Survival Prediction in Glioblastoma Using Combination of Deep Learning and Hand-Crafted Radiomic Features in MRI Images

Ying Zhuge, Holly Ning, Jason Y. Cheng, Erdal Tasci, Peter Mathen, Kevin Camphausen, Robert W. Miller, and Andra V. Krauze *

Radiation Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA;
Email: zhugey@mail.nih.gov (Y.Z.), hning@mail.nih.gov (H.N.), jason.cheng@nih.gov (J.Y.C.),
erdal.tasci@nih.gov (E.T.), peter.mathen@nih.gov (P.M.), camphauk@mail.nih.gov(K.C.),
robert.miller@nih.gov (R.W.M.)
*Correspondence: andra.krauze@nih.gov (A.V.K.)

***Abstract*—Glioblastoma (GBM) is the brain's most common malignant primary tumor. Survival prediction is crucial for risk stratification and impacts all aspects of care. However, previous attempts at predicting survival outcomes have relied on clinical parameters that suffer from uneven capture in available data sets, limiting the precision of transferable survival predictions. In this study, we propose a novel method for overall survival prediction in GBM patients that combines deep learning and hand-crafted radiomic features using brain Magnetic Resonance Imaging (MRI) images. The proposed method involves three main steps: 3D brain tumor segmentation using the nnU-Net model, patient classification into long, short, and mid-survivors using the Dense-Net model on segmented tumor regions, and the combination of deep learning features with hand-crafted radiomic features and patient age information. Feature selection is performed using the Least Absolute Shrinkage and Selection Operator (LASSO) regression model and the DeepSurv is utilized for survival prediction. The proposed method was evaluated on the BraTS Benchmark 2020 training data sets using nested five-fold cross-validation. The resulting C-index values for the training, validation, and testing sets were 0.984, 0.821, and 0.821, respectively, outperforming the random survival forest method. Our findings suggest that the proposed method has the potential to serve as an imaging biomarker for predicting overall survival in GBM patients, with superior transferability compared to traditional machine learning-based methods.**

***Keywords*—survival prediction, deep learning, neural networks, Random Survival Forest (RSF), Magnetic Resonance Imaging (MRI) image**

I. Introduction

Glioblastoma (GBM) is the brain's most common malignant primary tumor [1], with a heterogeneous histological appearance and generally poor prognosis. The average survival time of a patient with GBM is 12–18 months [2], with only 25% of GBM patients surviving more than one year and only 5% surviving more than five years. Patients diagnosed with GBM are typically managed with maximal safe resection followed by radiation therapy and chemotherapy. The ability to estimate prognosis is crucial for patients and physicians to select the most appropriate treatment plan, such that it is sufficiently aggressive to allow for tumor control while minimizing adverse long-term normal tissue changes [3–5]. The goal of superior survival prediction is to appropriately de-escalate management when the prognosis is poor, emphasizing the patient`s quality of life and best supportive care and aggressively managing patients who may experience superior survival outcomes. Given that the Overall Survival (OS) time can vary significantly across individuals, survival prediction models could be employed to guide the treatment of GBM patients and are clinically crucial for customized treatment management. Previous attempts at predicting survival outcomes have relied on clinical parameters in the neuro-oncology field, which are known to suffer from uneven capture in available data sets, impairing transferable survival precision. Age as a clinical feature has been the most consistent predictor of OS and is robustly captured in all data sets [6]. Given the above limitations and unmet needs in the neuro-oncology field, we propose a method for OS prediction that employs brain MRI images to classify patients, given that this is the most significant source of data generated in the clinic. The most consistently captured and thus currently publicly available MRI sequences are T1-, T2-, post-contrast T1-, and FLAIR-weighted MRI images. To maximally leverage MRI images as a data source in GBM, we employ a combination of deep learning and hand-crafted radiomic features on the sequences mentioned above, as they will sustain the basis of most prediction models for years to come. In this study, a novel method is proposed for overall survival prediction in GBM by using the combination of deep learning features, hand-crafted radiomic features, and patient age information, where deep learning features are extracted from a deep learning-based image classification

Manuscript received February 27, 2023; revised June 28, 2023; accepted July 31, 2023; published December 26, 2023.

task on T1-, T2-, post-contrast T1-, and FLAIR-weighted MRI images. The proposed method consists of three steps: (1) 3D brain tumor segmentation based on the popular nnU-Net model; (2) Patients are classified into three groups as long survivors (e.g., >900 days), short survivors (e.g., <300 days), and mid-survivors (between 300 and 900 days) using the Dense-Net model on segmented brain tumor regions with deep learning features extracted through the last fully connected layer of the classification model [7]; (3) Four hundred hand-crafted radiomic features are calculated on three tumor subregions (edema, enhancing tumor and necrosis) by using the pyradiomics toolbox, which includes 14 shape features, 18 first-order statistics features, 22 Gray-Level Co-occurrence Matrix (GLCM) features, 16 Gray-Level Run Length Matrix (GLRLM) features, 16 Gray-Level Size Zone Matrix (GLSZM) features, and 14 Gray-Level Difference Matrix (GLDM) features, for a total of 100 hand-crafted radiomic features for each MRI modality image [8]. Deep learning features are combined with these hand-crafted radiomic features and patient age information. After feature selection using the Least Absolute Shrinkage and Selection Operator (LASSO) regression model, a Cox proportional hazards neural network (DeepSurv [9]) is utilized for survival prediction on the combination of features. Three models, one in each step, have been built accordingly. The proposed method has three advantages over the previously mentioned methods. (1) Imaging features are implicitly extracted from a DenseNet-based image classification task, combining fine details in high image resolution levels and deep abstract semantic information in low image resolution levels. These features represent the most significant patterns in classifying patients into long-, mid-, and short-term survival categories; (2) Operations of complicated feature dimension reduction are not required, only one step of the LASSO feature reduction is necessary; (3) Instead of using a linear Cox regression model [10] or a radiomics nomogram [11], a subsequent Cox proportional hazards-based neural network model has been utilized for survival prediction with the combination of both deep learning features and hand-crafted radiomic features wherein the importance of each feature is weighted during the model training.

II. Literature Review

Multiple attempts have been made to design robust scoring systems predictive of outcomes in GBM [12, 13]. Mostly, these methods have relied on clinical parameters such as age, Karnofsky Performance Status (KPS), gender, and resection/methylation status. However, significant tumor heterogeneity, both in spatial and temporal space, limits the ability of these clinical prognostic factors to capture GBM characteristics fully, thus undermining robust survival prediction [14]. KPS, often poorly captured or not captured, varies over time, while the extent of resection is subject to interpretation. Methylation status is often unavailable and, in some publicly funded systems, only carried out in some patients (e.g., over the age of 65) and, as a result, is often missing [15]. Advanced Magnetic Resonance Imaging (MRI) techniques have shown benefits in initial diagnosis, treatment planning, and treatment response assessment and have been increasingly used as a non-invasive tool in survival prediction. MRI can provide distinctive imaging information independent of pathologic and clinical data [16]. Studies have shown a significant correlation between imaging phenotypes and genomic signatures. Thus, imaging phenotypes can serve as non-invasive biomarkers of cellular gene expression. Various statistical models have been developed to utilize MRI image features for survival prediction. Primarily, the MRI image features are extracted from segmented tumor regions [17, 18]. Feng *et al.* [19] developed a simple linear regression model for overall survival prediction using only nine radiomic features calculated from the brain tumor segmentation based on an ensemble of 3D U-Nets. Yousaf *et al*. [20] proposed a random forest classifier-based approach for survival prediction on the Brain Tumor Segmentation (BraTS) 2019 data. Radiomic features were extracted from the provided ground-truth segmentation masks, in which Haralick texture features were shown to be more significant for the survival prediction task. Baid *et al.* [21] also built a radiomic model based on a random forest for OS prediction in GBM. First-order, image intensity-based volume, shape-based geometry, and texture features, with 678 radiomic features, were extracted from three subregions of the brain tumor in T1-contrast enhanced-weighted and Fluid-Attenuated Inversion Recovery (FLAIR) MRI images. Macyszyn *et al*. [22] proposed a Support Vector Machines-based approach to identify complex and reproducible imaging patterns predictive of overall survival and molecular subtype in GBM. The machine learning algorithm selected approximately sixty diverse features from conventional and advanced preoperative multiparametric MRI images to derive imaging predictors of patient survival. Asthana *et al.* [23] proposed a regression model to predict the survival rates of patients with high-grade brain tumors based on the information set extracted from a segmented brain tumor using the U-Net based semantic segmentation method. Tang *et al*. [24] presented the OS time prediction for glioblastoma using the multi-modal deep K-Nearest Neighbors (KNN) strategy. Each patient's final overall OS time is determined by its *K* nearest patients with known OS time in a learned metric space. They claimed that, compared with the typical end-to-end prediction method, it is more robust to noise and data inconsistency. To overcome the lack of interpretability of radiomic features, Pálsson *et al.* [25] introduced novel MRI features computed from the whole-brain and tumor segmentation and fed into a random survival forest model to predict patient survival. The calculated features have a direct anatomical-functional interpretation via measuring the deformation caused by the brain tumor on the surrounding brain structures, which improves the performance of survival models for both overall and progression-free survival. The method applies to both pre-and post-operative MRI.

Unlike traditional machine learning approaches, neural network-based methods have recently attracted attention in

survival analysis. Wang *et al.* [26] developed a fully connected neural network with two hidden layers for survival prediction. Seven features, including approximate surface areas of three tumor sub-regions, the ratios of the volume of each tumor sub-region to the size of the whole brain, and patient age information, were employed as inputs to the neural network to produce the survival days prediction. Banerjee *et al*. [27] designed a radiomic model of similar Multi-Layer Perception (MLP) for predicting overall survival. Thirty-three semantic and fifty agnostic features from segmented brain tumor regions were extracted and provided as input to the MLP to indicate the number of survival days.

Among these survival prediction methods, radiomic feature generation is a critical step. However, one common problem is that most radiomic features are explicitly calculated or handcrafted. These handcrafted radiomic features may include first-order intensity distribution of voxels in brain tumor subregions, shape-based geometry features, and second-order texture features like gray level co-occurrence matrix, gray level run length matrix, gray level dependence matrix, gray level size zone matrix, and neighboring gray-tone difference matrix. Although the number of handcrafted features generated in this manner can reach tens of thousands, they are mainly superficial and low-order imaging features, which may not represent deep, abstract semantic characteristics imaging features of brain tumors [10]. In addition, if the number of radiomic features is too high, a complicated step of feature selection or feature vector dimension reduction is typically required. Sun *et al.* [28] applied a decision tree regression model with gradient boosting to rank the importance of each feature. The optimal number of features is determined through cross-validation. Fourteen out of 4,524 radiomic features are finally selected and fed to a random forest model to predict the survival of a patient with GBM. In Osman's method [29], a set of 147 radiomic features were extracted from segmented tumor sub-regions on conventional multi-modality MRI images, followed by LASSO Cox regression applied to obtain the coefficients of each radiomic feature. A radiomic signature with 9 features was constructed and then trained/tested on eight machine learning classification models for stratifying GBM patients based on survival. In Baid's method [21], to reduce the number of radiomic features, Spearman's correlation coefficient was calculated for each pair of radiomic features. The features with high Spearman's correlation were discarded. The features were further reduced by excluding all variables with statistically insignificant relationships with survival groups.

To overcome the drawbacks of superficial and low-order imaging features in radiomics, Lao *et al*. [10] proposed a deep learning-based radiomics model for survival prediction in GBM. A total of 98,304 deep features were extracted from a Convolutional Neural Network (CNN) via transfer learning. These features were combined with 1403 handcrafted features calculated from multi-modality MRI images to form the radiomic features. However, the approach still required a complicated four-step feature selection procedure. The final radiomic nomogram having only six radiomics features was constructed on the validation data set and combined with the clinical risk factors, such as age and KPS, based on a multivariate cox regression model for survival prediction. In summary, in this study, the goal of the proposed method for OS prediction in GBM is to combine deep learning features, hand-crafted radiomic features, and patient age information, employing T1-, T2-, post-contrast T1-, and FLAIR-weighted MRI images.

III. Materials and Methods

The fundamental steps in the proposed method: (1) 3D brain tumor segmentation; (2) Deep learning feature extraction via image classification, and (3) Cox proportional hazards neural networks, as well as the data sets utilized in this study, are explained in detail in below sections. The overall diagram of the proposed method is shown in Fig. 1.

A. Data

The publicly available data set BraTS 2020 has been used in the study, which includes 293 High-Grade Gliomas (HGG/GBM) and 76 Low-Grade Gliomas (LGG) patients [30, 31] with a pathologically confirmed diagnosis and available overall survival. The clinical characteristics of GBM patients in BraTS 2020 are shown in Table I. All these data are preoperative, with a native T1-weighted scan (T1)**,** a native T2-weighted scan (T2), a post-contrast T1 weighted scan (T1-Gd), and a T2 Fluid Attenuated Inversion Recovery weighted scan (T2-FLAIR) and were acquired with different clinical protocols using various scanners from different institutions. All images in the BraTS 2020 dataset have been segmented manually by one to four raters, following expert neuroradiologists' approval of the annotation protocol and neuro-radiologist annotations. These annotations were taken as ground truths for model training and testing. Annotations comprise the post-contrast enhancing tumor, the peritumoral edema, the necrotic, and the non-enhancing tumor core. All images have been co-registered to the same anatomical reference using a rigid transform-based program implemented in ITK [32]. In addition, all images have been skull-stripped and interpolated to 1×1×1 mm^3 voxel resolution. The overall survival data, defined in days, are included in a Comma-Separated Value (.csv) file corresponding to 236 BraTS 2020 imaging data pseudo-identifiers. The .csv file also includes the age of patients, as well as the resection status. into allow for increased homogeneity, in this study, only subjects with resection status of Gross Total Resection (GTR) were evaluated (119 subjects), as described in the BraTS 2020 data set website.

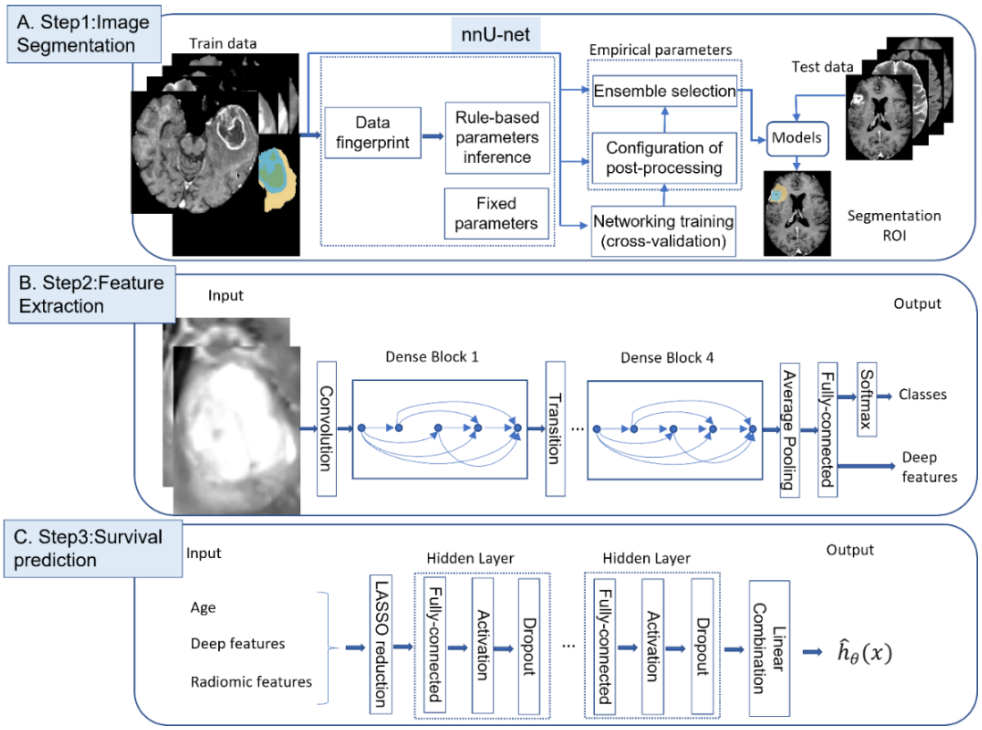

Figure 1. The diagram of the proposed method includes three steps: image segmentation via the nnU-net, deep imaging feature extraction, and survival prediction.

TABLE I. CLINICAL CHARACTERISTICS OF GBM PATIENTS IN BRATS 2020

Clinical characteristics	
No. of patients	236
No. of patients with GTR	119
Age Ranges (years)	18.98–86.65
Age Median (years)	61.47
OS Ranges (days)	5–1767
OS median (days)	371

B. Brain Tumor Segmentation

Thanks to the annual Multi-modal (BraTS challenge at the conference of the MICCAI since 2013. [31], methods of brain tumor segmentation in multi-modal MRI images have demonstrated a promising performance, particularly after the introduction of CNNbased deep learning techniques in the field of medical image segmentation [33]. The U-Net and its variants are the most popular methods among the winners of the annual BraTS challenge [34]. This study adopted a state-of-the-art nnU-Net ("no new U-Net"), a deep learning framework with automated self-configuration for brain tumor segmentation [7].

To configure a deep learning-based image segmentation method, we typically manually select a set of hyperparameters, and network architecture configurations are carried out with an iterative trial and error process in the model training and performance monitoring of the model on a validation set. When employing the automated configuration by nnU-Net, dataset properties such as imaging modality, image sizes, voxel spacings, class ratios, etc., are summarized in a "dataset fingerprint". A set of heuristic rules operates on this fingerprint to infer the data-dependent hyper-parameters of the pipeline. These are complemented by some fixed parameters, which do not require adaptation. Up to three architectures, a 2D U-Net, a 3D U-Net on full image resolutions, and a cascade 3D U–Net is trained based on these pipeline fingerprints in a 5–fold cross-validation. Finally, nnU-Net automatically selects the optimal ensemble of these architectures and performs postprocessing if required. An example of brain tumor segmentation of one patient is illustrated in Fig. 2.

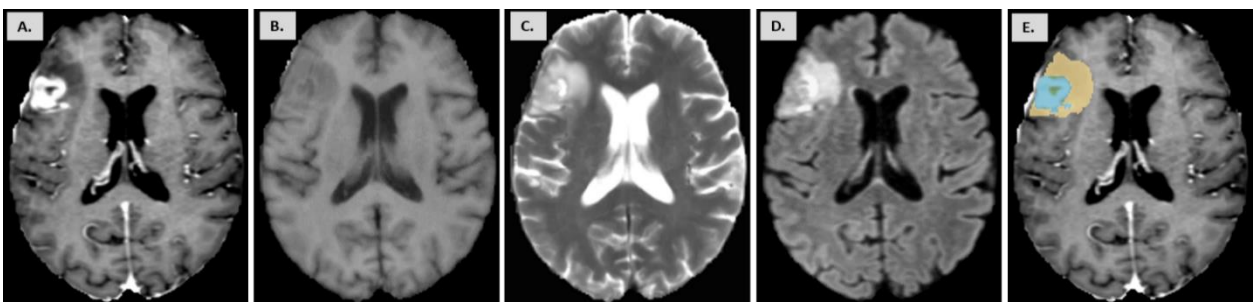

Figure 2. The brain tumor segmentation was illustrated using the nnU-Net method on a GBM patient. (a). T1-Gd. (b). T1. (c). T2. (d). T2-FLAIR. (e). Tumor segmentation result (yellow color on edema, green on tumor core, and blue on enhancing tumor).

C. Imaging Feature Extraction via Image Classification

We implicitly extract imaging features through image classification to alleviate the cumbersome procedure of hand-crafted feature extraction and feature dimension reduction. All patients are classified into three groups in terms of their survival days, short-survivors (<300 days), long-survivors (>900 days), and mid-survivors (between 300 and 900 days), as suggested by Baid *et al.* [21]. A survivor classification model based on the densely connected convolution neural networks (DenseNet) [35] was built for deep imaging feature extraction. The flowchart of the procedure is shown below in Fig. 1(b).

The architecture of the DenseNet takes the multi-modal image volumes of segmented brain tumor regions from Section III: B as inputs. The output includes classes from the softmax layer and imaging features from the last fully connected layer. In this study, we focus on the extracted imaging features that play the most significant role in classifying GBM patients into different survival groups. Four dense blocks are adopted in the architecture, which has 121 layers in total. In any dense block, each layer obtains additional features from all preceding layers and passes on its feature maps to all subsequent layers via concatenation operations. As a result, each layer has access to all the preceding feature-maps in its dense block and, therefore, can access the whole network's "collective feature knowledge. The network design has a strong gradient flow since the error signal can be more directly propagated to earlier layers. This is a kind of implicit deep supervision, as earlier layers can get direct supervision from the final classification layer. More diversified features are generated, with richer patterns for better classification. The transition layers between adjacent dense blocks include both convolution and pooling layers and are used to down-sample the feature-map size. The size of the final output imaging features is set as 1024 in the study.

D. Survival Prediction

Unlike in the BraTS challenge, where patient survival days are predicted and evaluated, we focus on predicting the probability of an event (risk of death) happening at a particular time. The Cox Proportional Hazard (CPH) is a linear regression model most widely used in survival studies of cancer patients. The CPH model is used to predict the risk (hazard) of an outcome (death) based on multiple prognostic variables [36]. The hazard at time *t* for an individual with covariates *x* is assumed to be

$$\lambda(t|x) = \lambda_0(t)e^{h(x)} \tag{1}$$

In the model, $\lambda_0(t)$ is a baseline hazard function, and $e^{h(x)}$, is the relative risk, a proportionate increase or reduction in risk associated with the set of covariates *x*. The log-risk function $h(x)$ is estimated by a linear function $\hat{h}_\beta(x) = \beta^T x$. To perform the CPH regression, parameters β are adjusted to optimize the cox partial likelihood, which is the product of the probability at each event time t_i until the event has occurred to individual i, given the set of individuals still at risk at time t_i. The cox partial likelihood is defined as

$$L(\beta) = \prod_{i:e_i=1} \frac{e^{\hat{h}_\beta(x_i)}}{\sum_{j\epsilon R(t_i)} e^{\hat{h}_\beta(x_j)}} \tag{2}$$

where the values t_i, e_i, and x_i are the respective event time, event indicator, and baseline data for the i^{th}observation. The risk set $R(t_i)$ is the set of patients still at risk of failure the event at time t_i [9].

However, it might be too simplistic to assume that the log-risk function $h(x)$ is a linear combination of covariates *x*, and it may not be appropriate in many applications.

The survival function is derived from the predicted hazard function as

$$S(t) = \exp\left\{-\int_0^t \lambda(t|x)dx\right\} \tag{3}$$

In this study, the DeepSurv, a neural network-based Cox proportional hazard model. Ref. [9] was utilized to estimate a nonlinear log-risk function $\hat{h}_\theta(x)$. The architecture of the DeepSurv is shown in Fig. 1(c), which is essentially a multi-layer perceptron. The Cox proportional hazard neural network input includes extracted deep learning features from Section III: C. hand-crafted radiomic features, and patient age information. Each hidden layer consists of a fully connected layer, a non-linear activation function, and a dropout layer to avoid over-fitting in model training. The output layer has a single node with a linear combination of the hidden features from the previous hidden layer, producing the output of the predicted log-risk hazard function $\hat{h}_\theta(x)$. The objective function for this network is set as the average negative log partial likelihood from the CPH regression model defined in Eq. (2), with an additional regularization, except that the linear function $\hat{h}_\beta(x)$ is replaced by a nonlinear log-risk function $\hat{h}_\theta(x)$:

$$l(\theta) = -\frac{1}{N_{e=1}}\sum_{i:e_i=1}\left(\hat{h}_\theta(x_i) - \log\sum_{j\epsilon R(t_i)} e^{\hat{h}_\theta(x_j)}\right) + \lambda \times \|\theta\|_2^2, \tag{4}$$

where λ is the l_2 regularization parameter, $N_{e=1}$ is the number of patients with an observable event e_i, and $R(t_i)$ is the set of patients that have not experienced the event at time t_i.

IV. Result and Discussion

For brain tumor segmentation, the implementation of nnU-Net (GitHub–MIC–DKFZ/nnUNet) in python has been utilized, which is an automated configuration method based on the PyTorch framework and covers the whole segmentation pipeline, including preprocessing, network architecture, training, and postprocessing. The segmentation model was trained on the BraTS 2020 dataset with 369 patients in five-fold cross-validation, which allows the nnU-Net to determine the optimal ensemble of three U-Net models and whether post-processing is required.

A modification of the implementation of 3D DenseNet-based image classification on the Monai framework (GitHub—Project-MONAI/MONAI: AI Toolkit for Healthcare Imaging) was employed to classify all patients

into three different survival groups. There was a total of 119 subjects with GTR. The data were divided into training (80%) and validation (20%). This step aims to identify the imaging features that play a key role in patient classification. The epoch training number is set as 1000, and the growth rate k = 32, which determines how many feature maps generated in each layer of a dense block is concatenated to the global state of the network.

The DeepSurv implementation was based on the pycox package. (GitHub—havakv/pycox: Survival analysis with PyTorch). The neural network architecture was a simple Vanilla MLP with three hidden layers with 60, 20, and 3 nodes in each layer, respectively. Batch normalization was utilized for reducing data noise and model training stabilization. The dropout rate was set as 0.1 among the hidden layers.

The concordance index (C-index) was used to measure the survival prediction accuracy. The C-index is a commonly used metric in survival prediction that compares the survival time recorded in a dataset to the predicted patient death time ranking. A score C-index of 0.5 is expected from random prediction, and 1.0 is expected if two rankings are in perfect concordance. The Random Survival Forest (RSF) method [37] for survival prediction was implemented based on the package of pysurvival (https://www.pysurvival.io/) for comparing the performance with the DeepSurv. To demonstrate the effectiveness of extracted imaging features from deep learning, we calculated four hundred hand-crafted radiomic features proposed by using the pyradiomics toolbox, [8] which include 14 shape features, 18 first-order statistics features, 22 gray-level co-occurrence matrix (GLCM) features, 16 gray-level run length matrix (GLRLM) features, 16 gray-level size zone matrix (GLSZM) features, and 14 gray-level difference matrix (GLDM) features, in total 100 radiomic features for each MRI modality image. These 400 hand-crafted radiomic features are integrated with the 1024 deep-learning features and patient age characteristics. After using the LASSO feature selection, 120 features are left and taken as the inputs to feed into both models of the DeepSurv and the RSF. The performances of these two models are evaluated by using nested five-fold cross-validation. All data are divided into 64% training, 16% validation, and 20% testing sets. The C-index values on the training, validation, and testing sets are listed in Table II. We also calculated the C-index values for the two models with separate image features and patient age characteristics for comparison purposes.

TABLE II. THE MEAN C-INDEX VALUES FOR THE TWO MODELS ON TRAINING, VALIDATION, AND TESTING SETS, USING HAND-CRAFTED RADIOMIC FEATURES, DEEP LEARNING FEATURES, AND THEIR COMBINATION

Model	Image features used	Training data	Validation data	Testing data
DeepSurv	Hand-crafted	0.9752	0.5872	0.5360
	Deep learning	0.9813	0.8130	0.8168
	Combination	0.9839	0.8210	0.8211
RSF	Hand-crafted	0.7398	0.6087	0.5626
	Deep learning	0.7978	0.7659	0.7738
	Combination	0.8166	0.7823	0.7847

To demonstrate if the feature reduction technique is helpful in the proposed method for survival prediction, we calculated the C-index values for the two models using all deep learning features, hand-crafted radiomic features, and patient age information on the same data sets without the LASSO feature selection. The mean C-index values are listed in Table III.

The DeepSurv method outperformed the RSF method (Table II) except when hand-crafted radiomic features were employed. This is consistent with similar findings in the literature [28, 38] wherein traditional machine learning methods have shown promising results while deep learning proved unstable in survival prediction when employing hand-crafted radiomic features. DeepSurv and RSF methods with deep learning features performed better than the corresponding methods with hand-crafted radiomic features. The methods with combination image features performed better than those with separate image features (hand-crafted or deep learning features).

TABLE III. THE MEAN C-INDEX VALUES FOR THE TWO MODELS WITHOUT LASSO FEATURE REDUCTION ON TRAINING, VALIDATION, AND TESTING SETS, USING HAND-CRAFTED RADIOMIC FEATURES, DEEP LEARNING FEATURES, AND THEIR COMBINATION

Model	Image features	Training data	Validation data	Testing data
DeepSurv	Hand-crafted	0.9698	0.5311	0.5039
	Deep learning	0.9707	0.7744	0.7592
	Combination	0.9724	0.7829	0.7727
RSF	Hand-crafted	0.7309	0.5838	0.5598
	Deep learning	0.8123	0.7864	0.7853
	Combination	0.8115	0.7919	0.7782

It is noteworthy that in the absence of the LASSO feature reduction (Table III), the RSF method continued to perform well for each of the three image features sets, across all training, validation, and testing datasets as compared to DeepSurv (Table III) which exhibits poorer C-index values in the absences on LASSO. The decline in performance likely reflects decreased stability secondary to an increasing number of image features in the absence of the LASSO feature reduction technique.

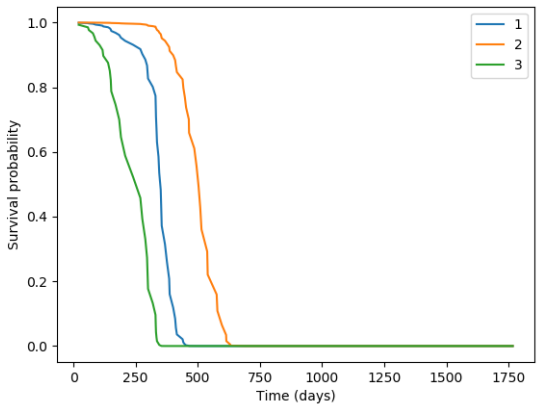

Figure 3. Predicted survival probability over time for three patients in the validation data set.

In addition to predicting the log-risk hazard function $\hat{h}_\theta(x)$ using the DeepSurv model, we also

implemented the survival probability prediction, as defined as Eq. (2). Fig. 3 shows the percent survival probability over time for three random patients in the validation data set. The abscissa represents the estimated survival time of a subject. The baseline hazard function $\lambda_0(t)$ is estimated via Breslow approximation [39].

Early attempts to employ neural networks for survival analysis date back to 1995 by Farragi and Simon [40]. A feed-forward neural network was built with only a single hidden layer, which was taken as a nonlinear extension of the Cox proportional hazards model to model the relationship between primary covariates and the corresponding hazard risk function. However, their performance had improved compared to the traditional CPH method. Recently, with the development of deep learning techniques and GPU computational capabilities, more sophisticated neural networks for survival analysis have improved over the CPH model when dealing with real-world non-linear data. One representative method is the DeepSurv as an addition to Faraggi-Simon's architecture with multiple hidden layers. Another similar method (DeepHit) was developed to process survival data with competing risks [41].

Suter *et al*. [42] compared the deep learning technique with the classical regression method for survival prediction of patients with high-grade brain tumors. Two CNN-based architectures were built, one for direct survival days regression and the other for extracting deep features which are combined with radiomics, shape, and atlas features, where the top 30 features are selected and fed into a Support Vector Classifier (SVC) for survival regression. Two CNNs show unstable results on BraTS 2018 data. In addition, in the second CNN architecture, a feature reduction technique was utilized to choose the top 30 features out of 1353 extracted features; however, none of the 120 deep features from the CNN architecture were retained. This may have been caused by the simple CNN architectures, which have only two/three convolutional layers and lack of capability to extract deeper and higher-order imaging features. Furthermore, both CNN architectures output the predicted survival in days. The model parameter selection in optimization was challenging since the accuracy performance was unstable. In a similar method proposed by Lao *et al.* [10] the final six features were picked out using the LASSO technique employing both 1403 handcrafted features and 98,304 deep features, all originating from a pre-trained CNN via transfer learning. In our analysis, both the DeepSurv and the traditional machine learning RSF methods based on deep learning features have shown promising results and outperformed the methods using hand-crafted features. The strength of this novel method is that in the procedure of deep imaging feature extraction, an elegant neural network architecture DenseNet has been utilized, and deep features are extracted through the task of survivor classification instead of the survival prediction in days. DenseNet was specially developed to improve accuracy caused by the vanishing gradient in very deep neural networks due to the long distance between input and output layers. It has been shown to have better feature use efficiency, outperforming the popular ResNet architecture with fewer parameters [35]. With the combination of both deep learning features and hand-crafted radiomic features, both DeepSurv and RSF methods outperform the methods with separate image features. Recently, transformers have attracted significant attention in medical image segmentation and classification [43, 44]. However, for brain tumor segmentation in MRI images, the nnU-Net is still a standard for comparison, enforcing the novelty of this study [45]. We note that the nnU-Net method won the MICCAI BraTS challenge 2020 and two extensions of nnU-Net ranked 1st and 3rd in the MICCAI BraTS challenge 2021 [46, 47]. Of note, Swin UNRTR, Swin Transformers for semantic segmentation of brain tumors in MRI images, published in 2022, only ranked 7th in the MICCAI BraTS challenge 2021 validation phase [48]. It should also be noted that for the MICCAI BraTS challenge 2023, ground truth labels are created by the fusion of nnU-Net results with those from earlier two methods: DeepScan and DeepMedic, with additional manual refinement by neuro-radiologists [49]. In the future, we will consider more advanced transformer-based image segmentation methods over the nnU-Net method, to investigate if there is any further performance improvement.

Future directions also include testing the replacement of the DenseNet architecture in the proposed method to improve survival prediction since the Self-attention mechanism in the transformers may result in superior image feature extraction.

Typically, the number of features often exceeds the number of training samples in machine learning. Feature selection is necessary for traditional machine learning methods to improve model interpretability, speed up computation, and improve model performance on unseen data. However, most feature selection methods are restricted to the linear estimation functions such as LASSO. Yamada et al. proposed a fully embedded feature selection method named Stochastic Gates (STG) for nonlinear functions by introducing the stochastic gates to the input layer of a neural network [50]. This method has outperformed other commonly used methods in predictive performance and feature selection in synthetic and real-life datasets. In future studies, it will be worth investigating a replacement of the DeepSurv with the STG to improve performance in survival prediction.

Limitations of our study include only utilizing patient age information as a clinical feature. Additional clinical features such as chemotherapy status, tumor histology, tumor site, and KPS need to be further combined with imaging features for survival prediction acknowledging limitations in both capture and accuracy of these features. In addition, in this study, only subjects with resection status of Gross Total Resection (GTR) were included to allow for testing of the novel method in a robust, more homogenous data set. We suspect that variable source documentation and interpretation of what constitutes Subtotal Resection (STR) or biopsy and how these relate to survival will add further complexity to survival prediction in the context of MRI images and in-depth analysis of all types of resection status will be required to

advance survival prediction in this space. Several studies have also demonstrated that incorporating MRI radiomics, clinical factors and genomics in traditional machine learning approaches can improve survival prediction [51, 52]. In a similar manner studies will need to be carried out using deep learning techniques based on integration of MRI radiomics, genetic and clinical risk factors to produce more effective and reliable survival prediction.

V. Conclusion

In this paper, we presented a survival prediction approach in GBM patients that employs a combination of deep learning features, hand-crafted radiomic features in MRI images, and patient age information. The deep learning features were extracted through the DenseNet-based model to classify patients into short, medium, and long survivors. We demonstrated that the proposed method performed better than the traditional machine learning based survival prediction method RSF. The method has the potential to be an imaging biomarker for prediction of the overall survival in patients with GBM facilitating care, decision making, and improving the outcomes of patients with GBM.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

YZ conceived the study, participated in ongoing data analysis of the data set, designed and optimized the analysis, and drafted and reviewed the manuscript; HN assisted in data organization and access and reviewed the manuscript; JYC assisted in data organization and access and reviewed the manuscript; ET assisted in study optimization and reviewed the manuscript; PM assisted in the concept review and reviewed the manuscript; KC assisted in the concept review and reviewed the manuscript; RWM provided resources, assisted in data organization and access and reviewed the manuscript; AVK conceived the study, participated in its design and coordination, co-drafted and reviewed the manuscript; all authors had approved the final version.

Funding

Funding was provided in part by NCI NIH intramural program (ZID BC 010990).

References

[1] F. E. Bleeker, R. J. Molenaar, and S. Leenstra, "Recent advances in the molecular understanding of glioblastoma," *J. Neurooncol,* vol. 108, no. 1, pp. 11–27, May 2012.

[2] E. G. V. Meir, C. G. Hadjipanayis, A. D. Norden, H. K. Shu, P. Y. Wen, and J. J. Olson, "Exciting new advances in neuro-oncology: The avenue to a cure for malignant glioma," *CA Cancer J. Clin.*, vol. 60, no. 3, pp. 166–193, 2010.

[3] R. A. Morshed, J. S. Young, S. L. Hervey-Jumper, and M. S. Berger, "The management of low-grade gliomas in adults," *J. Neurosurg Sci.,* vol. 63, no. 4, pp. 450–457, Aug. 2019.

[4] M. Geurts and M. J. V. D. Bent, "On high-risk, low-grade glioma: What distinguishes high from low?" *Cancer,* vol. 125, no. 2, pp. 174–176, 2019.

[5] M. J. V. D. Bent, M. Smits, J. M. Kros, and S. M. Chang, "Diffuse infiltrating oligodendroglioma and astrocytoma," *J. Clin. Oncol.,* vol. 35, no. 21, pp. 2394–2401, 2017.

[6] R. Zhao, J. Zeng, K. D. Vries, R. Proulx, and A. V. Krauze, "Optimizing management of the elderly patient with glioblastoma: Survival prediction online tool based on BC Cancer Registry real-world data," *Neuro Oncol. Adv.,* vol. 4, no. 1, 2022.

[7] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nat Methods,* vol. 18, no. 2, pp. 203–211, Feb. 2021.

[8] J. J. M. V. Griethuysen *et al.*, "Computational radiomics system to decode the radiographic phenotype," *Cancer Res,* vol. 77, no. 21, pp. e104–e107, Nov 1 2017.

[9] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network," *BMC Med Res Methodol,* vol. 18, no. 1, 24, 2018.

[10] J. Lao *et al.*, "A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme," *Sci. Rep.,* vol. 7, no. 1, 10353, 2017.

[11] X. Zhang *et al.*, "A radiomics nomogram based on multiparametric MRI might stratify glioblastoma patients according to survival," *Eur Radiol,* vol. 29, no. 10, pp. 5528–5538, Oct. 2019.

[12] C. D. Corso, R. S. Bindra, and M. P. Mehta, "The role of radiation in treating glioblastoma: Here to stay," *J. Neurooncol,* vol. 134, no. 3, pp. 479–485, Sep. 2017.

[13] R. Zhao and A. V. Krauze, "Survival prediction in gliomas: Current state and novel approaches," in *Gliomas*, W. Debinski Ed. Brisbane (AU), 2021.

[14] R. O. Mirimanoff *et al.*, "Radiotherapy and temozolomide for newly diagnosed glioblastoma: Recursive partitioning analysis of the EORTC 26981/22981-NCIC CE3 phase III randomized trial," *J. Clin. Oncol.,* vol. 24, no. 16, pp. 2563–2569, June 2006.

[15] P. Lambin *et al.*, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer,* vol. 48, no. 4, pp. 441–446, Mar. 2012.

[16] M. A. Mazurowski, A. Desjardins, and J. M. Malof, "Imaging descriptors improve the predictive power of survival models for glioblastoma patients," *Neuro Oncol.,* vol. 15, no. 10, pp. 1389–1394, Oct. 2013.

[17] M. Diehn *et al.*, "Identification of noninvasive imaging surrogates for brain tumor gene-expression modules," in *Proc. Natl. Acad. Sci. USA*, 2008, vol. 105, no. 13, pp. 5213–5218.

[18] Y. Li *et al.*, "Serial analysis of imaging parameters in patients with newly diagnosed glioblastoma multiforme," *Neuro Oncol.,* vol. 13, no. 5, pp. 546–557, May 2011.

[19] X. Feng, N. J. Tustison, S. H. Patel, and C. H. Meyer, "Brain tumor segmentation using an ensemble of 3D U-nets and overall survival prediction using radiomic features," *Front Comput Neurosci,* vol. 14, 25, 2020.

[20] S. Yousaf, S. M. Anwar, H. R. Prakash, and U. Bagci, "Brain tumor survival prediction using radiomics features," *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology*, Cham, pp. 284–293, 2020.

[21] U. Baid *et al.*, "Overall survival prediction in glioblastoma with radiomic features using machine learning," *Front Comput. Neurosc.*, vol. 14, 2020.

[22] L. Macyszyn *et al.*, "Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques," *Neuro Oncol.*, vol. 18, no. 3, pp. 417–425, Mar. 2016.

[23] P. Asthana, M. Hanmandlu, and S. Vashisth, "Brain tumor detection and patient survival prediction using U-Net and regression model," *International Journal of Imaging Systems and Technology*, vol. 32, pp. 1801–1814, 2022.

[24] Z. Tang, H. Cao, Y. Xu, Q. Yang, J. Wang, and H. Zhang, "Overall survival time prediction for glioblastoma using multimodal deep KNN," *Physics in Medicine and Biology*, vol. 67, 2022.

[25] S. Pálsson, S. Cerri, H. S. Poulsen, T. Urup, I. Law, and K. V. Leemput, "Predicting survival of glioblastoma from automatic whole-brain and tumor segmentation of MR images," *Scientific Reports*, vol. 12, no. 1, 19744, 2022.

[26] F. Wang, R. Jiang, L. Zheng, C. Meng, and B. Biswal, "3D U-Net based brain tumor segmentation and survival days prediction," *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 131–141, 2020.
[27] S. Banerjee, S. Mitra, and B. U. Shankar, "Multi-planar spatial-convnet for segmentation and survival prediction in brain cancer," *Lect Notes Comput Sc*, vol. 11384, pp. 94–104, 2019.
[28] L. Sun, S. T. Zhang, H. Chen, and L. Luo, "Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning," *Front Neurosci-Switz*, vol. 13, 2019.
[29] A. F. I. Osman, "A multi-parametric MRI-based radiomics signature and a practical ML model for stratifying glioblastoma patients based on survival toward precision oncology," *Front Comput Neurosc*, vol. 13, 2019.
[30] S. Bakas *et al.*, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Sci Data*, vol. 4, 170117, 2017.
[31] B. H. Menze *et al.*, "The multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
[32] W. S. L. Ibanez, *The ITK Software Guide 2.4*, Clifton Park, NY: Kitware, Inc., 2005.
[33] Y. Zhuge *et al.*, "Brain tumor segmentation using holistically nested neural networks in MRI images," *Med. Phys.*, vol. 44, no. 10, pp. 5234–5243, Oct. 2017.
[34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," presented at the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Cham, 2015.
[35] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.
[36] D. R. Cox, "Regression models and life-tables,"*Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 34, no. 2, pp. 187–220, 1972.
[37] I. Hemant, B. K. Udaya, H. B. Eugene, and S. L. Michael, "Random survival forests," *The Annals of Applied Statistics*, vol. 2, no. 3, pp. 841–860, Sep 2008.
[38] S. Yousaf, S. M. Anwar, H. RaviPrakash, and U. Bagci, "Brain tumor survival prediction using radiomics features," *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology*, pp. 284–293, 2020.
[39] N. Breslow, "Covariance analysis of censored survival data," *Biometrics*, vol. 30, no. 1, pp. 89–99, Mar. 1974.
[40] D. Faraggi and R. Simon, "A neural network model for survival data," *Stat. Med.*, vol. 14, no. 1, pp. 73–82, Jan. 1995.
[41] C. Lee, J. Yoon, and M. V. Schaar, "Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data," *IEEE Trans Biomed Eng*, vol. 67, no. 1, pp. 122–133, Jan. 2020.
[42] Y. Suter *et al.*, "Deep learning versus classical regression for brain tumor patient survival prediction," in *BrainLes@MICCAI*, 2018.
[43] S. Li, X. Sui, X. Luo, X. Xu, Y. Liu, and R. S. M. Goh, "Medical image segmentation using squeeze-and-expansion transformers," arXiv preprint, arXiv:2105.09511, 2021.
[44] J. Cheng, J. Liu, H. Kuang, and J. Wang, "A fully automated Multimodal MRI-based multi-task learning for glioma segmentation and IDH genotyping," *IEEE Transactions on Medical Imaging*, vol. 41, no. 6, pp. 1520–1532, Jun 2022.
[45] H. M. Luu and S. H. Park, "Extending nn-UNet for brain tumor segmentation," in *Proc. 7th International Conference on Brain Lesion Workshop*, 2021, pp. 173–186.
[46] RSNA. 2021-ai-challenge-winners. [Online]. Available: https://www.rsna.org/news/2021/november/2021-ai-challenge-winners
[47] V. Braunstein. VIDIA data scientists take top spots in MICCAI 2021 brain tumor segmentation challenge. [Online]. Available: https://developer.nvidia.com/blog/nvidia-data-scientists-take-top-spots-in-miccai-2021-brain-tumor-segmentation-challenge/
[48] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. Roth, and D. Xu, "Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images," *Lecture Notes in Computer Science*, 12962, 2021.
[49] H. Bran Li, G. M. Conte, S. M. Anwar, F. Kofler, I. Ezhov, K. V. Leemput, M. Piraud *et al.*, "The Brain Tumor Segmentation (BraTS) challenge 2023: Brain MR image Synthesis for tumor segmentation (BraSyn)," *Electrical Engineering and Systems Science*, vol. 5, 2023.
[50] Y. Yamada, O. Lindenbaum, S. Negahban, and Y. Kluger, "Feature selection using stochastic gates," in *Proc. the 37th International Conference on Machine Learning Research*, 2020.
[51] P. Mobadersany *et al.*, "Predicting cancer outcomes from histology and genomics using convolutional networks," in *Proc. Natl. Acad. Sci. USA*, 2018, vol. 115, no. 13, pp. 2970–2979.
[52] Y. Tan, W. Mu, X. C. Wang, G. Q. Yang, R. J. Gillies, and H. Zhang, "Improving survival prediction of high-grade glioma via machine learning techniques based on MRI radiomic, genetic and clinical risk factors," *Eur. J. Radiol.*, vol. 120, 108609, Nov 2019.