



Research article

Development of a nomograph integrating radiomics and deep features based on MRI to predict the prognosis of high grade Gliomas

Yutao Wang ¹, Qian Shao ^{2,*}, Shuying Luo ² and Randi Fu ²

¹ Affiliated Hospital of Medical School, Ningbo University, Ningbo 315020, China

² Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China

* **Correspondence:** Email: 243386393@qq.com; Tel: +8618758390460.

Abstract: The purpose of this study was to assess the overall survival of patients with HGG using a nomogram which combines the optimized radiomics with deep signatures extracted from 3D Magnetic Resonance Images (MRI) as well as clinical predictors. One training cohort of 168 HGG patients and one validation cohort of 42 HGG patients were enrolled in this study. From each patient's 3D MRI, 1284 radiomics features were extracted, and 8192 deep features were extracted via transfer learning. By using Least Absolute Shrinkage and Selection Operator (LASSO) regression to select features, the radiomics signatures and deep signatures were generated. The radiomics and deep features were then analyzed synthetically to generate a combined signature. Finally, the nomogram was developed by integrating the combined signature and clinical predictors. The radiomics and deep signatures were significantly associated with HGG patients' survival time. The signature derived from the synthesized radiomics and deep features showed a better prognostic performance than those from radiomics or deep features alone. The nomogram we developed takes the advantages of both radiomics and deep signatures, and also integrates the predictive ability of clinical indicators. The calibration curve shows our predicted survival time by the nomogram was very close to the actual time.

Keywords: high grade gliomas; radiomics; transfer learning; Magnetic Resonance Imaging; nomogram

1. Introduction

Glioblastoma multiforme is the most frequent malignant primary brain tumor in adults [1]. It is a deadly disease with a high mortality rate [2]. The median survival is only 12 to 14 months even with active treatment [3]. It is difficult to capture the comprehensive information of tumors in a non-invasive

way because of the spatial and temporal intra-tumor heterogeneity. Therefore, it is difficult to accurately predict the prognosis of patients.

In recent years, in the fields of health care and science, medical images can be analyzed at any time through radiomics [4]. Therefore, medical and scientific information can be provided repeatedly in detail, which enables the comprehensive understanding of the characteristics of the entire tumor. With the rapid development of MRI technology, repeated and non-invasive assessment of tissue characteristics is becoming one of the main methods in the field of gliomas, such as tumor diagnosis, staging, targeted therapy, and evaluation and monitoring of treatment response [5,6]. Conventional MRI mainly shows the difference in signal intensity and mass effect of glioma hemorrhage, necrosis, edema tissue and the extent of lesion invasion. Multimodal MRI can not only reflect the morphological characteristics of glioma, but also reflect the function and metabolic status of tumor tissue. However, the traditional medical imaging mode limits the analysis of images to a range of visual judgment. The resolution of the human eye is limited, so the fine and minute features in each image are not easily noticed by the naked eye [7]. Radiomics is a developing research field that aims to extract a large amount of complex information from traditional medical images to form a high-dimensional developable feature space, including features that are not easily being visible or quantifiable. This is not only a diagnostic method, but also a method to extract more useful information from tumor phenotypes, which can be used for personalized medicine.

In the past few years, a number of radiomics models have been proposed for survival prediction, distant metastasis prediction [8], molecular characteristics classification [9], etc. The high-throughput feature extraction is a critical task in radiomics. In previous studies, most extracted features are explicitly designed, or handcrafted. These handcrafted features include tumor shape, intensity, texture and wavelet textures. Although the number of handcrafted features can reach tens of thousands, these features are shallow and low-order image features. These features may not fully characterize the heterogeneity of the tumor, and therefore limit the potential of survival prediction models. In this case, it is necessary to extract deeper and higher-order features, which may improve the prediction performance of the survival prediction model.

Recently, the application of deep learning has been intensively demonstrated in computer vision [10,11]. Convolutional Neural Network (CNN) is a typical artificial neural network in deep learning [12], which has achieved great success in image and video recognition and segmentation [13,14]. When the data sets are large enough, the deep learning algorithms often perform better compared to traditional algorithms. However, when it comes to medical image analysis, the data sets are often inadequate to reach the full potential of deep learning. In computer vision, transfer learning and fine tuning are often used to solve the problem of a small data set [15]. Transfer learning can also be incorporated into current radiomics model, with the help of transfer learning, extracting a large number of deep features from hidden layers of CNN becomes possible [16]. These deep features contain more abstract information of medical images.

The main contributions of this work are as follows:

- (1) A new method for the prediction of prognosis of high grade gliomas based on MRI is proposed.
- (2) Radiomics and deep features are extracted based on multimodal MRI.
- (3) Transfer learning is used to solve the problem of insufficient sample size when extracting deep features.
- (4) Radiomics and deep features are combined to construct a nomogram in order to predict the

prognosis of high grade gliomas.

2. Materials and Method

2.1. Participants

In this study, pre-treatment MR images of 259 HGG patients were collected, and 210 of them met the screening criteria. The data set is from Multimodal Brain Tumor Segmentation Challenge 2019 (BraTS 2019) [17–19]. As BraTS 2019 are publicly available database without patient identifier, no institutional review board approval is required for the data set. The inclusion criteria were that patients with newly diagnosed and treatment-naive HGG and survival information and pre-treatment MR imaging including T1-weighted, T1-weighted Gadolinium contrast-enhanced, T2-weighted, and T2-weighted Fluid Attenuated Inversion Recovery (short for T1, T1ce, T2, and FLAIR). The exclusion criteria are patients with a history of surgery or chemoradiation therapy and patients missing survival information. Overall survival is calculated from the initial pathologic diagnosis date to death or censure point if still alive. We randomly divided patients into training cohort ($n = 168$) and a validation cohort ($n = 42$) according to the ratio of 4:1.

2.2. Data analysis and preprocessing

Background information occupies a large proportion of images which is not helpful for extracting radiomic features and deep features. Therefore, it is necessary to remove the background information around the brain area. The slice size after cutting is 128×128 . Each sequence of the original data contains 155 slices, and after removing 27 all-black slices, 128 slices can be obtained for each sequence. After the above processing, the size of the MRI is $128 \times 128 \times 128$. Then, the three-dimensional tumor subregions were manually drawn slice by slice by an MRI radiologist with 15-year experience and a radiologist with 10-year experience using ITK-SNAP. Both radiologists were blind to the pathological results. The segmentation results were also evaluated by three independent reviewers to make sure they agree. Three tumor subregions include necrosis area, enhancement area and edema area as shown in Figure 1.

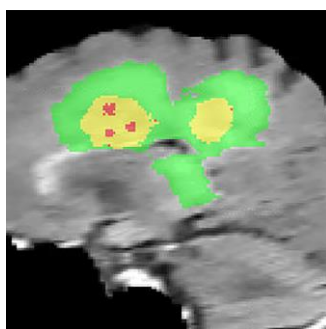


Figure 1. Segmentation labels.

As shown in Figure 1, the green part is the edema area, the yellow part is the enhancement area and the red part is the necrosis area. The necrosis area was the low intensity necrotic structures within the enhancing rim in T1ce and had hyper-intense signal in T2 and FLAIR. The enhancement area was confirmed as the Gadolinium enhancing rim excluding the necrotic center and hemorrhage with both

T1ce and T1 images. The edema area may include both peritumoral edema and any non-enhancing tumor. The segmented tumor subregions were then used for subsequent experiments. The follow-up experimental process includes feature extraction and selection, construction of radiomics signature, deep signature and combined signature, and nomogram modeling to obtain the survival prognosis of HGG patients. The overall framework is shown in Figure 2.

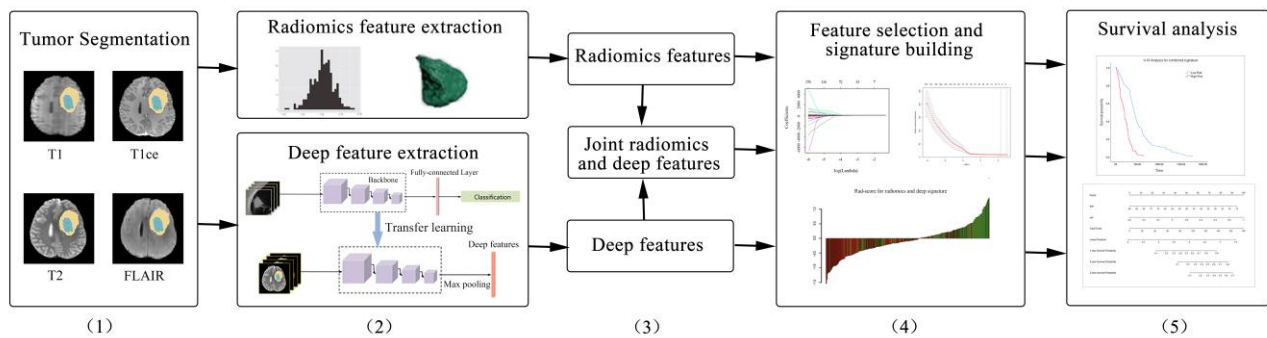


Figure 2. The framework for developing a nomogram.

As shown in Figure 2, we first segment the tumor region for each case. From each case's 3D MRI, 1284 radiomics features were extracted, and 8192 deep features were extracted via transfer learning. By using LASSO regression to select features, the radiomics signatures and deep signatures were generated. The radiomics and deep features were then analyzed synthetically to generate a combined signature. Finally, the nomogram was developed by integrating the combined signature and clinical predictors.

2.3. Feature extraction

In this paper, the radiomics features were extracted from three subregions and four MR modalities. The feature extraction subregions include necrosis, enhancement and edema. The radiomics features can be divided into three groups: (I) geometry, (II) intensity, (III) texture. The geometry features describe the three-dimensional shape characteristics of the tumor. The intensity features describe the first-order statistical distribution of the voxel intensities within the tumor. The texture features describe the patterns, or the second- and high-order spatial distributions of the intensities. A total of 1284 ($107 \times 4 \times 3$) radiomics features were extracted. Details of the radiomics features can be found in Appendix A1. All the features were extracted through the pyradiomics package version 3.0 [20].

Radiomics features are explicitly designed or handcrafted. Although the number of handcrafted features can reach tens of thousands, these features are shallow and low-order image features. These features may not fully characterize the heterogeneity of the tumor, and therefore may limit the potential of survival prediction models. In this case, it is necessary to extract deeper and higher-order features. In this study, deep features were extracted from pre-trained CNN via transfer learning. 3D-ResNet50 was chosen as the pre-trained CNN model [21]. The hyper-parameters of 3D-ResNet50 were weight decay 0.001, momentum 0.9, initial learning rate 0.001. Deep feature extraction includes 3 steps: pre-training of 3D-ResNet50, fine-tuning and feature extraction.

1. Pre-training: The 3D-ResNet50 model is pre-trained on 23 magnetic resonance data sets. The model had the ability to recognize the basic contours and details in the magnetic resonance image after pre-

training which could accelerate the gradient descent in the subsequent fine-tuning process.

2. Fine-tuning: The HGG patients in the BraTS 2019 data set were divided into 3 categories, which were postoperative survival time of less than 2 years, 2~3 years and more than 3 years. For four kinds of magnetic resonance sequences (including T1, T1ce, T2, and FLAIR), the model was fine-tuned by classification tasks. A total of four fine-tuned models could be obtained, and these models have the ability to identify the length of survival of patients after surgery.
3. Feature extraction: First, the gray values were normalized to range [0, 255] using linear transformation. Then, according to the segmentation results, the whole tumor (necrosis, enhancement and edema) area was used as the input of 3D-ResNet50. Finally, the deep features that could be computed by only forward propagation were extracted from the fully connected layer before the softmax layer. In total, 8192 (2048×4) deep features could be extracted for each patient. This procedure was accomplished by using the deep learning toolkit Pytorch. The structure of 3D-ResNet50 is shown in Figure 3. The details of 3D-ResNet50 can be found in Appendix A2.

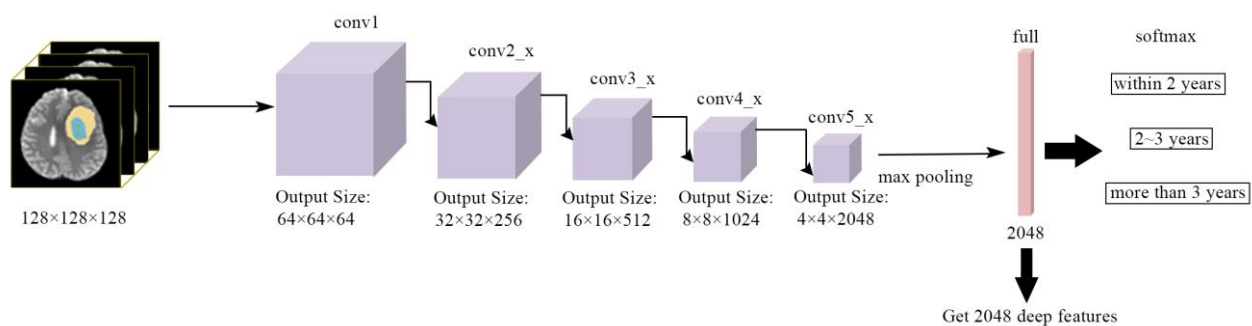


Figure 3. Illustration of deep features extraction.

2.4. Feature selection and signature construction

LASSO is a computationally attractive alternative to standard covariance selection for sparse high-dimensional graphs and an effective approach for the biomarker selection of high-dimensional data [22]. We used LASSO to select a subset of the most significant radiomics features and deep features from the training cohort.

For each patient, we use the selected linear combination of each radiomics feature and its corresponding weight as the radiomics score (Rad-score) [23], that is, the radiomics signature. The calculation method of the deep signature and the combined signature is the same as the radiomics signature. Without loss of generality, both deep signature and combined signature are represented by Rad-score.

2.5. Assessment of signatures

The association of the three signatures with patients' survival time was evaluated in the training cohort and then validated in the validation cohort. A univariate cox proportional hazards model was applied in the training and validation cohorts by using radiomics, deep and combined signature to calculate C-index and hazard ratio (HR) [24,25], enabling the evaluation of three signatures' predictive accuracy [26]. In addition, Kaplan-Meier survival analysis was performed in three cases to analyze the

correlation between the three signatures and patients' survival. We used the median Rad-score as the cut-off line to divide all patients into low-risk and high-risk groups. The significant differences in survival rates between two groups were accessed using a log-rank test.

2.6. Development and performance of the nomogram

We used combined signature and clinical information (age) to draw a nomogram of the survival probability of COX regression in the training cohort, considering that age has a greater impact on the prognosis of patients. The nomogram was evaluated in the training cohort and validated in the validation cohort. The discrimination performance of nomogram was assessed by the Harrell's C-index. The calibration curves of the radiomics nomogram for 1-, 2-, and 3-year survival time were evaluated by plotting the actual survival against predicted survival probabilities.

2.7. Statistical analysis

The statistical analysis of this research was based on R software (R: a free software environment for statistical computing and graphics. URL: <https://www.r-project.org/>). The details of the packages used are described in Appendix A3. All statistical tests were with a significance level at 0.05.

3. Results

3.1. Feature extraction and radiomics signature construction

In this paper, a total of 1284 radiomics features and 8192 deep features were extracted from the 3D images. In order to eliminate redundant or low-relevant features, we used the LASSO regression model to select the features of radiomics, deep and combined features.

As shown in Figure 4, a, b and c represent the coefficient distribution of each feature, and a coefficient profile plot was produced against the log (λ) sequence. And d, e and f are to use the 10-fold cross-validation to adjust the parameters in the LASSO model to get the minimum standard. The partial likelihood deviance was plotted versus log (λ). Dotted vertical lines were drawn at the optimal values by using the minimum criteria, and the dotted line indicated the number of selected features. As can be seen in d, e and f, seven, nine, and fifteen features with non-zero coefficients were chosen. They are used as the predictive features of radiomics, deep and combined feature groups, respectively.

We linearly combine the selected features of each group with their corresponding coefficients to obtain radiomics signature, deep signature and combined signature. Without loss of generality, radiomics signature, deep signature and combined signature were all represented by Rad-score.

3.2. Assessment of signatures in the training cohort

In this section, we assessed the prognosis performance of radiomics, deep and combined signatures using cox hazard regression model. In order to compare the prognostic performance of the three categories of signatures, we calculated the HR of each label, as shown in Table 1.

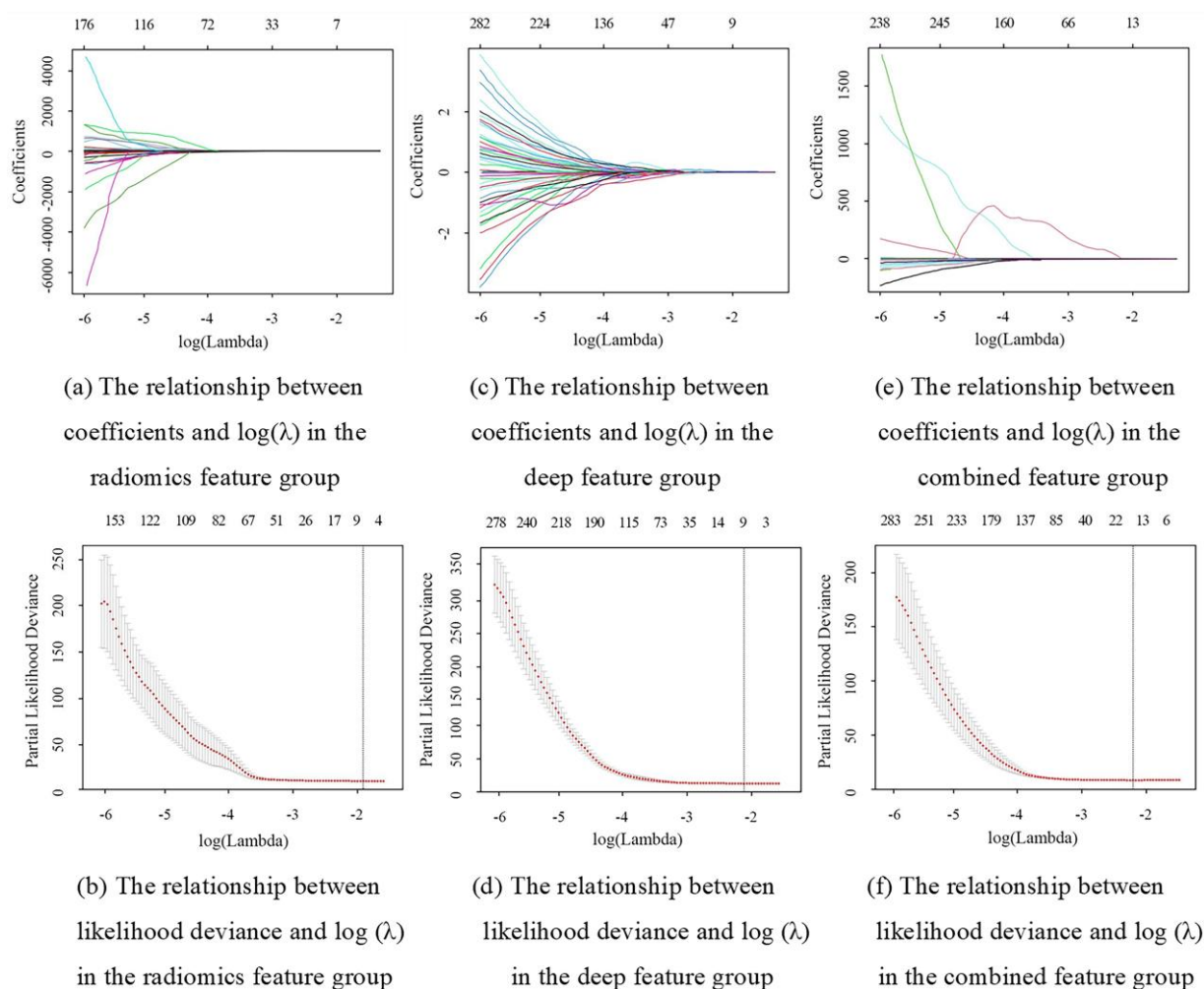


Figure 4. Feature selection using the LASSO cox regression model.

Table 1. HR analysis for the different signature groups.

	Training cohort		Validation cohort					
	HR	p value	95% CI for HR		HR	p value	95% CI for HR	
			Lower	Upper			Lower	Upper
Radiomics signature	6.430	0.007	2.587	15.984	4.636	0.014	1.365	15.747
Deep signature	9.712	< 0.0001	5.080	18.567	8.302	< 0.0001	4.273	16.129
Combined signature	12.328	< 0.0001	6.578	23.106	10.965	< 0.0001	5.959	20.178

It can be seen from the table that the HR values of the three signatures are all greater than 1, indicating that the larger the signature value, the higher the patient's risk of death and the shorter the survival time.

In addition, Kaplan-Meier survival analysis was performed in three cases to analyze the correlation between the three signatures and patients' survival. We used the median Rad-score as the cut-off line to divide all patients into low-risk and high-risk groups. The significant differences in survival rates between two groups were accessed using a log-rank test. Figure 5 shows the survival probability of the patients in the high-risk or low-risk cohort. The results of the log-rank test indicate the significant discrimination between two groups.

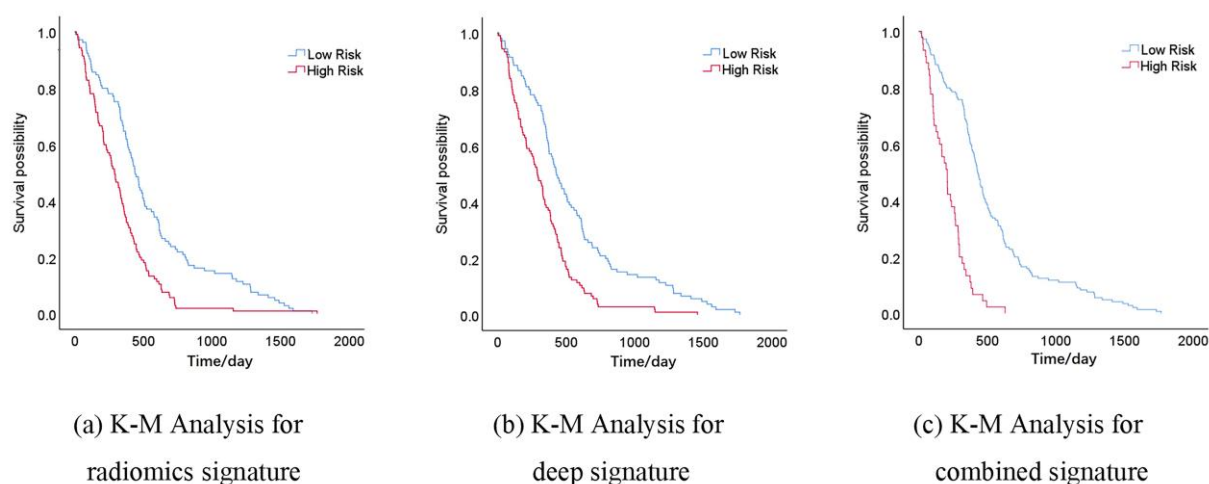


Figure 5. K-M analysis of the patients in the high- and low-risk groups in the training cohorts.

3.3. Validation of the signatures

In order to further evaluate the relationship between the three features and patient survival time, a univariate cox proportional hazards model was applied in the training and validation cohorts by using radiomics, deep and combined signature to calculate C-index. The higher the C-index, the better the prediction performance of the model. In the validation, the radiomics signature yielded a C-index of 0.679. The deep signature yielded a C-index of 0.707. The combined signature yielded a C-index of 0.718. Table 2 shows the C-index calculated for the training and validation cohorts.

Table 2. The Harrell concordance index of different radiomics signature and nomogram.

	Training cohort		Validation cohort	
	C-index	95% CI	C-index	95% CI
Radiomics signature	0.688	0.646-0.730	0.679	0.636-0.722
Deep signature	0.722	0.697-0.747	0.707	0.683-0.731
Combined signature	0.736	0.711-0.761	0.718	0.693-0.743
Nomogram	0.741	0.716-0.766	0.720	0.660-0.780

3.4. The nomogram and its performance

Considering that age has a greater impact on the prognosis of patients, we integrated the combined signature with clinical information (age) to generate a nomogram in the training cohort (Figure 6).

In the training cohort, the C-index of nomogram is 0.741, and in the validation cohort, the C-index of nomogram is 0.720. Figure 7 shows the calibration curve of nomogram. It can be seen that the predicted probability is very close to the actual survival time of the patient.

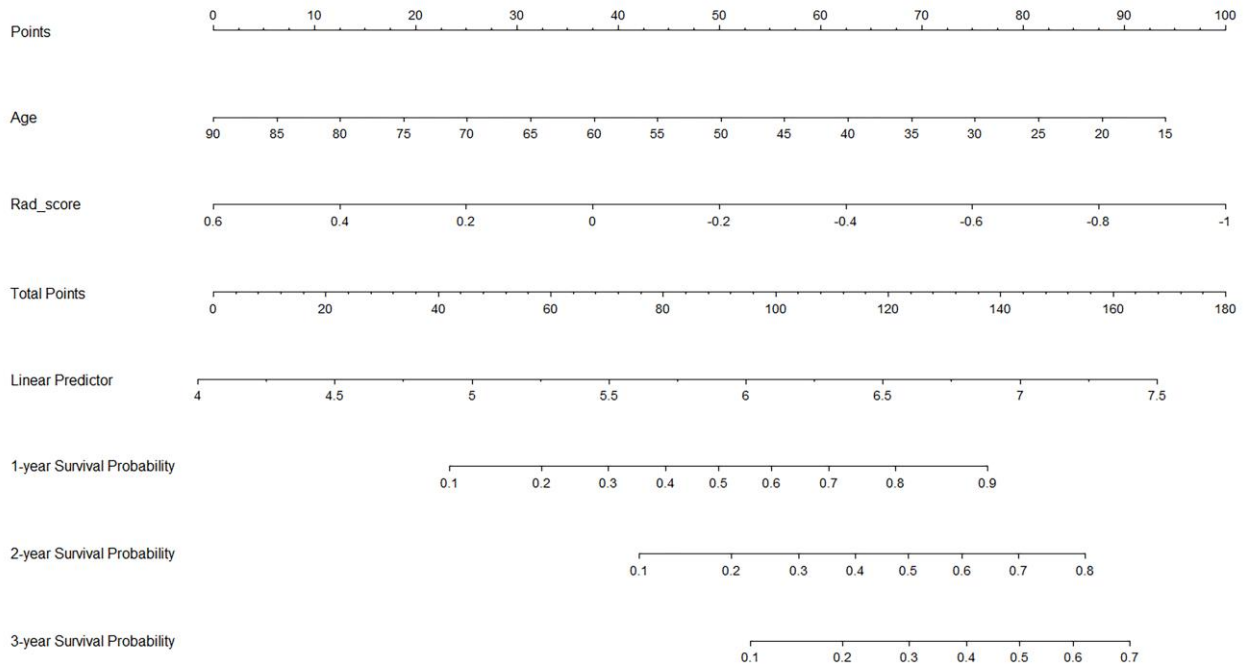


Figure 6. Nomogram.

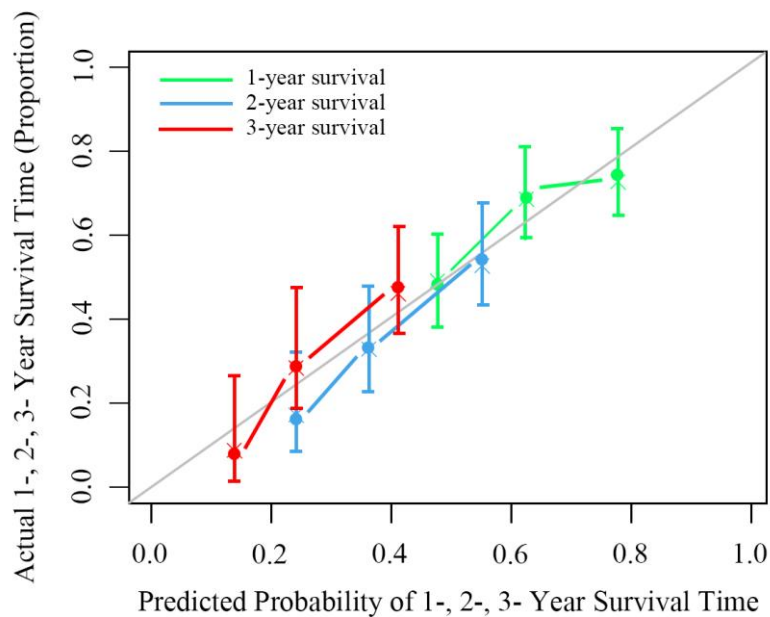


Figure 7. The calibration curve of the nomogram.

4. Conclusion

In conclusion, we proposed a prognostic model whose feature extraction is no longer limited to radiomics features. High-level deep features were extracted and integrated into our prognostic model. The results show that the deep features extracted through transfer learning are better than traditional radiomics features in predicting the survival rate of HGG patients. The results are as expected, because

deep features reflect higher-order imaging modes and capture more imaging heterogeneity than low-level shape, intensity, and texture features. According to the hypothesis of radiomics, the prognosis is poor because imaging heterogeneity of intra-tumor may be an expression of potential genetic heterogeneity, and it may cause the tumor to become resistant to treatment. However, how to explain the association between deep features and genetic features is still challenging. It is related to complex biological processes. Further work is needed to establish the principles of genomics to explain the correlation between deep features and genetic heterogeneity.

In the validation cohort, combined signatures can successfully predict the survival rate of patients, and they perform better than radiomics and deep signatures. From the statistical perspective, multivariate model is statistically robust in survival analysis [27]. Moreover, the intra-tumor genetic heterogeneity suggests that tumor subregions could be genetically different and may comprise multiple subclones. This could be better reflected by multiple high-order deep features extracted from multi-subregions in multi-modalities rather than individual feature. Similar to the genomic studies of exploring biomarkers from high-throughput genomic data, it is also regarded as a common “-omics” approach to construct a multi-factor radiomics signature for outcome prediction.

In this study, the performance of the combined signature is superior to clinical factors, such as age. None of these clinical factors successfully divided patients into groups with different prognostic risks. Based on the combined signature and clinical risk factors (age), we drew a nomogram that can intuitively predict the likelihood of survival. According to the calibration curve, we can see that the nomogram is of good prediction performance.

Despite encouraging results, this study still has some limitations. First, this is a retrospective study with a relatively small sample size. In the future, large-scale multi-center research is needed to fully evaluate the generalization ability of prognostic models. Second, due to the limitation of sample size, this study uses transfer learning to extract deep features. By fine-tuning on the pre-trained network or training from scratch, further work is required to train a dedicated feature extractor.

Acknowledgments

This work was supported by the Key Talents of Ningbo City Health Technology under Grant 2020SWSQNGG-06.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. J. M. Fahey, A. W. Girotti, Nitric Oxide antagonism to anti-glioblastoma photodynamic therapy: Mitigation by inhibitors of nitric oxide generation, *Cancers*, **11** (2019), 1–15.
2. G. Chi, F. Yang, D. Xu, W. Liu, Silencing hsa_circ_PVT1 (circPVT1) suppresses the growth and metastasis of glioblastoma multiforme cells by up-regulation of miR-199a-5p, *Artif. Cell. Nanomed. B*, **48** (2020), 188–196.
3. T. A. Dolecek, J. M. Propp, N. E. Stroup, C. Kruchko, CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005-2009, *Neuro-Oncology*, **14** (2012), 1–49.

4. M. Li, B. Li, J. Luo, J. Liang, F. Pan, Y. Zheng, et al., Ultrasound-based radiomics model in predicting efficacy of neoadjuvant chemotherapy in breast cancer, *Chin. J. Med. Imag. Tech.*, **35** (2019), 1331–1335.
5. H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, et al., Corrigendum: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, *Nat. Commun.*, **5** (2014), 4006.
6. C. Parmar, P. Grossmann, D. Rietveld, M. M. Rietbergen, P. Lambin, H. J. W. L. Aerts, Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer, *Front. Oncol.*, **5** (2015), 1–10.
7. L. He, Y. Q. Huang, Z. L. Ma, C. S. Liang, C. H. Liang, Z. Y. Liu, Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule, *Rep.*, **6** (2016), 34921.
8. T. P. Coroller, P. Grossmann, Y. Hou, E. R. Velazquez, R. T. H. Leijenaar, G. Hermann, et al., CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma, *Radiother. Oncol.*, **114** (2015), 345–350.
9. P. Kickingereder, D. Bonekamp, M. Nowosielski, A. Kratz, M. Sill, S. Burth, et al., Radiogenomics of glioblastoma: Machine learning–based classification of molecular characteristics by using multiparametric and multiregional MR imaging features, *Radiology*, **281** (2016), 907–918.
10. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, F. F. Li, Large-scale video classification with convolutional neural networks, *IEEE Conf. Comput. Vis. Pattern Recog.*, (2014), 1725–1732.
11. G. E. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science*, **313** (2006), 504–507.
12. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE*, **86** (1998), 2278–2324.
13. K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, Deep residual learning for image recognition, *IEEE Conf. Comput. Vis. Pattern Recog.*, (2016), 770–778.
14. L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected CRFs, *OALib J.*, **4** (2014), 357–361.
15. S. J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Engin.*, **22** (2010), 1345–1359.
16. A. Kensert, P. J Harrison, O. Spjuth, Transfer learning with deep convolutional neural networks for classifying cellular morphological changes, *Slas Discov.*, **24** (2019), 466–475.
17. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kerby, et al., Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features, *Sci. Data*, **4** (2017), 170117.
18. S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al., Identifying the best machine learning algorithms for brain tumor segmentation, *Prog. Assess. Overall Surv. Predic. BRATS Chall.*, (2018), 1–49.
19. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), *IEEE Trans. Med. Imag.*, **10** (2015), 1993–2024.

20. J. J. M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, et al., Computational radiomics system to decode the radiographic phenotype, *Cancer Res.*, **77** (2017), 104–107.
21. S. Chen, K. Ma, Y. Zheng, Med3D: Transfer learning for 3D medical image analysis, preprint, arXiv: 1904.00625.
22. N. Meinshausen, P. Bühlmann, High-dimensional graphs and variable selection with the Lasso, *Ann. Statist.*, **34** (2006), 1436–1462.
23. L. Yang, J. Yang, X. Zhou, L. Huang, W. Zhao, T. Wang, et al., Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients, *Eur. Radiol.*, **29** (2019), 2196–2206.
24. T. Sato, G. Berry, A comparison of two simple hazard ratio estimators based on the logrank test, *Stats. Med.*, **10** (2010), 749–755.
25. J. R. Figueira, S. Greco, B. Roy, ELECTRE methods with interaction between criteria: An extension of the concordance index, *Eur. J. Oper. Res.*, **199** (2009), 478–495.
26. T. M. Therneau, P. M. Grambsch, Modeling survival data: Extending the Cox Model, Springer Science & Business Media, Des Moines, 2013.
27. K. G M Moons, D. G Altman, J. B Reitsma, J. P. A. Ioannidis, P. Macaskill, E. W Steyerberg, et al., Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and Elaboration, *Ann. Intern. Med.*, **162** (2015), 1–73.



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)