Deep Learning Radiomics Algorithm for Gliomas (DRAG) Model: A Novel Approach Using 3D UNET Based Deep Convolutional Neural Network for Predicting Survival in Gliomas

Ujjwal Baid[1], Sanjay Talbar[1], Swapnil Rane[2], Sudeep Gupta[2],
Meenakshi H. Thakur[2], Aliasgar Moiyadi[2], Siddhesh Thakur[1],
and Abhishek Mahajan[2(✉)]

[1] Shri Guru Gobind Singhji Institute of Engineering and Technology,
Nanded, India
ujjwalbaid0408@gmail.com
[2] Tata Memorial Hospital, Tata Memorial Centre, Mumbai, India
drabhishek.mahajan@yahoo.in

Abstract. Automated segmentation of brain tumors in multi-channel Magnetic Resonance Image (MRI) is a challenging task. Heterogeneous appearance of brain tumors in MRI poses critical challenges in diagnosis, prognosis and survival prediction. In this paper, we present a novel approach for glioma tumor segmentation and survival prediction with Deep Learning Radiomics Algorithm for Gliomas (DRAG) Model using 3D patch based U-Net model in Brain Tumor Segmentation (BraTS) challenge 2018. Radiomics feature extraction and classification was done on segmented tumor for overall survival (OS) prediction task. Preliminary results of DRAG model on BraTS 2018 validation dataset demonstrated that the proposed method achieved a good performance with Dice scores as 0.88, 0.83 and 0.75 for whole tumor, tumor core and enhancing tumor, respectively. For survival prediction, 57.1% accuracy was achieved on the validation dataset. The proposed DRAG model was one of the top performing models and accomplished third place for OS prediction task in BraTS 2018 challenge.

Keywords: Brain Tumor Segmentation · Gliomas ·
Convolutional Neural Networks · Radiomics · MRI · Radiogenomics ·
Survival prediction

1 Introduction

Glioma is the most frequent primary brain tumor. It originates from glial cells and can be classified in High Grade and Low Grade depending upon the aggressiveness. Gliomas may have different degrees of aggressiveness, variable prognosis and several heterogeneous histological sub-regions. These are described by varying intensity profiles across different Magnetic Resonance Imaging (MRI) modalities, which reflect

diverse tumor biological properties [9]. De-spite of recent advances in automated algorithms for brain tumor segmentation in multimodal MRI scans, the problem is still a challenging task in medical imaging analysis [6, 7, 12].

Prior to BRATS challenge, researchers tested their proposed algorithms on local datasets and there was no gold standard available for fair evaluation of methods globally. BraTS challenge provided global platform for researchers to evaluate their proposed algorithms on publically available dataset with leaderboard. This year BraTS challenge was divided in two parts 1. Segmentation of brain tumor with intra-tumor parts 2. Prediction of overall survival of the patients in number of days based on imaging features.

Many computational methods based on texture analysis, probabilistic models, active contours, random forests are proposed for tumor segmentation over decades [10]. Several advances were made in active contours where either an initial seed point was mentioned which would grow till the boundaries of the tumor or a bounding box was drawn across the abnormal region which would further confine to tumor boundaries. Figure 1 shows FLAIR, T1, T1ce, and T2 images with intra-tumor parts- Green for Edema, Blue for Enhancing tumor and Red for Tumor Core. Researchers had proposed methods based on Non-Negative Matrix Factorization where a data matrix was generated from MR data which acted as a feature representation. This data matrix was further clustered with Fuzzy C-means clustering algorithm for brain tumor segmentation [14].
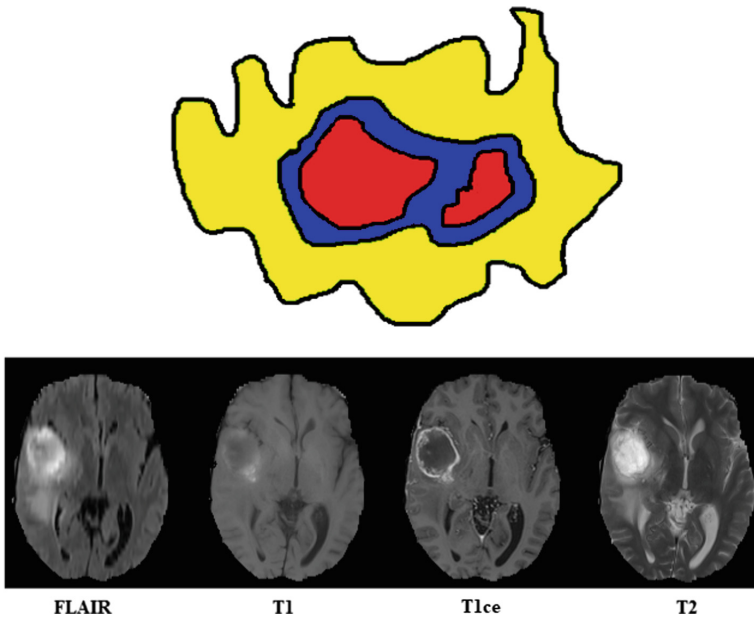
Fig. 1. MRI modalities with intra-tumor parts. Edema in yellow, enhancing tumor in blue and necrotic tumor is shown in red color (Color figure online)

Deep Learning algorithms have been outperforming over all other the state of the art methods for segmentation, classification and detection applications. Researchers have proposed application specific models for which Convolutional Neural Networks (CNN) is the basic building block of the architecture. The advantage of CNN is that it is computationally cheaper compared to Fully Convolutional Networks (FCN). The implementation of CNN is successful because of the advancement in the computational power of the machines. This enables to de-sign neural network models with deep architecture to extract features in an image. Researchers had proposed several pixel classification based approaches for segmentation task where a window was considered around a pixel and the class of the window was the class of center pixel. U-Net based models have outperformed over traditional machine learning methods in bio-medical image segmentation [11]. Recently, there has been an increase in popularity of 3D CNNs which are effective in segmentation task at the expense of additional computational complexity compared to other state of the art algorithms [5].

2   Method

We developed patch based 3D U-Net model for tumor segmentation and evaluated efficiency of radiomic features for OS prediction named as 'Deep Learning Radiomics Algorithm for Gliomas (DRAG) Model'. There was high class imbalance between tumor pixel and rest of the normal brain pixels in the BraTS datasets. This led to biased training of the model as the loss function was affected by normal brain pixels as compared to the tumor pixels. The problem became more challenging during intra-tumor segmentation. To overcome this issue, we adopted a patch-based training approach. Fixed sized 3D patches were extracted from the BraTS dataset which were used for training the network. Details of our approach are given in the section below.

2.1   Dataset

This proposed method was trained and validated on BraTS 2018 training dataset and validation dataset [1–3]. The training dataset included 210 High Grade Glioma (HGG) cases and 75 cases with Low Grade Glioma (LGG) while validation set consisted of 66 cases. For each case, there were four MRI sequences viz. the T1-weighted (T1), T1 with gadolinium enhancing contrast (T1ce), T2-weighted (T2) and FLAIR. All cases had been segmented manually, by four raters and marked annotations were approved by experienced neuro-radiologists into intra-tumor parts like tumor core, enhancing tumor and edema. The MRI data was collected from various institutions and acquired with different protocols, magnetic field strengths and MRI scanners. Furthermore, to pinpoint the clinical relevance of this segmentation task, BraTS 2018 also focused on the prediction of patient overall survival via analysis of radiomic features. For this purpose, the survival data (in days) of 163 cases was provided in training set and 54 cases in validation set. Reference segmentation and OS for validation set was hidden and evaluation was carried out via online evaluation portal.

2.2   Pre-processing

The MRI data in BraTS challenge dataset was already pre-processed which included skull stripping and the data was co-register and re-sampled to 1 mm × 1 mm × 1 mm resolution. The dimensions of each volume were 240 × 240 × 155. The intensity in-homogeneity was addressed with N4ITK tool [13]. All four MRI channel data was normalized to zero mean and unit variance.

2.3   Patch Extraction and Training

The proposed model is modified version of 3D U-Net with 3 down-sampling and 3 up-sampling branches with two back to back convolution layers with kernel size 3. 3D voxels with size 64 × 64 × 64 were extracted randomly from the training data and given as an input to the first layer of the model. Four patches extracted from FLAIR, T1, T2, T1c were concatenated together to form 64 × 64 × 64 × 4 and fed for training to the input layer along with corresponding Ground Truth. Patch extraction was challenging because of the high class imbalance between tumor area and normal brain tissues. During patch extraction, care was taken to include significant tumor area to avoid bias to background and non-tumor pixels. This was done for all the four modalities and ground truth as well. Each layer was followed by ReLU activation and Batch Normalization. No data augmentation was performed during the training of model.

At output 4 probability maps were generated for Necrosis, Edema, Enhancing Tumor and Background (including non-tumor brain pixels). The label was assigned to the map with highest probability amongst all. It was observed that there were some False Positives present in the segmentation output.

3D Connected Component Analysis was done to identify all the segmented components pre-sent in the segmented volume. Threshold value in terms of number of pixels was identified and insignificant small components which false positives were assigned to background label. This reduced false positives significantly. Similarly, to reduce over-segmentation in certain cases a binary brain mask was generated from brain volume and logical AND operation was performed on segmentation output. This improved the accuracy of the segmentation significantly.

2.4   Radiomic Feature Extraction and Training

After segmentation of intra-tumor parts, the next task in BraTS 2018 was to predict the over-all survival of the patients in number of days. For this task, organizers had provided only age details and OS in days which made the task challenging. From the last few years, researchers are working actively on Radiomic Feature extraction for tumor analysis and survival prediction task [4]. In our approach, we computed Radiomic features on FLAIR and T1c volume with different combination of intra-tumor parts as (Figs. 1 and 2):

- Edema, Enhancing tumor and tumor core i.e. Whole tumor (WT)
- Tumor Core and Enhancing tumor (TC+ET)
- Enhancing tumor (ET)
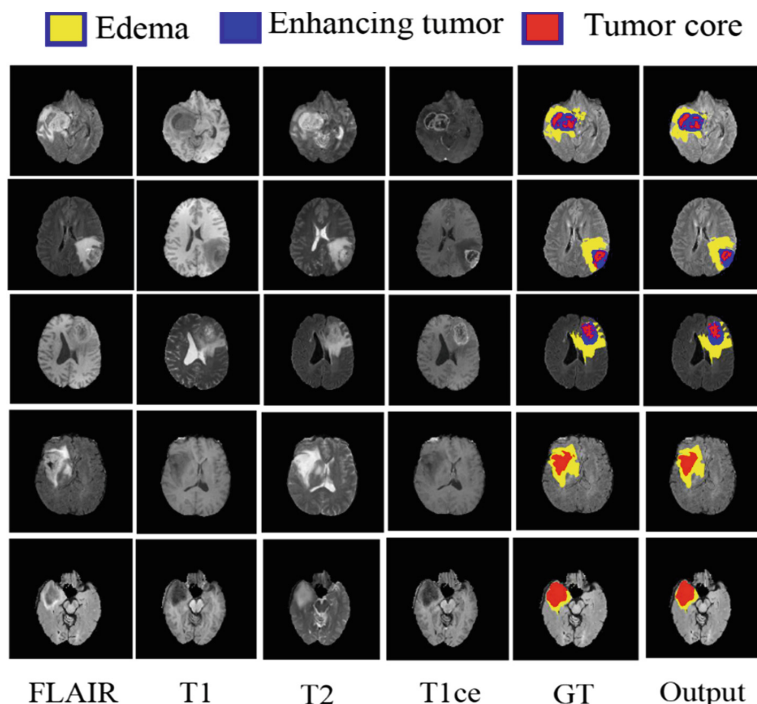
Fig. 2. Sample segmentation results. Each row represents one case. Columns from left to right: FLAIR, T1, T2, T1c, GT and Output. Segmentation labels: Yellow for edema, Blue for enhancing tumor and Red for tumor core. (Color figure online)

We computed first order statistics, shape features, Gray Level Co-occurrence Matrix and Gray Level Run Length Matrix features. We computed 468 features for edema, tumor core and enhancing tumor. These features were used to train the regression model for survival prediction task. We started with 679 variables [678 radiomic variables (113 from each of the different tumor parts for both FLAIR and T1c images) and Age]. The radiomic variables with near perfect correlation (Spearman's correlation coefficient 0.95 or higher, p 0.05 or lower) with each other were excluded with only one of the variables in each set being retained (N = 117). Age and all radiomics variables with no significant autocorrelations (N = 117) were assessed for relationship with survival.

Multi-Layer perceptron was used to train the neural network. Variables which had a statistically significant correlation (N = 56, including age) with survival were included for training the neural network. Results were replicated by setting a random seed. To assess the efficacy of the neural network and to correct over training, if any, we divided the BraTS 2018 training dataset (N = 163) into Training (51.5%), Validation (14.7%) and Testing (33.7%) subsets, randomly using Bernoulli variates. The neural network had two hidden layers with the number of units per layer set to auto, sigmoid activation functions for both the hidden as well as output layers. The variables were re-scaled

using adjusted normalization with a correction of 0.2. The neural network was designed to predict survival in days as well as two broad categories viz. survival <300 days and survival >=300 days. All statistical procedures for survival prediction were performed using SPSS for Windows v24 on a standard computer running Windows 10.

3   Result and Discussion

The performance of the proposed method was evaluated on BraTS 2018 training data with 285 cases and validated on 66 cases for segmentation. The validation leader-board gave interesting information about the performance of the different teams' algorithms. Average performance of proposed method on training data and validation data is given in Tables 1 and 2 respectively in terms of Dice Similarity Index and Sensitivity. The model was trained for 50 epochs and needed 48 h for training on NVIDIA P100 GPU with 128 GB system RAM. The framework was developed in Tensorflow [8].

Table 1. Performance of proposed method on BraTS 2018 training dataset for segmentation.

Evaluation metrics	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.8002	0.9324	0.9197	0.8951	0.9508	0.9359
Std. Dev.	0.2746	0.1056	0.1327	0.1397	0.0859	0.1005
Median	0.9062	0.9613	0.9564	0.9313	0.9645	0.9529
25 quantile	0.8421	0.9405	0.9303	0.8901	0.9451	0.9328
75 quantile	0.9422	0.9728	0.9687	0.9622	0.9777	0.9697

**Table 2.** Performance of proposed method on BraTS 2018 validation dataset for segmentation.

Evaluation metrics	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.7480	0.8780	0.8266	0.8266	0.9058	0.8186
Std. Dev.	0.2659	0.1345	0.1828	0.2306	0.1413	0.2136
Median	0.8527	0.9179	0.8985	0.9035	0.9436	0.9167
25 quantile	0.7325	0.8665	0.7771	0.8238	0.8953	0.7637
75 quantile	0.8853	0.9419	0.9444	0.9514	0.9720	0.9571

Overall, our approach reached a superior result in the whole tumor segmentation task with an average dice coefficient of 93% over training dataset and 87% over validation dataset. Sample segmentation results for intra-tumor parts are given in Fig. 2 The performance of the proposed approach is given in Table 3 in terms of Dice Coefficient and Hausdorff95 distance.

Table 3. Performance of proposed method on BraTS 2018 test dataset for segmentation.

Evaluation metrics	Dice			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Mean	0.6677	0.8474	0.7687	9.0554	17.2184	14.5728
Std. Dev.	0.3120	0.1699	0.2786	19.8975	28.9190	26.1504
Median	0.8013	0.9049	0.8946	2.2360	3.4641	3.3166
25 quantile	0.6556	0.8336	0.7519	1.4142	2.2360	2.0000
75 quantile	0.8656	0.9404	0.9328	3.6055	9.4604	8.4844

For the prediction of survival categories, the neural network demonstrated an accuracy of 70.2% in the training subset and 62.5% and 63.6% in the validation and testing subsets, respectively. The accuracy was 69.5% for the entire training dataset. The Area Under Curve (AUC) was 0.799 (Figs. 3 and 4). For prediction of survival in days, the proposed model performed better for values in the middle, with lower performance for the values at both extremes. The relative error was 0.842 for the training subset, 0.774 for the validation subset and 0.910 for the testing dataset.
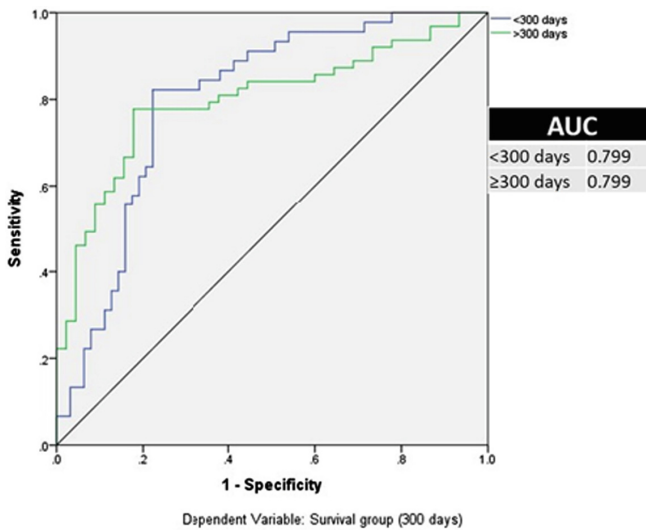
Fig. 3. ROC curve depicting the accuracy of the model for categorizing into <300 and >=300 days survival.

The performance of proposed OS prediction approach is given in Table 5 for 77 cases. The proposed approach stood third for overall Survival Prediction Task in BraTS 2018 Challenge (Table 4).
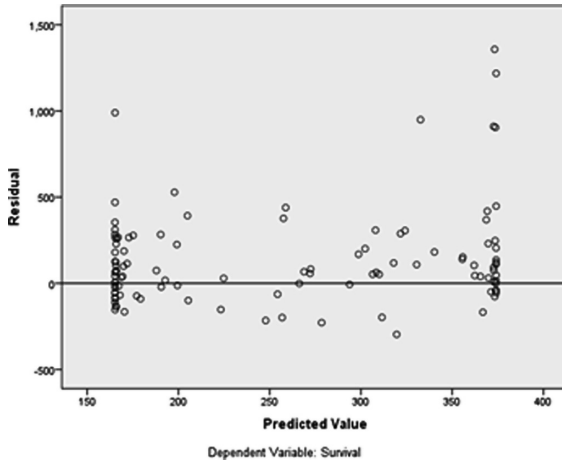
Fig. 4. Residual to predicted scatter plot showing the good fit of the model for survival values in the middle between 200–350 days.

Table 4. Performance of Multi-layer perceptron for OS prediction on validation dataset.

Method	Accuracy	MSE	Median SE	Std. Deviation	Spearman R
MLP	0.571	59550213.1	113611.616	128250465.8	0.427

Table 5. Performance for OS prediction on test dataset.

Cases	Accuracy	MSE	Median SE	Std. Deviation	Spearman R
77	0.558	338219.366	38408.16	939986.796	0.222

Individual variable importance analysis revealed that the Age is one of the most significant variables in this neural network. The other variables are shown in Table 6.

Table 6. Importance of the independent variable in descending order (F = FLAIR)

Variable name	Channel	Region	Importance	Normalized importance (%)
Age	–	–	0.07	100.0
Entropy	F	TC+ET	0.049	70.3
Variance	F	TC+ET	0.04	57.4
Enhance count	F	TC+ET	0.038	54.4
Core count	T1ce	WT	0.034	49.10
Cluster shade	T1ce	TC+ET	0.031	44.7
Edema count	F	TC+ET	0.03	42.8

(*continued*)

Table 6. (*continued*)

Variable name	Channel	Region	Importance	Normalized importance (%)
Dissimilarity	T1ce	TC+ET	0.029	42.3
Difference in entropy	F	ET	0.029	41.2
Variance	T1ce	TC+ET	0.027	39.0
Maximum probability	T1ce	ET	0.026	36.9
Sum of variance	T1ce	ET	0.025	36.3
Homogeneity	T1ce	TC+ET	0.024	34.7
Minimum	T1ce	WT	0.023	32.4
Correlation	T1ce	TC+ET	0.022	31.0
Inverse difference	T1ce	TC+ET	0.021	30.0
Contrast	F	TC+ET	0.021	29.9
Cluster shade	T1ce	ET	0.019	26.7
Correlation	T1ce	TC+ET	0.016	23.0
Variance	T1ce	ET	0.015	21.5
Maximum probability	F	ET	0.013	19.0
Cluster prominence	T1ce	ET	0.013	18.5
Dissimilarity	F	ET	0.013	18.5
Auto-correlation	T1ce	TC+ET	0.013	18.4
Inverse difference	T1ce	ET	0.013	18.4
Sum of squares variance	T1ce	ET	0.012	17.3
Difference in entropy	F	TC+ET	0.012	17.0
Average	F	TC+ET	0.012	16.6
Maximum probability	T1ce	TC+ET	0.011	15.4
Homogeneity	F	TC+ET	0.01	14.7
Difference in entropy	T1ce	TC+ET	0.009	13.1
Mean	F	TC+ET	0.009	12.4
Cluster prominence	T1ce	TC+ET	0.007	10.3
Sum average	T1ce	TC+ET	0.007	10.0
Inverse difference	F	TC+ET	0.006	9.2
Minimum	F	WT	0.005	7.8
Contrast	T1ce	TC+ET	0.005	6.7
Sum of intensities	F	ET	0.004	6.3
Contrast	F	ET	0.003	4.4
Homogeneity	F	ET	0.002	2.4
Contrast	T1ce	ET	0.001	1.9
Dissimilarity	T1ce	ET	0.001	1.7

4   Conclusion

In this study, we proposed a Deep Learning Radiomics Algorithm for Gliomas (DRAG) Model based on 3D U-Net network for brain tumor segmentation. 3D patches were extracted from multi-channel MRI data to train the proposed model. Radiomic features were extracted from FLAIR and T1ce channels for OS prediction task. MLP is trained with these radiomic features to predict the OS in days. The proposed approach achieved third place for OS prediction task in BraTS 2018 challenge [15].

The difference between mean and median in Table 2 indicates that for some cases, our pro-posed approach achieved poor accuracy, which is very close to zero and more analysis is required on this. Prediction of survival without more clinical data and treatment information is challenging and the same is reflected through accuracy of the participants in the leader-board. As the number of cases for OS prediction are less there is a need to develop an efficient feature selection algorithm which will select potential features for accurate OS prediction. Our future goal is to design radiomic features extraction pipeline with deep neural networks.

References

 1. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection (2017)
 2. Bakas, S. et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection (2017)
 3. Bakas, S., et al.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data **4**, 1–13 (2017)
 4. Kickingereder, P., et al.: Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. Radiol. **280**(3), 880–889 (2016)
 5. Konstantinos, K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2017)
 6. Mahajan, A., et al.: Bench to bedside molecular functional imaging in translational cancer medicine: to image or to imagine? Clin. Radiol. **70**(10), 1060–1082 (2015)
 7. Mahajan, A., Moiyadi, A.V., Jalali, R., Sridhar, E.: Radiogenomics of glioblastoma: a window into its imaging and molecular variability. Cancer Imaging **15**(Suppl. 1), 5–7 (2015)
 8. Martin, A., et al.: TensorFlow: large-scale machine learning on heterogeneous distributed systems. CoRR (2016)
 9. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BraTS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015)
10. Nelly, G., Eduard, M., Pilar, S.: State of the art survey on MRI brain tumor segmentation. Magn. Reson. Imaging **31**(8), 1426–1438 (2013)

11. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

12. Seow, P., Wong, J.H.D., Ahmad-Annuar, A., Mahajan, A., Abdullah, N.A., Ramli, N.: Quantitative magnetic resonance imaging and radiogenomic biomarkers for glioma characterisation: a systematic review. Br. J. Radiol. **91**, 20170930 (2017)

13. Tustison, N.J., et al.: N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging **29** (6), 1310–1320 (2010)

14. Baid, U., Talbar, S., Talbar, S.: Brain tumor segmentation based on non negative matrix factorization and fuzzy clustering. In: Fifth International Conference on Bio-Imaging (2017)

15. 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 preprint arXiv:1811.02629 (2018)