Glioma Prognosis: Segmentation of the Tumor and Survival Prediction Using Shape, Geometric and Clinical Information

Mobarakol Islam[1,2], V. Jeya Maria Jose[2,3], and Hongliang Ren[2(✉)]

[1] NUS Graduate School for Integrative Sciences and Engineering (NGS),
National University of Singapore, Singapore, Singapore
[2] Department of Biomedical Engineering,
National University of Singapore, Singapore, Singapore
mobarakol@u.nus.edu, jeyamariajose7@gmail.com, ren@nus.edu.sg
[3] Department of Instrumentation and Control Engineering,
National Institute of Technology, Tiruchirappalli, India

Abstract. Segmentation of brain tumor from magnetic resonance imaging (MRI) is a vital process to improve diagnosis, treatment planning and to study the difference between subjects with tumor and healthy subjects. In this paper, we exploit a convolutional neural network (CNN) with hypercolumn technique to segment tumor from healthy brain tissue. Hypercolumn is the concatenation of a set of vectors which form by extracting convolutional features from multiple layers. Proposed model integrates batch normalization (BN) approach with hypercolumn. BN layers help to alleviate the internal covariate shift during stochastic gradient descent (SGD) training by zero-mean and unit variance of each mini-batch. Survival Prediction is done by first extracting features (Geometric, Fractal, and Histogram) from the segmented brain tumor data. Then, the number of days of overall survival is predicted by implementing regression on the extracted features using an artificial neural network (ANN). Our model achieves a mean dice score of 89.78%, 82.53% and 76.54% for the whole tumor, tumor core and enhancing tumor respectively in segmentation task and 67.9% in overall survival prediction task with the validation set of BraTS 2018 challenge. It obtains a mean dice accuracy of 87.315%, 77.04% and 70.22% for the whole tumor, tumor core and enhancing tumor respectively in the segmentation task and a 46.8% in overall survival prediction task in the BraTS 2018 test data set.

Keywords: Brain tumor segmentation · Glioma ·
Convolutional neural network · Hypercolumn · PixelNet ·
Magnetic resonance imaging · Survival prediction

1   Introduction

Gliomas are the most frequent brain tumor with the highest mortality rate which develops from glial cell [7]. Early detection, accurate segmentation and estimation of the relative volume are very crucial for overall survival (OS) prediction, treatment and surgical planning. In addition, manual segmentation of tumor tissue is tedious, time consuming and requires strong supervision by a human expert. It is also prone to inter and intra-rater variability. So it is highly necessary to develop an automatic segmentation system to diagnose and estimate the volume, size, shape and location of the tumor. Overall survival prediction along with automatic segmentation would be a very useful tool that would help in better clinical diagnosis.

In recent years, the success of deep learning in this field is huge as it shows state of art performance in the applications of segmentation, classification, regression and detection. Iftekharuddin et al. [9] exploits convolutional neural network (CNN) for glioma segmentation and extracts handcrafted features like histogram, co-occurrence matrix, neighbourhood gray tone difference, run length, volume and areas to predict OS using random forest regression model. Jungo et al. [15] uses the residual convolutional neural network with Bayesian dropout to segment tumor and calculates the geometric features (e.g. volume, heterogeneity, rim width, surface irregularity etc.) from the segmented tumor. Later, a simple artificial neural network (ANN) is utilized to predict the exact days of OS. 3D U-net and linear regression approaches are also exploited to segment and predict OS [1].

In this paper, we propose a batch normalized CNN architecture with hypercolumn features inspired by multi-modal PixelNet [6,11,12] where a modest number of features are extracted from multiple convolution layers and trained with a multi-layer perceptron (MLP) to predict segmentation classes. We discuss about the various features (Geometric, Fractal, Image and Clinical) extracted from the segmentation output. We propose a new method of overall survival prediction by combining all the meaningful features that contribute to the number of days of survival left for the patient. Details about how the features are selected and various experiments that are run for finding the best regression technique has been discussed. The problems of generalizing a network over its performance on a validation data set is also discussed.

2   Methods

In this section, we discuss about the dataset provided, the model and methodologies to process data for segmentation and about the extraction of features, feature selection and regression techniques that were exploited for overall survival prediction.

2.1   Dataset

BraTS 2018 (Brain Tumor Image Segmentation Benchmark) training database [2–4,19] consists in total 285 cases of patients out of which the overall

survival prediction data was provided for 163 cases. BraTS 2018 Validation dataset [2–4,19] consists of 53 cases. All the data are a multi-modal MRI scan of 210 high-grade glioma (HGG) and 75 low-grade glioma (LGG) and 4 different modalities including T1 (spin-lattice relaxation), T1c (T1-contrasted), T2 (spin-spin relaxation) and FLAIR (fluid attenuation inversion recovery). Each scan is a continuous 3D volume of 155 2D slices of size 240 × 240. The volume of the various modalities is already skull-stripped, aligned with T1c and interpolated to 1 mm voxel resolution.

The provided ground truth with manual segmentation includes three labels: GD-enhancing tumor (ET—label 4), the peritumoral edema (ED—label 2), and the necrotic and non-enhancing tumor (NCR/NET—label 1). The predicted labels are evaluated by merging three regions: whole tumor (WT: all four labels), tumor core (TC: 1, 2) and enhancing tumor (ET: 4). Figure 1 illustrates all the four modalities for one of the samples of the training data set of BraTS 2018. Figure 2 illustrates the provided ground truth for the same sample. The green label corresponds to GD-enhancing tumor, yellow label corresponds to peritumoral edema and the red label corresponds to necrotic and non-enhancing tumor. ITK-SNAP [24] is the tool that was used to visualize the data.
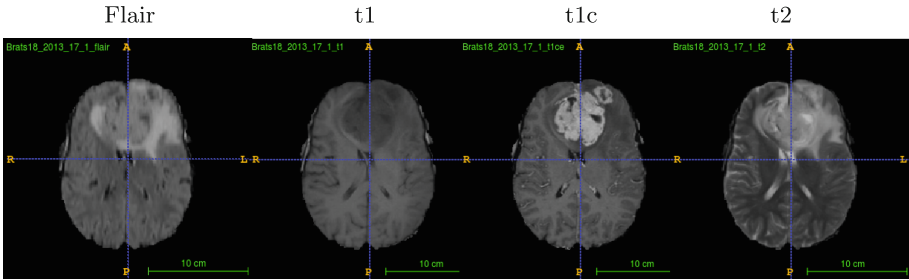

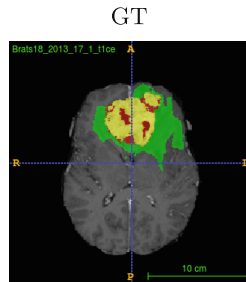**Fig. 1.** Visualization of the different modalities in the BraTS 2018 Training data set.

Fig. 2. Ground Truth with different segmentation labels as given in BraTS 2018 Training data set. Red, Yellow and Green colors represent Necrotic, Enhancing and Edema respectively. (Color figure online)
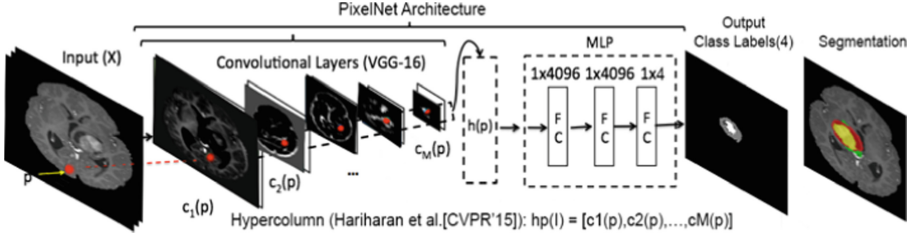
Fig. 3. Batch Normalized PixelNet architecture.

2.2   Segmentation

Proposed Model. Our proposed model is inspired from PixelNet [6,11] where we integrate additional 3 convolution layers and batch normalization after all the convolution layers. The model consists of 18 pixel-block and a hypercolumn layer followed by a multi-layer perceptron (MLP) of 3 fully connected layers as in Fig. 3. A pixel-block contains convolution, batch normalization (BN) [10] and ReLu layers sequentially. Hypercolumn layer extracts the features from multiple convolution layers and concatenates them into a feature vector which propagates to the MLP for pixel-wise classification.

2.3   Survival Prediction

Following the extraction of tumor from the MRI scans, the segmented tumor along with certain other parameters are used for survival prediction. The following paragraphs elucidate the features those are extracted along with the regression model that is built for predicting survival.

Feature Extraction. The tumor geometry and its location hold a very important role in deciding the number of days of survival [21]. Figure 4 visually illustrates how the features such as location or centroid of tumor, size and shape of tumor affect the overall survival of the patient. It is evident that more the proximity of tumor to the centre of brain, the lower is the overall survival of patient. Also, lesser the size or smaller the shape of tumor, higher is the overall survival of patient. So, we extract geometrical features which include First axis coordinates, Second axis coordinates, Third axis coordinates, Eigen Values, First axis length, Second axis length, Third axis Length, Centroid coordinates, meridional eccentricity and equatorial eccentricity for individual tumor types as well as whole tumor. Figure 5 illustrates the tumor extracted from the brain and gives an intuition of how the geometric features like centroid, eccentricity and axis lengths are calculated. Also, the volume of tumor and its ratio with respect to the total volume was calculated. Features of the tumor image including mean, variance, standard deviation, entropy, skewness [18], Kurtosis [18], entropy and histogram feature intensities were extracted. Fractal dimension of the necrotic
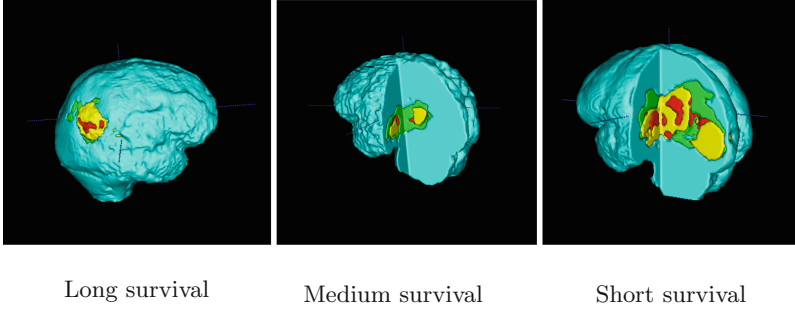
Long survival   Medium survival   Short survival

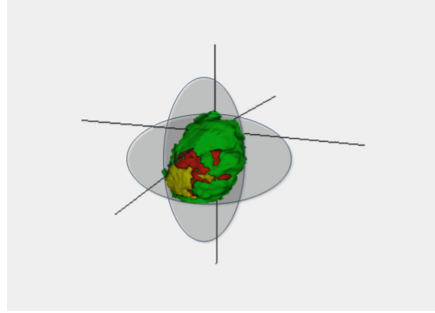**Fig. 4.** Visualization of tumor with different survival rate.

Fig. 5. Visualization of the extracted tumor from brain.

tumor has been found to play a pivotal role in the survival prediction, according to [22]. So, fractal dimension and fractal ratio were extracted for the necrotic part of the segmented tumor core. In addition, the age of the subject provided by the BraTS 2018 training data set was also included.

Feature Selection. Several experiments were conducted using different combinations of extracted features. After analyzing the cross-validation errors of the experiments, the most informative features are alone retained and others are neglected. Features like eigen values, eccentricity, skewness, mean and variance were not found to have an important role in survival prediction. Also, geometric features of the GD-enhancing tumor (ET—label 4) was found to be only increasing the cross-validation errors and hence was removed. So finally, a total of 50 features that were found to be the most informative are used in the regression model. These features are First axis coordinates, Second axis coordinates, Third axis coordinates, First axis length, Second axis length, Third axis Length, Centroid coordinates for part wise non-enhancing tumor core (NCR/NET—label 1), peritumoral edema (ED—label 2) as well as for the whole tumor without including GD-enhancing tumor (ET—label 4) in addition to Kurtosis, Entropy, Histogramic intensity, Fractal dimension and age.

Regression Model. With the selected features, we train a fully connected artificial neural network (ANN) with one hidden layer [23] and ReLu activation function. Figure 6 shows the ANN model that takes 50 features as input and gives the number of days of survival as the output. We run an experiment to find which configuration of the hidden layers gives the lowest mean squared error (MSE). The best configuration was found to be 100 neurons in the hidden layer. Then, we find the best epoch by finding the epoch for which the MSE is minimum by using a cross-validation data set while training. However, after analyzing many experiments, we find that when MSE is minimum, the accuracy is low and vice-versa. So, we find an epoch where the accuracy and the MSE of the model are balanced. Adam optimizer [16] with MSE loss function has been used to conduct all experiments with ANN. We also tune the hyper parameters like learning rate and batch size. The best result for BraTS 2018 Validation Data was acquired for the 900th epoch with a batch size of 10.
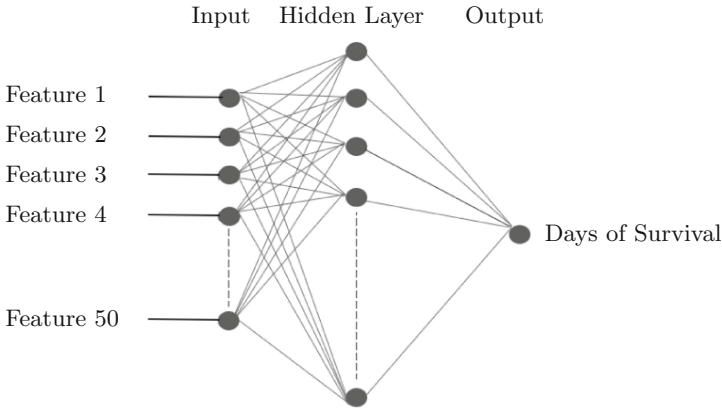

Fig. 6. Regression model using ANN.

3 Experiments and Results

3.1 Segmentation

We convert the 3D voxel of $240 \times 240 \times 155$ into 2D slices of 240×240 by ignoring blank slices of scan and ground-truth. We choose a sample of 2000 pixels per image and batch size of 10 in training time. We normalize data to zero mean and standard deviation and augment by flipping left-right. In testing phase, hyper-column has been formed with all the pixels inside brain region and predict slice by slice to form MRI. Finally, we adopt largest component analysis to remove false positive as a post-processing technique. We utilize Caffe [13] framework with a single Nvidia GPU 1080Ti GPU to perform all the experiments. Table 1 represents the Dice and Hausdorff performance of our model. It obtains dice

accuracy of 89.78%, 82.53% and 76.54% of whole tumor (WT), tumor core (TC) and enhance tumor (ET) respectively. Table 2 shows the performance metrics that the segmentation method has achieved on the test data set. It obtains dice accuracy of 87.315%, 77.04% and 70.22% of whole tumor (WT), tumor core (TC) and enhance tumor (ET) respectively.

Table 1. Dice and Hausdorff for BraTS 2018 validation dataset

	Dice			Hausdorff		
	WT	TC	ET	WT	TC	ET
Mean	**89.78**	82.53	76.54	5.09	7.11	3.60
StdDev	8.35	17.80	23.23	7.04	8.04	5.58
Median	90.51	86.87	83.15	3.08	4.53	2.23

Table 2. Dice and Hausdorff for BraTS 2018 Test dataset [5]

	Dice			Hausdorff		
	WT	TC	ET	WT	TC	ET
Mean	**87.315**	77.044	70.22	7.15	8.0024	4.594
StdDev	11.33	26.082	27.96	13.08	11.937	8.802
Median	90.36	87.89	80.93	3.6	4.24	2.23

The following Fig. 7 shows the visualizations of output from the segmentation network along with the input for BraTS 2018 [5] Test data set. The different modalities of the input namely- Flair, t1c and t2 are also illustrated. The prediction shows the labelling of tumor done by the network. It consists of 3 labels, green corresponds to GD-enhancing tumor, yellow corresponds to peritumoral edema and red corresponds to necrotic and non-enhancing tumor core.

3.2   Survival Prediction

Evaluation of the survival prediction model is performed on BraTS 2018 survival validation dataset (which is a subset (28 out of 53 subjects) of the segmentation validation dataset) [2–4,19]. Quantitative details for the ANN that gave the best accuracy of 67.9% has been listed in Table 3. This is the best accuracy that was achieved in the Validation Leader board till the time this paper is written. However, the same ANN was found to give relatively less accuracy for the test data set. Table 4 shows the quantitative results for BraTS 2018 [5] Test data set.
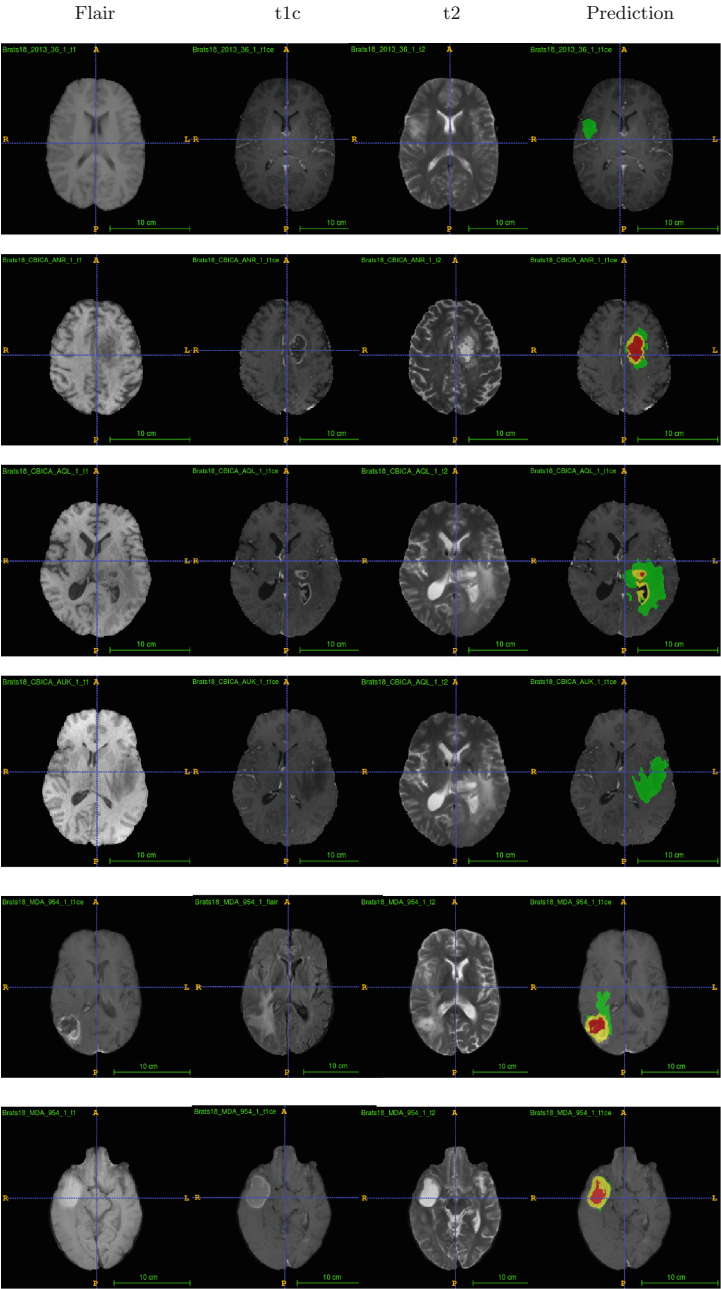
Fig. 7. Visualization of the predicted segmentation labels for BraTS 2018 Test data set. (Color figure online)

Table 3. Quantitative results of survival prediction on the BraTS 2018 validation data set using ANN

Cases	Accuracy	MSE	MedianSE	stdSE	SpearmanR
28	**67.9%**	96161.713	59473.481	117207.189	0.496

Table 4. Quantitative results of survival prediction on the BraTS 2018 Test data set using ANN

Cases	Accuracy	MSE	MedianSE	stdSE	SpearmanR
77	**46.8%**	341387.439	92892.2261	788491.731	0.148

4   Discussion

4.1   Experiments for Finding the Best Regression Technique

For finding out the best regression model and the meaningful features, several experiments are conducted. Experiments with different combinations of available features and regression models are conducted on the 33 subjects of BraTS17 validation dataset [2,3,19] to find the best possible combination of features and regression model. The details of the experiments on the entire features and the best 50 features are listed in Tables 4 and 5 respectively. When compared to ANN, other regression models such as Support Vector Machine (SVM) with Radial basis function (RBF) kernel [14], Random Forest [17], Linear Regression [20], Logistic Regression [8] are investigated but resulted in inferior performance. Also, using only the best 50 features showed improvement in accuracy and MSE for every models.

4.2   Issues with Generalizing a Network on Validation Data Set

The quantitative results of survival prediction on the BraTS 2018 test data set can be found in Table 4. When we use the same algorithm with the same 50 features that gave the best results in the validation data set for the test data set of survival prediction task, we find that the accuracy of the network has reduced drastically. Also, the mean squared error obtained increases. So, it is evident that the network does not perform well with the test data set as good it performs for the validation data set. The reason for this behaviour is that the model is not generalized. It is over-fitted to the train data.

The same architecture, with the same number of neurons in the hidden layers, with the same hyper parameters; trained for a certain number of epochs that gave the best accuracy for validation data set was used for the test data set. The number of epochs was fixed by experimenting which epoch gave the least mean squared error for the validation data set. This is where the network was over-fitted. Generalizing a network architecture and the number of epochs till which it should be trained should not be done by comparing the results with a

Table 5. Performance comparison with all the features in different machine learning models for BraTS 2017 validation data set.

Models	Accuracy	MSE	MedianSE	stdSE	SpearmanR
Linear Regression	50.5%	252353.061	95419.102	**429879.191**	0.263
SVM	33.3%	242147.277	62044.450	563941.194	0.142
Random Forest	42.4%	**208660.63**	**33367.111**	502312.762	0.213
Logistic Regression	39.4%	286470	40401	540201.363	**0.479**
ANN	**54.5%**	211967.681	53967.305	540221.112	0.206

Table 6. Performance comparison with the best 50 features in different machine learning models for BraTS 2017 validation data set.

Models	Accuracy	MSE	MedianSE	stdSE	SpearmanR
Linear Regression	50.5%	237148.501	78915.034	362705.604	0.221
SVM	42.4%	233367.604	127524.752	374037.279	0.218
Random Forest	39.4%	262224.703	**42507.088**	506754.007	**0.324**
Logistic Regression	36.4%	**181509.182**	58564	**249213.006**	0.13
ANN	**60.6%**	214207.487	60832.523	354332.371	0.293

few data. The data with which a network is generalized can be of a very meagre amount and might contain similar characteristics. So generalizing a network over the validation data set here has caused the network to perform badly in a fresh test data set.

Looking at Table 6, models like Logistic regression or Random Forest which would not have over fitted the data could have been used. Thus, selection of validation data is very crucial for any neural network as well as deep learning model where the data set is really small.

5 Conclusion

Batch Normalized pixelnet is found to give quality segmentation results for the BraTS 2018 Validation data set. The main advantage of the pixelnet is that it has freedom of sampling pixel during training phase. The background of the scan is removed during training and this helps the network to converge faster. For the survival prediction task, a lot of features were studied and extracted. The features that were found to increase the error in the regression problem of survival prediction problem were removed. Various regression models were experimented and ANN was found to give the best results for the BraTS 2018 Validation data set. However, due to overfitting the same network was not able to give good results for the BraTS 2018 test data set.

References

1. Amorim, P.H.A.: 3D U-Nets for brain tumor segmentation in MICCAI 2017 BraTS challenge. In: International MICCAI Brainlesion Workshop, pp. 9–14. Springer (2017)

2. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data **4**, 170117 (2017)

3. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. Cancer Imaging Arch. **286** (2017). https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q

4. 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/10.7937/K9/TCIA.2017.GJQ7R0EF

5. 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)

6. Bansal, A., Chen, X., Russell, B., Gupta, A., Ramanan, D.: PixelNet: representation of the pixels, by the pixels, and for the pixels. arXiv preprint arXiv:1702.06506 (2017)

7. Holland, E.C.: Progenitor cells and glioma formation. Curr. Opin. Neurol. **14**(6), 683–688 (2001)

8. Hosmer Jr., D.W., Lemeshow, S., Sturdivant, R.X.: Applied Logistic Regression, vol. 398. Wiley, New York (2013)

9. Shboul, Z.A., Vidyaratne, L., Alam, M., Iftekharuddin, K.M.: Glioblastoma and survival prediction. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 358–368. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_31

10. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)

11. Islam, M., Ren, H.: Multi-modal PixelNet for brain tumor segmentation. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 298–308. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_26

12. Islam, M., Ren, H.: Class balanced PixelNet for neurological image segmentation. In: Proceedings of the 2018 6th International Conference on Bioinformatics and Computational Biology, pp. 83–87. ACM (2018)

13. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678. ACM (2014)

14. Joachims, T.: Making large-scale SVM learning practical. Technical report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund (1998)

15. 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
16. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
17. Liaw, A., Wiener, M., et al.: Classification and regression by randomforest. R News **2**(3), 18–22 (2002)
18. Mardia, K.V.: Measures of multivariate skewness and kurtosis with applications. Biometrika **57**(3), 519–530 (1970)
19. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993 (2015)
20. Neter, J., Wasserman, W., Kutner, M.H.: Applied linear regression models (1989)
21. Pérez-Beteta, J., et al.: Glioblastoma: does the pre-treatment geometry matter? A postcontrast T1 MRI-based study. Eur. Radiol. **27**(3), 1096–1104 (2017)
22. Reishofer, G., Studencnik, F., Koschutnig, K., Deutschmann, H., Ahammer, H., Wood, G.: Age is reflected in the fractal dimensionality of MRI diffusion based tractography. Sci. Rep. **8**(1), 5431 (2018)
23. Schalkoff, R.J.: Artificial Neural Networks, vol. 1. McGraw-Hill, New York (1997)
24. Yushkevich, P.A., Gao, Y., Gerig, G.: ITK-SNAP: an interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: 2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), pp. 3342–3345. IEEE (2016)