Survival Prediction of Malignant Brain Tumor Patients

Rajput Snehal[a], Rupal K. Kapdi[b], Kanhaiya Sharma [c], Mehul S. Raval[d] and Mohendra Roy[e,*]

[a]Computer Science and Engineering Pandit Deendayal Energy University Gandhinagar, India; [b]Institute of Technology Nirma University Ahmedabad, Gujarat, India;[c]Computer Science and Information Technology Symbiosis institute of Technology Pune constituent of Symbiosis International University Pune, Maharashtra, India ;[d]School of Engineering and Applied Science Ahmedabad University Ahmedabad, India and [e]Department of Information and Communication Technology Pandit Deendayal Energy Gandhinagar, India

ABSTRACT
Glioblastoma (GBM) is an fatal form of malignant tumor, and patients have a very low survival tenure. The patient's survival is very much dependent on the physiology of the tumor. However, predicting survival days based on manual inspection of MRI images is exceedingly difficult and pertain to qualitative error. Alternatively, an automated method may help the medical professional to diagnose GBM and to predict the overall survival (OS) days, which can further help expects planning treatments. In this regard, segmentation of the tumor cells from the whole brain MRI and OS prediction is very crucial. Researchers can use an end-to-end method, which can automatic segments tumor using radiological images and further extracts features to predict survival days accurately. The proposed work predict comparable OS days with current top performing methods. Furthermore, we observed the role and impact of dataset on the performance of model. Also, we examined and reason out the performance impact when targeting survival days prediction as classification problem. The accuracy, MSE, Spearman ranking coefficient on the BraTS-2020 training set were 53.8%, 60668.61, 0.754, and on the validation set they were 55.2%, 79826.24, 0.711 respectively. This is consistent with the top performing approaches in the BraTS-2020 competition on the validation dataset.

1. Introduction

An automatically generated brain-tumor segmentation and OS prediction of Glioblastoma or Glioblastoma-Multiforme (GBM) patients have recently received widespread attention from the research fraternity due to their critical nature (1). These tasks are regarded as among the most difficult in the medical domains (2). Accurate delineation of malignant tumor cells and survival days prediction are crucial and directly impact the cycle of treatment and post-treatment planning.

GBM is an exceptionally invasive type of brain tumour found in adults with a highly infiltrative and diffusive nature. MRI has traditionally been the most basic

---

*Corresponding Authors Email: mohendra.roy@ieee.org

imaging technology used to examine these types of cancers due to its non-invasive, high resolution and contrast nature.

Deep learning-based segmentation methods have consistently outperformed traditional methods in recent years. Specifically, 3D UNet (3) based approaches have been proposed to generate robust segmentation results. The Brain Tumor Segmentation (BTS) comprises of dissecting tumor into following regions of interest (ROI): Enhancing-tumor (ET), Tumor-core (TC) and Whole-tumor (WT). A Dice score (DS) is used to evaluate segmentation result. DS measures overlapping pixels of predicted and ground truth maps, whereas Hausdorff-Distance - 95% (HD) determines the 95% percentile of the distance from the set of points from the predicted map, which is the closest point from the ground truth. In general, various 3D Unet based single, cascaded (4) and/or ensemble (2) (5) (4) of multiple models are used for BTS. In the recent BraTS-2020 (Brain-Tumor segmentation) challenge, Isensee et al. (2) implemented an ensemble of multiple 3D Unet based model for segmenting brain-tumor. They incorporated region-based training, post-processing, and a wide range of data augmentation techniques. Further, In BraTS-2021 Luu et al. (5) and Futrega et al. (6) further enhanced the performance of the model with minimal modifications(2).

The above segmentation techniques suggest that automatic tumor segmentation is not only highly computation expensive but also extremely challenging problem due to high variance in shape, structure, location, texture of tumorous tissues, lack of sufficient standard dataset and an imbalance between tumorous/lesion and healthy/background areas.

After brain tumor segmentation, survival prediction is another pivot sub-task which has gained wide attention (7). In comparison to BTS task, it is equally challenging yet crucial in medical domain. The survival rates of GBM patients are poor, most of them succumb to death within two years of diagnosis (8). The Overall-survival (OS) prediction of BraTS competition consists of predicting survival days of tumor patients. For prediction, various features were extracted from the segmentation-map obtained from the BTS task and further classifying it into short-term survival (if the survival days is less than 300 days), mid-term survival (if the survival days is in between 300 and 450 days), and long-term survival (if survival days greater than 450 days).

Since, both the tasks are conjugal in nature, it is essential to develop an end-to-end model to target both the BTS and OS prediction collectively. Some of the end-to-end methods are : Huang et al. (7) proposed a V-net based encoder-decoder architecture with includes Squeeze-and-Excitation and attention module. Further, for OS prediction various radiomics (such as intensity, texture and wavelet) and deep CNN features were extracted. Finally, the reduced set of features was used as input to train random-forest regressor (RFR) model. Ali et al. (9) implemented an ensemble of both 3D and 2D models for segmentation, while image-based features and radiomics-based features were extracted from the input images and segmentation result for OS prediction. Further to reduce dimensions of features and overfitting, recursive feature elimination (RFE) was performed. At the end, these reduced features set were fed as input to train a RFR model for predict OS in numbers of days.

Literature review suggests that for OS prediction, geometrical, statistical, location, and texture features were extracted from each tumor regions. Further, feature selection was performed on the datasets. They were trained on different regressor – RFR, Support-Vector Machine (SVM), and Decision-Trees (DT) to predict survival days of GBM patients'. As discussed priorly that OS prediction is also a difficult task due various reasons such as: incomplete data of patients, small dataset, less clinical information on gender, health condition, treatment, capturing biological characteristics,

and qualitative image properties from radiographic images. Also, when targeting survival prediction as a classification problem, minor changes in prediction can misclassify the sample, which will hugely impact the classification accuracy.

The proposed paper used permutation importance (PI) (10) based feature selection techniques and RFR model for prediction. The OS prediction of the BraTS competition (11; 12; 13) focuses on predicting survival days and classifying the survival days into Long-Term, Mid-Term, and Short-Term categories. Here, accuracy and and Spearman ranking coefficient (Spearman Ranking) are mainly used to rank and access the model performance. In the suggested approach, the selected features performance is comparable to the current state-of-art methods. The following are the paper's primary contributions:

- Accurate and comparable OS prediction with current top performing methods, on validation BraTS-2020 dataset.
- In-depth analysis of the impact of dataset on model performance.
- In-depth analysis of model performance in-terms of OS prediction.
- In-depth performance impact analysis, when mapping survival days prediction into different categories.

This paper has the following structure: The BraTS dataset is detailed in Section 2, the proposed O.S prediction methodology in Section 3. Section 4 includes the results and discussions. And at last, the conclusion and future work to enhance the performance of the proposed methodological approach is presented in Section 5.

2. BraTS-2020 Dataset

The BraTS2020 (11; 12; 13) training dataset comprises of 369 samples images and meta-data (Age, resection status and survival days) of 236 patients for OS prediction. Whereas validation dataset contains 125 isamples images for segmentation and meta-data information of 29 sample for OS prediction. All the sample cases have GTR resection status. There are four MRI modalities in each sample: T1-weighted, contrast enhancing (T1-ce), T2-weighted, and (T2-FLAIR), and manually-labelled result. The class labels of segmentation results are: Label-1 represents Non-enhancing tumor (NET-ROI) and necrotic tumor (NCR-ROI), label-2 represents edema (ROI), label-4 represents enhancing-tumor (ET-ROI), and label 0 represents background pixels. The dimensions of each image are : $240 \times 240 \times 155$ (width, height, and channels).

3. O.S Prediction Methodology

The proposed O.S prediction methodology can be seen in Figure 1.

Since, tumor segmentation is pre-requisite for O.S prediction, we have implemented a 3D network for segmenting brain tumor, which had U-Net-like architecture (14). It has the most straightforward architecture and was one of the best-performing segmentation models of the BraTS-2017 challenge. The details about the structure of model can be found here (14). The dice-scores obtained on the BraTS 2020 training-set are 0.819 for WT, 0.766 for TC, and 0.702 for ET. For validation-set DS are: 0.880(WT), 0.858(TC) and 0.759(ET).

Further, for predicting OS days, we extracted : image, radiomic-based features and trained Random forest regressor (RFR) model.
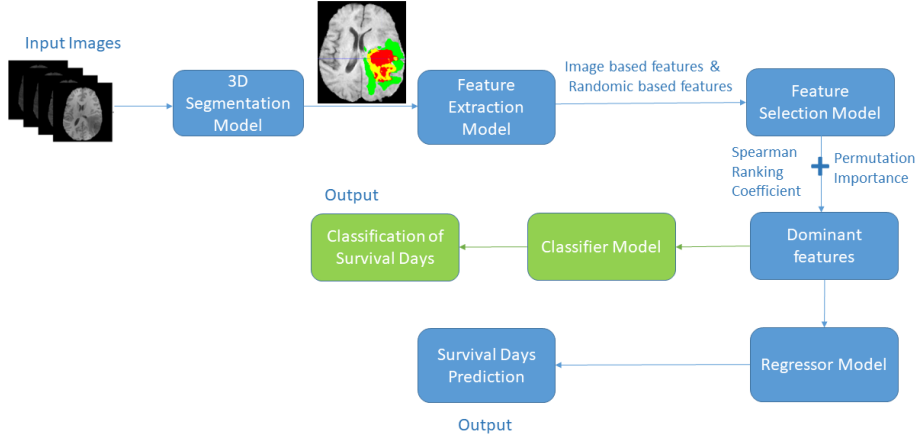
Figure 1.: The diagrammatic representation of the proposed framework for OS prediction.

3.1. Feature-Extraction and Feature-Selection

3.1.1. Feature-Extraction : (15; 16; 17)

Feature-extraction is a method to derive new features space from the original feature set. We extracted 1265 features which can be categorized into : 1) image-based features - 39, 2) radiomics-based features using LoG and wavelet filters (18) - 1226. Numbers beside feature categories show total numbers of features extracted.

Table 1.: Features extraction set

Image-based features	Shape based features (such as surface area, volume, proportion of tumor regions..), location based features (centroid of tumor regions..)
Radiomics features	Shape features 3D (such as flatness, elongation..), first-order statistical features (such as entropy, energy..), gray-level features (such as gray-level size-zone(GLSZ), gray-level co-occurrence matrix(GLCM)..)

3.1.2. Feature Selection

The fundamental purpose of feature-selection methods is to identify a subset of input features that have unique information to distinguish the target feature. We used permutation importance (PI) as feature selection techniques (19) to obtain 29 dominant features.

4

3.2. Regression and Classification Framework

We have trained RF regressor model to predict OS days and RF classifier to classifying into these three categories (long-term, mid-term and small-term survivor). RFR has proved to be more successful because of the following reasons: 1. The output prediction is the mean of the prediction of all the individual trees, and 2. The introduction of randomness in the growth of the trees and the splitting of the trees (20) reduces generalized errors and overcomes overfitting. Also, in the studies mentioned in (21), where the authors assessed 179 different classifiers on 121 different datasets, they found that RF outperforms all the other classifiers. The datasets also include survival prediction of breast cancer patients.

We have used only those samples whose resection status is GTR and five cross-validations to train the model in the training phase. Furthermore, we have used grid search to find set of optimal hyper-parameters.

4. Results and Discussion

All the results were obtained through BraTS online evaluation platform (22).

4.1. Quantifying Performance of the model

In the dominant feature set obtained through PI (feature selection method), it has been observed that significant features are based on wavelet filter and the Laplacian of Gaussian filter. The wavelet transform can capture spatial and global informations (23). Whereas, LoG filter which is widely used in biomedical image analysis, can enhance structural or edge information (24).

The accuracy of the RFR model on training are 53.5% and on validation sets are 55.2%, respectively. Table 3 presents the outcomes.

Table 2.: Performance metrics on training and validation BraTS-2020 dataset for OS prediction.

Dataset-2020	Accuracy	MSE	Spearman ranking
Training	53.8%	60668.60	0.75
Validation	55.2%	79826.20	0.71

4.2. Evaluation of RFR model for OS prediction

Since the BraTS OS prediction is a classification task, we classified the survival days from the training dataset into small-term (class 0), mid-term (class 1), long-term (class 2) categories. For short-term survival, the survival days are less than 300 days; for mid-term survival, the survival days are in between 300 to 450 days; and for long-term survival, the survival days are more than 450 days. The BraTS dataset don't include these categories but only survival days. Figure 2 shows a classification of survival days into categories and the distribution for the training dataset. The box plots for these categories are shown in Figure 3.

We can observe from Figure 2, that for small-term category the distribution of samples are very sparse for e.g. under 50 survival days, there are only two samples with vast difference in survival days. These can also be justify through the Figure 3 (a), where median survival days for small-term category is 143.5 days. Whereas for
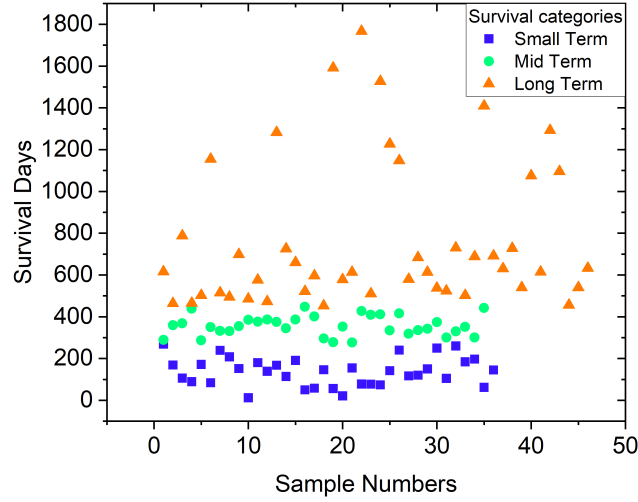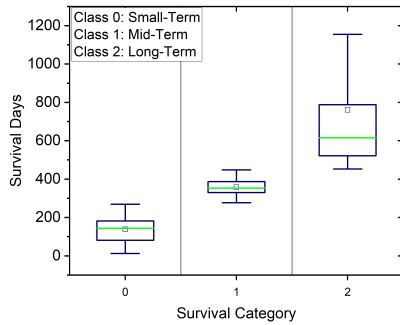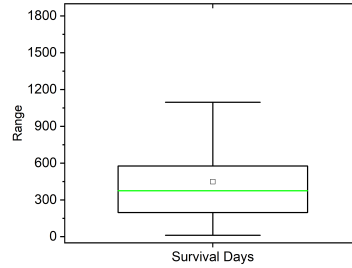
Figure 2.: Distribution of survival days categories i.e., Small-Term, Mid-Term and Long-Term from the training dataset. Samples with GTR resection status are plotted. $N = N_0 + N_1 + N_2 = 36 + 35 + 46 = 117$, where N_0, N_1, N_2 are the number of samples in class 0, 1, 2 respectively and the overall number of samples is N.

(a) Box-plot depicts survival days into different categories

(b) Box-plot for Survival days

Figure 3.: (a) A box-plot showing the distribution of long-term, mid-term and short-term survival categories. Green line of the whisker shows the median value of the respective classes. Box in the whisker shows the mean value of the respective classes. The median value of long-term, mid-term and short-term are 615.5, 353 and 143.5 respectively. (b) A box-plot showing the overall distribution of survival days from training data.

mid-range samples have even distribution and range value of classification is also less compare to other two categories. Whereas, sample distribution and range of survival days without categorisation can be seen in Figure 3 (b). Similarly, observing mean value of each categories validate the distribution of samples. Hence for the same reason, the prediction of our model is more accurate for the mid-term category, followed by small-term and long-term categories. A comparison of the predicted and ground-truth

survival days for the 30 test samples from the training samples is shown in Figure 4. We recognise that the suggested model exhibits a good degree of generalisation for survival days ranging from 180 to 950 days. The reason is that 1. The maximum number of data points of survival days falls in this range (cf. Figure 3(b)). 2. There are not enough samples to train the model for better approximation in any other range.
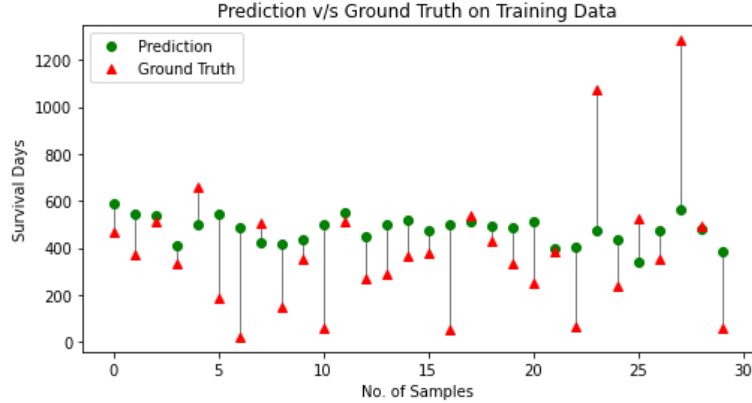

Figure 4.: RFR prediction on test samples taken randomly from the training dataset.

The overall accuracy of the OS prediction depends on how accurately the model predicts survival days for each of the three categories. Figure 5 depicts comparison plots between predicted and ground-truth survival days for each category to demonstrate the proposed model's classification performance. The quantitative evaluation of the proposed model on all the three categories on the training dataset is shown in Table 4. The MSE for the prediction of survival days in the small-term, mid-term, long-term categories are: 52762.20, 10850.27, and 101711.33, respectively. We observe that MSE is maximum for the long-term survival category compared to other survival categories. The reason is that for misclassified training samples, the difference between the predicted and ground-truth survival days is more for the long-term survival category compared to the other two categories. Whereas, the obtained accuracy is more for long-term survival category, followed by mid-term and least for the small-term survival category. The reasons for that are: 1. Since the sample distribution i.e., range is more for long term and so is the error margin, it results in the possibility of the large variation between the predicted and ground-truth falling in the same category. 2. The number of samples is more for long-term category compared to the other two categories.

Table 3.: Performance metrics on training and validation BraTS-2020 dataset for OS prediction.

Dataset-2020	Accuracy	MSE	Spearman ranking
Training	53.80%	60668.60	0.75
Validation	55.20%	79826.20	0.71

Further, to validate the results, we have used receiver operating characteristic (ROC) curve which is shown in Figure 6. It displays both true-positive rate (TPR) and false-positive rate (FPR). From the figure, we can observe that the predictability of the model for the mid-term and long-term survival category is better than short-
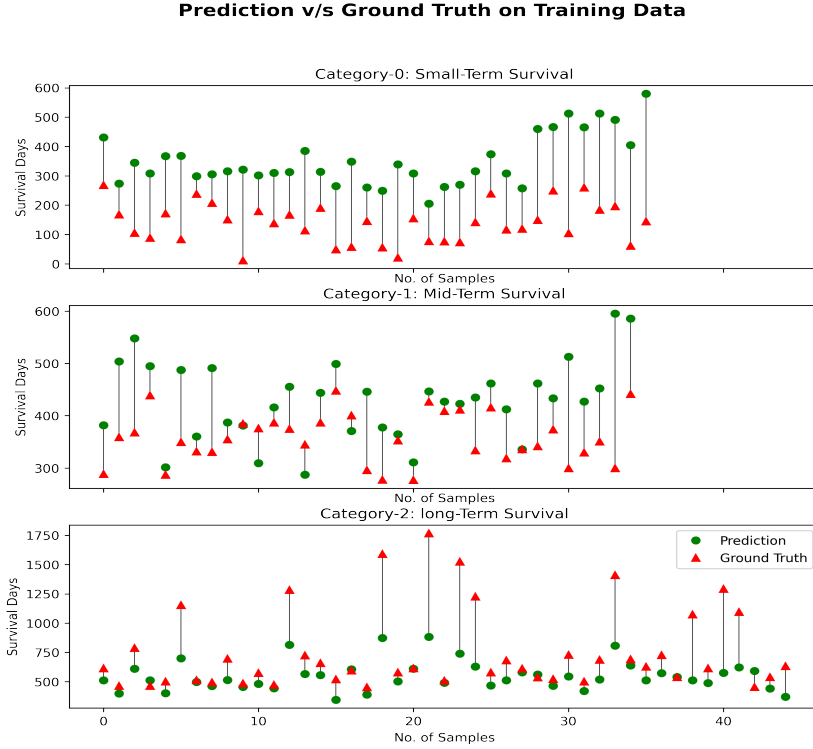
Figure 5.: Comparison plots between predicted and ground-truth survival days for small-term, mid-term and long-term survival.

term survival, which supports the results we are getting from the RFR regressor. The area underneath the curve is 0.57 for mid-term and long-term whereas the area for the small-term is 0.48 which shows our model has a reasonable discriminatory ability for the mid-term and long-term categories and very little discriminatory capability for the small-term category. The reason for the lower discriminatory capability is: 1. The discontinuity in the spread of samples is least for this category. 2. A smaller number of training samples.

Finally, a performance comparison of the RFR model with top-ranking models of BraTS-2020 competition on the training and validation datasets has been noted in Table 5. The performance metrics of the top-ranking models was obtained through the BraTS validation leaderboard (25).

Our model has shown a robust performance since it has performed significantly better in all the mentioned evaluation metrics. Also, the training and validation per-

Table 4.: Category-wise performance evaluation on training dataset BraTS-2020 challenge.

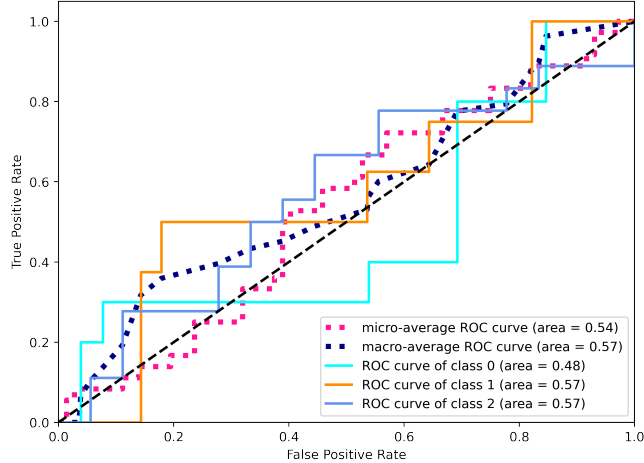| Category	Accuracy	MSE	Spearman ranking
Small-Term	25.00%	052762.20	0.31
Mid-Term	45.70%	010850.30	0.28
Long-Term	82.20%	101711.00	0.75

8

Figure 6.: ROC curve to multi-class using RFR classifier model (class-0 Small-Term survival, class-1 Long-Term survival and class-2 represents Mid-Term survival).

Table 5.: On the training and validation BraTS-2020 dataset, performance was compared to top-ranking models.

Dataset 2020	Team-name	Accuracy	Mean squared error (MSE)	Spearman-Ranking
Training	SCAN (26)	NA	NA	NA
	Redneucon (27)	82.20%	55499.71	0.833
	COMSATS-MIDL (9)	64.10%	62305.61	0.632
	Proposed	53.80%	60668.61	0.754
Validation	SCAN (26)	41.40%	098704.65	0.253
	Redneucon (27)	52.00%	122515.80	0.130
	COMSATS-MIDL (9)	48.30%	105079.40	0.134
	Proposed	**55.20%**	**079826.24**	**0.711**

Note: Team names and ranking were taken from BraTS challenge leaderboard (25) and ranking platform (28).

formance of our model is close to each other, signifying that the model is robust and generalizing well.

5. Conclusion and Future work

In this work, we proposed an in-depth analysis of the impact of dataset distribution on the performance of the model. Further, we inspect the performance of model on each instance and analysed it to reason its impact on instances in-terms of error margins. Also, we examine the performance of model when targeting O.S prediction as category problem. We explained the reason for the true-positive and false-positive classifications of the model. For regression and classification, we have trained random forest model with the dominant feature set obtained through PI feature selection

method. This feature selection methods quantified the role of each feature selected and hence unravelled the reason for performance of the model. This RF regressor model performed well for mid-term O.S prediction followed by small-term and long-term whereas RF classifier performed well for long-term and mid-term. The performance of the model can be increased by: 1. including more location-based features 2. larger dataset 3. improved segmentation results, which are used for feature extraction.

Acknowledgments

References

[1] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, and A. Crimi, "arxiv [preprint]. arxiv: 1811.02629," 2018.

[2] F. Isensee, P. F. Jaeger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, "nnu-net for brain tumor segmentation," *arXiv preprint arXiv:2011.00848*, 2020.

[3] 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*. Springer, 2015, pp. 234–241.

[4] H. Jia, W. Cai, H. Huang, and Y. Xia, "H2nf-net for brain tumor segmentation using multimodal mr imaging: 2nd place solution to brats challenge 2020 segmentation task," *arXiv preprint arXiv:2012.15318*, 2020.

[5] H. M. Luu and S.-H. Park, "Extending nn-unet for brain tumor segmentation," in *International MICCAI Brainlesion Workshop*. Springer, 2022, pp. 173–186.

[6] M. Futrega, A. Milesi, M. Marcinkiewicz, and P. Ribalta, "Optimized u-net for brain tumor segmentation," in *International MICCAI Brainlesion Workshop*. Springer, 2022, pp. 15–29.

[7] H. Huang, W. Zhang, Y. Fang, J. Hong, S. Su, and X. Lai, "Overall survival prediction for gliomas using a novel compound approach," *Frontiers in Oncology*, vol. 11, p. 724191, 2021.

[8] S. Witthayanuwat, M. Pesee, C. Supaadirek, N. Supakalin, K. Thamronganantasakul, and S. Krusun, "Survival analysis of glioblastoma multiforme," *Asian Pacific journal of cancer prevention: APJCP*, vol. 19, no. 9, p. 2613, 2018.

[9] M. J. Ali, T. Akram, B. R. Hira Saleem, and A. R. Shahid, "Glioma segmentation using ensemble of 2d/3d u-nets and survival prediction using multiple features fusion."

[10] M. K. MIT and K. Lopuhin, "permutation_importance," Aug 1965. [Online]. Available: https://eli5.readthedocs.io/en/latest/blackbox/permutation_importance.html

[11] 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.

[12] 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.

[13] 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.

[14] 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*. Springer, 2017, pp. 287–297.

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

[16] 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 Workshop*. Springer, 2019, pp. 131–141.

[17] 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 *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2021, pp. 189–199.

[18] 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.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. VanderPlas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *CoRR*, vol. abs/1201.0490, 2012.

[20] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, M. S. Lauer *et al.*, "Random survival forests," *Annals of Applied Statistics*, vol. 2, no. 3, pp. 841–860, 2008.

[21] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" *The journal of machine learning research*, vol. 15, no. 1, pp. 3133–3181, 2014.

[22] "Onlineevaluation," https://ipp.cbica.upenn.edu/, accessed: 2021-06-12.

[23] S. P. Singh and S. Urooj, "Wavelets: biomedical applications," *International Journal of Biomedical Engineering and Technology*, vol. 19, no. 1, pp. 1–25, 2015.

[24] H. Kong, H. C. Akakin, and S. E. Sarma, "A generalized laplacian of gaussian filter for blob detection and its applications," *IEEE transactions on cybernetics*, vol. 43, no. 6, pp. 1719–1733, 2013.

[25] "Validation survival leaderboard 2020," https://www.cbica.upenn.edu/BraTS20//lboardValidationSurvival.html, accessed: 2021-06-12.

[26] 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," *arXiv preprint arXiv:2012.06436*, 2020.

[27] J. M. Asenjo and A. M.-L. Solís, "MRI brain tumor segmentation using a 2d-3d u-net ensemble," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, 2021, pp. 354–366.

[28] "Brats-2020 ranking on survival prediction," https:https://www.med.upenn.edu//cbica/brats2020/rankings.html, accessed: 2021-06-12.