

Original paper

An attention-based multimodal approach for predicting overall survival in patients with glioblastoma using clinical and imaging data

Saman Mohammadpour^{1,A,B,C,D,E,F,G}, Hassan Emami^{1,A,C,D,E,G}, Reza Rabiei^{1,A,C,D,E,G}, Arash Roshanpoor^{2,A,C,D,E},
Morteza Tabatabaeefer^{3,A-B,D,E}, Faeghi Fariborz^{4,D,E}

¹Department of Health Information Technology and Management, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

²Department of Computer, Yl.C., Islamic Azad University, Tehran, Iran

³Department of Radiation Oncology, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

⁴Department of Radiology Technology, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Abstract

Purpose: To develop and evaluate an attention-based multimodal model for improving the prediction of overall survival in patients with glioblastoma by integrating clinical and imaging features.

Methods: Data from 358 glioblastoma patients were obtained from The Cancer Imaging Archive. The dataset included clinical data and three magnetic resonance imaging (MRI) sequences. Three categories of features – clinical, radiomics, and deep features – were used for survival prediction. Radiomics features were extracted using PyRadiomics, while deep features were extracted from MRI images using a Swin UNet Transformers. An attention-based multimodal predictive model was designed to integrate these three feature sets. The DeepHit framework was used for survival modeling. Model performance was evaluated using the concordance index (C-index) and the Brier score. In addition, accuracy, precision, recall, area under the curve (AUC), and F1-score were calculated to predict mortality risk at different time intervals.

Results: The proposed model achieved a C-index of 0.91 and a Brier score of 0.14. For identifying mortality risk within 6 months, the model achieved an accuracy of 0.93, precision of 0.87, recall of 0.92, AUC of 0.96, and an F1-score of 0.89. For mortality risk within 12 months, the corresponding values were 0.91, 0.87, 0.85, 0.92, and 0.88, respectively. For mortality risk within 24 months, the model achieved an accuracy of 0.75, precision of 0.71, recall of 0.70, AUC of 0.94, and an F1-score of 0.80.

Conclusions: The proposed model, leveraging a combined radiomics-based approach, demonstrated the capability to predict overall survival in patients with glioblastoma. However, to ensure generalizability, further validation using multicenter datasets is required.

Key words: glioblastoma, survival prediction, radiomics, deep learning, Swin UNETR, DeepHit.

Introduction

Glioblastoma is one of the most common and aggressive primary tumors of the central nervous system [1]. Overall survival among patients with this disease is highly

variable [2]. Overall survival, defined as the time from diagnosis to death or last follow-up, is a key indicator for treatment selection and evaluation of therapeutic outcomes [3]. The overall survival of patients with glioblastoma is influenced by multiple factors, including tumor morphological

Correspondence address:

Dr. Hassan Emami and Dr. Reza Rabiei, Department of Health Information Technology and Management, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran, e-mails: haemami@sbmu.ac.ir; R.Rabiei@sbmu.ac.ir

Authors' contribution:

A Study design · B Data collection · C Statistical analysis · D Data interpretation · E Manuscript preparation · F Literature search · G Funds collection

and biological characteristics, patient performance status, treatment response, and genetic factors [4]. A subset of these characteristics can be extracted and analyzed using advanced imaging techniques such as magnetic resonance imaging (MRI) and computed tomography [5].

Various approaches have been proposed to extract relevant features from medical images. In the simplest approach, experts manually derive parameters such as maximum tumor cross-sectional area, the T2-weighted (T2) and fluid-attenuated inversion recovery (FLAIR) mismatch sign, or apparent diffusion coefficient ratios from imaging data. Although this approach is reliable, it is limited to a small set of imaging features and is highly dependent on expert interpretation [6,7].

Another widely used approach is radiomics. In this method, the region of interest is first segmented, after which a predefined set of quantitative statistical, textural, and geometric features is extracted. These include intensity-based features such as mean, variance, and skewness; textural features such as the gray-level co-occurrence matrix (GLCM), gray-level dependence matrix (GLDM), and gray-level size zone matrix (GLSZM); and shape-based features [8,9]. Despite its advantages, radiomics features are characterized by a relatively low level of abstraction and are susceptible to inter-center MRI heterogeneity, overfitting, and reduced generalizability [10].

An alternative approach for feature extraction from images is deep learning. Deep learning models, particularly convolutional neural networks (CNNs), are capable of automatically learning meaningful representations from imaging data without the need for predefined features, and can capture complex and nonlinear relationships within the data [11–13]. In this paradigm, raw images are provided as input to the network and are processed through successive convolutional, activation, and pooling layers. At each convolutional layer, filters are automatically learned to identify increasingly complex patterns within the images. The features extracted from intermediate or final layers are high-dimensional vectors that represent hierarchical and abstract image characteristics.

In addition to medical images, valuable information for predicting disease outcomes is embedded in histopathological features and molecular biomarkers. However, integrating these high-dimensional and heterogeneous data into conventional predictive models poses significant challenges. Traditional statistical models, such as the Cox proportional hazards model, are limited in their ability to capture complex nonlinear relationships among variables, and their predictive performance may deteriorate as the volume and diversity of molecular and imaging data increase [10]. To overcome these limitations, machine learning – particularly deep learning – approaches have been increasingly adopted. By learning multiple levels of abstraction and capturing both local and global patterns, these models have demonstrated improved performance in survival prediction. Nevertheless, deep learning

models also suffer from certain limitations, including difficulty in modeling long-range dependencies among features and challenges in simultaneously integrating multimodal data sources such as imaging, clinical information, and pathological data [14,15].

In this context, attention-based multimodal models have shown potential for integrating heterogeneous data and identifying complex interactions among different feature modalities [10]. This capability has positioned multimodal approaches as promising alternatives to conventional statistical and single-modality deep learning models in survival prediction tasks.

Therefore, in the present study, an attention-based multimodal model was employed to improve the accuracy and generalizability of overall survival prediction in patients with glioblastoma. In this framework, radiomics and deep features extracted from multiple MRI sequences were jointly processed and integrated with clinical patient information to generate a comprehensive survival prediction model.

Material and methods

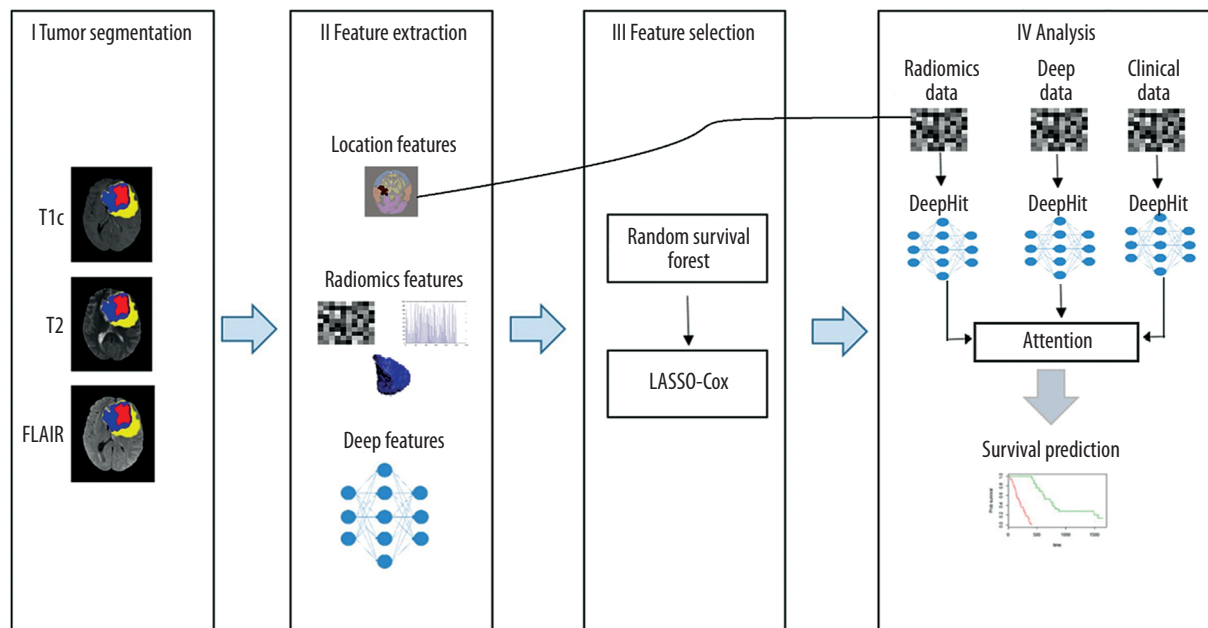
Study design

This study is applied developmental research with a retrospective design, conducted using data obtained from The Cancer Imaging Archive (TCIA) and performed in accordance with the principles of the Declaration of Helsinki. The primary objective of this study was to develop a multimodal deep learning model for predicting survival in patients with glioblastoma by integrating clinical data, radiomics features, and deep features extracted from MRI images (Figure 1).

Data

In this study, data from 358 patients diagnosed with glioblastoma were obtained from TCIA. The dataset comprised clinical information provided in CSV format, MRI scans including three sequences (T2, FLAIR, and contrast-enhanced T1-weighted images [T1c]), as well as tumor segmentation mask files in neuroimaging informatics technology initiative format. All acquired data were made available in a standardized and harmonized manner.

During the preprocessing of clinical data, missing values in categorical variables were imputed using the mode, while missing values in continuous variables were replaced with the mean. Variables with no variability (zero variance) were removed from the dataset. Subsequently, categorical variables were transformed into numerical representations using one-hot encoding. To ensure balanced feature contributions during model training and to prevent bias caused by differences in measurement scales, the RobustScaler method was applied.



FLAIR – fluid-attenuated inversion recovery, LASSO-Cox – least absolute shrinkage and selection operator Cox proportional hazards model, T1c – contrast-enhanced T1-weighted imaging, T2 – T2-weighted imaging

Figure 1. Architecture of the proposed model for glioblastoma overall survival prediction

For image preprocessing, all imaging data, including MRI sequences and tumor masks, were resampled to a standardized size of $96 \times 128 \times 128$ with uniform spatial resolution using the Medical Open Network for AI library. To normalize signal intensities, histogram rescaling was first applied to constrain voxel intensity values to the range (0, 1). This was followed by statistical normalization within the tumor region to reduce inter-scan variability. Subsequently, three distinct categories of features were extracted from these images:

Atlas-based feature extraction

To extract spatial tumor features, the automated anatomical labeling (AAL) atlas was employed. Using the symmetric normalization image registration algorithm, the AAL atlas was registered to the T1c image space of each patient. The overlap between the tumor mask and atlas regions was then analyzed to determine the percentage of involvement of different brain lobes and to identify the dominant lobe. In addition, the tumor center of mass was computed, and its Euclidean distance from the geometric center of the image was measured. These procedures resulted in the extraction of features such as tumor volume, dominant lobe, and distance from the brain center.

Radiomics feature extraction

Radiomics features were extracted from the T1c, FLAIR, and T2 sequences and the corresponding tumor regions using the PyRadiomics library. After resampling the images to a uniform resolution and generating binary tumor masks, a diverse set of features was extracted, including

shape-based features, first-order statistical features, and texture features such as GLCM, gray-level run-length matrix (GLRLM), GLSZM, GLDM, and neighboring gray-tone difference matrix (NGTDM).

Deep learning-based feature extraction

Deep features were extracted using a three-dimensional Swin UNet Transformers (Swin UNETR) model with pre-trained weights. By applying the segmentation masks to the MRI images, the model was constrained to focus on the tumor region. During training, various data augmentation transformations were applied to enhance model robustness. Features were extracted from the network and subsequently transformed into fixed-length compact vectors using three-dimensional adaptive average pooling.

Ground truth data

The overall survival time for each patient, as provided in the TCIA dataset, was used as the ground truth. These data served as the reference for deep feature extraction and for training the predictive models.

Data partitioning

To evaluate model performance, five-fold cross-validation ($K = 5$) was employed. In this framework, the dataset was divided into five independent and equally sized subsets (folds). The evaluation procedure was conducted over five separate iterations; in each iteration, four folds were used as the training set, while the remaining fold served as

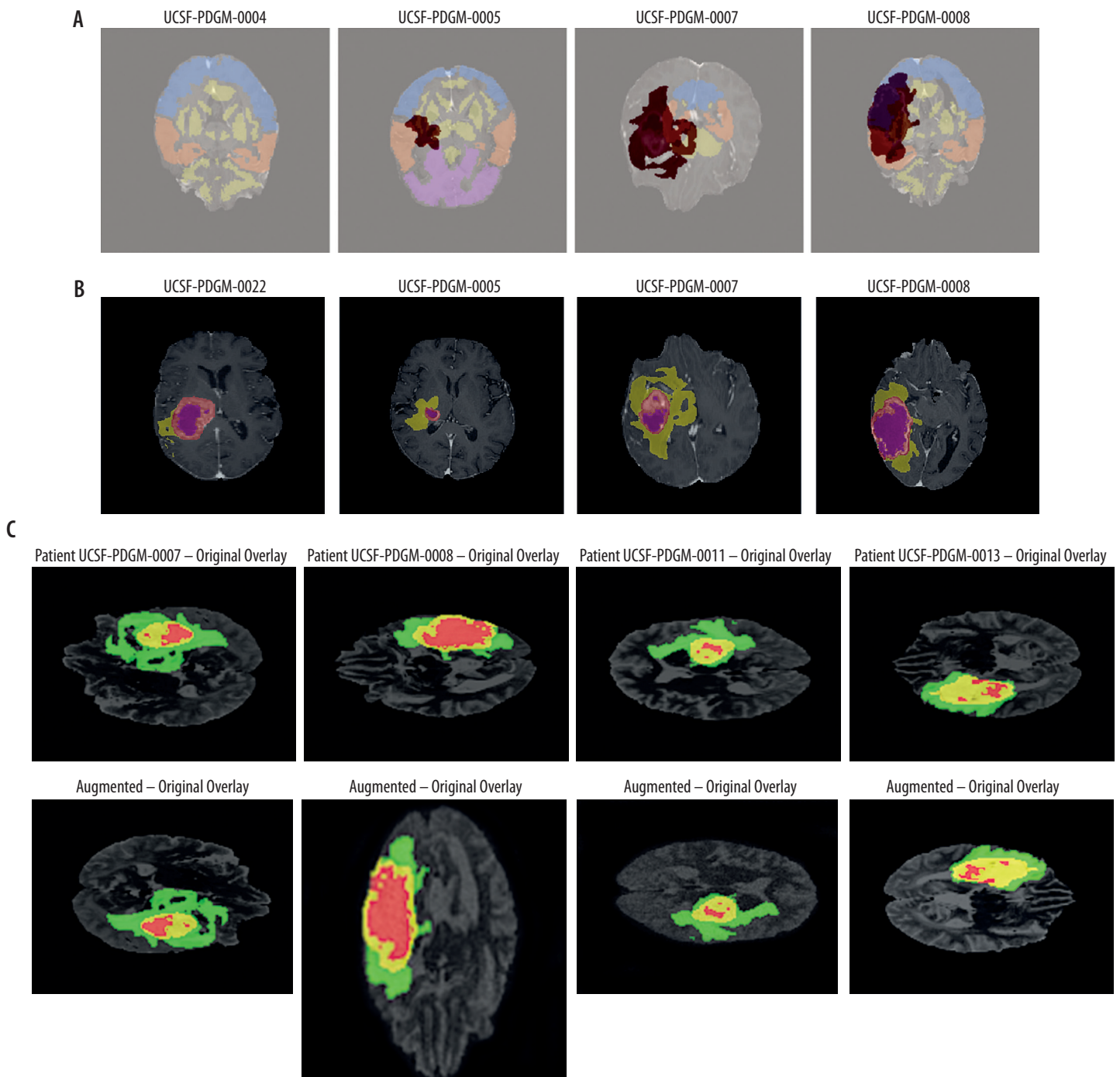


Figure 2. An example of the feature extraction pipeline: A) anatomical region localization of the tumor; B) tumor delineation and its corresponding labels; C) image augmentation for training the deep feature extraction model

the validation set. The sequence of iterations was designed such that each of the five folds was used exactly once as the validation set. This strategy ensured that all samples in the dataset contributed to both the training and evaluation phases.

Partitioning was performed at the sample level, ensuring that no individual sample appeared simultaneously in both the training and validation sets. This approach prevented overlap between the training and validation samples.

Modeling

The model was trained using data from 358 patients diagnosed with glioblastoma. The overall architecture consisted of three main branches: a deep feature branch, a radiomics branch, and a clinical data branch. Each branch incorporated a DeepHit network, employing batch normalization, a rectified linear unit activation function, and a dropout layer. The outputs of the three branches were subsequently fused and passed to an attention layer, which

dynamically learned the relative importance of each modality. Details of the model hyperparameter settings are provided in Table 1.

Training

Model training was performed using five-fold cross-validation ($K = 5$) in conjunction with a learning rate scheduling strategy. To facilitate survival time modeling, continuous survival times were discretized into discrete time intervals using the Lab-Trans-Discrete-Time method. Details of the model hyperparameter configurations are presented in Table 1.

Evaluation

Model performance was assessed using the concordance index (C-index) and the Brier score. In addition, the system was evaluated for predicting mortality risk at 6-, 12-, and 24-month time horizons. For these time-based predictions, performance was quantified using accuracy, precision, recall, area under the receiver operating characteristic curve (AUC), and the F1-score.

To further analyze model behavior and interpretability, calibration curves, ROC curves, precision–recall curves, and confusion matrices were generated.

Results

Data characteristics

The study population consisted of 358 patients diagnosed with glioblastoma whose data were recorded in The Cancer Imaging Archive. The dataset included clinical information and three types of pre-treatment MRI scans for each patient, namely T1c, T2, and FLAIR sequences. All imaging studies were acquired using a 3-T MRI scanner with a dedicated eight-channel head coil. Raw imaging data were preprocessed and spatially aligned using the FM-RIB software library, followed by skull stripping with the brain_mask algorithm. Tumor regions were initially segmented automatically and subsequently reviewed and refined by an experienced radiologist. Three primary tumor subregions were considered in the segmentation process: enhancing tumor, necrotic region, and abnormal (non-enhancing) region. As the data were obtained from a preprocessed public repository, no missing values were present in the dataset. The minimum, maximum, mean, and standard deviation of patient age were 21, 94, 61.77, and 12.02 years, respectively. Approximately 59% of the patients were male, with the remaining patients being female.

Anatomical features, including tumor volume (measured in number of voxels), the dominant brain lobe involved by the tumor, and the Euclidean distance between the geometric center of the tumor and the center of the brain, were extracted from MRI images and added to

Table 1. Model hyperparameter settings

Component	Parameter	Value/setting
Network architecture	Branches	3 (deep, radiomics, clinical)
	Hidden layer sizes	(256, 32, 16)
	Activation function	ReLU
	Normalization	Batch normalization
	Dropout rate	0.2
Attention module	Type	Sequential attention mechanism
	Output range	(0, 1)
Output layer	Survival model	DeepHit
	Predicted intervals	12, 24, 60 months
Training configuration	Optimizer	Adam
	Initial learning rate	5e-5
	LR scheduler	Reduce-LR-On-Plateau (factor = 0.1, patience = 10)
	Early stopping	Patience = 25
	Max epochs	350
	Batch size	32
	Cross-validation	5-fold

LR – learning rate, ReLU – rectified linear unit

the set of clinical features. The mean $\pm$ standard deviation of tumor volume and distance from the brain center were $90,693 \pm 59,979$ voxels and 45.28 ± 10.28 mm, respectively. The parietal lobe was reported as the most frequently involved dominant lobe among patients. The final number of clinical features used for model training and evaluation was 25. A complete list of these features is provided in the Appendix.

Across the three MRI sequences and multiple tumor regions, including the whole tumor, edema, and necrosis, a total of more than 3,072 quantitative features were initially extracted. These features comprised shape-based, first-order statistical, and texture features derived from GLCM, GLRLM, GLSZM, GLDM, and NGTDM matrices. Following the feature selection process, 35 radiomics features were ultimately retained in accordance with the list provided in the Appendix. This subset of features was used for training and evaluating both the baseline models and the glioblastoma-specific model.

For deep feature extraction, the Swin UNETR model achieved convergence after 25 training epochs with fine-tuning applied to the upper layers of the encoder, reaching a concordance index of 0.72. Figure 3 illustrates the trend of the concordance index during the training process of the feature extraction model. After training, a total of 864 deep features were extracted and stored for each patient. Through feature selection, 47 final deep features

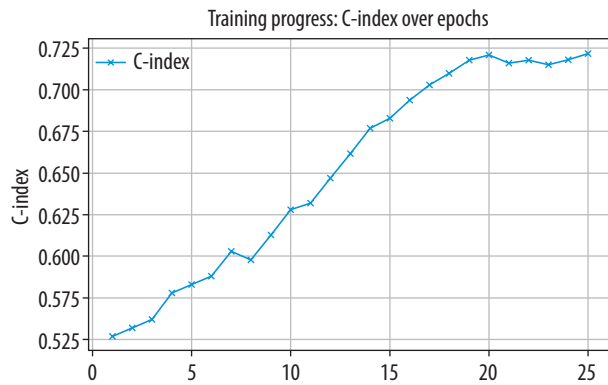


Figure 3. Trend of the concordance index during fine-tuning of the deep feature extraction model

C-index – concordance index

Table 2. Mean and standard deviation (SD) of the concordance index (C-index) and Brier score across 5 folds for the baseline and glioblastoma-specific models

Metric	Mean	SD
C-index	0.91	0.09
Brier score	0.14	0.04

were selected based on the list provided in the Appendix, all of which were used in the training and evaluation of both the baseline models and the glioblastoma model.

The mean and standard deviation of the concordance index and Brier score across the five validation folds, separated by the baseline model and the glioblastoma-specific model, are presented in Table 2.

To apply this model for survival prediction, three time horizons – under 6 months, under 12 months, and under 24 months – were designated as very high risk, high risk, and medium risk, respectively. Table 3 reports the mean evaluation metrics, including accuracy, precision, AUC, recall, and F1-score, for each risk group across the five validation folds. Figure 4 presents the confusion matrices of the model, stratified by risk group.

Discussion

In this study, an attention-based multimodal deep learning model was designed and evaluated for predicting overall survival in patients with glioblastoma. The primary aim was to improve prediction accuracy by inte-

grating clinical and imaging data, as recent studies have demonstrated that combining multi-source data can provide a deeper understanding of tumor characteristics and patient prognosis [16].

The dataset was divided into three feature categories – clinical, radiomics, and deep features – each contributing to the improved performance of survival prediction models. In the clinical data branch, numerical features were standardized using the RobustScaler method to mitigate the influence of outliers. Compared to Standard-Scaler, which is more sensitive to extreme values, RobustScaler has demonstrated more stable performance for clinical datasets [17].

Anatomical features, including tumor volume, distance from the brain center, and relative tumor location within brain lobes, were extracted using the AAL atlas. This approach aligns with the findings of Isensee *et al.* [18], who reported a negative correlation between tumor volume and patient survival (Pearson correlation = -0.42). Similarly, Bakas *et al.* [19] demonstrated that the tumor's distance from critical brain regions can significantly influence survival prognosis.

For radiomics feature extraction, tumor regions were segmented into three labels: whole tumor, edema, and necrosis. This segmentation enabled a more detailed analysis of tissue patterns, consistent with the approach reported by Pereira *et al.* [20], who highlighted its effectiveness for improving survival prediction. However, the present study also incorporated additional advanced features, such as tissue entropy and wavelet-based features, resulting in greater comprehensiveness and predictive accuracy. Strengths of this approach include the large number of extracted features and the use of advanced normalization techniques.

Over 3,072 radiomics features were extracted from tumor regions across different MRI sequences. To enhance feature accuracy, signal intensity normalization was performed, and geometric consistency between images and tumor masks was assessed using image registration algorithms. This procedure was intended to reduce errors caused by spatial misalignment and, compared to the study by Aerts *et al.* [8], which paid less attention to normalization, may have improved the quality of input data. GLSZM features, such as Large-Area-Emphasis and Zone-Variance, frequently appeared in the feature selection process, consistent with Aerts *et al.* [8], who reported a correlation coefficient of 0.58 between these features and patient survival.

Table 3. Mean evaluation metrics for risk prediction across 5 folds, stratified by risk group, for the baseline and glioblastoma-specific models

Risk group	AUC	Accuracy	Precision	Recall	F1-score
Very high (< 6 months)	0.93	0.87	0.91	0.96	0.89
High (< 12 months)	0.91	0.87	0.85	0.92	0.88
Medium (< 24 months)	0.75	0.71	0.70	0.94	0.80

AUC – area under the curve

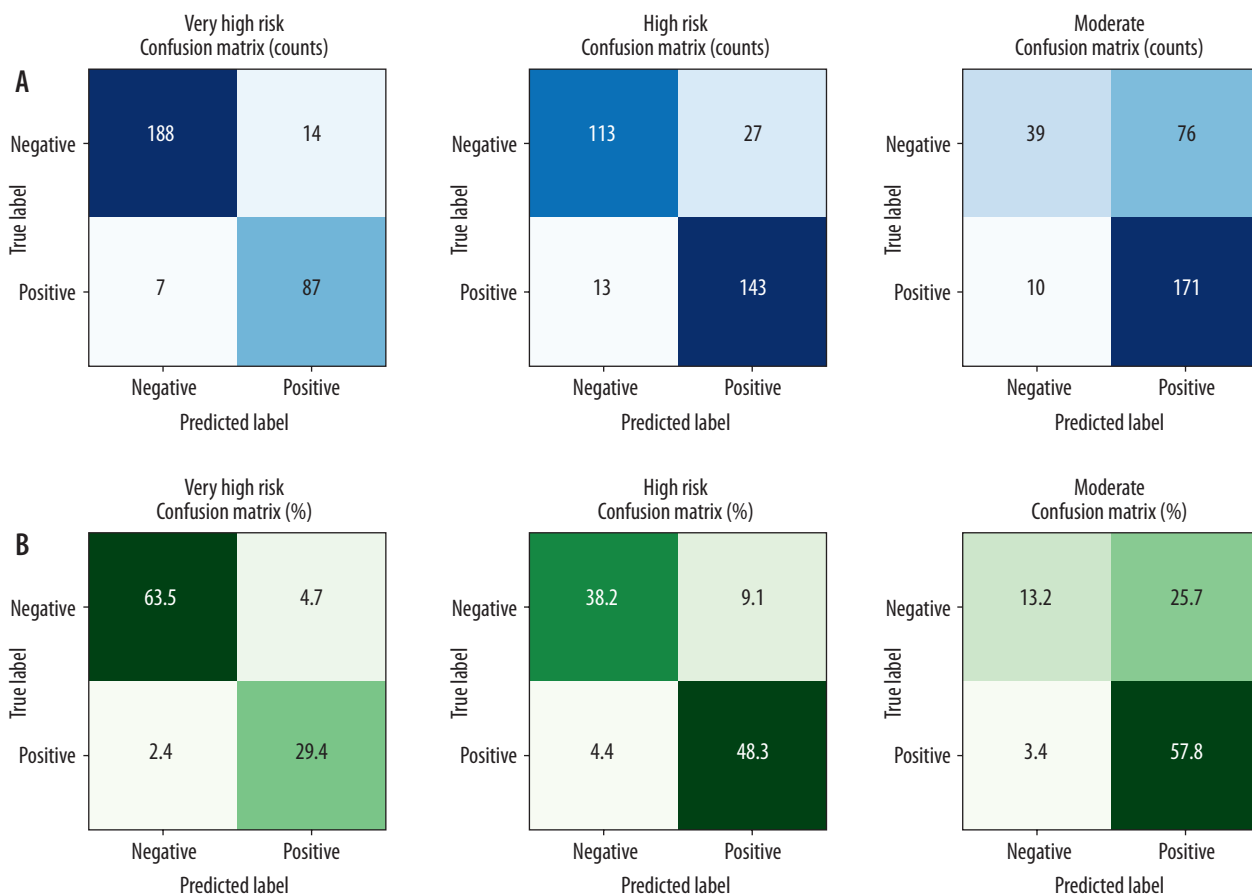


Figure 4. Confusion matrix of the glioblastoma-specific model: A) raw counts, B) percentages

Deep feature extraction in this study was performed using the Swin UNETR model, applied for the first time to glioblastoma survival prediction. Previously, Hatamizadeh *et al.* [21] used this model for brain tumor segmentation in MRI images, achieving an accuracy of 0.87 and a concordance index of 0.72 after fine-tuning the upper layers. Their study demonstrated that Swin UNETR significantly outperformed traditional models and is highly effective in capturing complex and hierarchical patterns from medical images.

In this study, deep feature extraction with Swin UNETR included data augmentations such as rotation, scaling, and Gaussian noise, which were intended to improve the model's generalizability. A total of 864 deep features were extracted from the images. The increase in the concordance index from 0.55 to 0.72 over 25 epochs indicated model convergence. The model's performance in this study was higher than that reported by Yan *et al.* [22], who used simpler CNNs and achieved a concordance index of 0.65. This improvement may be related to the advanced Swin UNETR architecture and more precise optimization procedures.

Feature selection was performed using a random survival forest, ultimately retaining 35 radiomics features and 47 deep features. This approach can be useful in survival analysis due to its ability to handle censored data and prioritize informative features. Compared to the study

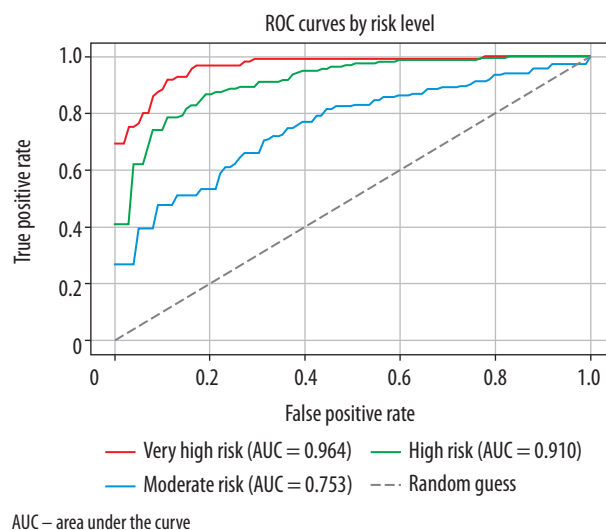


Figure 5. Performance receiver operating characteristic (ROC) curves of the model

by Lao *et al.* [23], which relied on simpler methods such as variance-based filters and the least absolute shrinkage and selection operator Cox proportional hazards model (LASSO-Cox) for key feature selection, the present study achieved superior performance. Similarly, Gillies *et al.* [24] used LASSO-Cox to identify survival-related features, highlighting its potential impact on model accuracy.

In this study, the model was designed as an attention-based multimodal deep learning framework to jointly analyze clinical, radiomics, and deep imaging features. This approach, building on the attention mechanism described by Vaswani *et al.* [25], allows the model to focus on key data features, thereby potentially improving predictive accuracy. Attention-based multimodal deep learning models have been widely used across medical and non-medical domains due to their capability to integrate multiple data sources or models.

For example, Mienye *et al.* [26] employed attention-based multimodal models to predict cardiovascular disease risk, achieving a concordance index of 0.88 by combining clinical and imaging data, whereas single-modality models such as Cox regression achieved only 0.75. In oncology, Lynch *et al.* [27] used a combined radiomics model for predicting survival in lung cancer patients, reporting an AUC of 0.93 when integrating genomic and clinical data, compared to 0.82 for single-modality support vector models. Additionally, Li *et al.* [28] applied a deep learning-based multimodal radiomics model for predicting breast cancer recurrence, achieving a concordance index of 0.92, while single-modality logistic regression models reached only 0.80.

The primary advantage of attention-based multimodal deep learning models lies in their ability to integrate information from multiple data sources, which may improve predictive performance.

DeepHit is a deep learning approach for modeling competing risks, particularly designed for discrete-time survival data [29]. Unlike traditional methods such as the Cox model, DeepHit directly models the survival probability distribution using deep neural networks and does not require any assumption about the underlying data distribution. This enables DeepHit to capture complex and nonlinear relationships among variables. Li *et al.* [29], who introduced DeepHit, demonstrated that it outperformed the Cox model and random survival forests in predicting survival among patients with cardiovascular diseases, achieving a concordance index of 0.85 compared to 0.78 and 0.81, respectively. Similarly, Katzman *et al.* [30] applied DeepHit to liver cancer survival prediction and reported a concordance index of 0.89, outperforming traditional Cox and parametric models, which achieved 0.80 and 0.83, respectively.

A key advantage of DeepHit is its ability to model competing risks and predict survival probabilities over specific time intervals, which is highly valuable in medical scenarios where multiple events may occur. Moreover, DeepHit can effectively handle multimodal data and provides greater flexibility for modeling heterogeneous datasets compared to traditional methods. However, it requires a relatively large sample size and careful hyperparameter tuning.

In the present study, the model achieved a concordance index of approximately 0.91. This level of performance is higher than the range reported for some previously

published models, which typically fall between 0.65 and 0.87 [16], indicating the competitive performance of the current model. Notably, the lower Brier score in the glioblastoma-specific model reflects more accurate survival probability estimates and greater prediction stability, aspects often overlooked in previous models.

Recent reviews indicate that most survival prediction models have used architectures such as CNNs, 3D U-Net, or hybrid radiomics–deep learning models, with performance varying depending on the data type and modality combination. These models generally perform better on homogeneous datasets, with average concordance indices ranging from 0.65 to 0.87 [16]. In contrast, the current model, by integrating attention mechanisms and combining imaging with clinical data, achieved a C-index above this reported range. This may reflect the value of multimodal integration, although the contribution of the attention mechanism requires separate evaluation.

Compared to classical radiomics models, which are typically based on algorithms such as support vector machine, random forest, or Cox and generally report concordance indices between 0.6 and 0.7 [16], the proposed model achieved a higher C-index in this study. This difference may be related to the deep network's ability to automatically learn complex imaging and clinical features, whereas classical models rely on manual feature extraction and often suffer performance degradation when faced with tumor heterogeneity and data variability. Furthermore, the integration of multimodal data and the use of attention layers in the current model may improve its flexibility in distinguishing risk groups and predicting survival intervals, while many review-reported models lack this level of data integration.

Overall, the findings of this study are consistent with existing reviews [16] and suggest that attention-based multimodal architectures may be useful for improving survival prediction performance. The higher C-index observed in this glioblastoma cohort supports further investigation of tumor-specific models, but does not establish that tumor-grade-specific models are generally more efficient than general models. Thus, the model developed in this study shows promising predictive performance, but its accuracy, stability, and interpretability require further evaluation in independent datasets.

The results also revealed that model validation metrics tend to decrease over longer time horizons. One possible factor responsible for this decline is data imbalance, particularly the low number of long-term survivors, which limits the model's ability to learn patterns for these groups and reduces the accuracy of long-term predictions. Additionally, extending the prediction horizon introduces greater complexity and inherent uncertainty, as environmental factors, treatments, and biological changes can influence patient survival and are not fully captured by the model. For instance, accuracy and AUC noticeably declined for low-risk patients in long-term predictions,

whereas short-term predictions for high-risk patients remained precise. This observation may be consistent with previous studies, such as Aerts *et al.* [8], which reported that heterogeneity in long-term data and uneven patient distributions reduce model accuracy. Therefore, although the proposed models perform well in short-term predictions in this dataset, caution is advised when applying them to long-term survival forecasts.

Conclusions

The findings of this study demonstrate the suitability of developing an overall survival prediction model for glioblastoma patients using a combined radiomics approach. By leveraging a multimodal strategy, the study was able to construct a predictive model. A concordance index of 0.91 and an AUC above 0.90 for high-risk groups indicate promising predictive performance. This performance may be related to the use of techniques such as Swin UNETR for deep feature extraction, random survival forests for feature selection, and the integration of multimodal data.

The model has the potential to serve as a complementary tool alongside other diagnostic and therapeutic ap-

proaches, assisting clinicians in planning personalized treatment strategies. However, before clinical implementation, further validation using more diverse datasets across different centers and settings is necessary.

Limitations

This model was developed and evaluated using data from TCIA. For practical clinical application, it requires further training and evaluation using larger and more diverse datasets. Additionally, before deployment in real-world settings, the generalizability of the model should be assessed on datasets from multiple centers to evaluate its performance across different populations and imaging protocols.

Disclosures

1. Institutional review board statement: Not applicable.
2. Assistance with the article: None.
3. Financial support and sponsorship: This study was financially supported by Shahid Beheshti University of Medical Sciences.
4. Conflicts of interest: None.

References

1. Luo C, Yang J, Liu Z, Jing D. Predicting the recurrence and overall survival of patients with glioma based on histopathological images using deep learning. *Front Neurol* 2023; 14: 1100933. DOI: 10.3389/fneur.2023.1100933.
2. Kesler SR, Harrison RA, Rao V, Dyson H, Petersen M, Prinsloo S. Predicting overall survival in diffuse glioma from the presurgical connectome. *Sci Rep* 2022; 12: 18783. DOI: 10.1038/s41598-022-22387-7.
3. Lobato-Delgado B, Priego-Torres B, Sanchez-Morillo D. Combining molecular, imaging, and clinical data analysis for predicting cancer prognosis. *Cancers (Basel)* 2022; 14: 3215. DOI: 10.3390/cancers14133215.
4. Whitfield BT, Huse JT. Classification of adult-type diffuse gliomas: impact of the World Health Organization 2021 update. *Brain Pathol* 2022; 32: e13062. DOI: 10.1111/bpa.13062.
5. Agravat RR, Raval MS. A survey and analysis on automated glioma brain tumor segmentation and overall patient survival prediction. *Arch Comput Methods Eng* 2021; 28: 4117-4152. DOI: 10.1007/s11831-021-09559-w.
6. Wen PY, van den Bent M, Youssef G, Cloughesy TF, Ellingson BM, Weller M, et al. RANO 2.0: update to the response assessment in neuro-oncology criteria for high- and low-grade gliomas in adults. *J Clin Oncol* 2023; 41: 5187-5199.
7. Patel SH, Poisson LM, Brat DJ, Zhou Y, Cooper L, Snuderl M, et al. T2-FLAIR mismatch, an imaging biomarker for IDH and 1p/19q status in lower-grade gliomas: a TCGA/TCIA project. *Clin Cancer Res* 2017; 23: 6078-6085.
8. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; 5: 4006. DOI: 10.1038/ncomms5006.
9. Chato L, Latifi S. Machine learning and radiomic features to predict overall survival time for glioblastoma patients. *J Pers Med* 2021; 11: 1336. DOI: 10.3390/jpm11121336.
10. Goma A, Huang Y, Hagag A, Schmitter C, Höfler D, Weissmann T, et al. Comprehensive multimodal deep learning survival prediction enabled by a transformer architecture: a multicenter study in glioblastoma. *Neurooncol Adv* 2024; 6: vdae122. DOI: 10.1093/noajnl/vdae122.
11. Truhn D, Schrading S, Haarbuerger C, Schneider H, Merhof D, Kuhl C. Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast MRI. *Radiology* 2019; 290: 290-297.
12. Vale-Silva LA, Rohr K. Long-term cancer survival prediction using multimodal deep learning. *Sci Rep* 2021; 11: 13505. DOI: 10.1038/s41598-021-92799-4.
13. Babaei Rikan S, Sorayaie Azar A, Naemi A, Bagherzadeh Mohasefi J, Pirnejad H, Wiil UK. Survival prediction of glioblastoma patients using modern deep learning and machine learning techniques. *Sci Rep* 2024; 14: 2371. DOI: 10.1038/s41598-024-53006-2.
14. Chen RJ, Lu MY, Weng WH, Chen TY, Williamson DF, Manz T. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. *Proceedings of the IEEE/CVF International Conference on Computer Vision* 2021; 3995-4005. DOI: 10.1109/ICCV48922.2021.00398.
15. He K, Gan C, Li Z, Rekik I, Yin Z, Ji W, et al. Transformers in medical image analysis. *Intell Med* 2023; 3: 59-78.

16. Mohammadpour S, Emami H, Rabiei R, Roshanpoor A, Tabatabaeifar M, Faeghi F. Scoping review of image-based overall survival prediction in glioma using machine learning. *Pol J Radiol* 2025; 90: e571-e581. DOI: 10.5114/pjr/209963.
17. Shortliffe EH, Cimino JJ (eds.). *Biomedical informatics: computer applications in health care and biomedicine*. 3rd ed. Springer; 2006.
18. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021; 18: 203-211.
19. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging* 2015; 34: 1993-2024.
20. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans Med Imaging* 2016; 35: 1240-1251.
21. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. *Brainlesion* 2022; 12962. DOI: 10.1007/978-3-031-08999-2_22.
22. Yan J, Zhang B, Zhang S, Cheng J, Liu X, Wang W, et al. Quantitative MRI-based radiomics for noninvasively predicting molecular subtypes and survival in glioma patients. *NPJ Precis Oncol* 2021; 5: 72. DOI: 10.1038/s41698-021-00205-z.
23. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multi-forme. *Sci Rep* 2017; 7: 10353. DOI: 10.1038/s41598-017-10649-8.
24. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278: 563-577.
25. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Advances in Neural Information Processing Systems* 2017; 30: 5998-6008.
26. Mienye ID, Sun Y, Wang Z. An improved ensemble learning approach for the prediction of heart disease risk. *Inform Med Unlocked* 2020; 20: 100402. DOI: 10.1016/j.imu.2020.100402.
27. Lynch CM, Abdollahi B, Fuqua JD, De Carlo AR, Bartholomai JA, Balgemann RN, et al. Prediction of lung cancer patient survival via supervised machine learning classification techniques. *Int J Med Inform* 2017; 108: 1-8.
28. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* 2018; 15: 81-94.
29. Lee C, Zame W, Yoon J, van der Schaar M. DeepHit: a deep learning approach to survival analysis with competing risks. *Proceedings of the AAAI Conference on Artificial Intelligence* 2018; 32. DOI: 10.1609/aaai.v32i1.11842.
30. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol* 2018; 18: 24. DOI: 10.1186/s12874-018-0482-1.