Survival prediction of Glioblastoma patients using multi-modal MRI radiomic features

Report for the Bachelor's Thesis Project

by

Anamitra Sengupta

20IE10004

Under the supervision of

Professor Jayanta Mukhopadhyay

Department of Electrical Engineering

Indian Institute of Technology, Kharagpur

April 2024

Abstract

Machine Learning models have become integral tools in medical practice, facilitating tasks ranging from treatment recommendation to disease detection and survival prediction. In the context of brain tumor diagnosis, MRI scans are commonly employed for tumor assessment. Traditionally, machine learning models analyze these scans based on predefined features such as tumor shape, size, and color. Notably, radiomic features have emerged as a popular feature set specifically tailored for this purpose, capturing crucial information encoded within MRI images.

However, relying solely on radiomic features may not yield optimal performance, particularly in challenging research scenarios such as the Brain Tumor Segmentation (BraTS) Challenge. To address this, our thesis endeavors to explore both radiomic features and the most effective models for predicting survival outcomes. We have conducted a comprehensive study involving multiple transformations of the four modalities present in MRI scans, extracting a diverse spectrum of radiomic features from each transformation.

Subsequently, we have evaluated various feature selection and pruning algorithms to effectively reduce the dimensionality of our dataset. This reduction aims to strike a balance between model performance and computational efficiency, ensuring that the selected features not only optimize model performance but also remain manageable in terms of computational resources.

Contents

1 Introduction

The burgeoning field of radiomics is rapidly altering the landscape of medical imaging, where crucial information can be meticulously extracted from images to inform and support diagnostic and treatment modalities. Radiomics, spanning diverse realms of medicine, is profoundly transforming our approach to analyzing medical images, offering a spectrum of applications ranging from the development of treatment recommender systems and disease diagnosis to the precise prediction of prognoses and beyond.

Survival prediction emerges as a particularly compelling application within this paradigm, involving the forecasting of both the likelihood and duration of a patient's survival predicated upon an array of clinical features, including intricate medical histories and discernible biological markers.

Central to this discussion are gliomas, the predominant primary malignancies affecting the brain, whose prevalence underscores the paramount importance of accurate and resilient tumor segmentation techniques and the prognostication of overall patient survival. These crucial factors are central to achieving accurate diagnoses, developing personalized treatment strategies, and identifying important risk factors.

In this endeavor, we undertake the task of survival prediction among patients afflicted with Glioblastoma, harnessing the potential of multi-modal Magnetic Resonance Imaging (MRI) radiomic features. The dataset used in our thesis is the Brain Tumor Segmentation (BraTS) Challenge 2020 dataset. MRI images exist in four distinct modalities—native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (T2-flair). To harness an ample set of features, extensive transformations are applied to each modality, facilitating the extraction of radiomic features from diverse perspectives. We leverage the Pyradiomics toolkit, a robust and open-source Python package, for these tasks. The dataset contained 26 transformed images for each modality, and each image yielded 120 extracted radiomic features. Additionally, we included the age of the patient as a feature in our analysis, recognizing its significance in predicting survival time.

This approach resulted in a substantial dataset, with each patient's data comprising a total of 12,481 distinct features. By incorporating such a comprehensive range of features, our models

can better identify intricate patterns within the data, leading to a more comprehensive understanding and potentially improving predictive accuracy.

Given the extensive feature set available, our next focus shifts to dimensionality reduction. Various techniques such as Fischer score-based pruning, recursive feature elimination using random forests, feature selection using LASSO regression, mRMR (minimum redundancy maximum relevance), and mRMR-ensemble methods (that include the exhaustive and bootstrap variants) are employed to refine the feature set to a more manageable size that strikes a balance between computational efficiency and retaining key predictive elements.

With our refined feature set in hand, we proceed to the critical stage of model selection and evaluation. We assess a diverse range of models, including Random Forest Regression, Cox-proportional hazards models, and the innovative DeepSurv framework (Katzman, J. et. al.), to determine their effectiveness in predicting survival outcomes. DeepSurv represents the integration of deep learning principles into the traditional Cox proportional hazards model, implemented using the Theano and Lasagne libraries in Python.

We also implement the Local Spatial Relationship Awareness Network (LSRAN), an innovative framework designed to capture the intricate spatial relationships embedded within the image data (M.-T. Tran et al., Patacchiola, M. et. al.). LSRAN adopts relational reasoning as a pretext task for learning salient representations within the neural network backbone. Following training, the relation head is discarded, and the refined backbone is deployed in downstream tasks such as classification and image retrieval, survival prediction in our case.

2 Literature Review

2.1 State-of-the-art results

For the survival prediction challenge, Ali et al. (2021) extracted radiomic and image-based parameters such as volume ratio, surface area, etc. The Laplacian of Gaussian filters was used to extract the radiomic characteristics. Another detail that was introduced was the patient's age. They employed Random Forest - Recursive Feature Elimination (RF-RFE) to choose the top-performing features. They used a Random Forest regressor with a grid search after identifying the important features to predict survival. On the validation dataset, they obtained an accuracy of 48.3% and a Spearman R coefficient of 0.134.

Patel et al. (2021) trained a Cox Proportional Hazards Model using 2048 deep features that were extracted from their segmentation network. After processing all inputs and ensemble models, 40 feature versions were produced. To avoid overfitting, dimensionality was then decreased using PCA. Predicted sub-region sizes and age at diagnosis were also included. Ten main components were used to train the Cox model on 118 patients with gross total resection (GTR) in order to get the best validation set performance. In this instance, a 65.5% accuracy and a 0.479 Spearman R coefficient were attained.

For better segmentation results and optimized performance with smaller batch sizes, S. R. González et al. (2021) used a 3D DenseNet CNN model, which was a modification of the 2D model of Huang et al. 3D convolutions replaced 2D filters, and instance normalisation replaced batch normalisation. They obtained a Spearman R coefficient of 0.442 and an accuracy of 55.2%.

C. Russo et al. (2020) used automatically extracted MRI-based features from the novel LesionEncoder (LE) framework (Feng Y-Z et al., 2021) in place of predefined imaging and radiomic features like volumetric parameters, intensity, morphological characteristics, histogram-based data, and textural properties. Principal Component Analysis (PCA) was used to reduce the dimensionality of the LE characteristics, which were then used as input for a generalized linear model (GLM) to predict patient OS. Their best sub-model achieved a Spearman R-value of 0.412 and an accuracy of 58.3%.

Minh-Trieu Tran et al. proposed a framework for predicting the survival of brain tumor patients using magnetic resonance images. Given the challenge of limited datasets, they introduced a

self-supervised learning approach to mitigate overfitting. This method involves identifying image patches from the same or different images to learn intra- and inter-image relationships. Additionally, a global structure awareness network was incorporated to capture global information. Their approach demonstrated a strong correlation between local spatial relationships and survivor class prediction in FLAIR MRI brain images, achieving an accuracy of 62.1% and a Spearman R correlation of 0.576. Importantly, their method outperformed other state-of-the-art methods.

Table 1 presents a compilation of the results obtained by models that used radiomic features, while Table 2 presents the ones that used other frameworks.

Paper	Description	Accuracy	MSE	medianSE	SpearmanR
Ali et al. (2021)	Random Forest Regressor using image-based and radiomic features	0.483	105079.4	37004.93	0.134
R. R. Agravat and M. S. Raval (2020)	Random Forest Regressor using statistical and radiomic features	0.517	116083.477	43974.090	0.217
R. Miron et al. (2021)	Extra Trees	0.414	87744.14	37636	0.321
V. K. Anand et al. (2021)	Random Forest Regressor using radiomic features	0.448	110677.443	22874.178	0.169

Table 1: Results of Models using Radiomic Features

2.2 mRMR Algorithm

Minimum redundancy maximum relevance (mRMR) is a standard feature selection algorithm that produces a set of relevant and complementary features (Ding and Peng, 2005). It involves defining mutual information between two features *x* and *y* as:

$$I(x;y) = \int\int p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy \quad (1)$$

Paper	Model	Accuracy	MSE	medianSE	SpearmanR
Patel et al. (2021)	Cox Proportional Hazards Model	0.655	152467	39601	0.479
S. R. González et al. (2021)	3D DenseNet 121	0.552	87581	51529	0.442
I. S. Han (2021)	2D U-net with neuromorphic attention	0.552	147898.5	52900	0.333
C. Russo et al. (2020)	GLM on features extracted from LesionEncoder framework	0.586	88311.58	27114.54	0.412
Minh-Trieu Tran et al. (2023)	Local Spatial Relationships and Global Structure Awareness in MRI Brain Images	0.621	97216.345	-	0.576

Table 2: Results of Models using features other than Radiomic features

First, we rank the input features $X = \{x_1, \dots, x_n\}$ by maximizing their Mutual Information (MI) with the output variable *y* and minimizing redundancy with previously selected features. Starting with the feature x_i that has the highest MI with *y*, we initialize the selected features set *S*:

$$x_i = argmax_{x_i \in X} I(x_i, y) \tag{2}$$

Then, we add features to *S* by choosing those with the highest relevance to *y* and the lowest redundancy with the existing selections, maximizing the score *q* at step *j*:

$$q_j = I(x_j, y) - \frac{1}{|S|} \sum_{x_k \in S} I(x_j, x_k) \tag{3}$$

We repeat this process until reaching the desired solution length.

2.3 mRMRe: Ensemble Approach

As justified by Jay et al. (2013), the mRMR algorithm has several disadvantages. There is no guarantee of discovering a globally optimal solution, and it's possible that alternative feature subsets of equal or better quality exist. Furthermore, a single mRMR run's feature selection is unlikely to comprehensively represent the range of biological processes associated with the studied phenotype.

To address these issues, two ensemble approaches were implemented, enabling the generation of multiple mRMR solutions in parallel. These approaches termed exhaustive and bootstrap ensemble mRMR were employed. The exhaustive variant extends the classical mRMR heuristic by initiating multiple feature selection procedures with the top $k > 1$ relevant features. Subsequently, *k* mRMR solutions are produced in parallel, ensuring that the initially selected feature is different.

In the bootstrap variant, the original dataset is resampled with replacement, creating *k* bootstraps. Classical mRMR feature selection is then conducted in parallel for each bootstrapped dataset, yielding *k* mRMR solutions.

They conducted a case study using two pharmacogenomic datasets: the Cancer Genome Project (CGP) [Garnett et al., 2012] and the Cancer Cell Lines Encyclopedia (CCLE) [Barretina et al., 2012]. In CGP, mRMRe.e and mRMRe.b outperformed classical mRMR by 1.7% and 1.9% for common cell lines, and by 2.1% and 3.4% for new cell lines in CCLE. These results highlight the superior generalization ability of the ensemble methods, particularly mRMRe.b, which identifies a more diverse panel of mRMR solutions.

2.4 Cox-proportional hazards models

The Cox proportional-hazards model (Cox, 1972) is a widely used statistical tool in medical research for analyzing the relationship between patients' survival time and various predictor variables. In clinical investigations, where multiple factors may influence patient prognosis, this model allows for the simultaneous assessment of these factors. It provides effect size estimates for each factor, aiding in the interpretation of results.

The Cox model evaluates how specified factors impact the rate of a particular event (e.g., infection, death) occurring at a given time. This rate, known as the hazard rate, is determined by a set of covariates, with coefficients measuring their effects. The model's hazard function ($h(t)$) accounts for changes in hazard over time, with the baseline hazard (h_0) representing the value when all covariates are zero. The coefficients ($b_1, b_2, \ldots, b_p$) measure the impact of the covariates ($x_1, x_2, \ldots, x_p$).

$$h(t) = h_0(t) \times e^{(b_1 x_1 + b_2 x_2 + \ldots + b_p x_p)} \quad (4)$$

The quantities e^{b_i} are called hazard ratios. Hazard ratios (HR) quantify the impact of covariates on the hazard rate, with HR > 1 indicating an increased hazard and HR < 1 indicating a reduced hazard. In cancer studies, covariates with HR > 1 are termed bad prognostic factors, while those with HR < 1 are termed good prognostic factors.

An essential assumption of the Cox model is proportionality, meaning hazard ratios remain constant over time and hazard curves do not intersect. This ensures that the hazard of an event in one group is consistently proportional to that in another group, irrespective of time. This assumption enables meaningful interpretation of hazard ratios and facilitates comparisons between different patient groups.

2.5 DeepSurv

DeepSurv (Katzman, J. et. al.), introduced as a modern Cox proportional hazards deep neural network, represents a significant advancement in survival analysis and personalized treatment recommendation systems. Traditional survival models, such as the Cox proportional hazards model (CPH), have been widely used in medical research to evaluate prognostic variables. However, they may oversimplify risk functions by assuming linearity, limiting their ability to capture complex relationships in survival data. To address this limitation, researchers have explored the application of deep learning techniques, including neural networks, to model nonlinear survival data.

DeepSurv, as a risk neural network, learns intricate and nonlinear relationships between prognostic features and individuals' risk of failure without requiring prior feature selection or

domain expertise. By leveraging deep neural networks, DeepSurv can provide personalized treatment recommendations based on computed risk, potentially improving median survival times for patients.

The concordance index (C-index) was utilized to assess the predictive accuracy of the models on survival data, as outlined by Frank E Harrell et. al.. The C-index is a widely used metric in survival analysis, quantifying the model's ability to predict the ordering of patients' death times. A C-index of 0.5 represents the performance of a random model, while a value of 1 indicates the perfect ranking of death times. To estimate the confidence intervals, bootstrapping was employed by sampling the test set with replacement. Table 3 presents the results of their evaluation.

Experiment	CPH	DeepSurv	RSF
Simulated Linear	0.773677 (0.772,0.775)	0.774019 (0.772,0.776)	0.764925 (0.763,0.766)
Simulated Non-linear	0.506951 (0.505,0.509)	0.648902 (0.647,0.651)	0.645540 (0.643,0.648)
WHAS	0.817620 (0.814, 0.821)	0.862620 (0.859,0.866)	0.893623 (0.891,0.896)
SUPPORT	0.582870 (0.581,0.585)	0.618308 (0.616,0.620)	0.613022 (0.611,0.615)
METABRIC	0.630618 (0.627,0.635)	0.643374 (0.639,0.647)	0.624331 (0.620,0.629)
Simulated Treat-ment	0.481540 (0.480,0.483)	0.582774 (0.580,0.585)	0.569870 (0.568,0.572)
Rotterdam & GBSG	0.657750 (0.654, 0.661)	0.668402 (0.665,0.671)	0.651190 (0.648, 0.654)

Table 3: Experimental Results for All Experiments: C-index (95% Confidence Interval)

2.6 Local Spatial Relationship Awareness Framework

Self-supervised relational reasoning for representation learning, as introduced by Jing and Tian (2020), involves utilizing relational reasoning as a pretext task to acquire valuable representations in neural network backbones. This approach has garnered attention due to its

potential to enhance downstream tasks such as classification and image retrieval. Relational reasoning, traditionally defined as understanding the connections between entities to accomplish higher-order goals, has been redefined in this context to include intra-reasoning (how entities relate to themselves) and inter-reasoning (how entities relate to each other) (Patacchiola, M. et. al.). This innovative formulation aims to enable learners to differentiate between objects from various scenes, enhancing their ability to generalize across diverse datasets. By incorporating random augmentations and leveraging intra-reasoning and inter-reasoning signals, self-supervised relational reasoning offers a promising avenue for learning useful representations without relying on labeled data.

Consider an unlabeled dataset $D = \{x_n\}_{n=1}^{N}$ and a non-linear function $f_\theta(\cdot)$ that is represented as a neural network (backbone) and parameterized by a vector of learnable weights θ. A vector $f_\theta(x_n) = z_n$ (representation) is produced by a forward pass and can be gathered into a set $Z = \{z_n\}_{n=1}^{N}$. The probability distribution of instances produced by applying stochastic data augmentation to x_n is expressed by the notation $A(x_n)$. The i-th set of random augmentations over all instances is represented by $D^{(i)} = \{x_n^{(i)}\}_{n=1}^{N}$, and the i-th sample from this distribution is represented by $x_n^{(i)} \sim A(x_n)$, which is a specific augmented version of the input instance. Similarly, $z_n^{(i)} = f_\theta(x_n^{(i)})$ is arranged within $Z^{(i)} = \{z_n^{(i)}\}_{n=1}^{N}$. The total number of augmentations $D^{(1)}, \ldots, D^{(K)}$ and their representations $Z^{(1)}, \ldots, Z^{(K)}$ should be indicated by K. The relation module $r_\phi(\cdot)$ will now be defined as a non-linear function approximator parameterized by ϕ. Its output will be a relation score y, and its input will be a pair of aggregated representations. Enumerating an aggregation function $a(\cdot,\cdot)$ and the loss between the score and a goal value $L(y,t)$, if $\backslash n$ is an index randomly picked from $\{1, \ldots, N\}\backslash\{n\}$, the entire learning objective may be expressed as:

$$argmin_{\theta,\phi} \sum_{n=1}^{N} \sum_{i=1}^{K} \sum_{j=1}^{K} L(r_\phi(a(z_n^{(i)}, z_n^{(j)})), t = 1) + L(r_\phi(a(z_n^{(i)}, z_{\backslash n}^{(j)})), t = 0) \qquad (5)$$

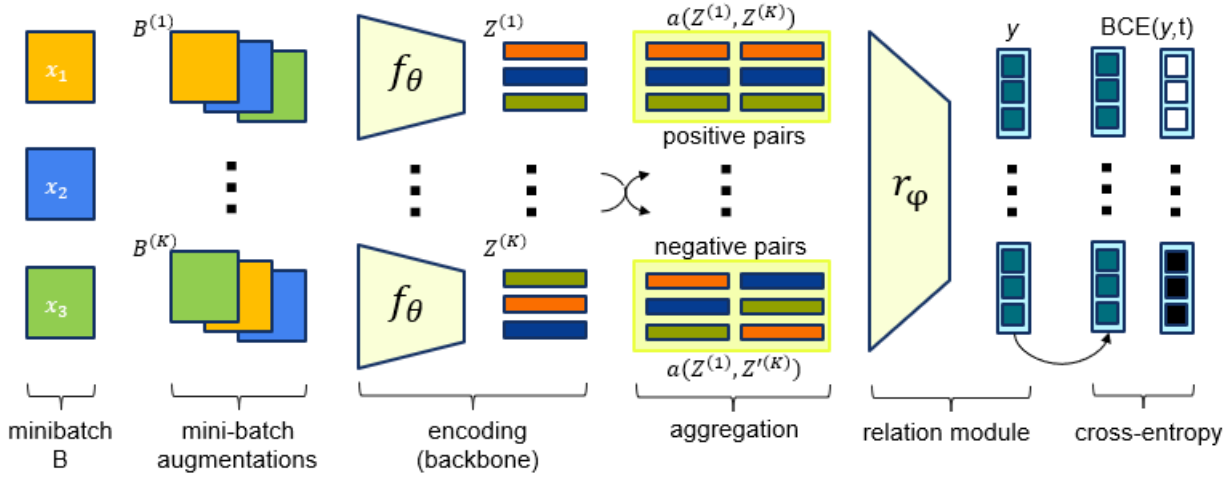

Figure 1: Overview of the training of the backbone

The augmentations used in the paper include random crop-resize and color distortion, while the aggregation strategy used is concatenation. The relation head has been modeled as a multi-layer perceptron (MLP). The relation score y represents a probabilistic estimate of representation membership, which can be induced through a sigmoid activation function. It follows that the objective reduces to the maximization of a Bernoulli log-likelihood, or similarly, the minimization of a binary cross-entropy loss:

$$L(y, t, \gamma) = \frac{1}{P}\sum_{i=1}^{P} - w_i\left[t_i \cdot logy_i + (1 - t_i) \cdot log(1 - y_i)\right] \quad (6)$$

with target $t_i = 1$ for positives and $t_i = 0$ for negatives. The weight w_i is a scaling factor:

$$w_i = \frac{1}{2}\left[(1 - t_i) \cdot y_i + t_i \cdot (1 - y_i)\right]^\gamma \quad (7)$$

where $\gamma \geq 0$ defines how sharp the weight should be. This loss is known as the focal loss. Figure 1 is an illustration of the process.

3 Scope & Objectives

The aim of this thesis is to investigate the utility of radiomics in the context of brain tumor survival prediction using MRI scans and to develop an accurate and clinically relevant predictive model.

Our objectives can be broadly divided into three categories:

1. Data Acquisition and Transformation
2. Feature Extraction and Dimensionality Reduction
3. Model Development for Survival Prediction

4 Work Progress

4.1 Data Acquisition and Transformation

We acquired data from the 2020 BraTS challenge, comprising abundant clinically-acquired pre-operative multimodal MRI scans of glioblastoma (GBM/HGG) and lower-grade glioma (LGG) with pathologically confirmed diagnoses and available overall survival (OS) information. We had 369 patients in the training dataset and 125 in the validation dataset, with each patient having 4 modalities of the MRI: native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (T2-flair).

The following transformations were independently applied to each of the modalities:

(a) Laplacian of Gaussian filter [*10*]

The Gaussian kernel is used to smooth the image and is defined as

$$G(x, y, z, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^3} e^{-\frac{x^2+y^2+z^2}{2\sigma^2}}$$

The variable σ has been chosen as $\sigma = \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$.

(b) Wavelet Filter (LLL, LLH, LHL, HLL, LHH, HLH, HHL, HHH) [*8*]

Wavelet filters break down data into different components at various scales and orientations. The components denoted as LLL, LLH, LHL, HLL, LHH, HLH, HHL, and HHH, represent the data's frequency and detail information at different levels. (L: low, H: high)

(c) Square of image intensities [*1*]

This computes the square of the image intensities.

$$f(x) = (cx)^2$$

Where *x* and *f(x)* are the original and filtered intensity, respectively.

(d) Square-root of image intensities [*1*]

This computes the square root of the absolute value of image intensities.

$$f(x) = \sqrt{cx}\ for\ x \geq 0,\ f(x) = -\sqrt{cx}\ for\ x < 0$$

$$where\ c = max(|x|)$$

Where *x* and *f(x)* are the original and filtered intensity, respectively.

(e) Logarithm of (the absolute value of original image + 1) [*1*]

This computes the logarithm of the absolute value of the original image + 1.

$$f(x) = c \cdot log(x + 1)\ for\ x \geq 0,\ f(x) = -c \cdot log(-x + 1)\ for\ x < 0$$

$$where\ c = \frac{max(|x|)}{log(max(|x|)+1)}$$

Where *x* and *f(x)* are the original and filtered intensity, respectively.

(f) Exponential of the original image [*1*]

This computes the exponential of the original image.

$$f(x) = e^{cx},\ where\ c = \frac{log(max(|x|))}{max(|x|)}$$

Where *x* and *f(x)* are the original and filtered intensity, respectively.

(g) Gradient Magnitude in the image [*1*]

It is calculated using gradient operators, such as the Sobel operator, which computes the rate of change of pixel values in both the horizontal and vertical directions.

(h) Local Binary Pattern (LBP) in 2D (image processing in a by-slice operation) [*1*]

LBP is a texture descriptor that characterizes the local patterns in an image by comparing the intensity of a central pixel with its neighboring pixels, encoding these patterns into binary values for texture analysis and object recognition.

(i) Local Binary Pattern (LBP) in 3D (using spherical harmonics) [*1*]

It computes the LBP in 3D using spherical harmonics.

The Pyradiomics 'imageoperations' module was used for each transformation. The italicized number beside each transformation indicates the number of images obtained. We have obtained a total of 26 images (including the original), per modality.

4.2 Feature Extraction and Dimensionality Reduction

The standard Pyradiomics features were extracted from each of the transformed and original images. These include:

(a) First Order Statistics (19 features)

(b) Shape-based (3D) (16 features)

(c) Shape-based (2D) (10 features)

(d) Gray Level Co-occurrence Matrix (24 features)

(e) Gray Level Run Length Matrix (16 features)

(f) Gray Level Size Zone Matrix (16 features)

(g) Neighbouring Gray Tone Difference Matrix (5 features)

(h) Gray Level Dependence Matrix (14 features)

A total of **120** features, were extracted for each image. The age of the patient was also added as a feature as it is an indicator of the patient's survival. Since, for each patient, we have **4** modalities, **26** transformations per modality, and 120 features per transformation, we obtain a total of 4*26*120 + 1= **12481** features.

For dimensionality reduction, multiple methods and statistics were considered. First, a univariate statistic, namely the Fisher score, was considered to reduce the number of features to **400**. It quantifies the separation between two or more groups of data points by assessing the variance between the group means relative to the variance within each group. The Fisher score ranking is an efficient algorithm, and this was necessary since the subsequent feature selection algorithms are computation-heavy.

Next, we employed four different methods for further dimensionality reduction. This includes Recursive Feature Elimination (RFE) with Random Forest Regressor, Minimum redundancy maximum relevance (mRMR), and mRMRe: ensemble approaches to better explore the feature space and build more robust predictors. The mRMRe method includes two implementations: mRMRe-Exhaustive (mRMRe.e) and mRMRe-Bootstrap (mRMRe.b). We obtained ten sets of features for the mRMRe methods. In each of these methods, the number of features was reduced to a variable ***nff***. We conducted the experiment for different values of *nff*.

Summing up, we extracted a total of 12481 features per patient and then applied a two-stage feature selection process: (a) Fischer-score based pruning to reduce feature pool size to 400, (b)

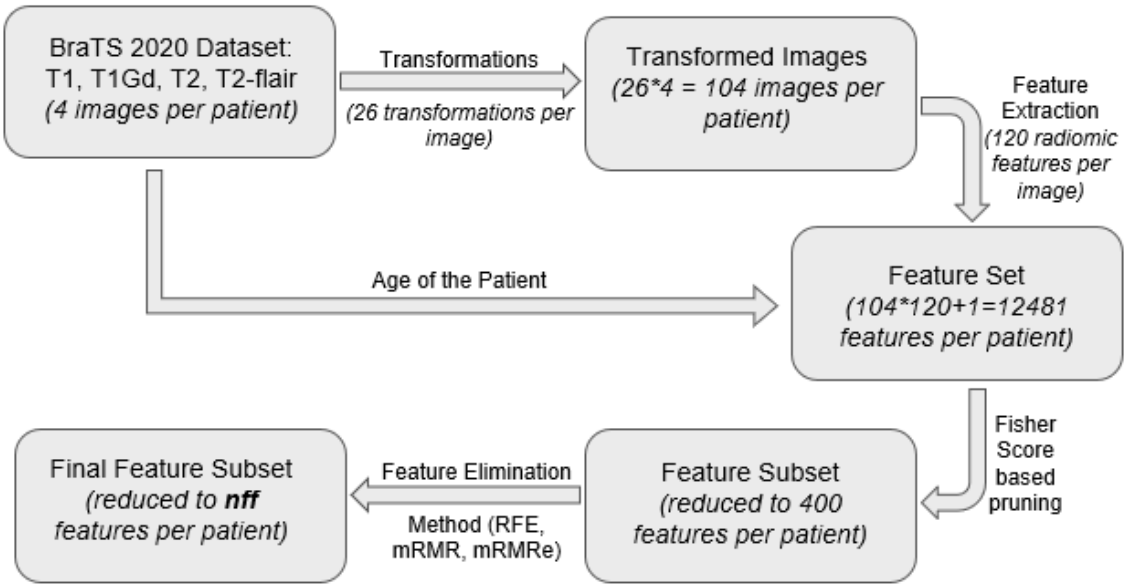

Figure 2: A flowchart explaining the feature extraction and dimensionality reduction process for Fischer score based pruning + {RFE, mRMR, mRMRe}

One of the four feature elimination techniques to reduce feature dimensionality to a smaller size, *nff*. Figure 2 is a summary of the entire process.

For feature selection using LASSO regression, an intermediate fischer score based filtering was not necessary since this is a computationally fast algorithm.

4.3 Model Development for Survival Prediction

We employed the following models for survival prediction after the feature selection stage: Random Forest Regressor, Cox-proportional Hazards model, and DeepSurv.

Only subjects with resection status of GTR (i.e., Gross Total Resection) were considered for training and validation. This yielded a training dataset of 117 samples and a validation dataset of 28 samples.

We also used the Local Spatial Relationship Awareness framework on the 2D slice images, the details of which have been mentioned below.

4.3.1 LSRAN model

Using the proposed framework, Minh-Trieu Tran et al. predict the survivor class within each slice image. Only the FLAIR MRI modality images were considered for classification. All the slices between index numbers 20 to 135 were taken into account.
The backbone was a modified ResNet-18 and the relation head comprised of two linear layers with batch-normalization in the first layer having activation functions of LeakyReLU and Sigmoid in the first and second layers, respectively. The feature maps produced in the initial training phase were of size 641. The number of augmentations per image were 32, batch size was 16 and the model was trained for 100 epochs. The augmentations used were random resized crop and random horizontal flip.
The final layer used for classification was a single linear layer with batch normalization and ReLU activation. The loss function for backbone training was focal loss (gamma = 2), while the loss function for classification was cross-entropy loss.
Employing 2D axial images for classification, a majority voting technique was applied to determine the survivor class for patient survival classification. Each patient was categorized as a short survivor (150 days), mid-survivor (375 days), or long survivor (525 days). Competitive results were obtained with an accuracy of 0.517 and a notably high SpearmanR correlation of 0.459 on the BraTS 2020 dataset validation sets.

4.3.2 Observations

We run the random forest regressor model for different values of the size of the final feature subset. The results are illustrated in Figure 3, where the different curves represent the accuracy vs final feature subset size for different models.
Table 3 illustrates the observations achieved on the validation dataset for the RFR model on the best final feature sub-sets produced using different techniques. The metrics used are mean and median square error, Spearman's rank correlation coefficient as well as accuracy, which was obtained by dividing the predicted value into three classes: long-survivors (e.g., >15 months), short-survivors (e.g., <10 months), and mid-survivors (e.g., between 10 and 15 months).

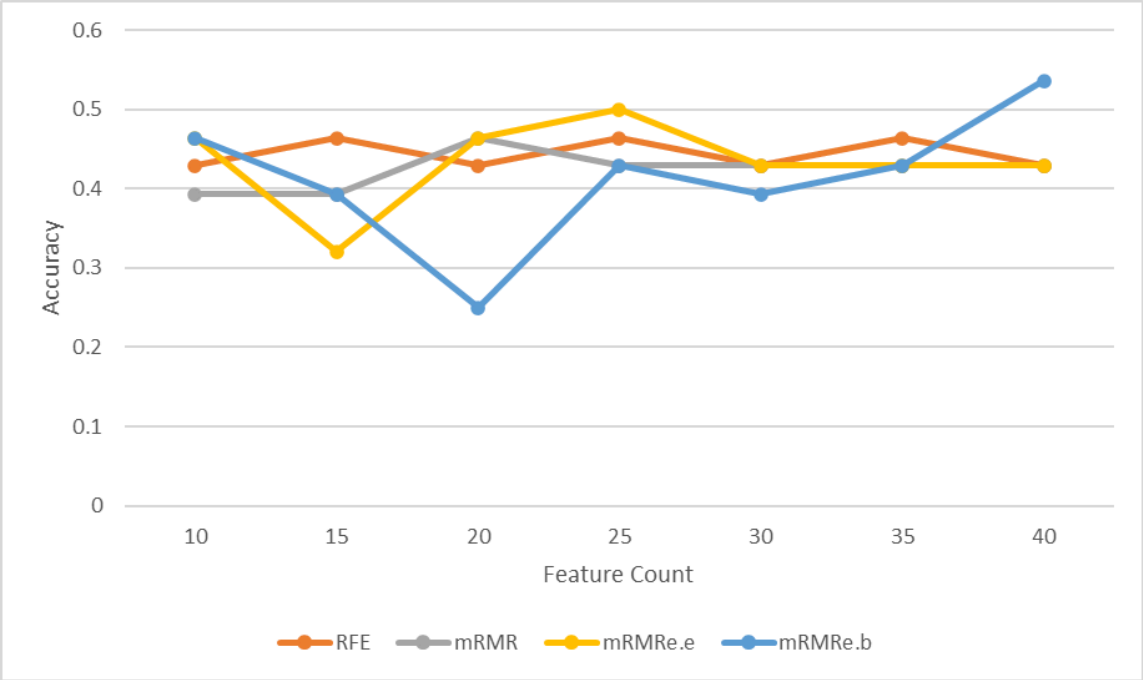

Figure 3: Classification accuracy vs feature subset size for different feature selection methods (model used is Random Forest Regressor)

Feature Selection Method	Final Feature Subset Size	Accuracy	MSE	medianSE	stdSE	SpearmanR
RFE	15	0.464	**99187.206**	29585.164	197125.954	0.142
mRMR	20	0.464	131324.487	58924.019	**153086.199**	0.151
mRMRe.e	25	**0.536**	106807.799	**21870.016**	225734.808	**0.504**
mRMRe.b	40	**0.536**	135138.072	81192.938	159172.243	0.124

Table 4: Observations on the Validation Dataset for RFR

Model	Accuracy	SpearmanR
Random Forest Regressor using image-based and radiomic features (Ali et al. (2021))	0.483	0.134
Random Forest Regressor using statistical and radiomic features (R. R. Agravat and M. S. Raval (2020))	0.517	0.217
Extra Trees (R. Miron et al. (2021))	0.414	0.321
Random Forest Regressor using radiomic features (V. K. Anand et al. (2021))	0.448	0.169
mRMRe.e Feature Selection, Random Forest Regression	**0.536**	**0.504**

Table 5: Comparison of our model against State of the Art models using radiomic features

For mRMRe.e and mRMRe.b, the feature set with the best training accuracy was chosen for the survival prediction task on the validation data. Table 4 presents a comparison of the models that use radiomic features against our best-performing model. The most favorable outcomes were obtained with the exhaustive variant of the mRMR ensemble approach, attaining an accuracy of 53.6%, surpassing the performance of models that rely on radiomic features (Table 4). Furthermore, it yielded the highest Spearman's rank correlation coefficient of 0.504 among all the models presented in Tables 1 and 2.

For feature selection using LASSO, we observed better accuracy for lower feature counts. Figure 4 illustrates the accuracy vs feature count plots for feature selection using LASSO and models: Random Forest Regressor (RFR) and Cox-proportional Hazards (CPH).

We observe excessive overfitting, especially in the CPH model. The best outputs obtained on applying RFR to the LASSO regression pruned feature subset is for a feature count of 5. We obtain a classification accuracy of 53.6% and a Spearman R coefficient of 0.369. The CPH model also yields the best accuracy for feature count 5, with an accuracy of 46.4% and Spearman R coefficient of 0.342.

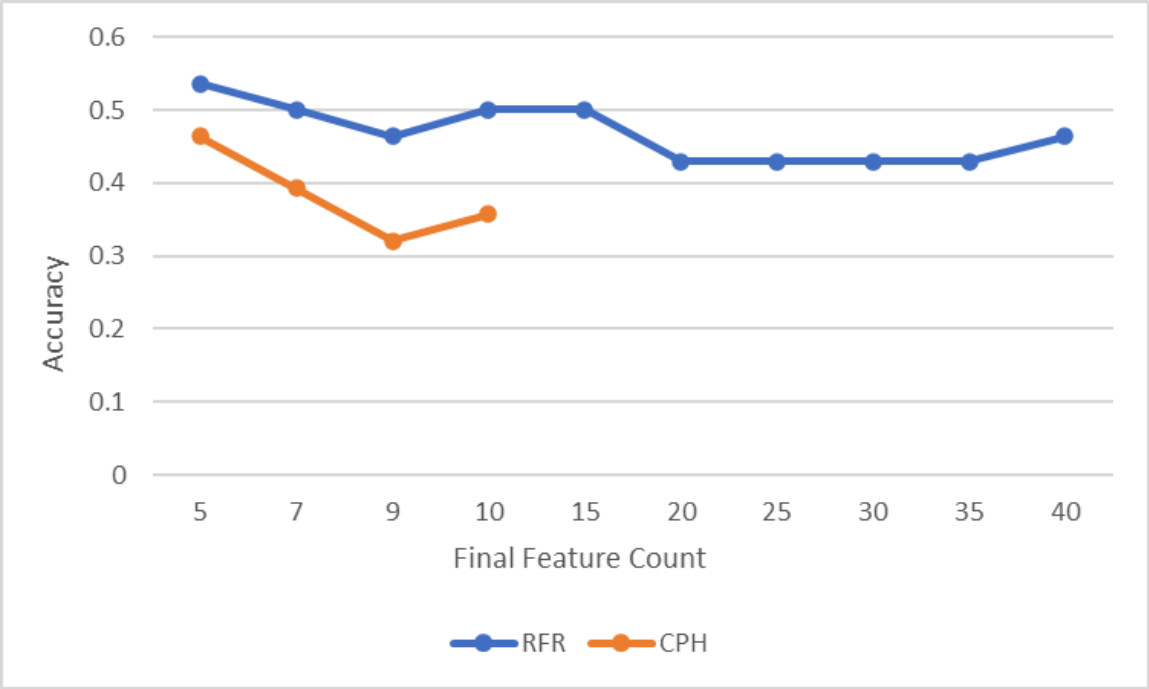

Figure 4: Classification accuracy vs feature subset size for models (RFR, CPH), pruned using LASSO regression

Final Feature Subset Size	RFR		CPH	
	Accuracy	SpearmanR	Accuracy	SpearmanR
5	**0.536**	0.369	**0.464**	**0.342**
7	0.5	0.35	0.393	-0.061
9	0.464	0.377	0.321	-0.301
10	0.5	0.458	0.357	-0.083
15	0.5	0.456		
20	0.429	0.44		
25	0.429	**0.496**		
30	0.429	0.454		
35	0.429	0.482		
40	0.464	0.215		

Table 6: Observations on the Validation Dataset for LASSO selected feature subsets using RFR & CPH (Note: The empty cells indicate unremarkable accuracy)

The DeepSurv model also exhibited overfitting. It was yielding low accuracy for feature count less than 100. The best results for different feature counts have been shown in Table 6.

Feature Count	Accuracy	MSE	medianSE	stdSE	SpearmanR
100	42.9%	565410.747	228286.232	676580.304	-0.1
500	42.9%	**186750.362**	**41600.618**	**337150.41**	0.219
1000	**46.4%**	255777.847	45129.561	453603.56	0.195
5000	39.3%	628744.364	180383.572	719573.763	**0.248**

Table 7: Observations on the Validation Dataset for DeepSurv

The Local Spatial Relationship Awareness Framework performed poorly as compared to the other models tried out in this thesis. We performed both averaging and majority voting to assimilate the results of all the slices. Table 8 illustrates the results of both.

Voting Technique	Accuracy	MSE	medianSE	stdSE	SpearmanR
Majority	0.286	141263.571	45162.5	237135.665	-0.306
Averaging	0.321	126440.766	28379.968	263429.646	-0.223

Table 8: Observations on the Validation Dataset for LSRAN

5 Conclusion

We successfully extracted a vast array of features using the Pyradiomics python package by performing various image transformations and then extracting the spectrum of standard radiomic features. Next, we performed a two-step feature pruning technique using the Fisher score statistic and multiple models to reduce the dimensionality down to a small figure.

The most favorable outcomes were obtained for a feature subset size of 25, with the exhaustive variant of the mRMR ensemble approach, attaining an accuracy of **53.6%**, surpassing the performance of models that rely on radiomic features (Table 4). Furthermore, it yielded the highest Spearman's rank correlation coefficient of **0.504** among all the models presented in Tables 1 and 2. We also obtained an accuracy of **53.6%** using mRMRe.b: the bootstrap variant of the ensemble approach, for a feature subset size of 40. We attained similarly impressive results with LASSO regression based feature pruning, achieving an accuracy of **53.6%** for a feature subset size of 5, using the RFR model. The cox-proportional hazards model displayed extensive overfitting, with training accuracy far succeeding the validation accuracy.

Both the DeepSurv model and the backbone trained in the LSRAN framework, failed to perform impressively with maximum accuracies of 46.4% and 32.1% respectively.

6 Future Works

Beyond radiomics features, one can delve into alternative feature extraction techniques such as statistical and image-based features to assess their impact on model performance. These features may encode essential information that can help increase the overall accuracy of the regression task. Statistical features can extract information such as the amount of edema, the extent of tumour, etc. (Agravat, R., & Raval, M. S. (2020)) while image-based features can include the position of the whole tumor center, the position of the enhancing tumor center, etc. (Ali et al.(2021)) that might have a high correlation with the expected survival time.

We can explore the integration of mRMR with other feature selection methods, such as wrappers, to develop a two-stage selection algorithm. This approach has the potential to demonstrate that the feature space identified by mRMR is more characterizing. (Peng et al. (2005))

Besides, alternative models beyond random forest regressor can be explored. As shown in Table 2, one can go beyond classical machine learning models to obtain better accuracy. Convolutional neural networks, with frameworks such as 3D DenseNet (S. R. González et al.) may be explored. The Swin Transformer UNet presents another viable option for extracting global features from MRI images. Leveraging a self-attention mechanism, the Swin Transformer architecture effectively captures the global dependencies between different parts of an image. Subsequently, the encoder outputs of this model can serve as deep features inputs to models like the cox-proportional hazards model or a random forest regressor. By integrating these global features with the local features obtained through an LSRAN framework and radiomic features, a holistic survival prediction approach may be achieved, amalgamating diverse feature types (M.-T. Tran et al.).

References

Ali, M.J., Akram, M.T., Saleem, H., Raza, B., Shahid, A.R. (2021). Glioma Segmentation Using Ensemble of 2D/3D U-Nets and Survival Prediction Using Multiple Features Fusion. *In: Crimi, A., Bakas, S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2020. Lecture Notes in Computer Science(), vol 12659. Springer, Cham.* *https://doi.org/10.1007/978-3-030-72087-2_17*

Patel, Jay & Chang, Ken & Hoebel, Katharina & Gidwani, Mishka & Arun, Nishanth & Gupta, Sharut & Aggarwal, Mehak & Singh, Praveer & Rosen, Bruce & Gerstner, Elizabeth & Kalpathy-Cramer, Jayashree. (2021). Segmentation, Survival Prediction, and Uncertainty Estimation of Gliomas from Multimodal 3D MRI Using Selective Kernel Networks. *10.1007/978-3-030-72087-2_20.*

Rosas Gonzalez, Sarahí & Zemmoura, Ilyess & Tauber, C.. (2021). 3D Brain Tumor Segmentation and Survival Prediction Using Ensembles of Convolutional Neural Networks. *10.1007/978-3-030-72087-2_21.*

Russo, C., Liu, S., & Di Ieva, A. (2020). Impact of Spherical Coordinates Transformation Pre-processing in Deep Convolution Neural Networks for Brain Tumor Segmentation and Survival Prediction. *ArXiv. /abs/2010.13967*

Feng, Y., Liu, S., Cheng, Z., Quiroz, J. C., Rezazadegan, D., Chen, P., Lin, Q., Qian, L., Liu, X., Berkovsky, S., Coiera, E., Song, L., Qiu, X., & Cai, X. (2021). Severity Assessment and Progression Prediction of COVID-19 Patients Based on the LesionEncoder Framework and Chest CT. *Information, 12(11), 471. https://doi.org/10.3390/info12110471*

Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol. 2005 Apr;3(2):185-205. doi: 10.1142/s0219720005001004. PMID: 15852500.*

Nicolas De Jay, Simon Papillon-Cavanagh, Catharina Olsen, Nehme El-Hachem, Gianluca Bontempi, Benjamin Haibe-Kains, mRMRe: an R package for parallelized mRMR ensemble feature selection, *Bioinformatics, Volume 29, Issue 18, September 2013, Pages 2365–2368, https://doi.org/10.1093/bioinformatics/btt383*

Barretina J, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature. 2012 Mar 28;483(7391):603-7. doi: 10.1038/nature11003. Erratum in: Nature. 2012 Dec 13;492(7428):290. Erratum in: Nature. 2019 Jan;565(7738):E5-E6. PMID: 22460905; PMCID: PMC3320027.*

Garnett MJ, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature. 2012 Mar 28;483(7391):570-5. doi: 10.1038/nature11005. PMID: 22460902; PMCID: PMC3349233.*

Agravat, R., & Raval, M. S. (2020). 3D Semantic Segmentation of Brain Tumor for Overall Survival Prediction. *ArXiv. /abs/2008.11576*

Miron, Radu & Albert, Ramona & Breaban, Mihaela. (2021). A Two-Stage Atrous Convolution Neural Network for Brain Tumor Segmentation and Survival Prediction. *10.1007/978-3-030-72087-2_25.*

Anand, V. K., Grampurohit, S., Aurangabadkar, P., Kori, A., Khened, M., Bhat, R. S., & Krishnamurthi, G. (2021). Brain Tumor Segmentation and Survival Prediction using Automatic Hard mining in 3D CNN Architecture. *ArXiv. /abs/2101.01546*

Han, I. S. (2021). Multimodal brain image analysis and survival prediction using neuromorphic attention-based neural networks. *In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6 (pp. 194-206). Springer International Publishing.*

Hanchuan Peng, Fuhui Long and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226-1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.*

Cox, D. R. (1971). Regression Models and Life-Tables. Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x*

Katzman, J., Shaham, U., Bates, J., Cloninger, A., Jiang, T., & Kluger, Y. (2016). DeepSurv: Personalized Treatment Recommender System Using A Cox Proportional Hazards Deep Neural Network. *ArXiv.* *https://doi.org/10.1186/s12874-018-0482-1*

Patacchiola, M., & Storkey, A. (2020). Self-Supervised Relational Reasoning for Representation Learning. *ArXiv. /abs/2006.05849*

Tran, Minh-Trieu & Yang, Hyung-Jeong & Kim, S.H. & Lee, Guee-Sang. (2023). Prediction of Survival of Glioblastoma Patients Using Local Spatial Relationships and Global Structure Awareness in FLAIR MRI Brain Images. *PP. 1-1. 10.1109/ACCESS.2023.3266771.*

Frank E Harrell, Kerry L Lee, Robert M Califf, David B Pryor, and Robert A Rosati. Regression modeling strategies for improved prognostic prediction. *Statistics in medicine, 3(2):143–152, 1984.*

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M. A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., Corso, J. J., … Van Leemput, K. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE transactions on medical imaging, 34(10), 1993–2024.* *https://doi.org/10.1109/TMI.2014.2377694*

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., & Davatzikos, C. (2017). Advancing The Cancer Genome Atlas glioma MRI

collections with expert segmentation labels and radiomic features. *Scientific data, 4, 170117.* *https://doi.org/10.1038/sdata.2017.117*

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R. T., Berger, C., Ha, S. M., Rozycki, M., Prastawa, M., Alberts, E., Lipkova, J., Freymann, J., Kirby, J., Bilello, M., Wiest, R., Kirschke, J., Wiestler, B., . . . Menze, B. (2018). Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. *ArXiv. /abs/1811.02629*