a feasibility study," *Advances in radiation oncology*, vol. 6, no. 5, p. 100746, 2021.

[167] S. Cepeda, A. Pérez-Nuñez, S. García-García, D. García-Pérez, I. Arrese, L. Jiménez-Roldán, M. García-Galindo, P. González, M. Velasco-Casares, T. Zamora, *et al.*, "Predicting short-term survival after gross total or near total resection in glioblastomas by machine learning-based radiomic analysis of preoperative mri," *Cancers*, vol. 13, no. 20, p. 5047, 2021.

[168] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Icml*, 2010.

[169] H. K. Bhuyan and V. Ravi, "Analysis of subfeature for classification in data mining," *IEEE Transactions on Engineering Management*, 2021.

[170] H. K. Bhuyan and C. M. Reddy, "Sub-feature selection for novel classification," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 477–482, IEEE, 2018.

[171] A. Sharma and P. K. Mishra, "State-of-the-art in performance metrics and future directions for data science algorithms," *Journal of Scientific Research*, vol. 64, no. 2, pp. 221–238, 2020.

[172] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," *arXiv preprint arXiv:1908.03265*, 2019.

[173] E. Pang, W. Shi, X. Li, and Q. Wu, "Glioma segmentation using encoder-decoder network and survival prediction based on cox analysis," in *International MICCAI Brainlesion Workshop*, pp. 318–326, Springer, 2020.

[174] M. J. Ali, M. T. Akram, H. Saleem, B. Raza, and A. R. Shahid, "Glioma segmentation using ensemble of 2d/3d u-nets and survival prediction using multiple features fusion," in *International MICCAI Brainlesion Workshop*, pp. 189–199, Springer, 2020.

[175] C. Russo, S. Liu, and A. Di Ieva, "Spherical coordinates transformation pre-processing in deep convolution neural networks for brain tumor segmentation in mri," *Medical & Biological Engineering & Computing*, vol. 60, no. 1, pp. 121–134, 2022.

[176] Y. Suter, U. Knecht, M. Alão, W. Valenzuela, E. Hewer, P. Schucht, R. Wiest, and M. Reyes, "Radiomics for glioblastoma survival analysis in pre-operative mri: exploring feature robustness, class boundaries, and machine learning techniques," *Cancer Imaging*, vol. 20, no. 1, pp. 1–13, 2020.

[177] N. Czarnek, K. Clark, K. B. Peters, and M. A. Mazurowski, "Algorithmic three-dimensional analysis of tumor shape in mri improves prognosis of survival in glioblastoma: a multi-institutional study," *Journal of neuro-oncology*, vol. 132, no. 1, pp. 55–62, 2017.

[178] Y. Choi, Y. Nam, J. Jang, N.-Y. Shin, Y. S. Lee, K.-J. Ahn, B.-s. Kim, J.-S. Park, S.-s. Jeon, and Y. G. Hong, "Radiomics may increase the prognostic value for survival in glioblastoma patients when combined with conventional clinical and genetic prognostic models," *European Radiology*, vol. 31, no. 4, pp. 2084–2093, 2021.

[179] P. C. De Witt Hamer, V. K. Ho, A. H. Zwinderman, L. Ackermans, H. Ardon, S. Boomstra, W. Bouwknegt, W. A. van den Brink, C. M. Dirven, N. A. van der Gaag, *et al.*, "Between-hospital variation in mortality and survival after glioblastoma surgery in the dutch quality registry for neuro surgery," *Journal of neuro-oncology*, vol. 144, no. 2, pp. 313–323, 2019.

[180] C. Henker, M. C. Hiepel, T. Kriesen, M. Scherer, Ä. Glass, C. Herold-Mende, M. Bendszus, S. Langner, M.-A. Weber, B. Schneider, *et al.*, "Volumetric assessment of glioblastoma and its predictive value for survival," *Acta Neurochirurgica*, vol. 161, no. 8, pp. 1723–1732, 2019.

[181] U. Nestler, K. Lutz, U. Pichlmeier, W. Stummer, K. Franz, H. Reulen, and A. Bink, "Anatomic features of glioblastoma and their potential impact on survival," *Acta neurochirurgica*, vol. 157, no. 2, pp. 179–186, 2015.

[182] N. Kudulaiti, Z. Zhou, C. Luo, J. Zhang, F. Zhu, and J. Wu, "A nomogram for individualized prediction of overall survival in patients with newly diagnosed glioblastoma: a

real-world retrospective cohort study," *BMC surgery*, vol. 21, no. 1, pp. 1–11, 2021.

[183] P. Whitmire, C. R. Rickertsen, A. Hawkins-Daarud, E. Carrasco, J. Lorence, G. De Leon, L. Curtin, S. Bayless, K. Clark-Swanson, N. C. Peeri, *et al.*, "Sex-specific impact of patterns of imageable tumor growth on survival of primary glioblastoma patients," *BMC cancer*, vol. 20, no. 1, pp. 1–10, 2020.

[184] F. Sfifou, E. M. Hakkou, E. A. Bouaiti, M. Slaoui, H. Errihani, A. Al Bouzidi, R. Abouqal, A. El Ouahabi, and N. Cherradi, "Correlation of immunohistochemical expression of hif-1alpha and idh1 with clinicopathological and therapeutic data of moroccan glioblastoma and survival analysis," *Annals of Medicine and Surgery*, vol. 69, p. 102731, 2021.

[185] I. E. Bennett, K. M. Field, C. M. Hovens, B. A. Moffat, M. A. Rosenthal, K. Drummond, A. H. Kaye, and A. P. Morokoff, "Early perfusion mri predicts survival outcome in patients with recurrent glioblastoma treated with bevacizumab and carboplatin," *Journal of neuro-oncology*, vol. 131, no. 2, pp. 321–329, 2017.

[186] E. H. Fyllingen, L. E. Bø, I. Reinertsen, A. S. Jakola, L. M. Sagberg, E. M. Berntsen, Ø. Salvesen, and O. Solheim, "Survival of glioblastoma in relation to tumor location: a statistical tumor atlas of a population-based cohort," *Acta neurochirurgica*, vol. 163, no. 7, pp. 1895–1905, 2021.

[187] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[188] D. Nandi, A. S. Ashour, S. Samanta, S. Chakraborty, M. A. Salem, and N. Dey, "Principal component analysis in medical image processing: a study," *International Journal of Image Mining*, vol. 1, no. 1, pp. 65–86, 2015.

[189] R. Jain, L. Poisson, J. Narang, D. Gutman, L. Scarpace, S. N. Hwang, C. Holder, M. Wintermark, R. R. Colen, J. Kirby, *et al.*, "Genomic mapping and survival prediction in glioblastoma: molecular subclassification strengthened by hemodynamic imaging biomarkers," *Radiology*, vol. 267, no. 1, p. 212, 2013.

[190] X. Feng, N. Tustison, and C. Meyer, "Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features," in *International MICCAI Brainlesion Workshop*, pp. 279–288, Springer, 2018.

[191] S. Park and N. Kwak, "Analysis on the dropout effect in convolutional neural networks," in *Asian conference on computer vision*, pp. 189–204, Springer, 2016.

[192] F. Kofler, J. C. Paetzold, I. Ezhov, S. Shit, D. Krahulec, J. S. Kirschke, C. Zimmer, B. Wiestler, and B. H. Menze, "A baseline for predicting glioblastoma patient survival time with classical statistical models and primitive features ignoring image information," in *International MICCAI Brainlesion Workshop*, pp. 254–261, Springer, 2019.

[193] A. S. Akbar, C. Fatichah, and N. Suciati, "Modified mobilenet for patient survival pre-

diction," in *International MICCAI Brainlesion Workshop*, pp. 374–387, Springer, 2020.

List of Publications

1. Gurinderjeet Kaur, Prashant Singh Rana, Vinay Arora, *"State-of-the-art techniques using pre-operative brain MRI scans for survival prediction of glioblastoma multiforme patients and future research directions"*, Clinical and Translational Imaging, Springer, 10(1):355-389, 2022. [SCIE, IF 2.507]

2. Gurinderjeet Kaur, Prashant Singh Rana, Vinay Arora, *"Extracting Radiomic Features from Pre-operative and Segmented MRI Scans Improved Survival Prognosis of Glioblastoma Multiforme Patients through Machine Learning: A Retrospective Study"*, Multimedia Tools and Applications, Springer, 1-36, 2022. [SCIE, IF 2.577]

3. Gurinderjeet Kaur, Prashant Singh Rana, Vinay Arora, *"Deep learning and machine learning-based early survival predictions of glioblastoma patients using pre-operative three-dimensional brain magnetic resonance imaging modalities"*, International Journal of Imaging Systems and Technology, Wiley,33(1): 340-361, 2022. [SCIE, IF 2.177]

Appendix

Table A1: Feature selection using ML-based models

Feature Category	Features	Feature selection method				
		LASSO	KNN	XGB	RF	DT
Clinical	Age					
Image Intensity Diagnostics	Image original Maximum					
	Image original Mean					
	Mask original Volume Num					
Exponential first-order	10 Percentile					
	Kurtosis					
	Mean					
	Skewness					
Exponential glcm	Cluster Shade					
	Correlation					
	Inverse Difference Moment Normalized					
	Informational Measure of Correlation1					
	Informational Measure of Correlation2					
	Inverse Variance					
	Joint Energy					
	Joint Entropy					
	Sum Entropy					
Exponential gldm	Dependence Entropy					
	Dependence Non Uniformity Normalized					
	Dependence Variance					
	Large Dependence High Gray Level Emphasis					
	Large Dependence Low Gray Level Emphasis					
	Small Dependence High Gray Level Emphasis					
	Small Dependence Low Gray Level Emphasis					
Exponential glrlm	Gray Level Non Uniformity Normalized					
	Run Entropy					
Exponential glszm	Large Area Low Gray Level Emphasis					
Logarithm first-order	10Percentile					
	Kurtosis					
	Minimum					
Logarithm glcm	Difference Variance					
	Inverse Difference Moment Normalized					
	Inverse Difference Normalized					
	Inverse Variance					
	Sum Entropy					
	Sum Squares					
Logarithm gldm	Dependence Entropy					
	Large Dependence High Gray Level Emphasis					
	Small Dependence Low Gray Level Emphasis					
Logarithm glszm	Large Area High Gray Level Emphasis					
	Zone Entropy					
Log sigma 1 mm 3D first-order	90 Percentile					
	Energy					
	Entropy					
	Maximum					
	Root Mean Squared					
	Skewness					
Log sigma 1 mm 3D glcm	Cluster Prominence					
	Cluster Shade					
	Difference Entropy					
	Inverse Difference Moment Normalized					
	Inverse Difference Normalized					
	Maximum Probability					
Log sigma 1 mm 3D gldm	Large Dependence High Gray Level Emphasis					
Log sigma 1 mm 3D glrlm	Long Run Low Gray Level Emphasis					
Log sigma 1 mm 3D glszm	Size Zone Non Uniformity					
Log sigma 3 mm 3D first-order	Kurtosis					
	Maximum					
	Median					
	Range					
	Skewness					
	Inverse Difference Moment Normalized					
	Inverse Difference Normalized					
	Informational Measure of Correlation1					

Table A1 continued from previous page

Feature Category	Features	Feature selection method				
		LASSO	KNN	XGB	RF	DT
Log sigma 3 mm 3D glcm	Informational Measure of Correlation2					
	Inverse Variance					
	Maximum Probability					
Log sigma 3 mm 3D gldm	Large Dependence High Gray Level Emphasis					
	Large Dependence Low Gray Level Emphasis					
Log sigma 3 mm 3D glrlm	Long Run Low Gray Level Emphasis					
	Short Run Low Gray Level Emphasis					
Log sigma 3 mm 3D glszm	Large Area High Gray Level Emphasis					
	Large Area Low Gray Level Emphasis					
	Small Area Low Gray Level Emphasis					
Log sigma 3 mm 3D ngtdm	Complexity					
	Contrast					
Log sigma 5 mm 3D first-order	90 Percentile					
	Energy					
	Interquartile Range					
	Kurtosis					
	Range					
	Skewness					
	Total Energy					
Log sigma 5 mm 3D glcm	Contrast					
	Correlation					
	Inverse Difference Normalized					
	Informational Measure of Correlation1					
	Informational Measure of Correlation2					
Log sigma 5 mm 3D gldm	Large Dependence High Gray Level Emphasis					
Log sigma 5 mm 3D glszm	Gray Level Non Uniformity Normalized					
	Large Area Emphasis					
	Large Area Low Gray Level Emphasis					
	Zone Entropy					
Log sigma 5 mm 3D ngtdm	Contrast					
	Strength					
Original first-order	10 Percentile					
	Kurtosis					
	Skewness					
Original glcm	Inverse Difference Normalized					
	Joint Energy					
Original gldm	Large Dependence High Gray Level Emphasis					
Original glrlm	Long Run Low Gray Level Emphasis					
Original ngtdm	Busyness					
Original shape	Elongation					
	Flatness					
	Least Axis Length					
	Major Axis Length					
	Maximum 2D Diameter Row					
	Maximum 2D Diameter Slice					
	Maximum 3D Diameter					
	Minor Axis Length					
	Sphericity					
	Surface Volume Ratio					
Square first-order	10 Percentile					
	Kurtosis					
Square glcm	Cluster Shade					
	Cluster Tendency					
	Correlation					
Square gldm	Large Dependence Low Gray Level Emphasis					
Square glrlm	Gray Level Non Uniformity Normalized					
	Long Run Low Gray Level Emphasis					
	Run Variance					
Square glszm	Zone Entropy					
Square ngtdm	Busyness					
Squareroot first-order	10 Percentile					
	Kurtosis					
Squareroot glcm	Inverse Difference Moment Normalized					
	Informational Measure of Correlation1					
Image-based	Tumor Volume					
Wavelet HH first-order	Kurtosis					
	Maximum					
	Mean					
	Median					
	Skewness					
	Cluster Shade					
	Correlation					
	Difference Variance					

Table A1 continued from previous page

Feature Category	Features	Feature selection method				
		LASSO	KNN	XGB	RF	DT
Wavelet HH glcm	Inverse Difference Normalized					
	Informational Measure of Correlation1					
Wavelet HH gldm	Dependence Entropy					
	Dependence Non Uniformity					
	Large Dependence High Gray Level Emphasis					
	Large Dependence Low Gray Level Emphasis					
	Small Dependence Emphasis					
	Small Dependence Low Gray Level Emphasis					
Wavelet HH glrlm	Long Run Low Gray Level Emphasis					
	Run Variance					
Wavelet HH glszm	Gray Level Variance					
	Large Area High Gray Level Emphasis					
	Low Gray Level Zone Emphasis					
	Size Zone Non Uniformity					
	Small Area Low Gray Level Emphasis					
Wavelet HH ngtdm	Strength					
Wavelet HL first-order	Median					
Wavelet HL glcm	Cluster Shade					
	Correlation					
	Inverse Difference Moment Normalized					
Wavelet HL glrlm	Long Run Low Gray Level Emphasis					
Wavelet LH first-order	Interquartile Range					
	Kurtosis					
	Maximum					
	Mean					
	Median					
	Skewness					
Wavelet LH glcm	Cluster Shade					
	Correlation					
	Inverse Difference Normalized					
	Informational Measure of Correlation1					
	Maximum Probability					
Wavelet LH gldm	Small Dependence Low Gray Level Emphasis					
Wavelet LH glrlm	Low Gray Level Run Emphasis					
Wavelet LH glszm	Size Zone Non Uniformity					
	Size Zone Non Uniformity Normalized					
Wavelet LH ngtdm	Contrast					
Wavelet LL first-order	10 Percentile					
	Kurtosis					
	Mean					
	Skewness					
Wavelet LL glcm	Cluster Prominence					
	Cluster Shade					
	Correlation					
Wavelet LL gldm	Dependence Non Uniformity Normalized					
	Large Dependence High Gray Level Emphasis					
Wavelet LL glszm	Large Area Low Gray Level Emphasis					
	Low Gray Level Zone Emphasis					
Wavelet LL ngtdm	Busyness					

Table A2: Cross-validation MAE of different regression models on training dataset before hyperparameter tuning

Fold	ridge	ada	br	llar	rf	et	knn	lightgbm	gbr	omp	huber	dt	lr	en	lasso	par
0	209.085	270.773	262.565	262.930	274.114	278.802	294.760	288.208	295.623	265.682	376.043	373.300	308.713	498.674	519.492	1774.980
1	359.378	411.340	320.054	288.817	432.125	392.386	276.960	321.633	361.411	297.515	280.802	411.500	401.974	1005.871	738.047	4822.944
2	451.096	234.874	122.786	265.781	223.850	210.673	165.540	281.947	282.680	363.822	260.812	309.600	390.537	531.001	712.352	880.399
3	289.276	212.211	236.700	230.375	220.786	218.204	270.420	208.865	199.553	298.606	380.007	273.100	231.885	403.136	566.134	1925.677
4	199.064	177.594	129.762	206.324	142.305	162.123	191.822	190.218	144.832	157.345	353.793	223.666	308.472	446.996	476.008	1121.799
5	204.423	381.888	358.068	645.210	354.965	402.040	353.022	324.692	360.310	735.097	328.247	445.111	692.356	879.818	974.881	616.868
6	175.75	395.148	429.831	332.516	450.388	433.843	440.644	389.273	401.148	425.940	453.617	391.888	985.225	1661.393	1249.044	410.641
7	176.863	214.742	256.328	326.527	238.022	302.701	207.866	231.228	285.376	431.599	434.721	231.888	317.235	763.834	1037.054	1633.822
8	237.113	131.020	173.084	166.366	163.346	146.238	129.844	101.462	179.333	256.723	333.574	242.777	174.336	274.374	454.467	1803.009
9	267.709	281.599	275.776	246.943	292.613	260.896	251.466	338.354	293.113	230.440	396.818	465.888	404.320	848.589	1388.539	2290.367
Mean	256.977	271.119	256.496	297.179	279.251	280.790	258.234	267.588	280.338	346.277	359.843	336.872	421.505	731.369	811.602	1728.051
Std	84.347	91.521	92.789	125.625	99.484	96.166	87.270	80.611	79.423	152.772	58.494	87.054	229.389	383.258	316.656	1183.498

Table A3: Cross-validation MAE of different regression models on training dataset after hyperparameter tuning

Fold	ridge	ada	br	llar	rf	et	knn	lightgbm	gbr	omp	huber	dt	lr	en	lasso	par
0	200.486	248.411	273.012	262.930	257.812	274.153	268.193	264.146	269.050	269.801	376.295	314.310	308.713	269.153	259.850	1771.854
1	360.891	363.241	320.996	288.817	333.203	321.664	285.891	307.651	287.808	310.071	280.765	257.552	401.974	290.310	298.455	4816.165
2	421.476	148.423	158.053	265.781	159.399	164.079	144.850	198.154	154.203	254.148	287.049	222.071	390.537	154.085	158.820	879.336
3	297.567	255.555	269.508	230.375	233.354	257.321	266.831	224.153	269.699	300.507	379.865	248.062	231.885	269.389	255.866	1924.279
4	194.211	155.294	148.638	206.324	140.594	137.918	137.463	136.547	149.175	172.950	306.876	205.836	308.472	148.700	164.700	1121.979
5	233.954	327.361	333.743	645.210	307.284	342.891	336.722	331.414	333.495	668.988	325.385	223.569	692.356	334.763	331.914	617.526
6	180.750	389.735	373.445	332.516	382.383	385.628	365.814	355.768	359.881	391.538	472.138	529.625	985.225	360.406	355.578	410.792
7	159.144	212.895	203.992	326.527	192.967	213.410	208.891	211.540	210.236	402.849	434.754	284.678	317.235	209.988	217.675	1634.080
8	236.829	154.974	164.223	166.366	163.642	156.103	162.324	110.018	149.892	177.773	333.637	211.097	174.336	150.289	109.670	1799.744
9	268.486	235.820	213.234	246.943	226.900	228.259	209.541	245.636	212.356	236.328	322.031	349.860	404.320	211.961	209.417	2283.507
Mean	255.379	249.171	245.884	297.179	239.754	248.143	238.652	238.503	239.580	318.495	351.880	284.666	421.505	239.904	236.195	1725.926
Std	79.411	82.617	75.886	125.625	76.446	79.583	75.022	75.721	72.458	137.818	59.934	92.998	229.389	72.920	75.102	1181.330

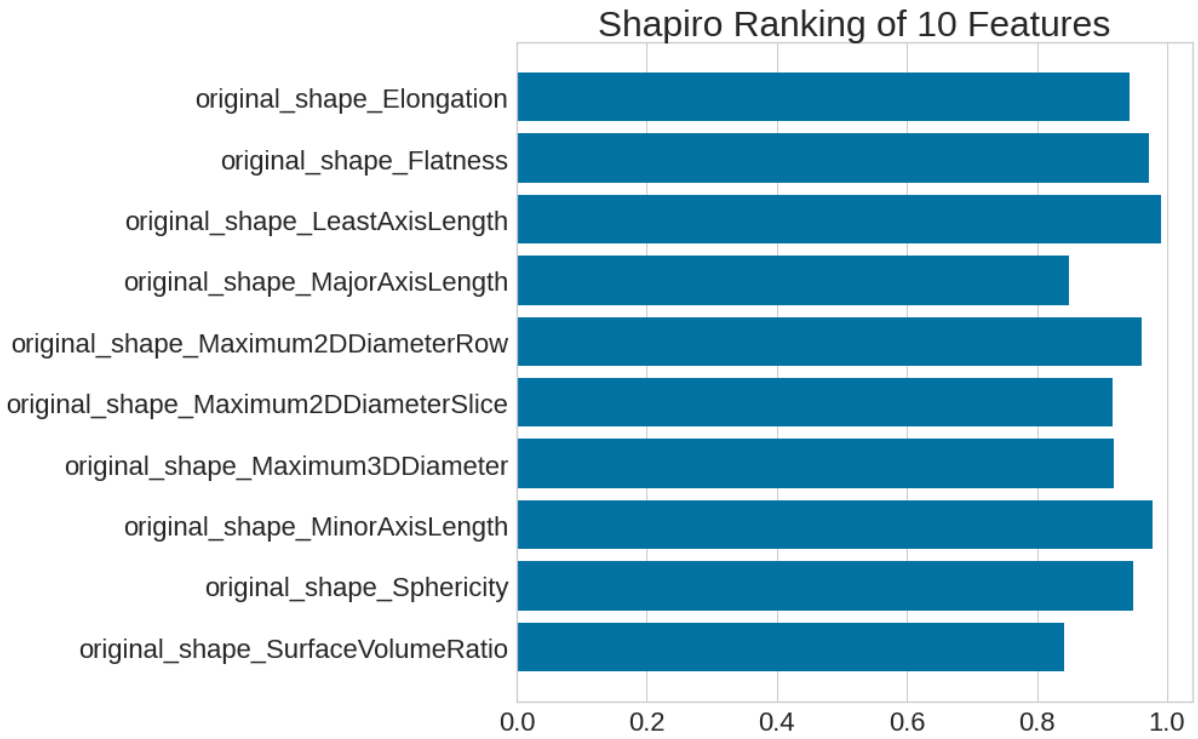


Figure F.1: Shapiro ranking of selected shape features

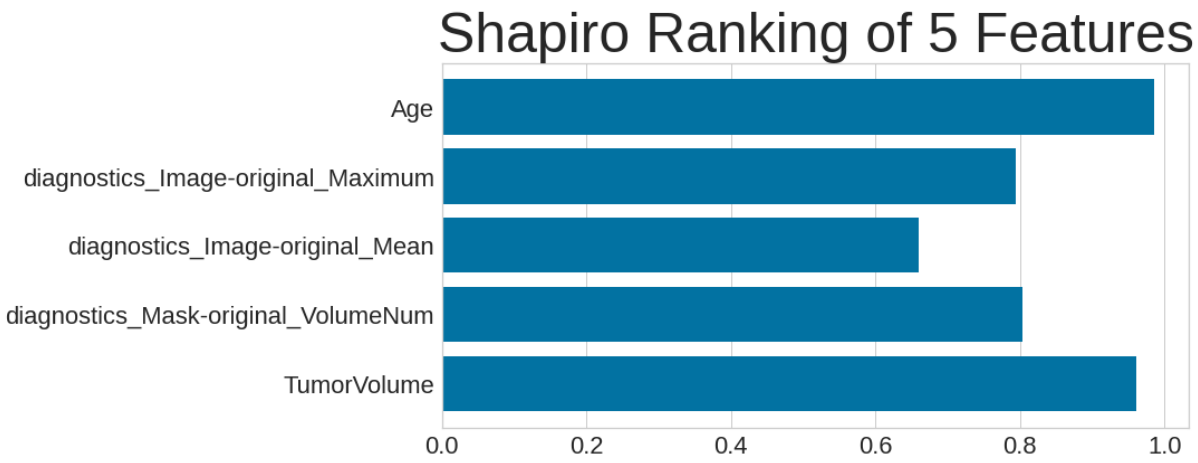


Figure F.2: Shapiro ranking of selected features: age, diagnostics and tumor volume

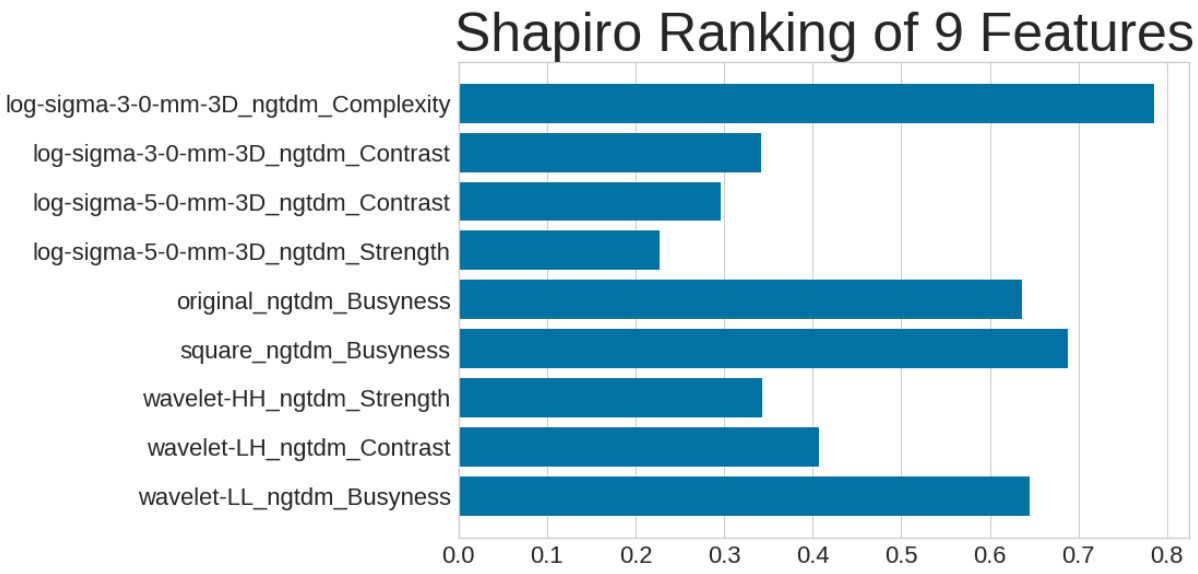


Figure F.3: Shapiro ranking of selected NGTDM features

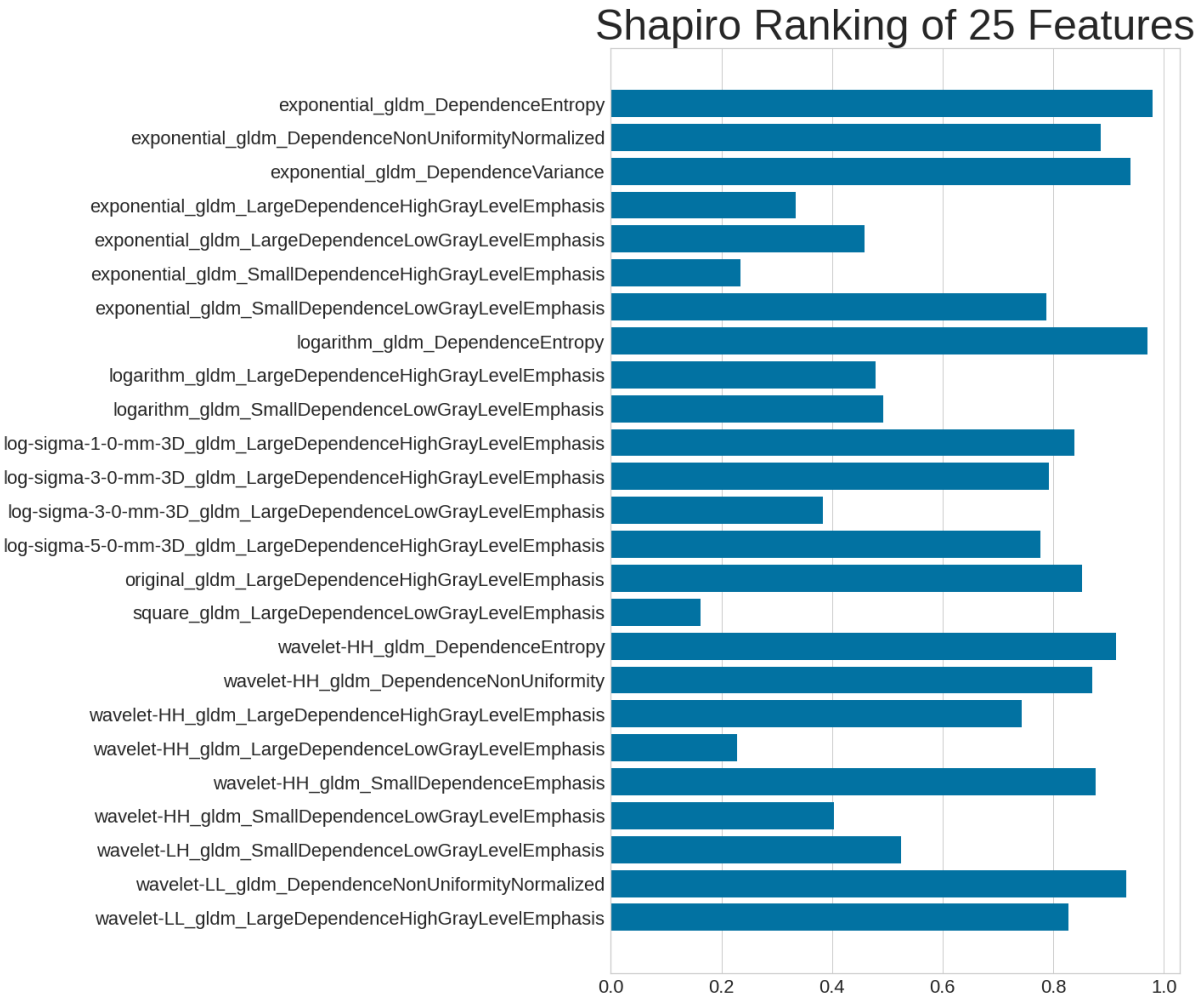


Figure F.4: Shapiro ranking of selected GLDM features

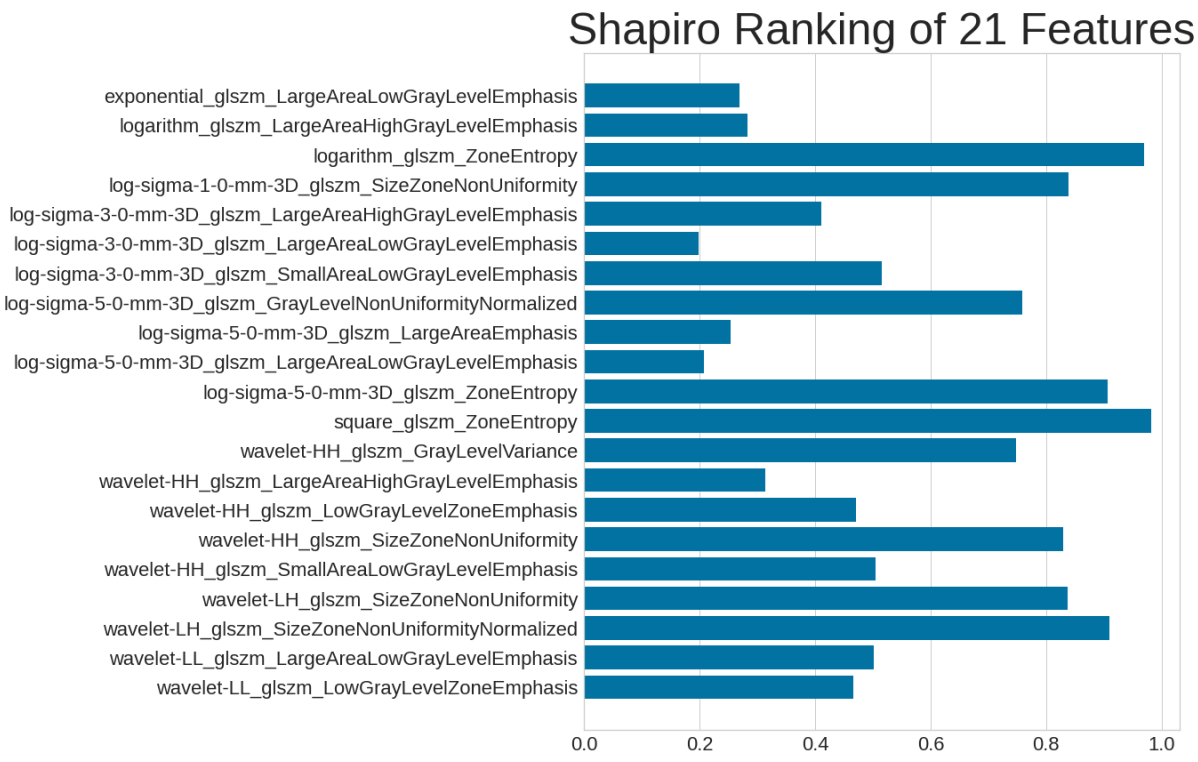


Figure F.5: Shapiro ranking of selected GLSZM features

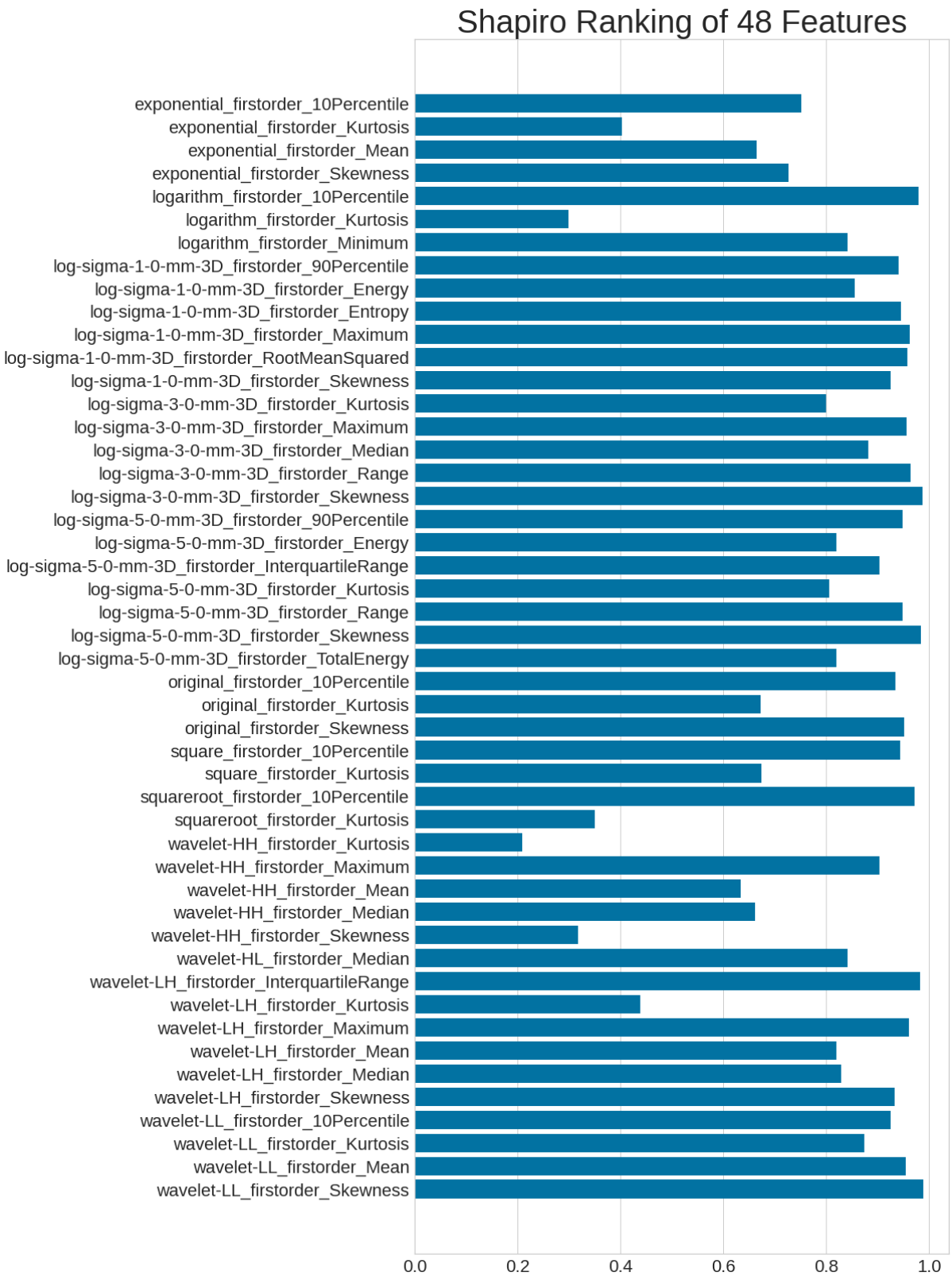


Figure F.6: Shapiro ranking of selected first order features

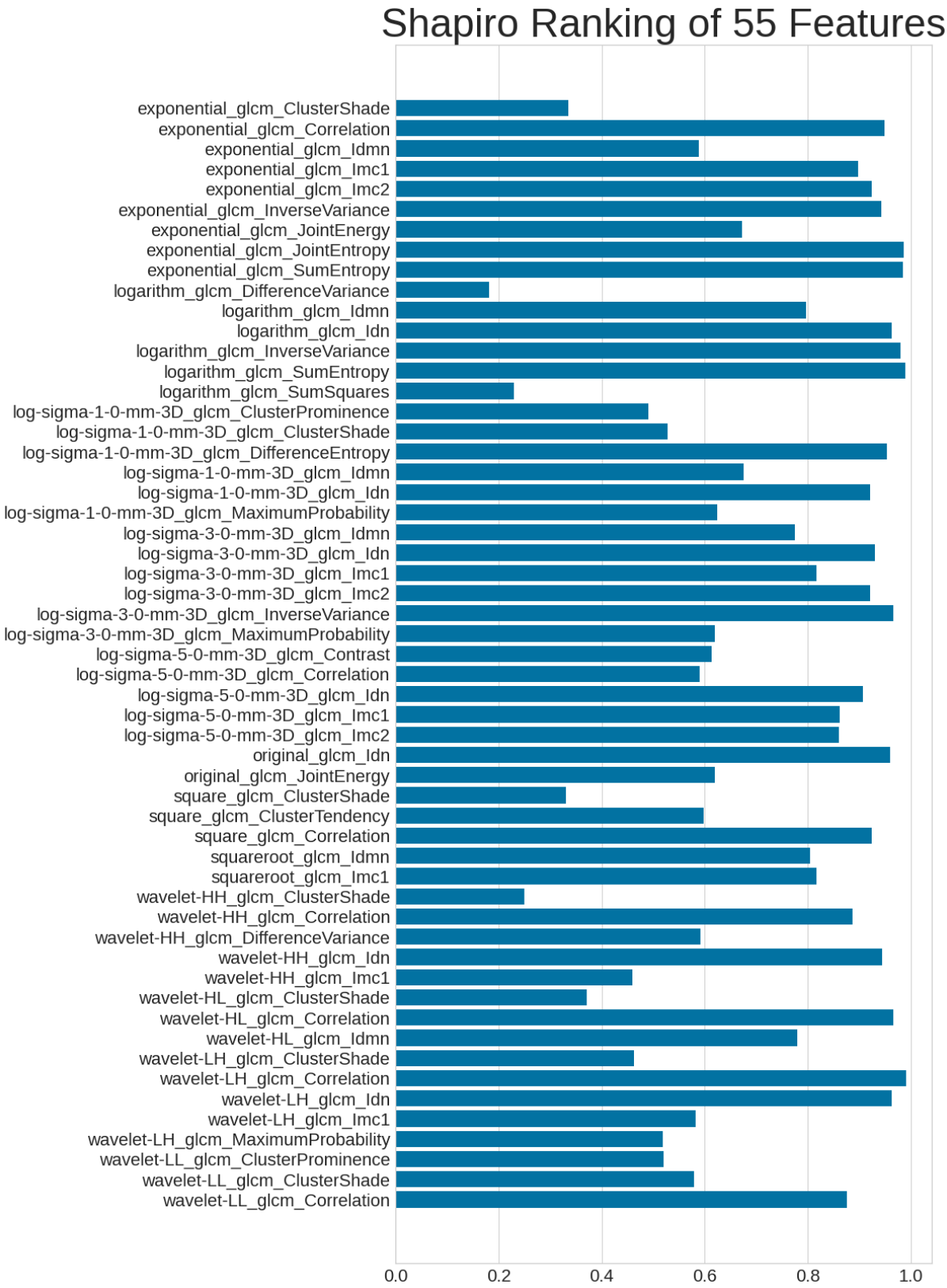


Figure F.7: Shapiro ranking of selected GLCM features

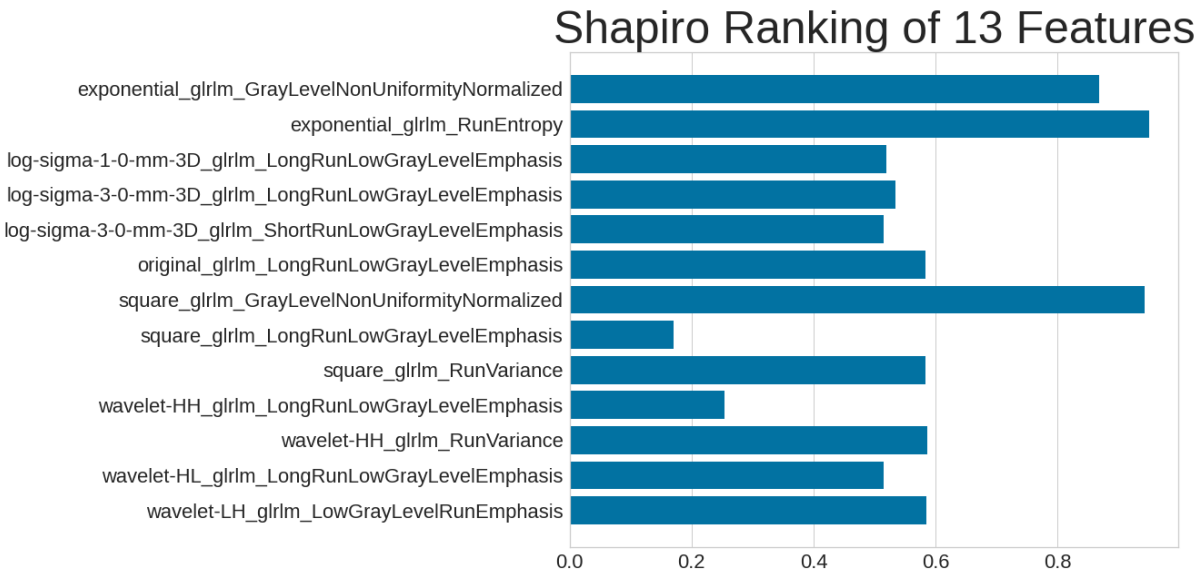


Figure F.8: Shapiro ranking of selected GLRLM features