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

October 2023

Abstract

Machine Learning models are commonly used in medicine - for treatment recommendation, disease detection, survival prediction, etc. For patients suffering from brain tumour, we typically use MRI scans to assess the tumour. For a machine learning model to analyse the same, we can define standard features such as tumour shape, size, and colour. A popular set of features, specifically used for this purpose, are radiomic features. These features encode essential information about the MRI-scanned image that can then be used in the machine learning model developed. However, these features alone might not perform the best in standard research problems such as the Brain Tumor Segmentation (BraTS) Challenge. A study of the radiomic features along with a study of the models best applicable to tackle the survival prediction problem has been taken up in this thesis. We have performed multiple transformations on the four modalities of the MRI scans and extracted the entire spectrum of radiomic features from each such transformation. Successively, we have experimented with various feature selection and pruning algorithms, to reduce the dimensionality of our dataset to a suitable number that also performs the best on the specified model, while not being computationally too expensive. A random forest regressor has been used for the final prediction.

Contents

1 Introduction **1**
2 Literature Review **2**
2.1 State-of-the-art results 2
2.2 mRMR Algorithm 3
2.3 mRMRe: Ensemble Approach 4
3 Scope & Objectives **5**
4 Work Progress **5**
4.1 Data Acquisition and Transformation 5
4.2 Feature Extraction and Dimensionality Reduction 7
4.3 Model Development for Survival Prediction 8
4.3.1 Observations 9
5 Conclusion **9**
6 Future Works **10**
References **11**

1 Introduction

Radiomics is currently reshaping the landscape of medical imaging, enabling us to extract critical information from images to enhance diagnosis and treatment decisions. Applied across diverse fields of medicine, radiomics is revolutionizing the way we analyze medical images, with applications ranging from treatment recommender systems and disease diagnosis to prognosis prediction and beyond. Survival prediction is one such application that involves forecasting the likelihood and duration of a patient's survival based on various clinical features, such as medical history and biological markers.

Gliomas, the prevalent primary brain malignancies, underscore the significance of precise and resilient tumor segmentation and the prediction of overall patient survival. These factors play crucial roles in diagnostic accuracy, treatment strategy development, and the identification of risk factors. In this project, we attempt the survival prediction of patients with Glioblastoma using multi-modal Magnetic Resonance Imaging (MRI) radiomic features. MRI images are of 4 modalities: native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (T2-flair). Consequently, to obtain a vast pool of features, we performed multiple transformations on each modality and extracted radiomic features from each transformed image. We used Pyradiomics, the open-source Python package for these tasks. The number of transformed images per modality was 26, and the number of extracted radiomic features per image was 120. This yielded a total of 12481 features per patient. Extracting a large number of features per data point can enhance the model's ability to capture complex patterns, providing a more comprehensive representation of the data and potentially improving its predictive accuracy.

Our next task was that of dimensionality reduction. Due to the vast quantity of extracted features, pruning was essential to save time complexity and eliminate irrelevant features. Various feature pruning techniques were employed and the number of features per patient was subsequently reduced to 25. A comparison of each feature elimination technique has been presented in this thesis. A Random Forest Regression model was then employed for survival prediction on the Brain Tumor Segmentation (BraTS) Challenge 2020 dataset.

2 Literature Review

2.1 State-of-the-art results

In Ali et al. (2021), the authors extracted the radiomic and image-based features like volume ratio, surface area, etc. for the survival prediction task. The radiomic features were extracted by applying the Laplacian of Gaussian filters. The age of the patient was also added as a feature. To select the best-performing features, they used Random Forest - Recursive Feature Elimination (RF-RFE). Following the identification of significant features, they employed a Random Forest regressor with a grid search to forecast survival. They achieved an accuracy of 48.3% and a Spearman R coefficient of 0.134 on the validation dataset.

A Cox Proportional Hazards Model was used by Patel et al. (2021), trained on 2048 deep features extracted from their segmentation network. 40 feature versions were generated by processing all inputs and ensemble models, and then dimensionality was reduced via PCA to prevent overfitting. Age at diagnosis and predicted sub-region sizes were also incorporated. The Cox model was trained on 118 patients with gross total resection (GTR), utilizing 10 principal components for optimal validation set performance. An accuracy of 65.5% and a Spearman R coefficient of 0.479 were obtained in this case.

S. R. González et al. (2021) employed a 3D DenseNet CNN model, a modification of the 2D model by Huang et al., with 3D convolutions replacing 2D filters and Instance normalization replacing Batch normalization for improved segmentation results and optimized performance with reduced batch size. They received an accuracy of 55.2% and a Spearman R coefficient of 0.442.

Instead of relying on predefined imaging and radiomic features like volumetric parameters, intensity, morphological characteristics, histogram-based data, and textural properties, C. Russo et al. (2020) utilized automatically extracted MRI-based features from the innovative LesionEncoder (LE) framework (Feng Y-Z et al., 2021). The LE features underwent dimensionality reduction through Principal Component Analysis (PCA) and were subsequently employed as input for a generalized linear model (GLM) to forecast patient OS. Their best sub-model achieved an accuracy of 58.3% and a Spearman R coefficient of 0.412.

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

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

Table 2: Results of Models using features other than 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:

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

First, we rank the input features $X = \{x_1, \ldots, x_n\}$ 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*:

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

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*:

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

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.

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.

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 on 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 [*I*]

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 [*I*]

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) [*I*]

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 [*I*]

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 [*I*]

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) [*I*]

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) [*I*]

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

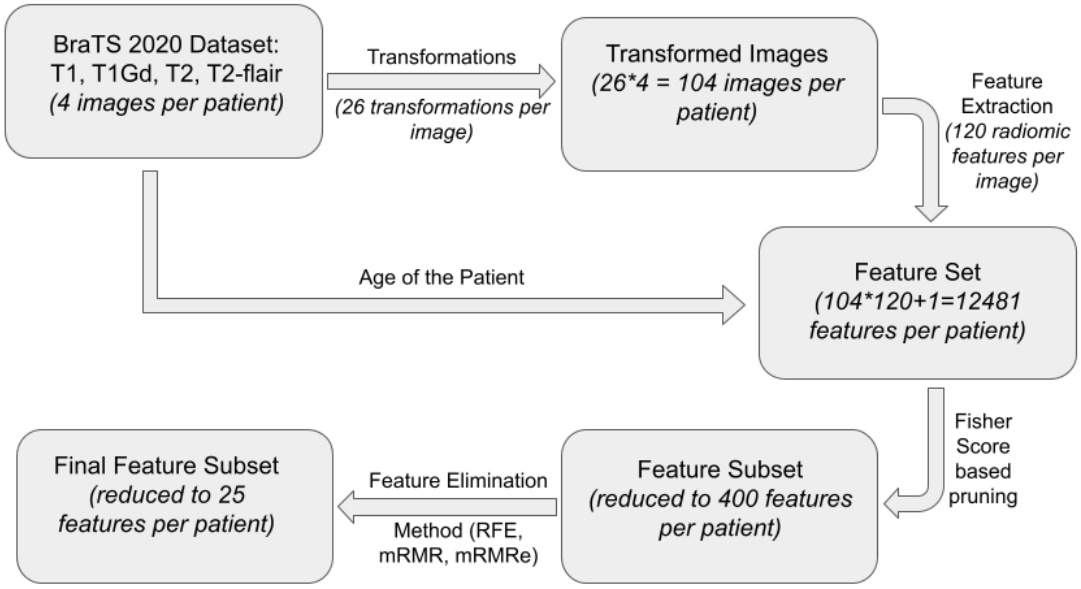


Figure 1: A flowchart explaining the feature extraction and dimensionality reduction process

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 25. Figure 1 is a summary of the entire process.

4.3 Model Development for Survival Prediction

A Random Forest Regressor model was deployed for survival prediction. For hyperparameter tuning, RandomizedSearchCV was used initially. For finer tuning, the values close to the outputs of RandomizedSearchCV were used as input to GridSearchCV. 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.

4.3.1 Observations

Table 3 illustrates the observations achieved for each model on the validation dataset. 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).

Feature Selection Method	Accuracy	MSE	medianSE	stdSE	SpearmanR
RFE	0.429	**97573.625**	31050.522	190274.716	0.037
mRMR	0.464	101398.232	30453.609	**153663.883**	0.144
mRMRe.e	**0.536**	106807.799	**21870.016**	225734.808	**0.504**
mRMRe.b	0.429	211166.272	109680.116	267685.771	-0.003

Table 3: Observations on the Validation Dataset for different models

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 4: 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 confusion matrices of the training dataset for the different models are shown in Figure 2.

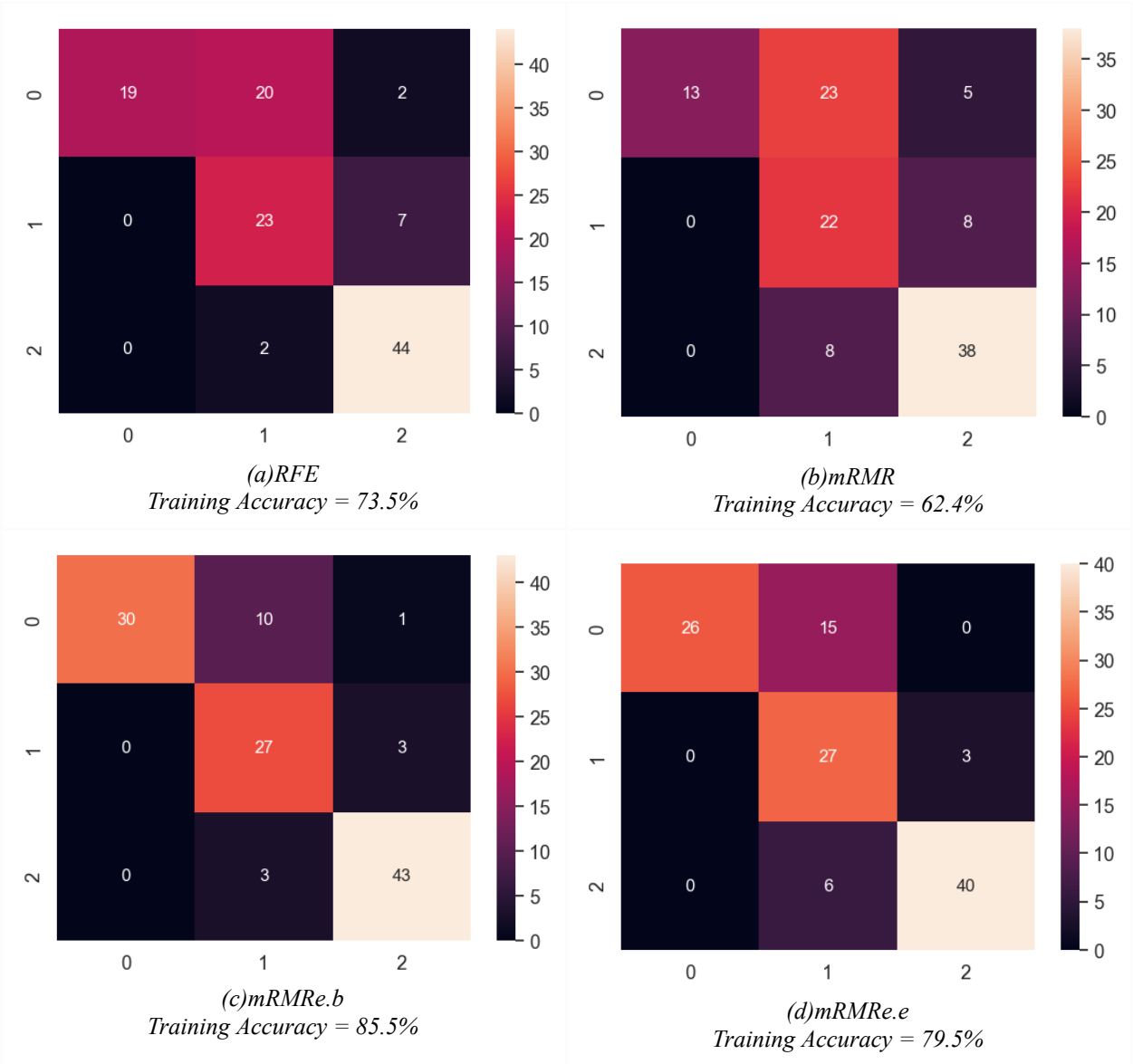
5 Conclusion

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 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.

Figure 2 also demonstrates that all the models have exhibited overfitting, as their training accuracy consistently surpasses their validation accuracy.



(a)RFE
Training Accuracy = 73.5%

(b)mRMR
Training Accuracy = 62.4%

(c)mRMRe.b
Training Accuracy = 85.5%

(d)mRMRe.e
Training Accuracy = 79.5%

Figure 2: Confusion Matrix for each model on the Training Dataset (true label along the x-axis and predicted label along the y-axis)

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))

The optimal number of features for survival prediction is another important factor in the model's predictive accuracy. This can be determined by constraining the feature elimination process to varying numbers of features and executing the model separately for each case.

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. Another such model is Deepsurv, a Cox proportional hazard model that presents a robust foundation for further investigations because of its ability to model and predict time-to-event data.

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.*