

Article

Multi-Parametric Deep Learning Model for Prediction of Overall Survival after Postoperative Concurrent Chemoradiotherapy in Glioblastoma Patients

Han Gyul Yoon ^{1,†} , Wonjoong Cheon ^{2,3,†}, Sang Woon Jeong ², Hye Seung Kim ⁴, Kyunga Kim ⁴, Heerim Nam ⁵, Youngyih Han ^{1,2,*} and Do Hoon Lim ^{1,*}

¹ Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea; hangyul.yoon@samsung.com

² Samsung Advanced Institute for Health Science & Technology (SAIHST), Sungkyunkwan University School of Medicine, Seoul 06351, Korea; wonjoongcheon@ncc.re.kr (W.C.); sharkj@g.skku.edu (S.W.J.)

³ Proton Therapy Center, National Cancer Center, Goyang 10408, Korea

⁴ Statistics and Data Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul 06351, Korea; hyeseung.kim@sbri.co.kr (H.S.K.); kyunga.j.kim@samsung.com (K.K.)

⁵ Department of Radiation Oncology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul 03181, Korea; heerim.nam@samsung.com

* Correspondence: youngyih.han@samsung.com (Y.H.); dh8.lim@samsung.com (D.H.L.); Tel.: +82-2-3410-2612 (D.H.L.)

† Co-first author (Equal contribution).

Received: 23 July 2020; Accepted: 10 August 2020; Published: 14 August 2020



Abstract: This study aimed to investigate the performance of a deep learning-based survival-prediction model, which predicts the overall survival (OS) time of glioblastoma patients who have received surgery followed by concurrent chemoradiotherapy (CCRT). The medical records of glioblastoma patients who had received surgery and CCRT between January 2011 and December 2017 were retrospectively reviewed. Based on our inclusion criteria, 118 patients were selected and semi-randomly allocated to training and test datasets (3:1 ratio, respectively). A convolutional neural network-based deep learning model was trained with magnetic resonance imaging (MRI) data and clinical profiles to predict OS. The MRI was reconstructed by using four pulse sequences (22 slices) and nine images were selected based on the longest slice of glioblastoma by a physician for each pulse sequence. The clinical profiles consist of personal, genetic, and treatment factors. The concordance index (C-index) and integrated area under the curve (iAUC) of the time-dependent area-under-the-curve curves of each model were calculated to evaluate the performance of the survival-prediction models. The model that incorporated clinical and radiomic features showed a higher C-index (0.768 (95% confidence interval (CI): 0.759, 0.776)) and iAUC (0.790 (95% CI: 0.783, 0.797)) than the model using clinical features alone (C-index = 0.693 (95% CI: 0.685, 0.701); iAUC = 0.723 (95% CI: 0.716, 0.731)) and the model using radiomic features alone (C-index = 0.590 (95% CI: 0.579, 0.600); iAUC = 0.614 (95% CI: 0.607, 0.621)). These improvements to the C-indexes and iAUCs were validated using the 1000-times bootstrapping method; all were statistically significant ($p < 0.001$). This study suggests the synergistic benefits of using both clinical and radiomic parameters. Furthermore, it indicates the potential of multi-parametric deep learning models for the survival prediction of glioblastoma patients.

Keywords: glioblastoma; survival prediction; deep learning; radiomics

1. Introduction

Glioblastoma is the most common primary malignant brain tumor in adults, which remains a fatal disease [1]. Even with surgery and postoperative concurrent chemoradiotherapy (CCRT), the five-year overall survival (OS) rate for glioblastoma is 9.8%, and the approximate median OS is only about 15 months [2,3]. However, the survival outcomes of glioblastoma are heterogeneous among patients and are known to be relevant to numerous clinical and genetic factors [4–8].

Recently, several studies have attempted to analyze this heterogeneity of prognosis in glioblastoma patients by using features from medical images [9]. The development and improvement of numerous computational algorithms (such as machine learning) have rendered these “radiomic” studies more feasible. Using machine learning, many recent studies have built prediction models that take radiomic features as inputs and predict several clinical variables for glioblastoma, including survival outcomes [10–18]. However, these studies have several limitations. Many studies used handcrafted features from a region of interest (ROI) delineated by humans or features selectively determined by researchers; as such, the reproducibility of these methodologies cannot be guaranteed [10,11,13,15,16]. This is mainly due to their use of machine learning algorithms that are not based on deep learning. In these studies, the feature extraction and prediction models are separated using conventional machine learning algorithms. Thus, researchers are required to either select features using evaluation methods or to redefine features using dimension-reduction techniques such as principal component analysis. In contrast, a deep learning model can automatically extract features and integrate the feature extraction and prediction processes into a single model. In addition, several studies did not include patients’ clinical information in their prediction model, using only image data as input parameters [15,16]. Considering that there have been already several studies where clinical metrics were married to radiological measures for the evaluation, this problem becomes more significant [19–21]. Furthermore, very few studies predicted survival outcomes as continuous—rather than categorical (long- or short-term)—variables in glioblastoma patients [10,12].

To summarize, although deep learning has several strengths in image processing (such as automated feature extraction), only a few studies have used deep learning frameworks to predict the survival outcomes of glioblastoma. Of these studies, models predicting survival outcomes as continuous variables were rarely used. Moreover, despite their impact on the prognosis of glioblastoma, significant clinical variables were not fully integrated into the radiomics models of previous studies. Therefore, in this study, we built neural network-based deep learning models that (1) do not require an ROI delineation and a pre-defined artificial feature extraction or manual selection process; (2) predict the survival outcomes of glioblastoma as OS by month, rather than with categorical variables; (3) use clinical features and radiomic data together; and (4) integrate feature extraction and an OS-prediction algorithm into one model. This study aimed to evaluate the predictive performances of the deep learning models and identify whether the addition of clinical features—such as personal, genetic, treatment, and radiomic factors—delivers synergistic benefits in survival predictions.

2. Results

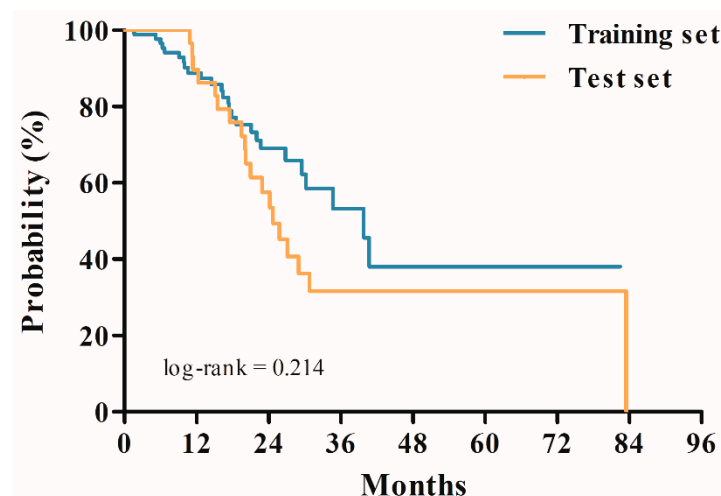
2.1. Patient Characteristics

The patient characteristics are summarized in Table 1. No significant differences were observed between the training and test sets in terms of baseline characteristics such as age, sex, Eastern Cooperative Oncology Group (ECOG) performance status, resection status, isocitrate dehydrogenase (IDH) mutation, O-6-methylguanine-DNA-methyltransferase (MGMT) hypermethylation, and additional adjuvant temozolomide (TMZ) cycles after CCRT. Furthermore, no difference was observed in OS between the two groups in terms of survival analysis using the Kaplan–Meier estimate (p -value = 0.214) (Figure 1).

Table 1. Baseline patient characteristics ($n = 118$).

Variables	Training Set ($n = 88$) ^a	Test Set ($n = 30$)	<i>p</i> -Value
Age (years)	Median 59 (IQR 50.75–64.25)	Median 54.5 (IQR 48–65.5)	0.410
Survival time (months) ^b	Median 17.60 (IQR 10.35–26.68)	Median 23.00 (IQR 17.03–33.59)	0.214 ^c
Sex			0.531
Male	44 (50.0%)	13 (43.3%)	
Female	44 (50.0%)	17 (56.7%)	
ECOG Performance Status			0.753
0–1	73 (83.0%)	27 (90.0%)	
2	15 (17.0%)	3 (10.0%)	
Resection			0.931
Gross total resection	36 (40.9%)	12 (40.0%)	
Subtotal resection	52 (59.1%)	18 (60.0%)	
IDH mutation			0.468
Yes	7 (8.0%)	4 (13.3%)	
No	81 (92.0%)	26 (86.7%)	
MGMT hypermethylation			0.921
Yes	42 (47.7%)	14 (46.7%)	
No	46 (52.3%)	16 (53.3%)	
Adjuvant TMZ cycles	Median 6 (IQR 4–6)	Median 6 (IQR 4.5–6)	0.300
Total radiotherapy dose			0.778 ^d
≥60 Gy	73 (83.0%)	26 (86.7%)	
<60 Gy	15 (17.0%)	4 (13.3%)	

ECOG, Eastern Cooperative Oncology Group; IDH, isocitrate dehydrogenase; MGMT, O-6-methylguanine-DNA-methyltransferase; TMZ, temozolomide; IQR, interquartile range. ^a Using the “createDataPartition” function from the “caret” R package, patient data were allocated to the training and test datasets while maintaining the ratio of IDH-mutants of each dataset. ^b Described as mean $\pm$ standard deviation. ^c Using Kaplan–Meier analysis. ^d Using Fisher’s exact test.

**Figure 1.** Kaplan–Meier survival curves for overall survival.

2.2. Model Performance Measured by C-Index and Integrated Area Under the Time-Dependent Receiver Operating Characteristic (ROC) Curve (iAUC)

Table 2 lists the root mean squared error (RMSE), the square root of the mean squared residuals, and Pearson’s correlation coefficient between the ground truth and predicted OS in each model. M_{CR} , the model which used both the clinical and radiomic features, showed the highest correlation coefficient (0.788) and the lowest RMSE (14.21 ± 23.07) among all the models in the study. However, RMSEs and correlation coefficients are not ideal evaluation metrics for survival-prediction models, mainly because of censored observations in survival data. Therefore, we calculated Harrell’s C-index and iAUC values of each model, and Table 3 lists the C-indexes and iAUC values of the models for OS prediction. Compared with the models using one type (M_{C1a} , M_{C1b} , M_{C1c}) or two types (M_{C2a} , M_{C2b} , M_{C2c}) of clinical features, the model using all three types of clinical features (M_{C3}) showed a higher C-index (0.693 (95% confidence interval (CI): 0.685, 0.701)) and iAUC (0.723 (95% CI: 0.716, 0.731)) in

most cases. Only the C-index of M_{C2b} (0.696 (95% CI: 0.688, 0.704)), which was slightly higher than that of M_{C3} , was an exception. Furthermore, M_{CR} showed a higher C-index (0.768 (95% CI: 0.759, 0.776)) and iAUC (0.790 (95% CI: 0.783, 0.797)) than M_{C3} (C-index = 0.693 (95% CI: 0.685, 0.701); iAUC = 0.723 (95% CI: 0.716, 0.731)) and M_R (C-index = 0.590 (95% CI: 0.579, 0.600); iAUC = 0.614 (95% CI: 0.607, 0.621)). To summarize, M_{CR} had the highest predictive ability of all models in the study (Figure 2).

Table 2. Root mean squared error and correlation coefficient of each model for overall survival prediction.

Model	Included Features	RMSE (Months) ^a	Correlation Coefficient
M_{C1a}	Personal only	16.96 ± 23.89	0.562
M_{C1b}	Genomic only	19.88 ± 30.40	0.194
M_{C1c}	Treatment only	25.18 ± 36.89	0.073
M_{C2a}	Personal + Genomic	17.19 ± 22.96	0.579
M_{C2b}	Personal + Treatment	16.64 ± 28.92	0.593
M_{C2c}	Genomic + Treatment	29.18 ± 38.57	−0.222
M_{C3}	Personal + Genomic + Treatment = Clinical	16.01 ± 26.54	0.712
M_R	Radiomic only	17.14 ± 25.47	0.499
M_{CR}	Clinical + Radiomic	14.21 ± 23.07	0.788

RMSE, root mean squared error. ^a Described as mean ± standard deviation.

Table 3. Harrell's concordance index (C-Index) and integrated area under the time-dependent receiver operating characteristic curve (iAUC) values of each model for overall survival prediction.

Model	Included Features	C-Index (95% CI)	iAUC (95% CI)
M_{C1a}	Personal only	0.644 (0.635, 0.653)	0.644 (0.636, 0.653)
M_{C1b}	Genomic only	0.664 (0.656, 0.671)	0.641 (0.634, 0.649)
M_{C1c}	Treatment only	0.562 (0.553, 0.570)	0.579 (0.572, 0.586)
M_{C2a}	Personal + Genomic	0.696 (0.688, 0.704)	0.675 (0.666, 0.684)
M_{C2b}	Personal + Treatment	0.665 (0.655, 0.675)	0.671 (0.663, 0.679)
M_{C2c}	Genomic + Treatment	0.640 (0.630, 0.650)	0.664 (0.657, 0.672)
M_{C3}	Personal + Genomic + Treatment = Clinical	0.693 (0.685, 0.701)	0.723 (0.716, 0.731)
M_R	Radiomic only	0.590 (0.579, 0.600)	0.614 (0.607, 0.621)
M_{CR}	Clinical + Radiomic	0.768 (0.759, 0.776)	0.790 (0.783, 0.797)

iAUC, integrated area under the time-dependent receiver operating characteristic curve; CI, confidence interval.

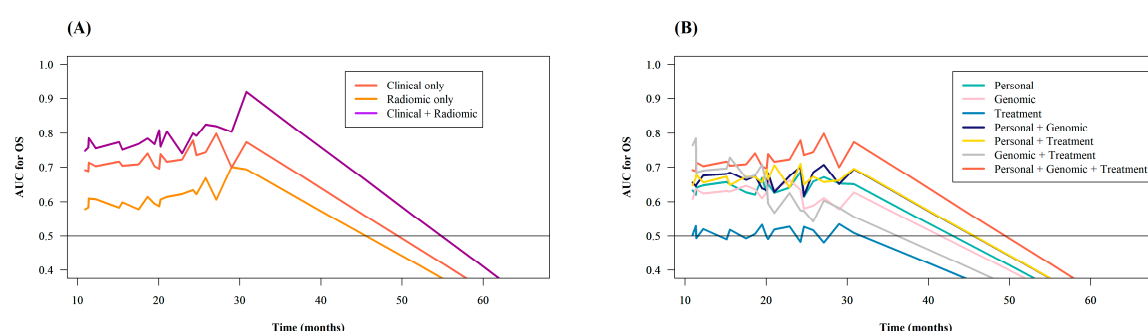


Figure 2. (A) and (B): Time-dependent receiver operating characteristic (ROC) curves of the models.

We statistically validated the synergistic effects of the clinical and radiomic features using 1000-times bootstrapping; the results are summarized in Table 4. The use of both clinical and radiomic features significantly improved the C-index compared with using the clinical (value difference = 0.074 (95% CI: 0.070, 0.078), $p < 0.001$) or the radiomic (value difference = 0.178 (95% CI: 0.174, 0.183), $p < 0.001$) features alone. The iAUC of the model using both features was also significantly higher than that of the model using clinical features alone (value difference = 0.067 (95% CI: 0.064, 0.070), $p < 0.001$) and the model using radiomic features alone (value difference = 0.176 (95% CI: 0.174, 0.179), $p < 0.001$).

Table 4. Comparison of prediction performances measured using Harrell's C-Index and iAUC value.

Index	Model 1	Model 2	Value Difference (95% CI) ^a	p-Value ^b
C-Index	Clinical only	Clinical + Radiomic	0.074 (0.070, 0.078)	<0.001
	Radiomic only	Clinical + Radiomic	0.178 (0.174, 0.183)	<0.001
iAUC	Clinical only	Clinical + Radiomic	0.067 (0.064, 0.070)	<0.001
	Radiomic only	Clinical + Radiomic	0.176 (0.174, 0.179)	<0.001

iAUC, integrated area under the time-dependent receiver operating characteristic curve; CI, confidence interval.

^a The value difference was calculated as (value of Model 2)–(value of Model 1) with 1000-times bootstrapping.

^b Using Bonferroni correction, the level of statistical significance in each comparison was set at $0.05/2 = 0.025$.

3. Discussion

This study investigated the predictive performance of multiparametric deep learning models for OS prediction in glioblastoma patients. The deep learning model employing both clinical and radiomic features showed a higher C-index and iAUC than any other model in the study. Furthermore, the benefits of utilizing both features rather than clinical or radiomic features alone were also evaluated using the bootstrapping method. The improvement in C-index and iAUC achieved by combining the two features was found to be statistically significant.

Several studies have designed radiomics models to predict the survival outcomes of glioblastoma [10–13,15,16]. Lao et al. [11] built a radiomics model with handcrafted and deep features from a dataset describing 75 patients. Compared with the Cox regression models of traditional risk factors (such as age and Karnofsky performance score [22]), the radiomics model showed improved performance in terms of OS prediction and the stratification of patients into low- and high-risk groups for a validation set containing 37 patients. Another study—also using a multiparametric radiomics model—demonstrated the predictive superiority of the radiomics model over models using conventional clinical factors for OS prediction [23]. Moreover, in the study of Bae et al. [10], radiomics models based on a random survival forest algorithm were used to predict the OS and progression-free survival of glioblastoma patients. A model incorporating radiomic features from the ROI—using clinical and genomic features—showed better predictive power than the models using each feature type alone.

However, as mentioned in the introduction, most previous studies use machine learning algorithms that are not based on deep learning. This has necessitated handcrafted features from the ROI or manual feature selection by clinicians, which is a major drawback of non-deep learning machine learning models. In contrast, this study used a model structure based on a convolutional neural network (CNN), a deep learning algorithm that has repeatedly shown promising ability to process and analyze image data [24–28]. Alongside its strengths in image analysis, CNN also presents other advantages such as automated feature extraction and selection. Owing to this benefit of CNNs, ROI delineation and artificial feature-selection processes could be omitted in this study. This characteristic of CNN and the fact that all MRI images in this study were obtained on the same machine may contribute to the reproducibility of our study results. Moreover, since it is relatively straightforward to adapt the network structure of models when using deep learning, we generated multiple prediction models with various combinations of input parameters. Thus, it became possible to identify the synergistic benefits obtained via the integrated use of clinical and radiomic features in this study.

CNN is a shift- or space-invariant deep learning model, which can automatically extract optimal features by itself from the given data to achieve the best performance. Thus, it is suitable for OS prediction considering positional relationship of features between MRI series [29]. In this study, we calculated the mean weight from the optimized weight of one-by-one convolution filter to analyze the weight of four pulse sequence MRI series. The one-by-one convolution network generated a weighted image ($512 \times 512 \times 1$) from the input image ($512 \times 512 \times 36$). As a result, MRI images of T1-weighted and apparent diffusion coefficient (ADC) images showed higher weight than T1-weighted

images with contrast-enhancement (CE), or T2-fluid-attenuated inversion recovery (FLAIR) pulse sequence images (Figure S1).

To the best of our knowledge, no studies except one have built deep learning-based models to predict OS as a continuous variable, using both clinical and radiomic data together [12]. In the exceptional study just referred to, the CNN-based model using image data and clinical/genomic features showed a lower RMSE and higher correlation coefficient (177.0 ± 130.0 and 0.4695 , respectively) than the random survival forest-based model (225.0 ± 136.0 and 0.1151 , respectively) or the CNN-based model using magnetic resonance imaging (MRI) data only (261.0 ± 175.0 and 0.0587 , respectively). However, owing to censored observations in survival data, RMSEs and correlation coefficients are not ideal evaluation metrics for survival-prediction models. Moreover, several critical clinical factors known to be significantly associated with the survival outcomes of glioblastoma (such as performance status and resection margin status) were not used in the study [4,5,30,31]. In contrast, our study calculated the C-index and iAUC values, which are more appropriate metrics for evaluating the predictive performance of survival-prediction models than the RMSE or correlation coefficient. The calculation of C-index and iAUC reflected individual survival status at each time point, which is not available in several statistical methods such as the standard ROC curve analysis [32,33]. Furthermore, our study included various clinical features known to be significantly associated with the survival outcomes of glioblastoma, which may have contributed to our more comprehensive analysis. Therefore, results from our study can overcome the limitations of the aforementioned study, and can address the clinical significance of multi-feature deep learning-based models in glioblastoma treatment.

The present study has several limitations. First, we retrospectively collected data from a single institution. Second, because our institution began to routinely acquire IDH mutation information in recent years, we excluded many patients due to a lack of genetic factors. Owing to this exclusion, we had to perform the study with a relatively small sample size. Additional studies with a large number of patients and a multi-center design are needed for the external validation of our results. Third, instead of Karnofsky performance score [22], which might have been a more idealistic criterion for evaluating patient's performance, our study used ECOG performance status. Fourth, only the limited number of patients received lower dose CCRT, and none of the included patients received tumor-treating fields (TTFs) in this study. Considering the modest sample size of the study, this might have contributed to the deterioration of our models' predictive performance. Fifth, the center images of each patient's MRI data (which included the largest proportion of suspicious lesions), were selected by clinicians in this study. However, this limitation might be overcome in future studies by using auto-segmentation algorithms such as U-net [34,35]. Sixth, several radiomic features including relative cerebral blood volume (rCBV) were not incorporated into our prediction models. With consideration of the recent advances, feature extraction reflecting the relationship of rCBV or the geographical relationship of radiographic features between different MRI series can be performed in future studies [36]. Finally, our study only used pre-operative MRI images as the radiomic features and predicted OS. Although this disadvantage could have been mitigated by using post-operative clinical features, further studies are required to validate the clinical utility of CNN-based models on post-operative images.

4. Materials and Methods

4.1. Patient Selection

The medical records of patients with glioblastoma were retrospectively reviewed. These patients had all received surgery followed by CCRT with TMZ between January 2011 and December 2017 at the Samsung Medical Center (Seoul, Korea). This study was approved by our Institutional Review Board (IRB #2019-02-070) and was performed under the guidelines of the Declaration of Helsinki. The inclusion criteria were as follows: (1) availability of pre-operative MRI data, including T1-weighted images, T1-weighted images with contrast-enhancement (CE), T2-fluid-attenuated inversion recovery (FLAIR) images, diffusion-weighted imaging (DWI), and apparent diffusion coefficient (ADC) images;

(2) completion of planned CCRT; and (3) availability of genetic profiles. Finally, 118 patients who met the inclusion criteria were included in the study. None of the patients received TTFs.

We obtained 14 clinical features of the included patients, and these features were classified into three categories: (1) “personal” factors ($n = 3$)—age, sex, and ECOG performance status; (2) “genetic” factors ($n = 2$)—IDH mutation status and MGMT hypermethylation status; (3) “treatment” factors ($n = 9$)—resection status, adjuvant TMZ cycles after CCRT (six cycles is the gold standard, according to the Stupp protocol [2] for glioblastoma), type of radiotherapy (RT), gross tumor volume (GTV), clinical target volume (CTV), and RT dose for the first RT, as well as GTV, CTV, and RT dose for the second RT planning (a boost RT plan after the first RT). GTV and CTV were used as continuous variables. Thereafter, patient data were allocated to the training and test datasets (in about 3:1 ratio, respectively) using the “createDataPartition” function from the “caret” R package. To avoid a significant difference in IDH mutation status between the two groups, patient data were allocated to the training and test datasets while maintaining the ratio of IDH-mutants of each dataset. No other variables were used for the stratification of the datasets.

4.2. Image Acquisition and Pre-Processing

All MRI images used in this study were taken at a single institution using identical 3.0 T MRI machines (Achieva, Philips Healthcare, Best, the Netherlands). The MRI images were reconstructed with matrix size of $512 \times 512 \times 22$ by using four pulse sequences. Four sequences of preoperative MRI data were used as “radiomic” features, including (1) T1-weighted images, (2) T1-weighted images with CE, (3) T2-FLAIR images, and (4) ADC images. Repetition times/echo times were 500.0/10.0 ms and 11,000.0/125.0 ms in the T1 and FLAIR images, respectively. These images shared the same acquisition parameters, with a section thickness of 5.0 mm and a reconstructed axial image of a 512×512 matrix with a pixel spacing of $0.469 \text{ mm} \times 0.469 \text{ mm}$. T1-CE images were captured 5 min after injecting a gadolinium-based agent. Moreover, DWI was performed with the following parameters: b -values = 1000 and 0 s/mm^2 , repetition/echo times = 3000.0/82.0 ms, and section thickness = 5.0 mm. We used a 256×256 matrix with a pixel spacing of $0.9375 \text{ mm} \times 0.9375 \text{ mm}$. Based on the DWI of two different b -values, ADC images were acquired using an EWS Workstation (Philips Healthcare, Amsterdam, Netherlands).

The preoperative MRI images used in this study were resized to a width and height of 256×256 pixels, to fix the size of the input for the deep learning models. Considering the memory capacity of the graphics processing unit (GPU), a total of nine images were selected, based on the longest slice of glioblastoma from one sequence of data ($\pm$ four slices). As a result, a total of 36 axial images (4 sequences $\times$ 9 images) from the four different image series were incorporated into our prediction model. The z-score normalization method was applied to the training data—including clinical and radiomic features—and the z-score parameters (mean and standard deviation) were determined. Then, the z-score parameters were independently applied to the test data.

4.3. Building Neural Network-Based Survival-Prediction Models

All procedures for building the deep learning-based OS-prediction models were performed using Google Tensorflow library version 1.14.0 (<http://tensorflow.org>) with an Nvidia GTX 1080Ti GPU. The OS prediction models were designed using clinical, radiomic, and both clinical and radiomic features. A total of nine neural network-based models were built in this study. These can be classified into five categories: (1) models with one type of clinical feature (M_{C1a} , M_{C1b} , and M_{C1c}), (2) models with two types of clinical feature (M_{C2a} , M_{C2b} , and M_{C2c}), (3) a model with all three types of clinical feature (M_{C3}), (4) a model using radiomic features (M_R), and (5) a model with both clinical and radiomic features (M_{CR}). The clinical features used in each M_C were as follows: M_{C1a} —personal only; M_{C1b} —genomic only; M_{C1c} —treatment only; M_{C2a} —personal and genomic; M_{C2b} —personal and treatment; M_{C2c} —genomic and treatment; M_{C3} —personal, genomic, and treatment.

Each model consisted of an input layer, a hidden layer, and an output layer. The model details are as follows: (1) in the models using clinical features alone (M_C), the size of the input layer was determined according to each model's input variable numbers. Thereafter, the input data passed through a hidden layer composed of four fully connected layers (256, 128, 64, and 32 nodes, respectively) and a single output node to predict the OS. (2) In the case of the radiomic model (M_R , Figure 3A), 36 images from the four MRI sequences (shaped $256 \times 256 \times 36$) were entered into the input layer of M_R . The hidden layer of M_R was based on a CNN composed of seven convolution layers and four fully connected layers. The CNN of the hidden layer performed two functions. The one-by-one convolution filter in the first convolution layer determined the importance of weight for each of the 36 images. The remaining six convolution layers (shaped from 28×28 to 7×7) automatically extracted the radiomic features (Radiomic Feature Extractor). Then, the 256 extracted radiomic features were inputted to the fully connected layer (Predictor) to predict the OS. (3) The prediction model employing both the clinical and radiomic features (M_{CR} , Figure 3B) had a similar CNN structure to M_R , except for a single additional fully connected layer containing 242 nodes. The 36 images from the four MRI sequences were entered into the input layer and passed through the CNN in a manner similar to that of M_R , and the Radiomic Feature Extractor in M_{CR} extracted 242 features using the additional fully connected layer. The 242 extracted radiomic features and 14 clinical features were then inputted to four fully connected layers to predict the OS. The architecture specifications of M_{CR} are listed in Table 5.

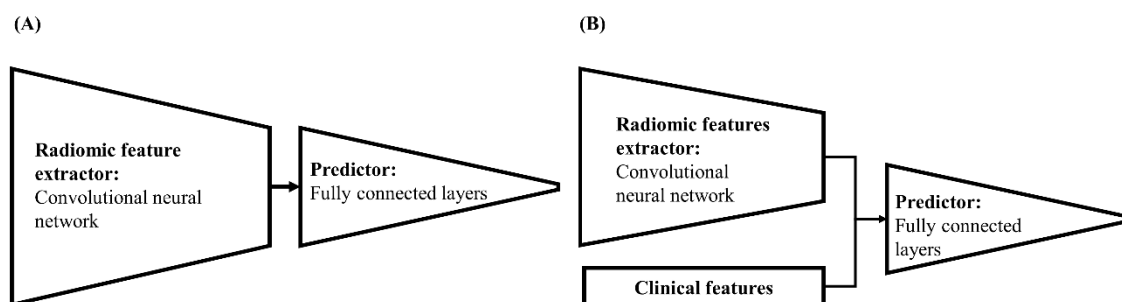


Figure 3. Diagrams of (A) the model using radiomic features and (B) the model using both clinical and radiomic features.

The batch size was empirically determined as 15; dropout and online augmentation techniques were applied to achieve generalizable performance. The online augmentation technique generated new images every epoch by using randomly selected parameters until the training procedure is over. The parameters of scaling, translation, rotation, and shear ranged from 95% to 100%, -3 to 3 pixels, -10 to 10 degrees, -3 to 3 degrees, respectively. The probability of image flip (left-right) was 50%. To manage the non-linearity, a leaky rectified linear unit activation function (LReLU) was located behind each layer [37]. After the input features had been passed from the input to the output layer, the loss was calculated (forward-propagation). The loss functions of the models were determined as the root mean squared error between the actual (ground-truth) and predicted OS; they were measured in months. In the back-propagation process, the hidden layer was optimized to minimize the loss using the adaptive moment estimation optimizer, with a learning rate of 0.0001. The training process was terminated when the loss became saturated on the smallest value, with a maximum of 5000 epochs. After the training was completed, the model structure and optimized parameters were saved, to be independently applied to the test set. The model performances were validated via 10-fold cross-validation.

Table 5. Layers and parameters of the prediction model employing both the clinical and radiomic features (M_{CR}).

Layer		Filter Shape	Shape	Activation/Pooling
Input layer	Input layer	-	$1 \times 256 \times 256 \times 36^{\dagger}$	None
Hidden layer: Extractor	Convolution layer	$1 \times 1 \times 36 \times 1$	$1 \times 256 \times 256 \times 1$	None/Max pooling
	Convolution layer	$1 \times 28 \times 28 \times 1$	$1 \times 128 \times 128 \times 1$	LReLU */Max pooling
	Convolution layer	$1 \times 14 \times 14 \times 1$	$1 \times 64 \times 64 \times 1$	LReLU/Max pooling
	Convolution layer	$1 \times 14 \times 14 \times 1$	$1 \times 32 \times 32 \times 1$	LReLU/Max pooling
	Convolution layer	$1 \times 7 \times 7 \times 1$	$1 \times 16 \times 16 \times 1$	LReLU/None
	Convolution layer	$1 \times 7 \times 7 \times 1$	$1 \times 16 \times 16 \times 1$	LReLU/None
	Flatten		$1 \times 1 \times 256$	None
	Fully connected layer	$1 \times 256 \times 242$	$1 \times 1 \times 242$	LReLU/None
Concatenate	Concatenate: clinical (14) and radiomic (242) features		$1 \times 1 \times 256$	None
Hidden layer: Predictor	Fully connected layer	$1 \times 242 \times 256$	$1 \times 1 \times 256$	LReLU/None
	Fully connected layer	$1 \times 256 \times 128$	$1 \times 1 \times 128$	LReLU/None
	Fully connected layer	$1 \times 128 \times 64$	$1 \times 1 \times 64$	LReLU/None
	Fully connected layer	$1 \times 64 \times 32$	$1 \times 1 \times 32$	None/None
Output layer	Fully connected layer	$1 \times 32 \times 1$	1×1	

[†] A total of 36 axial images from the four different image series were incorporated into our model. * Leaky rectified linear unit (LReLU) activation function.

4.4. Statistical Analysis

The primary objective of this study was OS prediction. The OS duration was calculated from the date of surgery to the date of the last follow-up (or death). Survival rates were calculated using the Kaplan-Meier method and were compared by log-rank tests. For a comparison of variables between groups, the Student's *t*-test was employed for continuous variables, and the Chi-square test or Fisher's exact test was used for categorical variables.

To evaluate the predictive performance of each model, Harrell's C-index and the iAUC were calculated. The iAUC is defined as the weighted mean of the area-under-the-curve over a follow-up period, and a higher iAUC suggests the superior predictive performance of the model [38,39]. The C-index and iAUC were calculated using the "survival" R package, and the "risksetAUC" function from the "risksetROC" R package, respectively. To calculate differences in the C-index and iAUC between two models, the 95% CIs of each value were computed using 1000-times bootstrapping. The difference was considered statistically significant if 95% CI of the difference did not include a zero. In each comparison, we used Bonferroni correction to compensate for multiple comparisons. In addition, we calculated the RMSE, the square root of the mean squared residuals, and Pearson's correlation coefficient between the ground truth and predicted OS in each model. All statistical analyses in this study were performed using the SAS software (version 9.4; Cary, NC, USA), and R (version 3.6.1; R Development Core Team, Vienna, Austria, <http://www.r-project.org>).

5. Conclusions

In conclusion, in this study, we proposed deep learning models for the prediction of OS in glioblastoma. In contrast to most previous studies, we predicted survival outcome as a continuous variable and automated the feature-extraction and selection processes by using CNN-based algorithms. Our findings suggest synergistic benefits of clinical and radiomic features; furthermore, they suggest future research directions in the building of multi-parametric deep learning models for the survival prediction of glioblastoma patients.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-6694/12/8/2284/s1>, Figure S1: The automatically determined weights for generating a weighted image to achieve high prediction accuracy for OS.

Author Contributions: Data curation, H.G.Y., W.C. and S.W.J.; formal analysis, H.G.Y., W.C., S.W.J., H.S.K. and K.K.; funding acquisition, Y.H.; project administration, Y.H. and D.H.L.; resources, H.N. and D.H.L.; software,

H.G.Y., W.C. and S.W.J.; supervision, Y.H. and D.H.L.; writing—original draft, H.G.Y. and W.C.; writing—review and editing, Y.H. and D.H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of the Korea, funded by the Ministry of Science, ICT and Future Planning (2019M2A2B4096537, 2019R1F1A1062775).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bleeker, F.E.; Molenaar, R.J.; Leenstra, S. Recent advances in the molecular understanding of glioblastoma. *J. Neurooncol.* **2012**, *108*, 11–27. [\[CrossRef\]](#)
2. Stupp, R.; Hegi, M.E.; Mason, W.P.; van den Bent, M.J.; Taphoorn, M.J.; Janzer, R.C.; Ludwin, S.K.; Allgeier, A.; Fisher, B.; Belanger, K.; et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol.* **2009**, *10*, 459–466. [\[CrossRef\]](#)
3. Koshy, M.; Villano, J.L.; Dolecek, T.A.; Howard, A.; Mahmood, U.; Chmura, S.J.; Weichselbaum, R.R.; McCarthy, B.J. Improved survival time trends for glioblastoma using the SEER 17 population-based registries. *J. Neurooncol.* **2012**, *107*, 207–212. [\[CrossRef\]](#)
4. Kim, N.; Chang, J.S.; Wee, C.W.; Kim, I.A.; Chang, J.H.; Lee, H.S.; Kim, S.H.; Kang, S.G.; Kim, E.H.; Yoon, H.I.; et al. Validation and optimization of a web-based nomogram for predicting survival of patients with newly diagnosed glioblastoma. *Strahlenther. Onkol.* **2020**, *196*, 58–69. [\[CrossRef\]](#)
5. Gittleman, H.; Lim, D.; Kattan, M.W.; Chakravarti, A.; Gilbert, M.R.; Lassman, A.B.; Lo, S.S.; Machtay, M.; Sloan, A.E.; Sulman, E.P.; et al. An independently validated nomogram for individualized estimation of survival among patients with newly diagnosed glioblastoma: NRG Oncology RTOG 0525 and 0825. *Neuro-oncology* **2017**, *19*, 669–677.
6. Lemeu, J.M.; Clavreul, A.; Menei, P. Intratumoral heterogeneity in glioblastoma: Don't forget the peritumoral brain zone. *Neuro-oncology* **2015**, *17*, 1322–1332. [\[CrossRef\]](#)
7. Soeda, A.; Hara, A.; Kunisada, T.; Yoshimura, S.; Iwama, T.; Park, D.M. The evidence of glioblastoma heterogeneity. *Sci. Rep.* **2015**, *5*, 7979. [\[CrossRef\]](#)
8. Inda, M.M.; Bonavia, R.; Seoane, J. Glioblastoma multiforme: A look inside its heterogeneous nature. *Cancers (Basel)* **2014**, *6*, 226–239. [\[CrossRef\]](#)
9. Juan-Albarracin, J.; Fuster-Garcia, E.; Garcia-Ferrando, G.A.; Garcia-Gomez, J.M. ONCOhabitats: A system for glioblastoma heterogeneity assessment through MRI. *Int. J. Med. Inf.* **2019**, *128*, 53–61. [\[CrossRef\]](#)
10. Bae, S.; Choi, Y.S.; Ahn, S.S.; Chang, J.H.; Kang, S.G.; Kim, E.H.; Kim, S.H.; Lee, S.K. Radiomic MRI phenotyping of glioblastoma: Improving survival prediction. *Radiology* **2018**, *289*, 797–806. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Lao, J.; Chen, Y.; Li, Z.C.; Li, Q.; Zhang, J.; Liu, J.; Zhai, G. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci. Rep.* **2017**, *7*, 10353. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Tang, Z.; Xu, Y.; Jin, L.; Aibaidula, A.; Lu, J.; Jiao, Z.; Wu, J.; Zhang, H.; Shen, D. Deep learning of imaging phenotype and genotype for predicting overall survival time of glioblastoma patients. *IEEE Trans. Med. Imaging* **2020**, *39*, 2100–2109. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Sanghani, P.; Ang, B.T.; King, N.K.K.; Ren, H. Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning. *Surg. Oncol.* **2018**, *27*, 709–714. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Liu, Y.; Xu, X.; Yin, L.; Zhang, X.; Li, L.; Lu, H. Relationship between glioblastoma heterogeneity and survival time: An MR imaging texture analysis. *AJNR Am. J. Neuroradiol.* **2017**, *38*, 1695–1701. [\[CrossRef\]](#)
15. Prasanna, P.; Patel, J.; Partovi, S.; Madabhushi, A.; Tiwari, P. Radiomic features from the peritumoral brain parenchyma on treatment-naïve multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: Preliminary findings. *Eur. Radiol.* **2017**, *27*, 4188–4197. [\[CrossRef\]](#)
16. Chaddad, A.; Sabri, S.; Niazi, T.; Abdulkarim, B. Prediction of survival with multi-scale radiomic analysis in glioblastoma patients. *Med. Biol. Eng. Comput.* **2018**, *56*, 2287–2300. [\[CrossRef\]](#)
17. Jang, B.S.; Jeon, S.H.; Kim, I.H.; Kim, I.A. Prediction of pseudoprogression versus progression using machine learning algorithm in glioblastoma. *Sci. Rep.* **2018**, *8*, 12516. [\[CrossRef\]](#)

18. Jeong, J.W.; Lee, M.H.; John, F.; Robinette, N.L.; Amit-Yousif, A.J.; Barger, G.R.; Mittal, S.; Juhasz, C. Feasibility of multimodal mri-based deep learning prediction of high amino acid uptake regions and survival in patients with glioblastoma. *Front. Neurol.* **2019**, *10*, 1305. [CrossRef]
19. Macdonald, D.R.; Cascino, T.L.; Schold, S.C., Jr.; Cairncross, J.G. Response criteria for phase II studies of supratentorial malignant glioma. *J. Clin. Oncol.* **1990**, *8*, 1277–1280. [CrossRef]
20. Wen, P.Y.; Macdonald, D.R.; Reardon, D.A.; Cloughesy, T.F.; Sorensen, A.G.; Galanis, E.; Degroot, J.; Wick, W.; Gilbert, M.R.; Lassman, A.B.; et al. Updated response assessment criteria for high-grade gliomas: Response assessment in neuro-oncology working group. *J. Clin. Oncol.* **2010**, *28*, 1963–1972. [CrossRef]
21. Kucharczyk, M.J.; Parpia, S.; Whitton, A.; Greenspoon, J.N. Evaluation of pseudoprogression in patients with glioblastoma. *Neurooncol. Pract.* **2017**, *4*, 120–134. [CrossRef] [PubMed]
22. Peus, D.; Newcomb, N.; Hofer, S. Appraisal of the Karnofsky Performance Status and proposal of a simple algorithmic system for its evaluation. *BMC Med. Inform. Decis. Mak.* **2013**, *13*, 72. [CrossRef] [PubMed]
23. Li, Q.; Bai, H.; Chen, Y.; Sun, Q.; Liu, L.; Zhou, S.; Wang, G.; Liang, C.; Li, Z.C. A fully-automatic multiparametric radiomics model: Towards reproducible and prognostic imaging signature for prediction of overall survival in glioblastoma multiforme. *Sci. Rep.* **2017**, *7*, 14331. [CrossRef] [PubMed]
24. Kahn, C.E., Jr. From images to actions: Opportunities for artificial intelligence in radiology. *Radiology* **2017**, *285*, 719–720. [CrossRef]
25. Yasaka, K.; Akai, H.; Kunimatsu, A.; Kiryu, S.; Abe, O. Deep learning with convolutional neural network in radiology. *Jpn. J. Radiol.* **2018**, *36*, 257–272. [CrossRef]
26. Bhandari, A.; Koppen, J.; Agzarian, M. Convolutional neural networks for brain tumour segmentation. *Insights Imaging* **2020**, *11*, 77. [CrossRef]
27. Wong, D.J.; Gandomkar, Z.; Wu, W.J.; Zhang, G.; Gao, W.; He, X.; Wang, Y.; Reed, W. Artificial intelligence and convolution neural networks assessing mammographic images: A narrative literature review. *J. Med. Radiat. Sci.* **2020**, *67*, 134–142. [CrossRef]
28. Bernal, J.; Kushibar, K.; Asfaw, D.S.; Valverde, S.; Oliver, A.; Marti, R.; Llado, X. Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: A review. *Artif. Intell. Med.* **2019**, *95*, 64–81. [CrossRef]
29. Qi, M.; Li, Y.; Wu, A.; Jia, Q.; Li, B.; Sun, W.; Dai, Z.; Lu, X.; Zhou, L.; Deng, X.J.M.P. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. *Med. Phys.* **2020**, *47*, 1880–1894. [CrossRef]
30. Lamborn, K.R.; Chang, S.M.; Prados, M.D. Prognostic factors for survival of patients with glioblastoma: Recursive partitioning analysis. *Neuro-Oncology* **2004**, *6*, 227–235. [CrossRef]
31. Li, J.; Wang, M.; Won, M.; Shaw, E.G.; Coughlin, C.; Curran, W.J., Jr.; Mehta, M.P. Validation and simplification of the Radiation Therapy Oncology Group recursive partitioning analysis classification for glioblastoma. *Int. J. Radiat. Oncol. Biol. Phys.* **2011**, *81*, 623–630. [CrossRef] [PubMed]
32. Kamarudin, A.N.; Cox, T.; Kolamunnage-Dona, R. Time-dependent ROC curve analysis in medical research: Current methods and applications. *BMC Med. Res. Methodol.* **2017**, *17*, 53. [CrossRef] [PubMed]
33. Han, X.; Zhang, Y.; Shao, Y. On comparing 2 correlated C indices with censored survival data. *Stat. Med.* **2017**, *36*, 4041–4049. [CrossRef] [PubMed]
34. Alom, M.Z.; Yakopcic, C.; Hasan, M.; Taha, T.M.; Asari, V.K. Recurrent residual U-Net for medical image segmentation. *J. Med. Imaging (Bellingham)* **2019**, *6*, 014006. [CrossRef] [PubMed]
35. Yang, T.; Song, J.; Li, L.; Tang, Q. Improving brain tumor segmentation on MRI based on the deep U-net and residual units. *J. Xray Sci. Technol.* **2020**, *28*, 95–110. [CrossRef]
36. Emblem, K.E.; Pinho, M.C.; Zollner, F.G.; Due-Tonnessen, P.; Hald, J.K.; Schad, L.R.; Meling, T.R.; Rapalino, O.; Bjornerud, A. A generic support vector machine model for preoperative glioma survival associations. *Radiology* **2015**, *275*, 228–234. [CrossRef]
37. Wang, S.H.; Phillips, P.; Sui, Y.; Liu, B.; Yang, M.; Cheng, H. Classification of Alzheimer’s Disease Based on Eight-Layer Convolutional Neural Network with Leaky Rectified Linear Unit and Max Pooling. *J. Med. Syst.* **2018**, *42*, 85. [CrossRef]
38. Heagerty, P.J.; Saha-Chaudhuri, P.; Saha-Chaudhuri, M.P. Package ‘RisksetROC’. 2012. Available online: <http://cran.rapporter.net/web/packages/risksetROC/risksetROC.pdf> (accessed on 26 September 2012).
39. Heagerty, P.J.; Zheng, Y. Survival model predictive accuracy and ROC curves. *Biometrics* **2005**, *61*, 92–105. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).