Deep Learning Versus Classical Regression for Brain Tumor Patient Survival Prediction

Yannick Suter[1]([✉]), Alain Jungo[1], Michael Rebsamen[1], Urspeter Knecht[1], Evelyn Herrmann[2], Roland Wiest[3], and Mauricio Reyes[1]

[1] Institute for Surgical Technology and Biomechanics,
University of Bern, Bern, Switzerland
`yannick.suter@istb.unibe.ch`
[2] University Clinic for Radio-Oncology, Inselspital, Bern University Hospital,
University of Bern, Bern, Switzerland
[3] Support Center for Advanced Neuroimaging,
University Institute of Diagnostic and Interventional Neuroradiology, Inselspital,
University of Bern, Bern, Switzerland

Abstract. Deep learning for regression tasks on medical imaging data has shown promising results. However, compared to other approaches, their power is strongly linked to the dataset size. In this study, we evaluate 3D-convolutional neural networks (CNNs) and classical regression methods with hand-crafted features for survival time regression of patients with high-grade brain tumors. The tested CNNs for regression showed promising but unstable results. The best performing deep learning approach reached an accuracy of 51.5% on held-out samples of the training set. All tested deep learning experiments were outperformed by a Support Vector Classifier (SVC) using 30 radiomic features. The investigated features included intensity, shape, location and deep features.

The submitted method to the BraTS 2018 survival prediction challenge is an ensemble of SVCs, which reached a cross-validated accuracy of 72.2% on the BraTS 2018 training set, 57.1% on the validation set, and 42.9% on the testing set.

The results suggest that more training data is necessary for a stable performance of a CNN model for direct regression from magnetic resonance images, and that non-imaging clinical patient information is crucial along with imaging information.

Keywords: Brain tumor · Survival prediction · Regression · 3D-Convolutional Neural Networks

1 Introduction

High-grade gliomas are the most frequent primary brain tumors in humans. Due to their rapid growth and infiltrative nature, the prognosis for patients with

gliomas ranking at grade III or IV on the Word Health Organization (WHO) grading scheme [17] is poor, with a median survival time of only 14 months. Finding biomarkers based on magnetic resonance (MR) imaging data could lead to an improved disease progression monitoring and support clinicians in treatment decision-making [10].

Predicting the survival time from pre-treatment MR data is inherently difficult, due to the high impact of the extent of resection (e.g., [18,23]) and response of the patient to chemo- and radiation therapy. The progress in the fields of automated brain tumor segmentation and radiomics have led to many different approaches to predict the survival time of high-grade glioma patients. Further, the introduction of the survival prediction task in the BraTS challenge 2017 [4,19] makes a direct performance comparison of methods possible. The current state-of-the-art approaches can roughly be classified into

1. Classical radiomics: Extracting intensity features and/or shape properties from segmentations and use regression techniques such as random forest (RF) regression [6], logistic regression, or sparsity enforcing methods such as LASSO [25].
2. Deep features: Neural networks are used to extract features, which are subsequently fed into a classical regression method such as logistic regression [7], support vector regression (SVR), or support vector classification (SVC) [14].
3. A combination of classical radiomics and deep features (e.g., [15]).
4. Survival regression from MR data using deep convolutional neural networks (CNNs) with or without additional non-imaging input (e.g., [16]).

Our experiments with 3D-CNNs for survival time regression confirmed observations made by other groups in last year's competition (e.g., [16]), that these models tend to converge and overfit extremely fast on the training set, but show poor generalization when tested on the held-out samples. The top-ranked methods of last year's competition were mainly based on RF. A reason for this may be the relatively few samples to learn from. Classical regression techniques typically have fewer learnable parameters compared to a CNN and perform better with sparse training data.

We present experiments ranging from simple linear models to end-to-end 3D-CNNs and combinations of classical radiomics with deep learning to benchmark new, more sophisticated approaches against established techniques. We believe that a thorough comparison and discussion will provide a good baseline for future investigations of survival prediction tasks.

2 Methods

2.1 Data

The provided BraTS 2018 training and validation datasets for the survival prediction task consist of 163 and 53 subjects, respectively. The challenge ranking is based on the performance on a test dataset with 77 subjects with gross total resection (GTR).

A subject contains imaging and clinical data. The imaging data includes images from the four standard brain tumor MR sequences (T1-weighted (T1), T1-weighted post-contrast (T1c), T2-weighted, and T2-weighted fluid-attenuated inversion-recovery (FLAIR)). All images in the datasets are resampled to isotropic voxel size $(1 \times 1 \times 1 \, \text{mm}^3)$, size-adapted to $240 \times 240 \times 155 \, \text{mm}^3$, skull-stripped, and co-registered. The clinical data comprises the subject's age and resection status. The three possible resection statuses are: (a) gross total resection (GTR), (b) subtotal resection (STR), and (c) not available (NA).

Segmentation: For our experiments, we rely on segmentations of the three brain tumor sub-compartments (i.e., enhancing tumor, edema, and necrosis combined with non-enhancing tumor). In the validation and testing dataset, the segmentation is not provided due to the overlap with the data of the BraTS 2018 segmentation task. To obtain the required segmentations, we thus employ the cascaded anisotropic CNN by Wang et al. [26]. Their method is publicly available[1] and contains pre-trained models on the BraTS 2017 training dataset, which is identical to the BraTS 2018 [2,3,5] training dataset. This enables us to compute the segmentations with the available models without retraining a new segmentation network.

2.2 Deep Survival Prediction and Deep Features

Two different CNNs are built for the survival regression task (see Fig. 1). CNN1 consists of five blocks with an increasing number of filters, each block has two convolutional layers and a max pooling operation. The last block is connected to two subsequent fully connected layers. CNN2 consists of three convolutional layers with decreasing kernel sizes with intermediary max-pooling, followed by fully-connected layers connected to the single value regression target. To include clinical information into the CNN2, the age and resection status were appended to the first fully-connected layers of CNN2, which we refer to as CNN2+Age+RS.

Both CNN variants take the four MR sequences and additionally the corresponding segmentation (see Sect. 2.1) as input, and output the predicted survival in days. We observed no performance gain by the additional segmentation input but it improved the training behavior of the network. Instead of regressing the survival days, we also tested direct classification in long-, mid-, and short-term survival, but without improvements.

We trained the CNNs with the Adam optimizer [13] and a learning rate of 10^{-5}, and performed model selection based on Spearman's rank coefficient on a held-out set. Batch normalization and more dropout layers did not lead to improvements, neither on the training behaviour nor the results.

Deep Feature Extraction: For the extraction of deep features, the size of the two last fully connected layers are decreased to 100 and 20 elements. The activations of these two layers serve as deep feature sets.

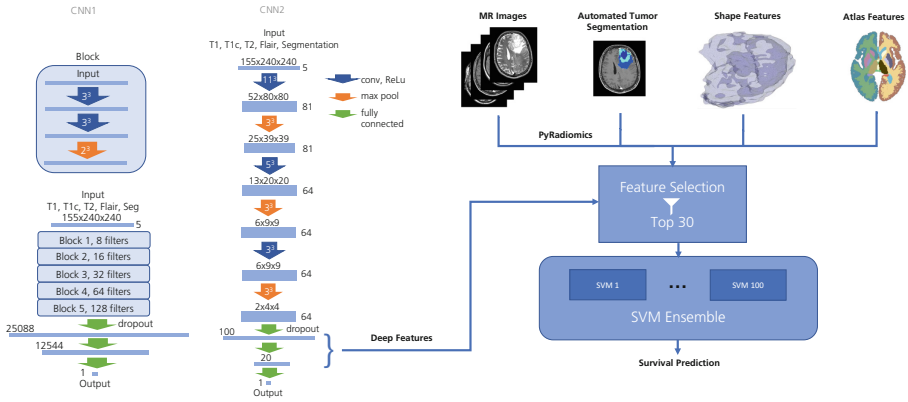[1] https://github.com/taigw/brats17.

Fig. 1. Summary of the tested methods for GBM patient survival predictions. Left: The architectures of our CNNs for direct survival regression. CNN2 was additionally used to extract deep features from the fully connected layers. For direct regression, the two last fully connected layers of CNN2 had 2048 and 384 elements. Right: Combination of classical radiomics, shape, and atlas features. The top 30 features were used to predict survival classes with a SVC.

2.3   Classical Survival Prediction

Feature Extraction: We extract an initial set of 1353 survival features from the computed segmentation together with the four MR images (i.e., T1, T1c, T2, and FLAIR).

Gray-Level and Basic Shape: 1128 intensity and 45 shape features are computed with the open-source Python package *pyradiomics*[2] version 2.2.0 [11]. It includes shape, first-order, gray level co-occurrence matrix, gray level size zone matrix, gray level run length matrix, neighbouring gray tone difference matrix, and gray level dependence matrix features. Z-score normalization and a Laplacian of Gaussian filter with $\sigma = 1$ is applied to the MR images before extraction. A bin width of 25 is selected and the minimum mask size set to 8 voxels. The features are calculated from all MR images and for all tumor sub-compartments provided by the segmentation (i.e., enhancing tumor, edema, necrosis combined with non-enhancing tumor).

Shape: 15 additional enhancing tumor shape features previously used as predictors for survival [12,21] complement the basic shape features from *pyradiomics*. These features are the rim width of the enhancing tumor, geometric heterogeneity, combinations of rim width quartiles and volume ratios of all combinations of the three tumor compartments.

---

[2] https://github.com/Radiomics/pyradiomics.

Atlas Location: Tumor location has previously been used for survival prediction (e.g., [1]), therefore atlas location features are included. Affine registration is used to align all subjects to FreeSurfer's [8] *fsaverage* subject and its subcortical segmentation (*aseg*) is used as the atlas. The volume fraction of each anatomical region occupied by the contrast enhancing tumor is used as a feature, resulting in 43 features in total.

Clinical Information: The two provided clinical features resection status and age are further added to the feature set.

Feature Selection: Since the number of extracted features (n = 1353) is much higher than the available samples (n = 163), a subset of features needs to be used. Apart from being necessary for many machine learning methods, a reduction of the feature space improves the interpretability of possible markers regarding survival [20].

We analyzed the following feature selection techniques to find the most informative features: (a) step wise forward/backward selection with a linear model, (b) univariate feature selection, and (c) model-based feature selection by the learned feature weights or importances. We observed a rather low overlap among the selected features by the different techniques, or even the parameterization of the techniques. Consequently, we chose the feature subsets according to their performance on the training dataset for different classical machine learning methods (e.g., linear regression, SVC, and RF). The best results were obtained by the feature subset produced by the model-based feature selection from a sparse SVC model, which consists of the features listed in Table 1.

Our model-based feature selection identified age by far as most important feature. Additionally, a majority of the 30 selected features are intensity-based, but the subset also contains shape and atlas features. We note that none of the 120 deep features was retained.

Feature-Based Models: Although the BraTS survival prediction task is set up as a regression task, the final evaluation is performed on the classification accuracy of the three classes: short-term (less than 10 months), mid-term (between ten and 15 months), and long-term survivors (longer than 15 months). As a consequence, we include classification models in addition to the regression models in our experiments. Since the prediction is required in days of survival, the output of the classifiers needs to be transformed from a class (i.e., short-term, mid-term, long-term) to a day scalar. We do this by replacing each class by its mean time of survival (i.e. 147, 376, 626 days).

For our experiments, we consider the following feature-based regression and classification models [9]:

– Linear and logistic regression
– RF regression and classification
– SVR and SVC
– SVC ensemble

Table 1. Selected feature set with feature category (Cat.), tumor sub-compartment (Comp.), and the MR image (Img.) in decreasing order of importance. ED: Edema, ET: Enhancing tumor, NCR/NET: Necrosis and non-enhancing tumor. Feature importance is decreasing from top to bottom.

Feature	Cat.	Comp.	Img.
Age	Clinical		
Sphericity	Shape	ED	
Optic chiasm	Atlas		
Small area low gray level emphasis	Intensity	ED	T2
Correlation	Intensity	CE	Flair
Cluster shade	Intensity	NCR/NET	T1c
Small dependence high gray level emphasis	Intensity	CE	T1c
Correlation	Intensity	NCR/NET	T1
Maximum	Intensity	ED	T1
Maximum	Intensity	ED	T1c
Left amygdala	Atlas		
Information measure of correlation 1	Intensity	NCR/NET	T1
Large dependence low gray level emphasis	Intensity	NCR/NET	T1c
Cluster shade	Intensity	ED	T2
Inverse variance	Intensity	ED	T1
Small dependence high gray level emphasis	Intensity	CE	T2
Median	Intensity	CE	T2
Busyness	Intensity	ED	T1
Correlation	Intensity	NCR/NET	T1c
Right-vessel	Atlas		
Large area low gray level emphasis	Intensity	ET	Flair
Right caudate	Atlas		
Difference variance	Intensity	ED	Flair
Right cerebellum cortex	Atlas		
Cluster prominence	Intensity	ED	T2
Maximum 2D diameter slice	Intensity	ED	
Inverse difference normalized	Intensity	CE	Flair
Skewness	Intensity	ED	Flair
Median	Intensity	ET	T1
Right ventral diencephalon	Atlas		

We use 50 trees and an automatic tree depth for the RF models and linear kernels for the support vector approaches, SVR, and SVC. To handle the multi-class survival problem we employ the *one-versus-rest* binary approach for SVC and logistic regression. The ensemble method consists of 100 SVC models that

are separately built on random splits of 80% of the training data. The final class prediction is performed by majority vote. We choose an ensemble to increase robustness against outliers or unrepresentative subjects in the training set. All classical feature-based models are implemented with *scikit-learn*[3] version 0.19.1.

2.4   Evaluation

We evaluated the classical feature-based approaches by 50 repetitions of a stratified five-fold cross-validation on the BraTS 2018 training dataset. These repetitions allowed us to examine the models' robustness besides their average performance. The CNN approaches were evaluated on a randomly defined held-out split of the training set, consisting of 33 subjects. This held-out set was also used to evaluate a subset of the feature-based methods in order to compare classical approaches to the CNN approaches. Moreover, the classical and CNN models were evaluated on the BraTS 2018 validation set. This dataset contains 53 subjects but only the 28 subjects with resection status GTR are evaluated. Finally, we selected the best-performing model to predict survival on the BraTs 2018 challenge test dataset, which consists of 77 evaluated subjects with GTR resection status (out of 130 subjects).

3   Results

In this section, we compare the performance of the CNN to the classical feature-based machine learning models on the BraTS 2018 training and validation datasets, and present the BraTS 2018 test set results. We introduced a reference baseline for the comparison of the different models. This baseline consists of a logistic regression model solely trained on the age feature. This minimal model provides us with a reference for the training and validation set.

Table 2 lists the results of the different models on the training dataset. To ensure a valid comparison, the table is subdivided by the two evaluation types, repeated cross-validation (CV) and hold-out (HO) (see Sect. 2.4). The results from the CV analysis highlights that by far the best results are achieved by the logistic regression, SVC, and ensemble SVC models, which performed very similarly. Except for the RF model, the classification models clearly outperformed their regression counterparts. The results from the HO analysis (Table 2, bottom) additionally reveals that well-performing classical methods (logistic regression and SVC) outperform all three CNN approaches (CNN1, CNN2, CNN2+Age+RS) by a large margin.

Table 3 presents the results obtained on the validation dataset. We can observe similar patterns as for the training set results: the classification models outperform the regression models with respect to the accuracy (except the RF), the SVC models (i.e., SVC ensemble and SVC) achieve the best performances,

---

[3] http://scikit-learn.org/stable/index.html.

Table 2. Results achieved on the BraTS 2018 training dataset by 100 stratified five-fold cross-validation (CV) runs (reported as mean ± standard deviation) and on one split with 33 held-out (HO) samples. The baseline consists of a logistic regression model with age as single feature. Best results per metric and evaluation type (Eval.) are presented in bold. Acc.: Accuracy, MSE: Mean squared error, r_S: Spearman's rank coefficient, RS: Resection status.

Eval.	Method	Acc.	MSE/days^2	r_S
CV	Baseline	0.489 ± 0.06	136323 ± 44378	0.300 ± 0.14
	Linear regression	0.552 ± 0.08	260706 ± 647176	0.573 ± 0.12
	SVR	0.554 ± 0.08	257542 ± 637062	0.574 ± 0.12
	RF regression	0.444 ± 0.08	117320 ± 42503	0.332 ± 0.17
	Logistic regression	0.721 ± 0.07	$\mathbf{93158 \pm 35665}$	$\mathbf{0.617 \pm 0.12}$
	SVC	$\mathbf{0.722 \pm 0.07}$	93571 ± 35861	0.612 ± 0.12
	RF	0.512 ± 0.07	136334 ± 47392	0.324 ± 0.15
	SVC ensemble	0.720 ± 0.07	93485 ± 35652	$\mathbf{0.617 \pm 0.12}$
HO	Logistic regression	0.697	30756	0.579
	SVC	**0.727**	**28226**	**0.616**
	CNN1	0.515	50598	0.298
	CNN2	0.424	56496	0.235
	CNN2+Age+RS	0.394	61798	−0.194

Table 3. Results achieved on the BraTS 2018 validation dataset (28 samples). The baseline consists of a logistic regression model with age as single feature. Best results per metric are presented in bold. Acc.: Accuracy, MSE: Mean squared error, r_S: Spearman's rank coefficient, RS: Resection status.

Method	Acc.	MSE/days^2	r_S
Baseline	0.464	128841	0.288
Linear regression	0.464	89059	0.426
SVR	0.464	89035	0.426
RF regression	0.393	80980	0.342
Logistic regression	0.5	90791	0.393
RF	0.357	169782	-0.058
SVC	0.536	85471	0.501
SVC ensemble	**0.571**	**79381**	**0.556**
CNN1	0.370	172821	0.104
CNN2	0.394	157617	−0.112
CNN2+Age+RS	0.444	137912	−0.005

and the CNNs remain behind the feature-based methods and the baseline. Additionally, we observe an overall decrease in performance compared to the training set results.

The results of CNN1 on our validation split (accuracy of 0.515) could not be replicated on the BraTS validation set, where it performed poorly with an accuracy of 0.37. CNN2 showed worse results on our validation split than the deeper CNN1, but performed better on the BraTS validation set.

Overall, the SVC ensemble performed best on the training and validation set and we consequently selected it for the challenge, where our method achieved an accuracy of 0.429, a mean squared error of 327725 days^2 and a Spearman's rank coefficient of 0.172.

4   Discussion

In this section, we discuss the presented results and highlight findings from the deep learning, and classical regression and classification experiments.

CNNs: The two CNNs overfit very fast on the training data, and showed highly variable performance between epochs. Model selection during training was therefore challenging, since both the accuracy and Spearman's rank coefficient were very unstable.

We postulate that more data would be needed to fully benefit from direct survival estimation with 3D-CNNs. When inspecting the filters of CNN1 and CNN2, most of the learning took place at the fully connected layers and almost none at the first convolutions layer. This effect and the fast overfitting of the CNN models indicate the lack of samples and are reasons for the poor performance on unseen data.

Classical Regression and Classification: Using classical regression techniques with hand-crafted features has the advantage of better interpretability. Models with fewer learnable parameters, such as the classical regression methods we tested, typically achieve more robust results on unseen data when only few training samples are available.

The atlas used for feature extraction most likely has too many regions for the number of training samples. Small anatomical structures, such as the optic chiasm, may not be accurately identified by simple registration to an atlas. Figure 2 shows the distribution of the contrast enhancing tumor segmentation per survival class. The short survivors with large contrast enhancing tumor loads contribute highly to the overall cumulative occurrence in the training data. The class-wise occurrence maps suggest that more training samples are needed to detect predictive location patterns (e.g. as reported in [22,24]). Additionally, a coarser atlas subdivision driven by clinical knowledge is in order. In the light of this caveat, the location features used here should be seen as approximate localization information with limited clinical interpretability.
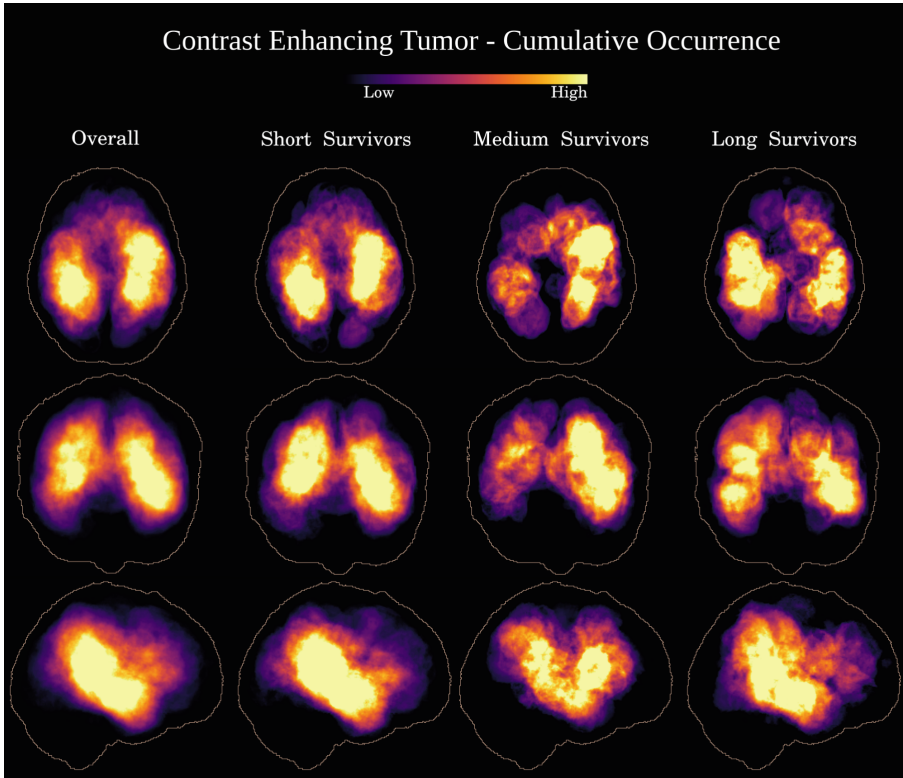
Fig. 2. Cumulative occurrence of the contrast enhancing tumor. Columns, from left to right: Overall across all three survival classes, short survivors (<10 months), medium survivors (≥10 months and ≤15 months), and long survivors (>15 months). Rows: Projection along axial, coronal and sagittal axes.

Performance on the Testing Data: The accuracy of 72.2% and 57.1% on the training and validation set could not be maintained on the testing data. The large performance drop might be caused by still too many features compared to the training set size. Other possible reasons may include a lack of feature robustness or different class distribution compared to the training data. Moreover, the survival time distributions within classes do not drop at the class boundaries, such that a small shift in the prediction can cause a large accuracy difference because ending up in a different class.

In conclusion, classical machine learning techniques using hand-crafted features still outperform deep learning approaches with the given data set size. The robustness of features regarding image quality and across MR imaging centers needs close attention, to ensure that the performance can be maintained on unseen data. We hypothesize that adding post-treatment imaging data and more clinical information to the challenge dataset would boost the performance of the survival regression.

References

1. Awad, A.W., et al.: Impact of removed tumor volume and location on patient outcome in glioblastoma. J. Neuro Oncol. **135**(1), 161–171 (2017). https://doi.org/10.1007/s11060-017-2562-1

2. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. Cancer Imaging Arch. (2017). https://doi.org/10.1038/sdata.2017.117

3. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. Cancer Imaging Arch. (2017). https://doi.org/10.1038/sdata.2017.117

4. Bakas, S., Reyes, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. ArXiv e-prints, November 2018

5. Bakas, S., et al.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data **4**, 170117 (2017). https://doi.org/10.1038/sdata.2017.117

6. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression trees (1984)

7. Cox, D.R.: The regression analysis of binary sequences. J. R. Stat. Society Ser. B (Methodol.), **20**(2), 215–242 (1958)

8. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012). https://doi.org/10.1016/j.neuroimage.2012.01.021

9. Hastie, T., Friedman, J., Tibshirani, R.: The Elements of Statistical Learning. SSS, vol. 1. Springer, New York (2001). https://doi.org/10.1007/978-0-387-21606-5

10. Gillies, R.J., Kinahan, P.E., Hricak, H.: Radiomics: images are more than pictures, they are data. Radiology **278**(2), 563–577 (2015). https://doi.org/10.1148/radiol.2015151169

11. van Griethuysen, J.J., et al.: Computational radiomics system to decode the radiographic phenotype. Cancer Res. **77**(21), e104–e107 (2017). https://doi.org/10.1158/0008-5472.CAN-17-0339

12. Jungo, A., et al.: Towards uncertainty-assisted brain tumor segmentation and survival prediction. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 474–485. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_40

13. Kinga, D., Adam, J.B.: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR), vol. 5 (2015)

14. Lampert, C.H., et al.: Kernel methods in computer vision. Found. Trends® Comput. Graph. Vis. **4**(3), 193–285 (2009). https://doi.org/10.1561/0600000027

15. Lao, J., et al.: A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. Sci. Rep. **7**(1), 10353 (2017). https://doi.org/10.1038/s41598-017-10649-8

16. Li, Y., Shen, L.: Deep learning based multimodal brain tumor diagnosis. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 149–158. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_13

17. Louis, D.N., et al.: The 2016 world health organization classification of tumors of the central nervous system: a summary. Acta Neuropathol. **131**(6), 803–820 (2016). https://doi.org/10.1007/s00401-016-1545-1

18. Meier, R., et al.: Automatic estimation of extent of resection and residual tumor volume of patients with glioblastoma. J. Neurosurg. **127**(4), 798–806 (2017). https://doi.org/10.3171/2016.9.JNS16146

19. Menze, B.H., Jakab, A., Bauer, S., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694

20. Pereira, S., et al.: Enhancing interpretability of automatically extracted machine learning features: application to a RBM-random forest system on brain lesion segmentation. Med. Image Anal. **44**, 228–244 (2018). https://doi.org/10.1016/j.media.2017.12.009

21. Pérez-Beteta, J., et al.: Glioblastoma: does the pre-treatment geometry matter? A postcontrast T1 MRI-based study. Eur. Radiol. (2017). https://doi.org/10.1007/s00330-016-4453-9

22. Rathore, S., et al.: Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond idh1. Sci. Rep. **8**(1), 5087 (2018). https://doi.org/10.1038/s41598-018-22739-2

23. Sanai, N., Polley, M.Y., McDermott, M.W., Parsa, A.T., Berger, M.S.: An extent of resection threshold for newly diagnosed glioblastomas. J. Neurosurg. **115**(1), 3–8 (2011). https://doi.org/10.3171/2011.2.JNS10998

24. Steed, T.C., et al.: Differential localization of glioblastoma subtype: implications on glioblastoma pathogenesis. Oncotarget **7**(18), 24899 (2016). https://doi.org/10.18632/oncotarget.8551

25. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Society Ser. B (Methodol.), **58**(1), 267–288 (1996)

26. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 178–190. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_16