cancers

MDPI

Article

Attention-Based Dual-Branch Deep Learning for Overall Survival Prediction in Patients with Brain Metastases Using MRI: A Multicenter Retrospective Study

Ji Zhang [1,†], Jiaqian He [1,†], Mingrou Guo [1], Zhenhua Zhang [2], Yiran Mu [1], Bowen Shi [3], Jianping Wu [4], Xiuling Lv [5], Wenliang Yu [4], Congying Xie [1,*] and Xiance Jin [1,*]

[1] Department of Radiotherapy and Medical Oncology, 1st Affiliated Hospital of Wenzhou Medical University, Wenzhou 325015, China
[2] Department of Radiology, 1st Affiliated Hospital of Wenzhou Medical University, Wenzhou 325003, China
[3] School of Public Health, Wenzhou Medical University, Wenzhou 325035, China
[4] Department of Radiotherapy, The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou 324000, China
[5] Department of Radiology, Cixi People's Hospital, Wenzhou Medical University, Wenzhou 315300, China
* Correspondence: wzxiecongying@163.com (C.X.); jinxc1979@hotmail.com (X.J.); Tel.: +86-577-88069370 (X.J.)
† These authors contributed equally to this work.

Simple Summary

Prognostic parameters and models are believed to be helpful in improving the treatment outcome for patients with BM. This multicenter retrospective study was set up to develop and validate an attention-based deep learning model to predict overall survival in patients with brain metastases using pre-treatment T1-CE MRI. For BM patients with multiple lesions, the results of different radiomics-based feature aggregation methods are not consistent. Moreover, compared with GPA indices and radiomics models, the prediction of OS for patients with BM was improved by developing a deep learning network. We found that considering the multi-lesion situation of brain metastases is helpful in improving the prediction of overall survival for patients with brain metastases. Our results may help improve risk stratification for patients with brain metastases. Specifically, the model may assist clinicians in identifying high-risk patients who could benefit from closer follow-up, although this study did not evaluate whether its use changes treatment decisions or improves patient outcomes.

Abstract

Objectives: To develop and validate an attention-based deep learning (DL) model to predict overall survival (OS) in patients with brain metastases (BM) using pre-treatment magnetic resonance image (MRI). **Methods**: A total of 216 BM patients treated with radiotherapy or neurosurgical resection were retrospectively enrolled from two hospitals and divided into training ($n = 128$), internal validation ($n = 56$), and external validation ($n = 32$) cohorts. Three feature aggregation methods and seven radiomics models were developed to predict OS. A DL model named DeepBMSurv, incorporating a cross-attention mechanism, was built using pre-treatment MR images. Its prognostic performance was compared with radiomics models and graded prognostic assessment (GPA). A nomogram integrating GPA, radiomics score (Rad-score) and DL score (DL_RiskScore) was also developed. Model performance was assessed by the concordance index (C-index). **Results**: Among the radiomics models, the $\text{RadScore}_{\text{mean}}$ model showed the most stable discrimination across cohorts, with C-index values of 0.685, 0.672, and 0.671 in the training, internal validation, and external validation cohorts. DeepBMSurv achieved numerically higher C-indices than both the radiomics model and GPA model, with C-indices of 0.806 vs. 0.685 vs. 0.620 in the training cohort,

Check for updates

Academic Editors: Michał Strzelecki, Adam Piórkowski, Karolina Nurzynska and Rafał Obuchowicz

Received: 13 October 2026
Revised: 16 November 2026
Accepted: 16 November 2026
Published: 18 November 2026

Copyright: © 2026 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.

0.755 vs. 0.672 vs. 0.628 in the internal validation cohort, and 0.733 vs. 0.671 vs. 0.579 in the external validation cohort. The nomogram integrating DL_RiskScore, Rad-score, and GPA achieved C-index values of 0.809, 0.764 and 0.733 in training, internal validation, and external validation cohorts, respectively. **Conclusions**: A DL model based on pre-treatment MRI can effectively predict survival outcomes in patients with BM, offering a potential tool for individualized prognostic risk stratification.

Keywords: brain metastases; overall survival; radiomics; deep learning; attention

1. Introduction

Brain metastasis (BM) is the most common neurological complication in patients with advanced cancer [1]. Approximately 20% to 40% of patients with solid tumors will develop BM during the course of their disease [2]. BM has been a significant contributor to mortality among cancer patients [3]. Despite the improvement in multimodal treatment and systemic therapeutic approaches, including the combined application of surgery, radiotherapy, chemotherapy, immunotherapy and targeted therapy, etc., the prognosis of BM patients is still poor and highly variable with a median overall survival (OS) ranging from 4 to 16 months [4]. Multiple risk factors, such as the number of metastatic lesions, molecular subtypes, and treatment response, influence survival outcomes [5,6]. Clinically, there is an urgent need for accurate survival prediction to support treatment planning and enable stratified, rational management of BM patients.

Prognostic factors, such as recursive partitioning analysis (RPA) [7], Graded Prognostic Assessment (GPA) [8], Basic Score for Brain Metastases (BSBM) [9], Brain Metastasis Velocity (BMV) score [10], etc., have been used to guide treatment decision. However, the rapid advancements in treatment options require further validation of these decade old factors for today's clinical standards [11]. The emergence of radiomics offers a promising avenue for decoding tumor phenotype–prognosis relationships by extracting high-throughput features from medical images [12,13]. MRI is regarded as the standard imaging modality for BM due to its high sensitivity in detecting small lesions and accurately defining their anatomical context [14,15]. Previous studies have demonstrated that MRI-based radiomic features can non-invasively predict response to stereotactic radiosurgery (SRS) [16], and that MRI-derived BM zone percentage serves as an independent prognostic factor of local tumor control for non-small cell lung cancer (NSCLC) treated by gamma knife radiosurgery (GKRS) [17]. Combining MRI radiomic features with RNA sequencing data are able to improve survival predictions and personalized treatment strategies in NSCLC patients with BM [16–18]. Nevertheless, the standalone utility of MRI-based radiomics for robust survival prediction remains limited by inherent challenges related to imaging variability, data quality, and modeling assumptions [19].

Deep learning (DL) has shown considerable promise in predicting treatment outcomes and stratifying risk in BM patients using pre-treatment contrast-enhanced MRI and clinical data [20,21]. Attention-based DL architectures, in particular, can improve performance by dynamically focusing on context-aware regions and capturing the global context relationships crucial for recognizing complex patterns in medical images [22,23]. Moreover, because BM patients often present with multiple metastatic lesions, incorporating features from multiple sites may improve OS prediction [24]. Therefore, the purpose of this study is to develop and validate an attention-based DL model and its integration with radiomics and clinical factors in OS prediction for patients with multiple BMs using pre-treatment MR images.

2. Materials and Methods

2.1. Patients and Follow-Up

We retrospectively analyzed data from patients with BM treated at two independent hospitals. Inclusion criteria were: (1) receipt of brain radiotherapy or neurosurgery, and (2) availability of T1-contrast enhanced (T1-CE) MR scans within 2 weeks prior to brain-directed treatment. Exclusion criteria were as follows: (1) missing T1-CE MR data or poor image quality, and (2) incomplete clinical or follow-up data. Patients with BM from hospital one (treated between January 2020 to December 2024) and hospital two (treated between January 2020 to December 2023) were reviewed. Clinical data were collected from electronic medical records, including age, gender, Karnofsky performance score (KPS), primary tumor site, presence of extracranial metastasis, and the number of BM lesions. The GPA score was calculated for each patient based on these clinical factors [8]. Patients from hospital one were randomly divided into a training cohort (70%) and an internal validation cohort (30%). Patients from hospital two were assigned to the external validation cohort. The overall study workflow is illustrated in Figure 1.

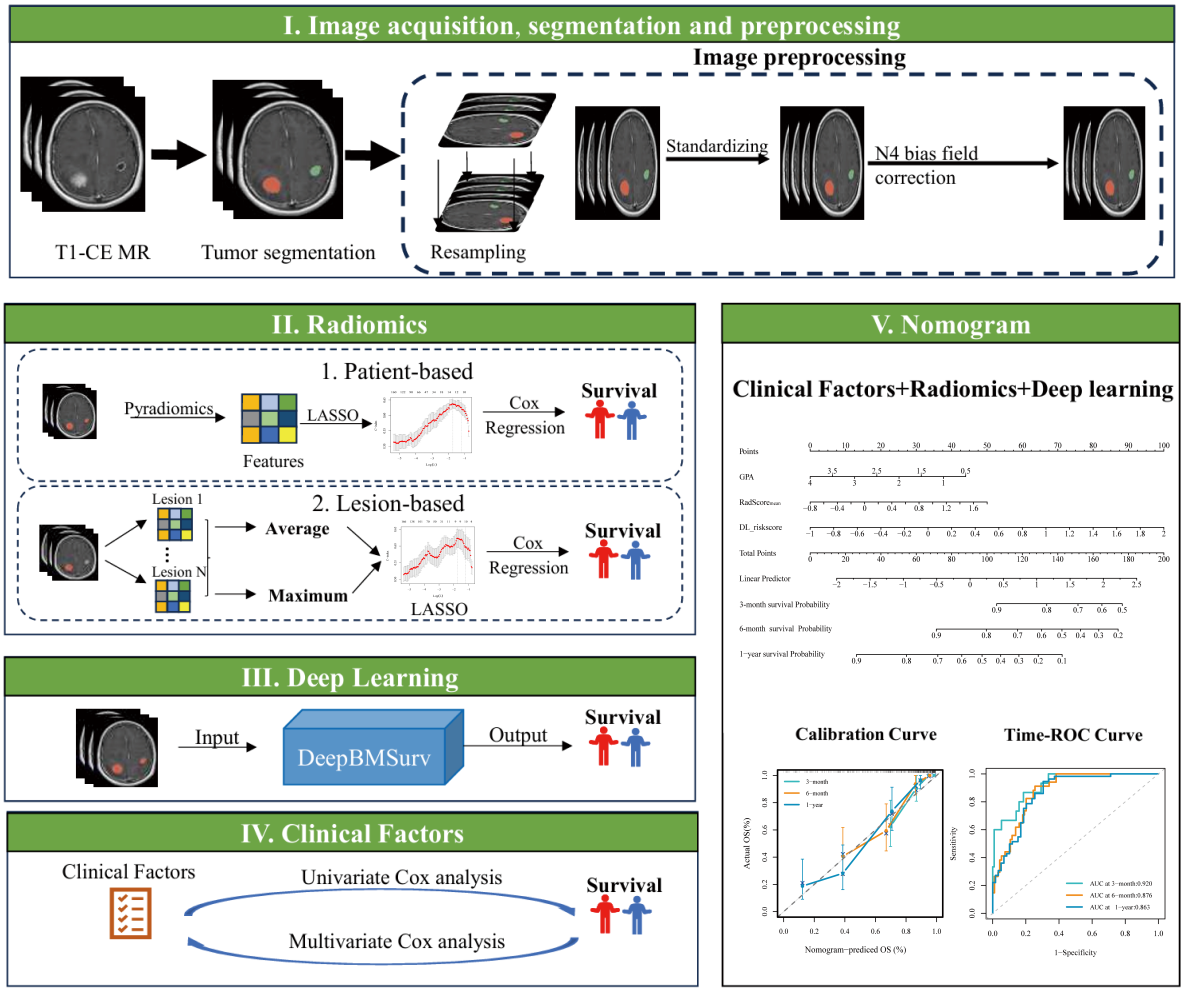


Figure 1. The overall workflow of this study. Workflow diagram for the development of the predictive models.

The primary endpoint was OS, defined as the time from initiation of brain treatment to death from any cause or the last follow-up. In this study, the follow-up closing date was 30 June 2026 and patients alive at that date were censored at their last contact. This study was approved by our institution's Clinical Research Ethics Committee (ECCR) (KY2025-

R151) and was conducted in accordance with the Declaration of Helsinki. The requirement for written informed consent was waived due to the retrospective nature of this study, with confirmation of patient data confidentiality.

2.2. Image Acquisition and Preprocessing

MRI images were retrieved from Picture Archiving and Communication Systems (PACS) in the original Digital Imaging and Communications in Medicine (DICOM) format. This study focused on T1-CE sequences. Detailed MRI acquisition parameters for each hospital are shown in Supplementary Table S1. BM lesions were manually segmented as regions of interest (ROIs) on MRI by a radiation oncologist with 5 years of experience via the 3D slicer (version 4.2.1, https://www.slicer.org). For patients with multiple lesions, each was delineated as a separate ROI. All segmentations were reviewed by a senior radiation oncologist with over 20 years of experience in neuroimaging.

All DICOM images were converted to Neuroimaging Informatics Technology Initiative (NIfTI) format for standardized processing. To address heterogeneity from multi-center protocols and hardware, a rigorous three-step preprocessing pipeline was implemented to all T1-CE sequences: (1) N4 bias field correction; (2) resampling to a uniform isotropic resolution of $1 \times 1 \times 1$ mm^3 using B-spline interpolation; (3) intensity normalization via Z-Score standardization [25,26].

2.3. Radiomic Features and Score

Radiomic features were extracted through two strategies: (1) a patient-pooled strategy, where features were derived from a single ROI encompassing all lesions per patient; (2) a lesion-level strategy, where features were extracted from each lesion independently. All features were extracted using PyRadiomics (version 3.0.1) with Image Biomarker Standardization Initiative (IBSI)-compliant definitions [27]. Extracted features included shape-based features ($n = 14$), first-order features ($n = 18$), texture features ($n = 73$), and wavelet features ($n = 728$). Among the texture features, there are gray level co-occurrence matrix (GLCM) features ($n = 22$), gray level run length matrix (GLRLM) features ($n = 16$), gray level size zone matrix (GLSZM) features ($n = 16$), gray level dependence matrix (GLDM) features ($n = 14$) and neighboring gray tone difference matrix (NGTDM) features ($n = 5$). The specific features extracted and the parameter settings for feature extraction using PyRadiomics are shown in Supplementary Table S2.

Features were standardized using z-score normalization in training cohort and the same fixed training-derived parameters from the training cohort were then applied to normalize the feature matrices of both the internal and external validation cohorts. Features were screened using a two-step process. First, Cox univariate regression ($p < 0.1$) identified candidate features. These were subsequently refined using multivariate least absolute shrinkage and selection operator (LASSO) Cox regression [28]. Three candidate radiomic signatures were constructed based on aggregation method. For $\text{RadScore}_{\text{mean}}$, features were averaged across all lesions at the feature level to yield a single patient-level feature vector, and LASSO Cox regression was then applied to this matrix in the training cohort. $\text{RadScore}_{\text{dominant}}$ used only the largest lesion, and $\text{RadScore}_{\text{pooled}}$ used a single ROI encompassing all lesions. Fusion models were not simple combinations of the constituent RadScore. The candidate features selected for the constituent signatures were pooled into a combined feature set, and LASSO Cox regression was re-run on this pooled set in the training cohort, yielding a single fused RadScore per patient. The final fused score was a linear combination of the retained features weighted by their re-estimated LASSO coefficients.

2.4. DL Model

We developed DeepBMSurv, a novel dual-branch deep learning architecture integrating a convolutional neural network (CNN) and a Vision Transformer (ViT) via a cross-attention mechanism (Figure 2) The first branch employs a ResNet-34 CNN backbone with Pyramid Squeeze Attention (PSA) modules in its final two residual blocks to extract multi-scale local features. The second branch uses a standard ViT to model global, long-range dependencies. A cross-attention mechanism enables multi-level feature interaction, allowing the ViT to query and integrate salient local features from the CNN. The fused feature vector passes through a global average pooling layer and a fully connected layer, outputting a prognostic risk score termed DL_RiskScore.

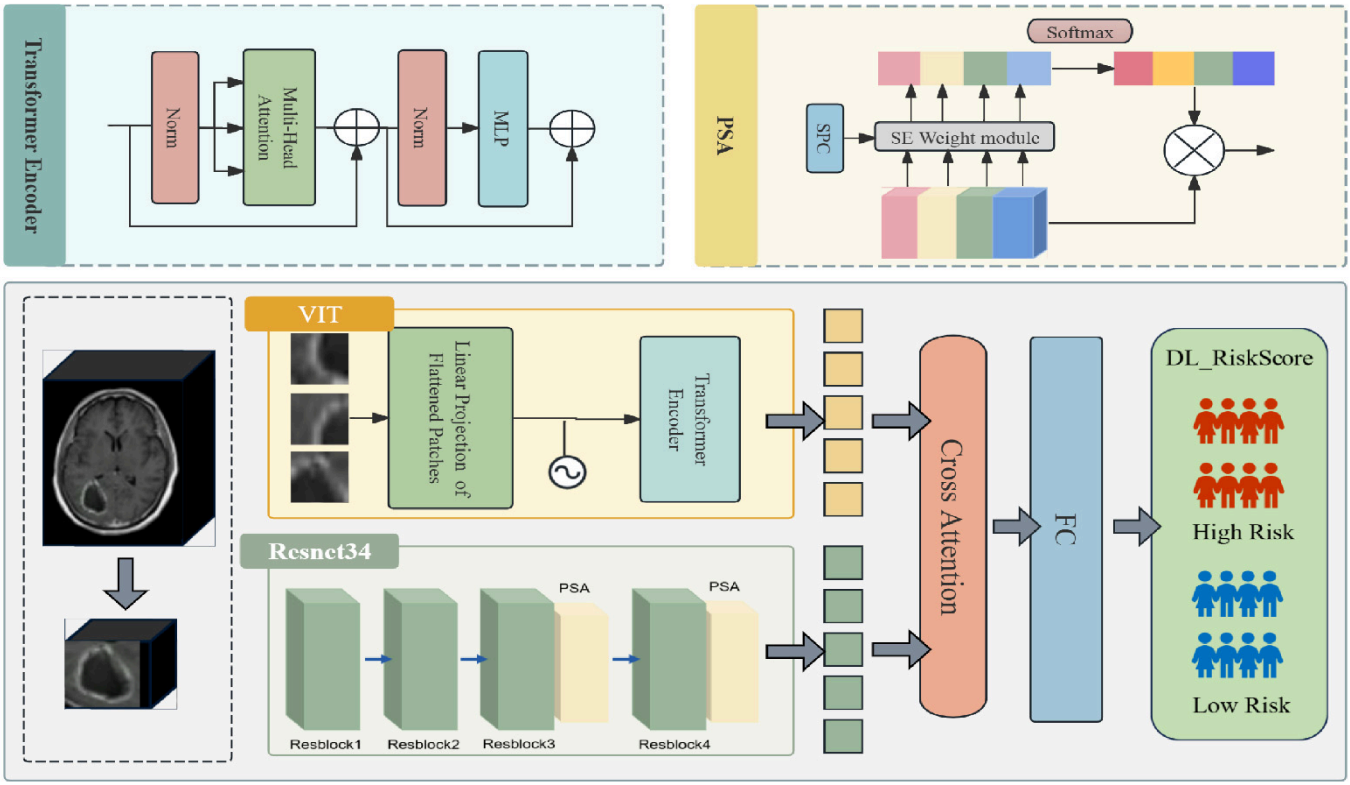


Figure 2. The schematic diagram of the DeepBMSurv architecture; the architecture of the PSA module is shown at the top left.

DeepBMSurv was implemented using PyTorch 2.1.2 on a hardware platform with a 12-core Intel Xeon Silver 4310 CPU and an NVIDIA GeForce RTX 3090 GPU. Preprocessed 3D MRI images were cropped to 124 × 124 × 124 voxels. The crop was centered on the centroid of the union of all manually delineated lesion masks for each patient, so that the patch was positioned to encompass all segmented lesions rather than a single lesion. Because all volumes were resampled to 1 × 1 × 1 mm^3 isotropic resolution before cropping, the patch corresponds to a fixed 124 mm field of view per axis. The model was trained for a maximum of 200 epochs with a batch size of 2 using the Adam optimizer (initial learning rate: 1×10^{-6}). The learning rate was halved if the C-index on the internal validation cohort did not improve for 10 consecutive epochs. Early stopping (patience = 20) was applied, and the model with the highest validation C-index was selected. The model was trained with Cox partial likelihood loss. Patients without an observed death were handled as censored observations. They contributed to the risk sets of patients who experienced an event but did not contribute an event term, which is the standard treatment under the Cox loss. All network weights were trained from scratch, without loading pre-trained weights.

2.5. Survival Assessment and Nomogram Development

Kaplan–Meier analysis assessed the association between model-derived risk score and survival. Patients were divided into high- and low-risk groups using optimal cut-points determined by X-tile software (version 3.6.1, Yale University School of Medicine, New Haven, CT, USA) on the training cohort; these thresholds were applied to the validation cohorts. Group differences were tested using a weighted log-rank test (G-rho rank test, rho = 1). To assess whether the model still retains its predictive performance in a more homogeneous clinical population, exploratory subgroup analysis was conducted based on the type of primary tumor. The patients were divided into lung cancer brain metastasis and breast cancer brain metastasis based on the recorded primary site. The C-index of the DeepBMSurv within each subgroup was evaluated in three cohorts. Clinical factors associated with OS were selected via univariate Cox analysis ($p < 0.1$), followed by multivariate Cox regression analysis. Age was dichotomized at 58 years, the optimal cut-point identified by X-tile software in the training cohort and the same cut-point was applied unchanged to the internal and external validation cohorts. GPA is a comprehensive index that includes age, KPS, extracranial metastasis, and the number of lesions. Therefore, these overlapping variables should be regarded as exploratory analyses. To examine the impact of this overlap, a sensitivity analysis was performed in which GPA was excluded. A nomogram integrating significant clinical factors, radiomic signature, and the DL_RiskScore was developed to quantify individual survival probability. The prognostic performance of the radiomic signature, DeepBMSurv, and the nomogram was evaluated using the C-index.

2.6. Statistical Analysis

Continuous variables are presented as median or mean, and categorical variables as frequencies and percentages. Between-group comparisons used the $\chi 2$ (categorical variables) or the Mann–Whitney U test (continuous variables). The median OS and the mean OS were calculated using the KM method with the "survival" package. Model performance was assessed via Harrell's C-index and time-dependent receiver operating characteristic (ROC) analyses across all cohorts. Nomogram calibration was visualized using calibration curves.

Radiomic feature selection employed the "glmnet" package. Nomogram construction used the "rms" and "survival" packages. Time-dependent ROC analysis was performed with the "timeROC" package. C-indices were calculated using the "survcomp" package, with 95% confidence intervals (95% CIs) derived using Noether's method. For all calibration metrics, 95% CIs were derived from 1000 patient-level bootstrap resamples using the percentile method. Decision curve analysis was performed with the "dcurves" package, which operates within a time-to-event framework and right-censoring was appropriately handled using Kaplan–Meier estimation of event probabilities, allowing all patients to be included without artificial dichotomization. All analyses were performed using R (version 4.4.2, https://www.r-project.org) and the SPSS software (version 27.0, IBM, Armonk, NY, USA). A two-sided p-value < 0.05 was considered statistically significant.

3. Results

3.1. Patient Characteristics

Following the enrollment criteria outlined in Figure 3, a total of 216 patients with BM were enrolled in this retrospective study. The baseline clinical characteristics of the cohorts are detailed in Table 1. The overall cohort was divided into a training cohort ($n = 128$), an internal validation cohort ($n = 56$), and an external validation cohort ($n = 32$). The median age was 61.0 years (range: 28−78), 61.0 years (range: 27−78), and 66.0 years (range: 36−83) for the training, internal validation, and external validation cohorts, respectively. The corresponding median OS durations were 418 days (range: 10−2294), 398 days (range: 3−2238),

and 304 days (range: 13–1526). The numbers of deaths and censored patients in each cohort are shown in Supplementary Table S3. Clinical characteristics were well-balanced between the training and internal validation cohorts, with no statistically significant differences observed (all $p > 0.05$).

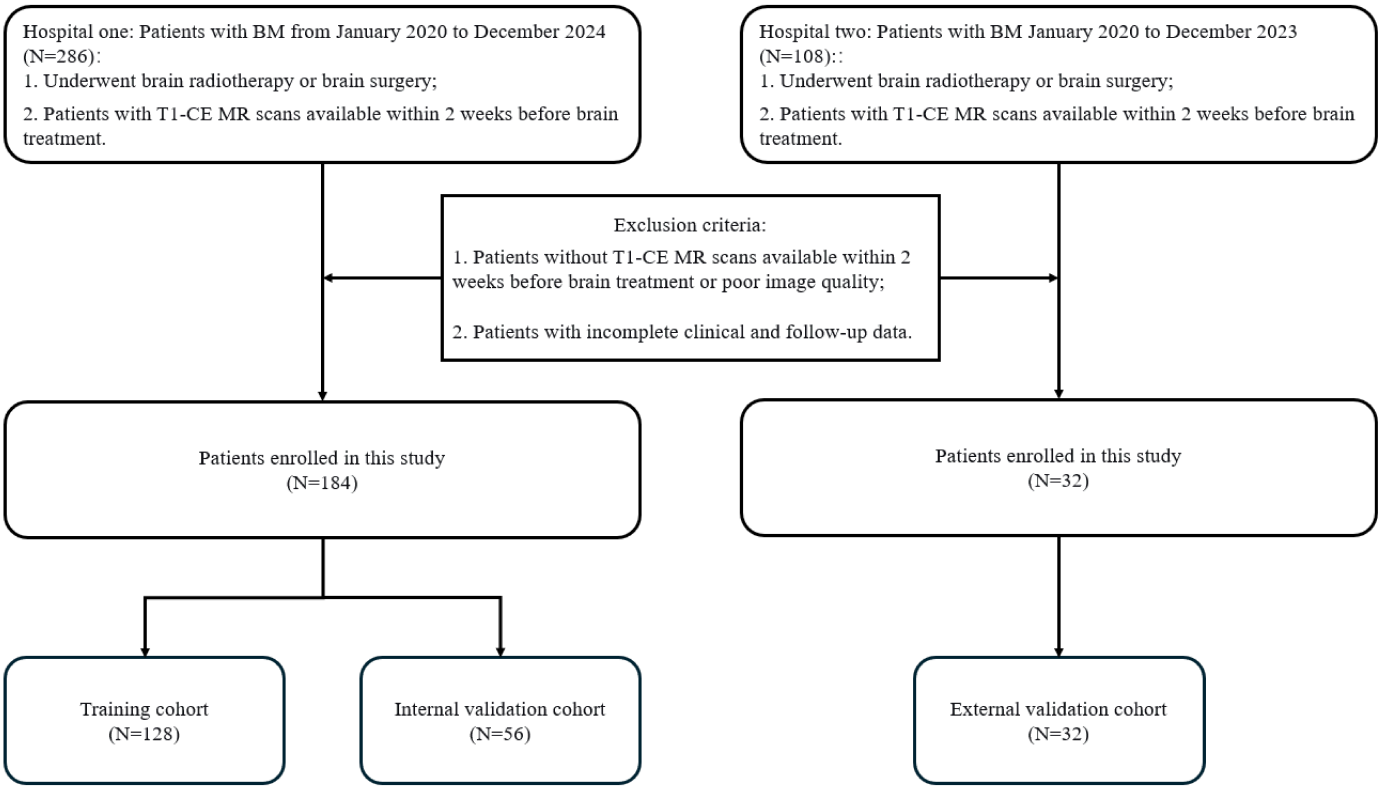


Figure 3. The detail flowchart of the patient recruitment process.

Table 1. Characteristics of patients in the training, internal validation, and external validation cohorts.

Characteristic	Training Cohort (N = 128)	Internal Validation Cohort (N = 56)	External Validation Cohort (N = 32)	p-Value
Gender, No. (%)				0.191 [#1]
Male	69 (53.9%)	36 (64.3%)	25 (78.1%)	
Female	59 (46.1%)	20 (35.7%)	7 (21.9%)	
Age				0.379 [#2]
Range	28–78	27–78	36–83	
Median	61	61	66	
Mean	58.48	60.25	63.91	
KPS				0.368 [#2]
Median	80	70	70	
Primary site, No. (%)				0.557 [#1]
Lung	93 (72.7%)	43 (76.8%)	26 (81.3%)	
Breast	15 (11.7%)	5 (8.9%)	3 (9.4%)	
Other	20 (15.6%)	8 (14.3%)	3 (9.4%)	
Extracranial metastasis, No. (%)				0.743 [#1]
Yes	42 (32.8%)	17 (30.4%)	19 (59.4%)	
No	86 (67.2%)	39 (69.6%)	13 (40.6%)	

Table 1. *Cont.*

Characteristic	Training Cohort ($N = 128$)	Internal Validation Cohort ($N = 56$)	External Validation Cohort ($N = 32$)	*p*-Value
No. of metastases lesions, No. (%)				0.107 [#1]
Single	103 (80.5%)	39 (69.6%)	12 (37.5%)	
Multiple	25 (19.5%)	17 (30.4%)	20 (62.5%)	
GPA, No. (%)				0.243 [#2]
$0 \leq GPA < 1.5$	21 (16.4%)	11 (19.6%)	14 (43.8%)	
$1.5 \leq GPA \leq 2.5$	70 (54.7%)	36 (64.3%)	17 (53.1%)	
$2.5 < GPA \leq 4$	37 (28.9%)	9 (16.1%)	1 (3.1%)	
Overall survival (d)				0.695 [#2]
Range	10−2294	3−2238	13−1526	
Median	418	398	304	
Mean	794.09	641.95	601.51	

p-value used to compare the training cohort and the internal validation cohort. [#1] used $\chi 2$ Tests, [#2] used Mann–Whitney U tests.

3.2. Performance of the Radiomics Model

A total of 833 radiomic features were initially extracted. For the three distinct radiomics models ($RadScore_{pooled}$, $RadScore_{mean}$ and $RadScore_{dominant}$), 375, 432, and 89 features, respectively, were preselected via univariate Cox regression ($p < 0.1$) for further analysis. Feature selection was then performed using LASSO Cox regression and the tuning parameter λ was selected by 5-fold cross-validation, with the cross-validated Harrell's C-index (Supplementary Figure S1). This process resulted in the selection of 15, 9, and 25 radiomic features for $RadScore_{pooled}$, $RadScore_{mean}$, and $RadScore_{dominant}$ models, respectively. For the fusion models, LASSO selected 29, 15, 17, and 18 features for the $RadScore_{mean}$ + $RadScore_{dominant}$, $RadScore_{pooled}$ + $RadScore_{mean}$, $RadScore_{pooled}$ + $RadScore_{dominant}$, and the comprehensive $RadScore_{pooled}$ + $RadScore_{mean}$ + $RadScore_{dominant}$ models, respectively. The final radiomic signature (RadScore) for each model was calculated as a linear combination of these selected features weighted by their respective LASSO coefficients (Supplementary Table S4).

Table 2 presents a comparative analysis of the discriminatory performance of these models. In the training cohort, the C-indices for the $RadScore_{pooled}$, $RadScore_{mean}$, and $RadScore_{dominant}$ models were 0.794, 0.685, and 0.691, respectively. In the internal validation cohort, the corresponding C-indices were 0.658, 0.672, and 0.613. In the external validation cohort, they were 0.538, 0.671, and 0.626. $RadScore_{pooled}$ achieved the highest C-index in the training cohort (0.794) but declined in internal (0.658) and external (0.538) validation. $RadScore_{mean}$ was more stable across cohorts (0.685, 0.672, and 0.671, respectively), and $RadScore_{dominant}$ was intermediate (0.691, 0.613, and 0.626, respectively). These three radiomics models showed comparable discrimination. Furthermore, although the fusion models showed a certain improvement in the C-index value, the generalization ability of the different cohorts is poor. Because $RadScore_{mean}$ showed the smallest drop from training to external validation and the most consistent performance across all three cohorts (from 0.685 to 0.671), it was selected as the optimal radiomics model for subsequent analysis. Using an optimal cut-point of −0.20 derived from the training cohort via X-tile software, patients were stratified into high- and low-risk groups. The distribution of the Rad-score across cohorts is shown in Supplementary Figure S2. Kaplan–Meier analysis confirmed significant survival differences between the risk groups in the training cohort (log-rank $p < 0.001$) and the internal validation cohort (log-rank $p = 0.005$) (Supplementary Figure S3A,B). In the external validation cohort, however, all 32 patients were classified into the high-risk

group under the training-derived cut-point, so a between-group survival comparison was not possible and only a single Kaplan–Meier curve is shown (Supplementary Figure S3C).

Table 2. Comparison of the discriminatory performance of clinical, different radiomics, deep learning, and their combination models in predicting OS.

Models	Training Cohort		Internal Validation Cohort		External Validation Cohort	
	C-Index	95% CI	C-Index	95% CI	C-Index	95% CI
$RadScore_{pooled}$	0.794	0.672−0.886	0.658	0.456−0.815	0.538	0.285−0.774
$RadScore_{mean}$	0.685	0.551−0.794	0.672	0.469−0.826	0.671	0.396−0.864
$RadScore_{dominant}$	0.691	0.557−0.799	0.613	0.413−0.781	0.626	0.357−0.835
$RadScore_{pooled}$ plus $RadScore_{mean}$	0.731	0.600−0.832	0.705	0.502−0.850	0.507	0.260−0.751
$RadScore_{pooled}$ plus $RadScore_{dominant}$	0.690	0.557−0.799	0.677	0.474−0.830	0.643	0.371−0.846
$RadScore_{mean}$ plus $RadScore_{dominant}$	0.782	0.654−0.872	0.655	0.453−0.814	0.651	0.377−0.851
$RadScore_{mean}$ plus $RadScore_{dominant}$ plus $RadScore_{pooled}$	0.726	0.594−0.828	0.705	0.502−0.805	0.643	0.371−0.846
GPA	0.620	0.475−0.747	0.628	0.411−0.803	0.579	0.292−0.821
DeepBMSurv	0.806	0.753−0.853	0.755	0.675−0.824	0.733	0.627−0.826
Nomogram	0.809	0.684−0.893	0.764	0.563−0.891	0.733	0.453−0.901

3.3. Performance of the DL Model

We developed the DeepBMSurv model to predict OS from pre-treatment T1-CE MR images. The model outputs a DL-based risk score (DL_RiskScore). The DL_RiskScore achieved C- indices of 0.806 [95% CI: 0.753−0.853], 0.755 [95% CI: 0.675−0.824], and 0.733 [95% CI: 0.627−0.826] in the training, internal validation, and external validation cohorts, respectively. Ablation studies were conducted to evaluate the contribution of each architectural component (Supplementary Table S5). The baseline Resnet-34 model yielded C-indices of 0.705 [95% CI: 0.645−0.761], 0.706 [95% CI: 0.633−0.777], and 0.635 [95% CI: 0.485−0.785] across the three cohorts. Incorporating PSA modules improved performance to 0.760 [95% CI: 0.707−0.810], 0.733 [95% CI: 0.643−0.813], and 0.605 [95% CI: 0.457−0.751]. The standalone ViT baseline achieved C-indices of 0.716 [95% CI: 0.660−0.765], 0.665 [95% CI: 0.564−0.768], and 0.611 [95% CI: 0.463−0.762]. Although training ResNet-34 solely with PSA or solely with ViT did not exceed the baseline of ResNet-34, incorporating a cross-attention mechanism between the CNN and ViT branches (Resnet34CrossViT) further enhanced performance to 0.766 [95% CI: 0.713–0.813], 0.744 [95% CI: 0.670–0.817], and 0.699 [95% CI: 0.596–0.789]. The full DeepBMSurv architecture, integrating all components, delivered the best overall prognostic performance.

Patients were stratified into high- and low-risk groups using a DL_RiskScore cut-point of 0.58, determined from the training cohort. The score distributions for both risk groups across all cohorts are shown in Figure 4A–C. Kaplan–Meier curves demonstrated significant survival separation between the risk strata in the training (Figure 4D), internal validation (Figure 4E), and external validation (Figure 4F) cohorts (log-rank $p < 0.05$). To further assess the robustness of the DeepBMSurv, we performed an exploratory subgroup analysis focusing on DeepBMSurv without retraining (Supplementary Table S6). In the lung cancer subgroup, DeepBMSurv retained consistent discrimination across all cohorts (C-index: 0.777, 0.737, and 0.762, respectively). In the breast cancer subgroup, the C-index was 0.942 in the training cohort and 0.714 in the internal validation cohort. Due to the small number of breast cancer cases, reliable estimation could not be conducted in the external validation cohort.

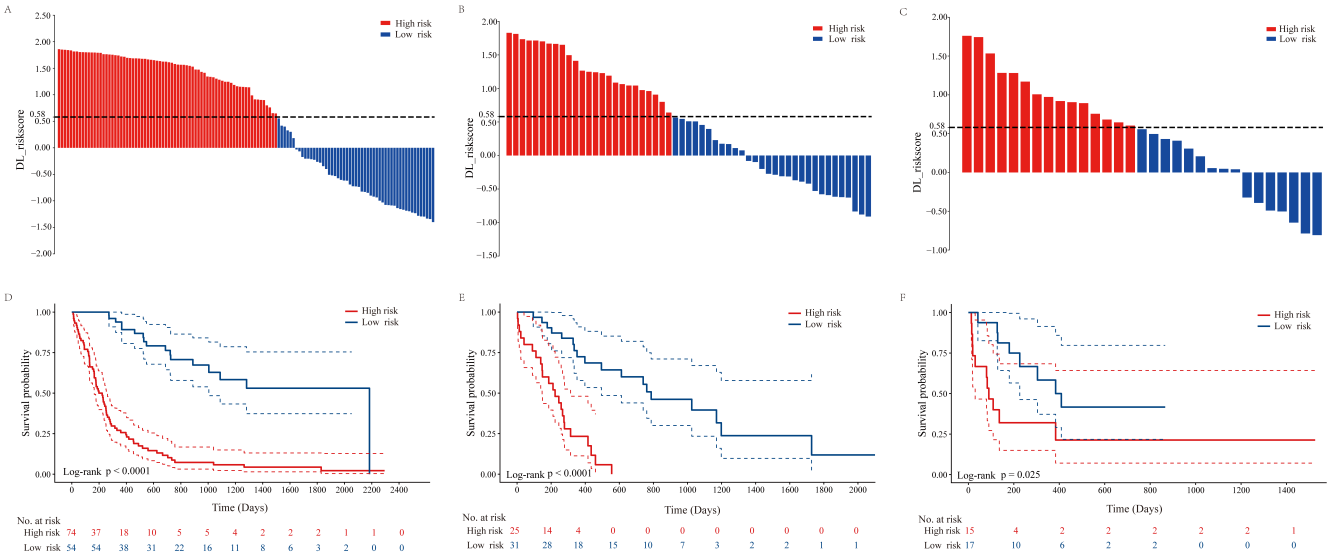


Figure 4. DL_riskscore for each patient in the (**A**) training cohort, (**B**) internal validation cohort, and (**C**) external validation cohort. Patients were classified into high- and low-risk groups with a threshold of 0.58. The Kaplan–Meier survival curves of the low- and high-risk groups according to the log-rank test for patients in (**D**) training cohort, (**E**) internal validation cohort, and (**F**) external validation cohort. The horizontal dashed lines in (**D**–**F**) indicate 95% CIs.

3.4. Prognostic Factors and Nomogram Development

Univariable and multivariable Cox regression analyses were performed on the training cohort to identify clinical predictors of OS (Table 3). In multivariable analysis, the GPA score remained an independent predictor of OS (Hazard Ratio [HR] = 0.755; 95% CI: 0.576–0.989; $p = 0.041$). In an additional sensitivity analysis excluding GPA (Supplementary Table S7), age was significantly associated with OS (HR = 0.572, 95% CI:0.352–0.930; $p = 0.024$), whereas KPS ($p = 0.152$) and number of lesions ($p = 0.062$) were not. Although there is collinearity between age and GPA, GPA is still retained as the indicator used to develop the nomogram because it is a validated, clinically applicable scoring metric and its use is supported, rather than being contradicted by sensitivity analysis. A clinical model based solely on GPA yielded C-indices of 0.620 [95% CI: 0.475–0.747], 0.628 [95% CI: 0.411–0.803], and 0.579 [95% CI: 0.292–0.821] in the training, internal validation, and external validation cohorts, respectively (Table 2). To provide an individualized prognostic tool, a nomogram was constructed by integrating the independent clinical predictor (GPA score), the optimal radiomics signature ($RadScore_{mean}$) and the DL_RiskScore (Figure 5A). This combined nomogram achieved C-indices of 0.809 [95% CI: 0.684–0.893], 0.764 [95% CI: 0.563–0.891], and 0.733 [95% CI: 0.453–0.901] in the training, internal validation, and external validation cohort, respectively (Table 2). Because the external cohort contributed to the selection of $RadScore_{mean}$ and the confirmation of the nomogram specification, this external C-index of 0.733 should be interpreted as an optimistic estimate. Calibration curves for 3-month, 6-month, and 1-year OS are shown in Figure 5B–D. To complement the calibration assessment, calibrations at 3 months, 6 months, and 1 year using the calibration intercept, calibration slope, and Brier score, with 95% CIs from 1000 patient-level bootstrap resamples were quantified (Supplementary Table S8). Calibration slopes were 1.00 (95% CI 0.81–1.28), 0.98 (0.70–1.42), and 0.56 (0.10–1.32) in the training, internal, and external cohorts, respectively. Calibration intercepts were positive in the training cohort (range 1.50–3.41) and internal validation cohort (range 1.71–3.62) with 95% CIs excluding 0, indicating a tendency to overestimate survival probability. The external validation cohort calibration intercept declined from 1.77 at 3 months to −0.14 at 1 year, with 95% CIs crossing 0 at 6 months and

1 year and excluding 0 at 3 months. Brier scores increased with the prediction horizon in all three cohorts. Time-dependent ROC analysis further validated the model's predictive accuracy, with AUCs for 3-month, 6-month, and 1-year OS detailed in Figure 5E–G.

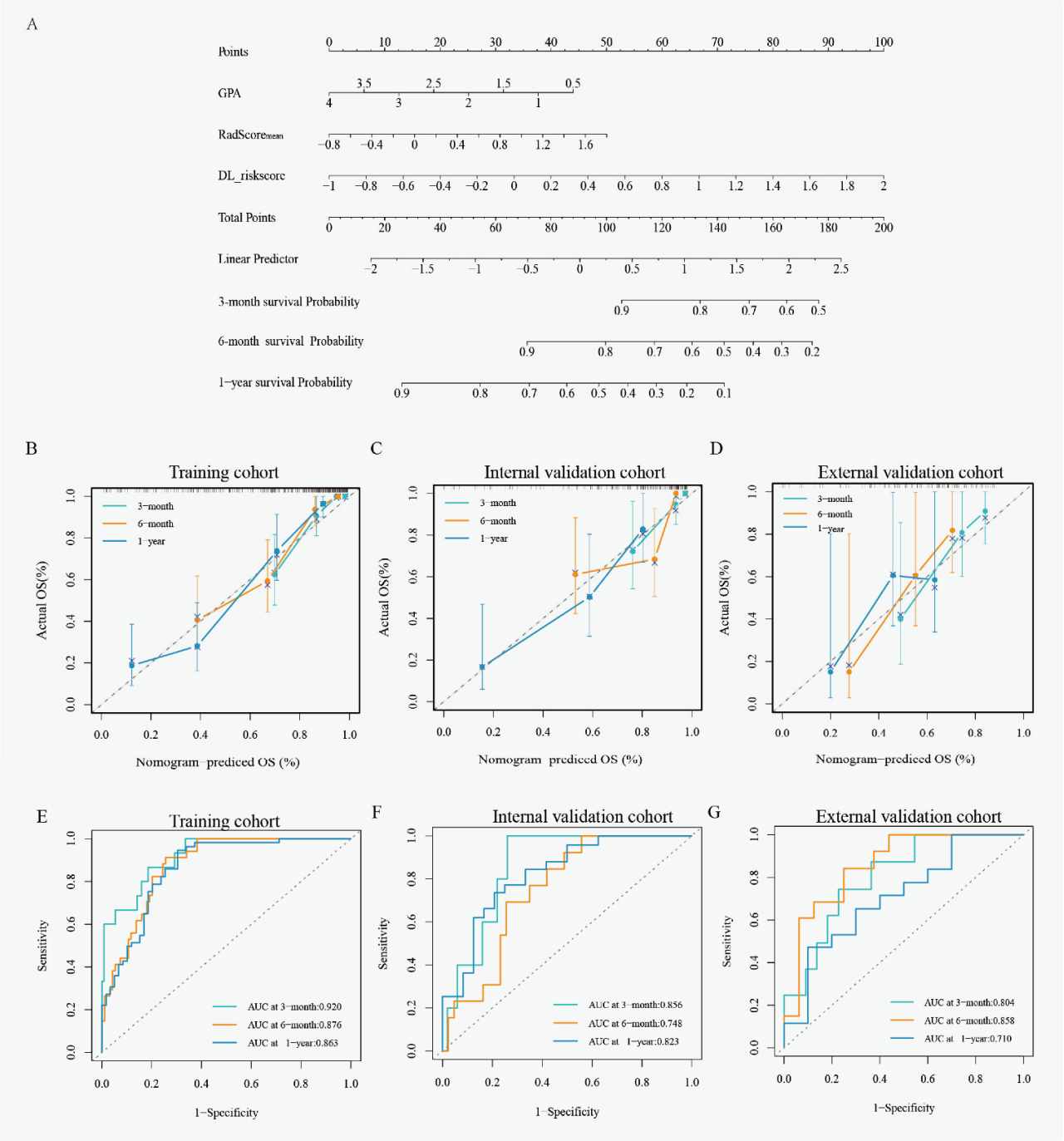


Figure 5. (**A**) Nomogram integrated with DL_riskscore, Rad-score and GPA; Calibration curves of the nomogram for 3-months survival, 6-months survival and 1-year survival in (**B**) training cohort, (**C**) internal validation cohort, and (**D**) external validation cohort. Time-dependent ROC curve at 3 months, 6 months and 1 year was used to assess the prognostic performance of the nomogram in (**E**) training cohort, (**F**) internal validation cohort, and (**G**) external validation cohort.

Table 3. Univariable and multivariable Cox regression analysis for risk factors associated with overall survival.

Variable	Univariate Analysis			Multivariate Analysis		
	HR	95% CI	*p*-Value	HR	95% CI	*p*-Value
Gender (female $^{+}$, male)	1.298	0.850–1.982	0.228			
Age (>58 $^{+}$, ≤58)	0.599	0.370–0.969	0.037	0.647	0.388–1.081	0.096
KPS	0.990	0.980–1.001	0.079	0.999	0.986–1.012	0.890
Primary site (no-lung $^{+}$, lung)	1.400	0.849–2.308	0.187			
Extracranial metastasis (yes $^{+}$, no)	0.816	0.530–1.257	0.356			
Number of lesion (multiple $^{+}$, single)	0.633	0.390–1.029	0.065	0.724	0.426−1.232	0.234
GPA	0.695	0.564–0.855	<0.001	0.755	0.576−0.989	0.041 *

$^{+}$: represents the reference category. * $p < 0.05$.

Decision curve analysis at 1 year demonstrated that within a wide range of clinically reasonable probability thresholds, both the nomogram and the DL model demonstrated higher net benefits than the GPA and radiomics models in the internal validation cohort (Figure 6A). In the external validation cohort, although the curve fluctuated significantly due to the limited sample size, the nomogram consistently showed positive net benefits at moderate probability thresholds, outperforming the GPA model (Figure 6B).

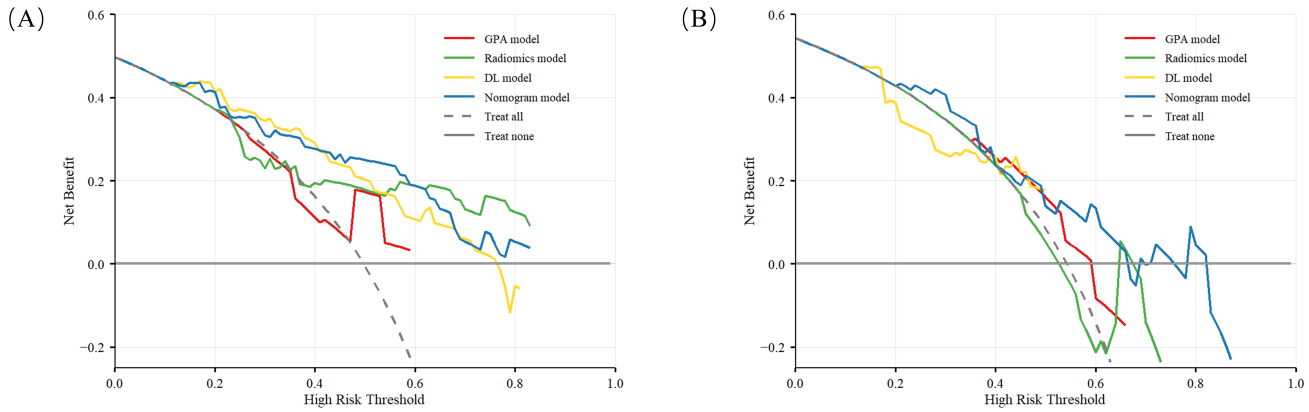


Figure 6. Decision curve analysis for 1-year survival in the validation cohorts. (**A**) Internal validation cohort. (**B**) External validation cohort. The curves show the net benefit of the Graded Prognostic Assessment (GPA), radiomics model, deep learning (DL) model, and integrated nomogram across high-risk threshold probabilities.

4. Discussion

In this study, we developed and validated DeepBMSurv, a novel cross-attention-based DL framework for predicting OS in patients with BM using pre-treatment MRI. The model demonstrated robust prognostic performance, achieving C-indices of 0.755 and 0.733 in internal and external validation cohorts, respectively, outperforming both GPA and radiomics models. Furthermore, we constructed a clinically practical nomogram by integrating the GPA score, the optimal radiomics signature ($RadScore_{mean}$), and the DL_RiskScore. The integrated nomogram yielded C-indices of 0.764 and 0.733 for OS prediction in the internal and external validation cohorts, respectively.

The clinical management of BM is inherently complex due to the varying degrees of inter- and intra- tumoral heterogeneity and the challenge of overcoming the blood–brain barrier for systemic therapies [29,30]. A further limitation concerns treatment heterogeneity. The management of brain metastases is highly variable, encompassing stereotactic radiosurgery, whole-brain radiotherapy, neurosurgical resection, and diverse systemic regimens, and treatment selection is itself influenced by prognosis. Consequently, local interventions

such as radiotherapy and neurosurgical resection remain cornerstone treatments, as reflected in our study cohort. The diverse primary origins and metastatic burdens observed in our study population underscore the urgent need for precise, individualized risk stratification. Our study was designed to test whether pre-treatment imaging carries a prognostic signal that is shared across this heterogeneous treated population rather than to estimate treatment-specific effects. Detailed treatment data including SRS dose and fractionation, WBRT schedule, extent of resection, and agent-level systemic therapy were not systematically collected, and we therefore could not adjust for treatment in the survival models. Consequently, the external C-index of 0.733 (95% CI 0.627–0.826) should be interpreted as discrimination across a heterogeneous treated population, not as performance within a uniform treatment stratum, and confounding by treatment selection cannot be excluded. The observation that prognostic discrimination persists across this heterogeneity is consistent with the model capturing tumor-intrinsic biological aggressiveness rather than acting as a proxy for treatment intensity. By providing an accurate survival estimate, tools like DeepBMSurv and the associated nomogram may assist in prognostic risk stratification and risk communication [6,31,32]. Specifically, the model may help identify high-risk patients who could benefit from closer follow-up. However, this study evaluated survival prediction only, it did not assess whether using the model changes treatment decisions or improves follow-up.

Accurately predicting survival in BM patients is challenging due to the multitude of influencing factors. Previous studies have investigated various elements, including age, KPS, number of metastases, tumor volume, and the presence of extracranial disease [33,34]. Among standardized indices, the GPA is widely adopted for its objectivity, quantifiability, and ease of use. In this study, the external validation cohort represents a higher-risk population and differed from the training cohort in several clinically relevant characteristics, including a higher proportion of multiple lesions, more frequent extracranial metastasis, and a GPA distribution shifted toward lower scores. These differences may limit the transportability of the model to populations with a higher burden of intracranial and extracranial disease. The diagnostic-specific GPA (DS-GPA) and its disease-specific updated version have enhanced the prognostic discrimination ability in specific primary tumor types. We intentionally adopted the original GPA to provide a unified assessment standard for heterogeneous primary tumor types in this study cohort, thereby achieving a single predictive map applicable to all patients [8]. In this study, while the GPA score remained a significant independent predictor (HR = 0.755, $p < 0.05$) in multivariate analysis, its standalone predictive power was limited (C-index: 0.579 in external validation). This result highlights that conventional clinical factors alone may be insufficient to capture the complex prognostic information embedded within medical images.

Radiomics has emerged as a promising non-invasive approach to decode such information by extracting high-dimensional quantitative features from medical images [21,35]. Liao et al. recently developed a DL model combining MRI radiomics, clinical features, and EGFR status to predict survival after radiosurgery in lung cancer brain metastases, reporting a C-index of 0.75 and time-dependent AUCs of 0.82–0.92. Unlike Liao et al., the patients in our study did not receive a single treatment modality; our study integrated a multi-lesion deep learning risk score covering various types of primary tumors and validated the model in an independent external cohort [21]. A key finding was the comparative performance of different feature aggregation strategies for multiple metastases. Contrary to some studies suggesting the largest lesion carries the most prognostic weight, our $\text{RadScore}_{\text{mean}}$ model (based on the average signature of all lesions) outperformed the $\text{RadScore}_{\text{dominant}}$ model (based on the largest lesion) in external validation. This suggests that the aggregate biological burden of all metastatic lesions may be more prognostically informative than the

characteristics of any single dominant lesion. Our optimal radiomics model demonstrated predictive performance (C-index: 0.67 in validation cohorts) comparable to previous studies. However, radiomics relies on handcrafted feature engineering, which may not capture the full spectrum of latent prognostic patterns within complex imaging data.

DL offers a powerful alternative by automatically learning hierarchical and comprehensive feature representations directly from images. In this study, the DeepBMSurv model not only successfully stratified patients into distinct risk groups but also achieved a C-index of 0.733 in the external cohort compared to the radiomics models. This suggests that MRI-based DL algorithms hold significant predictive value for BM outcomes. Exploratory subgroup analysis provided important evidence for the clinical application of this model. In the lung cancer subgroup that constitutes the majority of the multi-center cohort (72.7%, 76.8%, and 81.3% for the training, internal and external validation cohorts, respectively), DeepBMSurv demonstrated consistent and robust discriminatory ability in all three cohorts (C index: 0.777, 0.737, and 0.762). The subgroup analysis based on the type of primary tumor was limited due to the small sample size. Therefore, these exploratory results need to be verified in larger-scale cohorts specifically targeting the specific primary tumors. The architecture of DeepBMSurv synergistically combines a CNN (ResNet-34 with PSA attentional modules) for local feature extraction and a ViT for modeling global contextual dependencies, linked via a cross-attention mechanism. This design allows the model to focus adaptively on the most prognostically salient image regions across different scales. Similarly, previous studies suggested that attentional mechanisms, such as a PSA attention module and cross-attention, may improve the prediction performance of convolutional neural networks [36,37]. Our ablation studies validated the contribution of each component, particularly noting that the cross-attention mechanism boosted model generalizability (external validation C-index: 0.699). The integration of complementary prognostic information often yields the most robust predictive models. It is worth noting that the input size of the network (124 × 124 × 124) is not fixed and can be adjusted according to the characteristics of the target patient population. We selected this size because it accommodated the spatial distribution of all lesions in our cohort while remaining compatible with GPU memory constraints. In future studies involving larger brain volumes or more widely distributed lesions, a larger input size could be considered to ensure full coverage of the tumor burden. By combining the deep learning DL_RiskScore, the radiomics RadScore, and the clinical GPA score, our nomogram achieved numerically higher C-index in training and internal validation cohorts and the same C-index in external cohorts, demonstrating the additive value of each data modality to some extent. The nomogram achieved numerically higher C-indices than DeepBMSurv, with C-indices of 0.809 vs. 0.806 in the training cohort and 0.764 vs. 0.755 in the internal validation cohort. However, in the external validation cohort, the C values were 0.733 vs. 0.733. This may reflect the differences in baseline risk estimates and case mix and therefore should be taken into account when applying the nomogram. In addition, although the selection of the radiomics model was primarily based on its internal validation stability, its external performance was also reviewed during the final decision. This introduces a potential optimism bias to the reported external C-index of the nomogram. The decision curves suggest potential clinical net benefit for selected risk thresholds in the validation cohort. However, the ranking of the models changes across thresholds, and the nomogram does not show a consistent advantage over DeepBMSurv due to the relatively small sample size of external validation.

This study has several limitations. First, despite external validation, the sample size, particularly of the external cohort, remains limited. This may contribute to the observed shift in the calibration curve for external validation and introduces potential selection bias inherent to retrospective designs. Future validation with larger, multi-center prospec-

tive datasets is essential. Second, our analysis was restricted to T1-CE MRI sequences. Incorporating multi-parametric MRI (e.g., T2, FLAIR, perfusion) could provide a more comprehensive characterization of the tumor microenvironment. Third, radiomic features, particularly texture features, can be sensitive to discretization, interpolation, and segmentation parameters. Although identical preprocessing was applied across all cohorts and the signatures were validated externally across different scanners and slice thicknesses, a formal multi-parameter stability analysis was not performed and should be addressed in future prospective studies. In addition, detailed information regarding the brain treatment modalities (stereotactic radiotherapy, whole brain radiotherapy, or surgery) and systemic treatments (chemotherapy, targeted therapy, or immunotherapy) was inconsistent in this retrospective study. Due to the potential impact of treatment differences on survival rates and the variations across different centers, residual confounding factors resulting from the treatments cannot be ruled out. Fourth, the external cohort was used to select the radiomic aggregation strategy ($RadScore_{mean}$) and the nomogram specification. Therefore, the external C-index values for the selected radiomics model and the final nomogram should be regarded as optimistic estimates. A fully independent, pre-registered external validation in a prospective cohort is required to confirm the true generalizability of the model. Finally, while DeepBMSurv is designed for BM patients irrespective of primary origin, its performance within specific primary cancer subgroups (e.g., lung cancer vs. breast cancer) warrants dedicated investigation in future studies.

5. Conclusions

In conclusion, we have developed and validated DeepBMSurv, a preoperative MRI-based DL model for predicting OS in patients with BM. This model demonstrates higher C-index estimates than traditional clinical and radiomics approaches and may provide a robust tool for individualized prognostic risk stratification.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers18223693/s1, Table S1: The details of the MR image acquisition parameters for the different hospitals; Table S2: List of the radiomics features extracted by PyRadiomics and the extracted setting parameters; Table S3: The numbers of deaths and censored patients in each cohort; Table S4: Selected features by LASSO cox analysis in different models; Table S5: Ablation experiment of different deep learning models in predicting OS; Table S6: The exploratory subgroup analysis of the DeepBMSurv model based on the type of the primary tumor; Table S7: Sensitivity analysis excluding GPA: multivariable Cox model of GPA components for overall survival in the training cohort; Table S8: Quantitative calibration of the nomogram at 3 months, 6 months, and 1 year in the training, internal validation, and external validation cohorts; Figure S1: Selection of survival-associated radiomic features using LASSO Cox regression. The tuning parameter λ was selected by 5-fold cross-validation, with the cross-validated Harrell's C-index plotted against $\log(\lambda)$. (A) $RadScore_{pooled}$ model (B) $RadScore_{mean}$ model, (C) $RadScore_{dominant}$ model, (D) $RadScore_{mean}$ plus $RadScore_{dominant}$ model, (E) $RadScore_{mean}$ plus $RadScore_{pooled}$ model, (F) $RadScore_{dominant}$ plus $RadScore_{pooled}$ model and (G) $RadScore_{mean}$ plus $RadScore_{dominant}$ plus $RadScore_{pooled}$ model; Figure S2: Radiomics scores from $RadScore_{mean}$ model for each patient in the (A) training cohort and (B) internal validation cohort and (C) external validation cohort. Patients were classified into high- and low-risk groups with a threshold of −0.20. In the external validation cohort, all 32 patients were classified into the high-risk group under the training-derived cut-point, a between-group survival comparison was therefore not possible, and a single Kaplan–Meier curve is shown; Figure S3: The Kaplan–Meier survival curves of the low-and high-risk groups according to the log-rank test for patients in (A) training cohort, (B) internal validation cohort, and (C) external validation cohort. The vertical dashed line is the 95% confidence interval. In the external validation cohort, all 32 patients were classified into the high-risk group under the training-derived cut-point, a between-group survival comparison was therefore not possible, and a single Kaplan–Meier curve is shown.

Author Contributions: J.Z.: Conceptualization, data curation, methodology, visualization and writing—original draft; J.H.: Data curation, validation, writing—original draft; M.G.: Conceptualization and investigation; Z.Z.: Data curation and investigation; Y.M.: Data curation; B.S.: Data curation; J.W.: Data curation; X.L.: Validation; W.Y.: Supervision; C.X.: Conceptualization, resources, supervision and writing—review and editing; X.J.: Conceptualization, funding acquisition, resources, supervision and writing—review and editing. All authors were involved in the data curation and acquisition of resources. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported partially by the National Natural Science Foundation (12475352), a Key project of the Zhejiang Natural Science Foundation (LZ24A050008), a Key project of the Zhejiang Provincial Health Science and Technology Program (WKJ-ZJ-2437), a Major project of the Wenzhou Science and Technology Bureau (ZY2022016, ZY2020011), a project of the Wenzhou Science and Technology Bureau (YC20260001), Zhejiang Engineering Research Center for Innovation and Application of Intelligent Radiotherapy Technology, Zhejiang-Hong Kong Precision Theranostics of Thoracic Tumors Joint Laboratory, and Wenzhou Key Laboratory of Basic Science and Translational Research of Radiation Oncology, Zhejiang Key Laboratory of Intelligent Cancer Biomarker Discovery and Translation, Discipline Cluster of Oncology, Wenzhou Medical University (z3-2023037), Zhejiang Higher Education Association Project (KT2024088).

Institutional Review Board Statement: This study was approved by the institutional review board of the Review of Ethics Committee in Clinical Research (ECCR) of the First Affiliated Hospital of Wenzhou Medical University and conducted in accordance with the Declaration of Helsinki. (ECRR KY2025-R151, 7 May 2025).

Informed Consent Statement: The requirement for informed consent was waived due to the retrospective nature of the study.

Data Availability Statement: The datasets generated and analyzed during the current study are not publicly available due to privacy concerns but are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank all participants for their contributions to this study.

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

References

1. Nayak, L.; Lee, E.Q.; Wen, P.Y. Epidemiology of Brain Metastases. *Curr. Oncol. Rep.* **2012**, *14*, 48–54. [CrossRef] [PubMed]
2. Achrol, A.S.; Rennert, R.C.; Anders, C.; Soffietti, R.; Ahluwalia, M.S.; Nayak, L.; Peters, S.; Arvold, N.D.; Harsh, G.R.; Steeg, P.S.; et al. Brain Metastases. *Nat. Rev. Dis. Primers* **2019**, *5*, 5. [CrossRef] [PubMed]
3. Fecci, P.E.; Champion, C.D.; Hoj, J.; McKernan, C.M.; Goodwin, C.R.; Kirkpatrick, J.P.; Anders, C.K.; Pendergast, A.M.; Sampson, J.H. The Evolving Modern Management of Brain Metastasis. *Clin. Cancer Res.* **2019**, *25*, 6570–6580. [CrossRef] [PubMed]
4. Lamba, N.; Wen, P.Y.; Aizer, A.A. Epidemiology of Brain Metastases and Leptomeningeal Disease. *Neuro-Oncology* **2021**, *23*, 1447–1456. [CrossRef] [PubMed]
5. Cagney, D.N.; Martin, A.M.; Catalano, P.J.; Redig, A.J.; Lin, N.U.; Lee, E.Q.; Wen, P.Y.; Dunn, I.F.; Bi, W.L.; Weiss, S.E.; et al. Incidence and Prognosis of Patients with Brain Metastases at Diagnosis of Systemic Malignancy: A Population-Based Study. *Neuro-Oncology* **2017**, *19*, 1511–1521. [CrossRef] [PubMed]
6. Lamba, N.; Kearney, R.B.; Catalano, P.J.; Hassett, M.J.; Wen, P.Y.; Haas-Kogan, D.A.; Aizer, A.A. Population-Based Estimates of Survival among Elderly Patients with Brain Metastases. *Neuro-Oncology* **2021**, *23*, 661–676. [CrossRef] [PubMed]
7. Gaspar, L.; Scott, C.; Rotman, M.; Asbell, S.; Phillips, T.; Wasserman, T.; McKenna, W.G.; Byhardt, R. Recursive partitioning analysis (RPA) of prognostic factors in three Radiation Therapy Oncology Group (RTOG) brain metastases trials. *Int. J. Radiat. Oncol. Biol. Phys.* **1997**, *37*, 745–751. [CrossRef] [PubMed]
8. Sperduto, P.W.; Berkey, B.; Gaspar, L.E.; Mehta, M.; Curran, W. A New Prognostic Index and Comparison to Three Other Indices for Patients with Brain Metastases: An Analysis of 1,960 Patients in the RTOG Database. *Int. J. Radiat. Oncol. Biol. Phys.* **2008**, *70*, 510–514. [CrossRef] [PubMed]
9. Weltman, E.; Salvajoli, J.V.; Brandt, R.A.; de Morais Hardt, R.; Prisco, F.E.; Cruz, J.C.; de Oliveira Borges, S.R.; Wajsbrot, D.B. Radiosurgery for brain metastases: A score index for predicting prognosis. *Int. J. Radiat. Oncol. Biol. Phys.* **2000**, *46*, 1155–1161. [CrossRef] [PubMed]

10. Farris, M.; McTyre, E.R.; Cramer, C.K.; Hughes, R.T.; Ayala-Peacock, D.N.; Randolph, D.M.; Bourland, J.D.; Tatter, S.B.; Laxton, A.W.; McTyre, K.W.; et al. Brain metastasis velocity: A novel prognostic metric predictive of overall survival and freedom from whole-brain radiation therapy after distant brain failure following upfront radiosurgery alone. *Int. J. Radiat. Oncol. Biol. Phys.* **2017**, *98*, 131–141. [CrossRef] [PubMed]
11. Nieder, C.; Mehta, M.P. Prognostic Indices for Brain Metastases–Usefulness and Challenges. *Radiat. Oncol.* **2009**, *4*, 10. [CrossRef] [PubMed]
12. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; van Stiphout, R.G.P.M.; Granton, P.; Zegers, C.M.L.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis. *Eur. J. Cancer* **2012**, *48*, 441–446. [CrossRef] [PubMed]
13. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [CrossRef] [PubMed]
14. Pope, W.B. Brain Metastases: Neuroimaging. In *Handbook of Clinical Neurology*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 149, pp. 89–112, ISBN 978-0-12-811161-1.
15. Arbizu, J.; Domínguez, P.D.; Diez-Valle, R.; Vigil, C.; García-Eulate, R.; Zubieta, J.L.; Richter, J.A. Neuroimagen de los tumores cerebrales. *Rev. Esp. Med. Nucl.* **2011**, *30*, 47–65. [CrossRef] [PubMed]
16. Du, P.; Liu, X.; Xiang, R.; Lv, K.; Chen, H.; Liu, W.; Cao, A.; Chen, L.; Wang, X.; Yu, T.; et al. Development and Validation of a Radiomics-Based Prediction Pipeline for the Response to Stereotactic Radiosurgery Therapy in Brain Metastases. *Eur. Radiol.* **2023**, *33*, 8925–8935. [CrossRef] [PubMed]
17. Huang, C.-Y.; Lee, C.-C.; Yang, H.-C.; Lin, C.-J.; Wu, H.-M.; Chung, W.-Y.; Shiau, C.-Y.; Guo, W.-Y.; Pan, D.H.-C.; Peng, S.-J. Radiomics as Prognostic Factor in Brain Metastases Treated with Gamma Knife Radiosurgery. *J. Neurooncol.* **2020**, *146*, 439–449. [CrossRef] [PubMed]
18. Deng, F.; Xiao, G.; Tanzhu, G.; Chu, X.; Ning, J.; Lu, R.; Chen, L.; Zhang, Z.; Zhou, R. Predicting Survival Rates in Brain Metastases Patients from Non-Small Cell Lung Cancer Using Radiomic Signatures Associated with Tumor Immune Heterogeneity. *Adv. Sci.* **2025**, *12*, 2412590. [CrossRef] [PubMed]
19. Li, Y.; Qin, Q.; Zhang, Y.; Cao, Y. Noninvasive Determination of the IDH Status of Gliomas Using MRI and MRI-Based Radiomics: Impact on Diagnosis and Prognosis. *Curr. Oncol.* **2022**, *29*, 6893–6907. [CrossRef] [PubMed]
20. Chen, Y.; Yuan, Q.; Cramer, C.K.; Helis, C.A.; He, G.; Chen, C.; Choi, A.; Young, P.J.; Wang, Y.; Xing, F.; et al. A multimodal MRI-radiomics deep learning model for survival risk stratification after gamma knife radiosurgery in patients with brain metastases: A multicenter retrospective study. *Radiother. Oncol.* **2026**, *222*, 111686. [CrossRef] [PubMed]
21. Liao, C.-Y.; Lee, C.-C.; Yang, H.-C.; Chen, C.-J.; Chung, W.-Y.; Wu, H.-M.; Guo, W.-Y.; Liu, R.-S.; Lu, C.-F. Predicting Survival after Radiosurgery in Patients with Lung Cancer Brain Metastases Using Deep Learning of Radiomics and EGFR Status. *Phys. Eng. Sci. Med.* **2023**, *46*, 585–596. [CrossRef] [PubMed]
22. Mao, N.; Zhang, H.; Dai, Y.; Li, Q.; Lin, F.; Gao, J.; Zheng, T.; Zhao, F.; Xie, H.; Xu, C.; et al. Attention-Based Deep Learning for Breast Lesions Classification on Contrast Enhanced Spectral Mammography: A Multicentre Study. *Br. J. Cancer* **2023**, *128*, 793–804. [CrossRef] [PubMed]
23. Hu, C.; Chen, W.; Li, F.; Zhang, Y.; Yu, P.; Yang, L.; Huang, L.; Sun, J.; Chen, S.; Shi, C.; et al. Deep Learning Radio-Clinical Signature for Predicting Neoadjuvant Chemotherapy Response and Prognosis from Pretreatment CT Images of Locally Advanced Gastric Cancer Patients. *Int. J. Surg.* **2023**. *Publish Ahead of Print*. [CrossRef] [PubMed]
24. Masson-Grehaigne, C.; Lafon, M.; Palussière, J.; Leroy, L.; Bonhomme, B.; Jambon, E.; Italiano, A.; Cousin, S.; Crombé, A. Single- and Multi-Site Radiomics May Improve Overall Survival Prediction for Patients with Metastatic Lung Adenocarcinoma. *Diagn. Interv. Imaging* **2024**, *105*, 439–452. [CrossRef] [PubMed]
25. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-Based Phenotyping. *Radiology* **2020**, *295*, 328–338. [CrossRef] [PubMed]
26. Tustison, N.J.; Avants, B.B.; Cook, P.A.; Zheng, Y.J.; Egan, A.; Yushkevich, P.A.; Gee, J.C. N4ITK: Improved N3 Bias Correction. *IEEE Trans. Med. Imaging* **2010**, *29*, 1310–1320. [CrossRef] [PubMed]
27. Van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [CrossRef] [PubMed]
28. Tibshirani, R. The Lasso Method for Variable Selection in the Cox Model. *Stat. Med.* **1997**, *16*, 385–395. [CrossRef]
29. Muldoon, L.L.; Soussain, C.; Jahnke, K.; Johanson, C.; Siegal, T.; Smith, Q.R.; Hall, W.A.; Hynynen, K.; Senter, P.D.; Peereboom, D.M.; et al. Chemotherapy Delivery Issues in Central Nervous System Malignancy: A Reality Check. *JCO* **2007**, *25*, 2295–2305. [CrossRef] [PubMed]

30. Oh, Y.; Taylor, S.; Bekele, B.N.; Debnam, J.M.; Allen, P.K.; Suki, D.; Sawaya, R.; Komaki, R.; Stewart, D.J.; Karp, D.D. Number of Metastatic Sites Is a Strong Predictor of Survival in Patients with Nonsmall Cell Lung Cancer with or without Brain Metastases. *Cancer* **2009**, *115*, 2930–2938. [CrossRef] [PubMed]
31. Noh, T.; Walbert, T. Brain Metastasis: Clinical Manifestations, Symptom Management, and Palliative Care. In *Handbook of Clinical Neurology*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 149, pp. 75–88, ISBN 978-0-12-811161-1.
32. Lamba, N.; Cagney, D.N.; Brigell, R.H.; Martin, A.M.; Besse, L.A.; Catalano, P.J.; Phillips, J.G.; Pashtan, I.M.; Bi, W.L.; Claus, E.B.; et al. Neurosurgical Resection and Stereotactic Radiation Versus Stereotactic Radiation Alone in Patients with a Single or Solitary Brain Metastasis. *World Neurosurg.* **2019**, *122*, e1557–e1561. [CrossRef] [PubMed]
33. Barnholtz-Sloan, J.S.; Yu, C.; Sloan, A.E.; Vengoechea, J.; Wang, M.; Dignam, J.J.; Vogelbaum, M.A.; Sperduto, P.W.; Mehta, M.P.; Machtay, M.; et al. A Nomogram for Individualized Estimation of Survival among Patients with Brain Metastasis. *Neuro-Oncology* **2012**, *14*, 910–918. [CrossRef] [PubMed]
34. Sperduto, P.W.; Chao, S.T.; Sneed, P.K.; Luo, X.; Suh, J.; Roberge, D.; Bhatt, A.; Jensen, A.W.; Brown, P.D.; Shih, H.; et al. Diagnosis-Specific Prognostic Factors, Indexes, and Treatment Outcomes for Patients with Newly Diagnosed Brain Metastases: A Multi-Institutional Analysis of 4259 Patients. *Int. J. Radiat. Oncol. Biol. Phys.* **2010**, *77*, 655–661. [CrossRef] [PubMed]
35. Zhang, J.; Jin, J.; Ai, Y.; Zhu, K.; Xiao, C.; Xie, C.; Jin, X. Computer Tomography Radiomics-Based Nomogram in the Survival Prediction for Brain Metastases From Non-Small Cell Lung Cancer Underwent Whole Brain Radiotherapy. *Front. Oncol.* **2021**, *10*, 610691. [CrossRef] [PubMed]
36. Song, P.; Li, J.; Fan, H.; Fan, L. TGDAUNet: Transformer and GCNN Based Dual-Branch Attention UNet for Medical Image Segmentation. *Comput. Biol. Med.* **2023**, *167*, 107583. [CrossRef] [PubMed]
37. Xie, H.; Zeng, X.; Lei, H.; Du, J.; Wang, J.; Zhang, G.; Cao, J.; Wang, T.; Lei, B. Cross-Attention Multi-Branch Network for Fundus Diseases Classification Using SLO Images. *Med. Image Anal.* **2021**, *71*, 102031. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.