DIVER-Surv: Diffusion and Virtual-Node Graph Fusion for Cross-Cohort Glioma Survival

Minji Cho[1], Sinyoung Ra[1], Jiwon Chung[2], and Hyunjin Park[1,2(✉)]

[1] Department of Artificial Intelligence,
Sungkyunkwan University, Suwon 16419, South Korea
mjcho246@gmail.com
[2] Department of Electrical and Computer Engineering,
Sungkyunkwan University, Suwon 16419, South Korea
hyunjinp@skku.edu

Abstract. Accurate survival prediction for adult diffuse gliomas requires robust integration of multi-parametric MRI (mpMRI) and heterogeneous clinical variables across institutions. We propose a multimodal survival framework that combines diffusion-guided imaging representation learning, language-based clinical encoding, and interaction-aware fusion. For imaging, we learn robust representations from standard mpMRI sequences using dual-branch 3D diffusion-based self-supervision. Structured clinical variables are summarized into short patient-level text descriptions and encoded with a frozen biomedical language model, reducing the need for manual covariate harmonization across cohorts. Multimodal fusion is performed using an attention-based graph with a learnable virtual node to capture cross-modal interactions. The model is optimized end-to-end with a Cox proportional hazards loss, diffusion denoising loss, and a stop-gradient consistency regularizer. We evaluate our method on three public cohorts (UCSF-PDGM, UPENN-GBM, RHUH-GBM) under cross-cohort protocols. Our method consistently yields significant risk-group separation across internal and external tests, including the difficult small-sized RHUH cohort. Our code is available at https://github.com/minjicho246/DIVER-Surv.

Keywords: Multimodal survival analysis · Diffusion-guided representation learning · Virtual-node attention graph.

1 Introduction

Adult diffuse gliomas exhibit substantial heterogeneity in clinical outcomes, with glioblastoma representing the most aggressive subtype and a median overall survival of approximately 12–15 months [12]. This critical clinical landscape necessitates the development of advanced prognostic frameworks capable of delivering precise survival forecasts to support tailored therapeutic strategies [5,18]. Prognosis in diffuse gliomas is driven by a multifaceted set of predictors. Beyond basic patient demographics like age, survival is heavily influenced by molecular markers—such as isocitrate dehydrogenase (IDH) mutation and O6-methylguanine-DNA methyltransferase (MGMT) methylation status—which serve as primary

indicators of therapeutic sensitivity [10]. Moreover, multi-parametric MRI (mpMRI) provides complementary structural and phenotypic insights into tumor burden and progression [24,23]. Cox Proportional Hazards (CPH) model [7] and Random Survival Forests (RSF) [9] are widely used baselines due to their reliability and interpretability using multimodal inputs [29]. But they often rely on pre-defined, handcrafted features [17,19]. Consequently, these models may face challenges in fully capturing the intricate, non-linear interactions between high-dimensional mpMRI patterns and heterogeneous clinical variables without extensive feature engineering.

Deep survival models extend classical survival analysis by learning non-linear risk functions from high-dimensional inputs, with representative approaches such as DeepSurv [11] and DeepHit [14]. Yet robust multimodal survival prediction in neuro-oncology remains challenging in cross-institutional settings due to three limitations. First, discriminative imaging encoders trained on limited cohorts can be sensitive to acquisition variability (scanner/protocol), leading to unstable risk estimates under domain shift [30,1]. Second, clinical variables are heterogeneous across institutions with partial overlap and inconsistent naming, so prior work often requires manual harmonization or restricts inputs to a predefined covariate set, potentially discarding informative cohort-specific variables [21,25,20]. Third, multimodal fusion is commonly implemented via early/late concatenation, limiting the ability to capture cross-modal interactions between clinical context and complex imaging patterns that jointly drive prognosis [13,27,16].

These gaps motivate a framework that (i) learns acquisition-robust imaging features beyond purely discriminative supervision, (ii) incorporates heterogeneous clinical variables without manual alignment, and (iii) models cross-modal dependencies during fusion. Accordingly, we propose **DIVER-Surv** (***D**iffusion-guided **I**nteraction-aware **V**irtual-node **E**mbedding **R**efinement for Cross-Cohort Glioma **Surv**ival*), a multimodal survival framework that regularizes the imaging encoder using a denoising diffusion probabilistic model (DDPM)-style [8] denoising objective, converts structured clinical variables into natural-language patient summaries encoded by a frozen biomedical pre-trained language model (PLM) [15], and performs fusion via attention-driven message passing [28] on an implicit modality graph with a learnable virtual node. Our contributions are threefold: (i) diffusion-guided 3D imaging embeddings for improved robustness across sites; (ii) LLM-driven clinical text representations that alleviate manual covariate harmonization; and (iii) an attention-driven virtual-node graph for interaction-aware multimodal integration.

2 Method

The proposed framework consists of three stages aligned with our core contributions for robust multimodal survival analysis (Fig. 1). First, we obtain clinical text embeddings from structured clinical variables. Second, a dual-branch diffusion-guided 3D image encoder processes mpMRI. Two stochastic perturbations are used during training, together with a stop-gradient consistency regu-

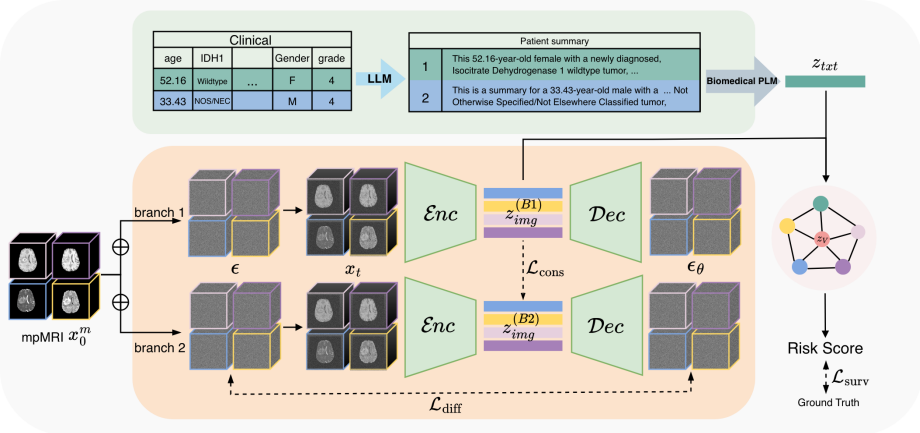

Fig. 1: Overview of the proposed multimodal survival framework. Structured clinical data are converted into text summaries and encoded via a biomedical PLM (z_{txt}). Multi-parametric MRI volumes (x_0^m) are processed through a dual-branch diffusion encoder under stochastic perturbations (x_t). The shared 3D U-Net is optimized using a noise-prediction loss ($\mathcal{L}_{\text{diff}}$) and a stop-gradient consistency loss ($\mathcal{L}_{\text{cons}}$). The resulting image and text embeddings are integrated via a virtual-node attention graph to predict risk scores ($\mathcal{L}_{\text{surv}}$).

larizer, to reduce embedding variance and promote acquisition-invariant representations. Finally, image and text embeddings are fused through an attention-driven virtual-node graph, and the updated virtual node is used to estimate a patient-specific risk score.

2.1 Text-Based Clinical Variable Encoding

To handle heterogeneous clinical variables and inconsistent naming across multi-institutional cohorts, we serialize each patient's observed structured data into a natural-language summary. Specifically, we format available indicators (e.g., demographics, molecular markers, and treatment- or surgery-related variables) as variable–value pairs, while omitting missing entries, using a fixed prompt template to instruct an LLM to (i) integrate the inputs into a coherent narrative, (ii) expand common abbreviations (e.g., IDH, MGMT), and (iii) avoid adding unverifiable information. The summary is then encoded using a frozen biomedical PLM to obtain a semantic clinical text embedding (z_{txt}), which is used as the clinical text node in our multimodal graph model. To prevent information leakage, survival outcomes are never provided to the LLM.

2.2 Diffusion-guided Representation Learning: Dual-branch

We employ a DDPM-style denoising objective as self-supervision to learn acquisition-robust 3D mpMRI representations, without synthesizing images at

inference. Given the four input sequences $\{x_0^m\}_{m\in\{\mathrm{T1,T1c,T2,FLAIR}\}}$, we generate two stochastic branches for each input (i.e., two independently noised views) by applying independent diffusion timesteps and noise perturbations. Each noisy branch is processed by a timestep-conditioned 3D U-Net denoiser [22,6] with shared parameters trained to predict the injected noise, which encourages learning representations that are less sensitive to inter-site variability (e.g., scanner/protocol differences). From the modality-specific bottleneck features, we obtain four image embeddings $\{z_{\mathrm{T1}}, z_{\mathrm{T1c}}, z_{\mathrm{T2}}, z_{\mathrm{FLAIR}}\}$ via projection heads; these embeddings are used as image nodes for multimodal fusion. We add a consistency term that penalizes discrepancies between embeddings from the two branches, further reducing embedding variance under mild perturbations.

2.3 Multimodal Fusion via Virtual Node Attention Graph

We fuse modality-specific image embeddings $\{z_{\mathrm{T1}}, z_{\mathrm{T1c}}, z_{\mathrm{T2}}, z_{\mathrm{FLAIR}}\}$ and the clinical text embedding z_{txt} using an implicit fully-connected modality graph with a learnable virtual node z_V. After projecting all nodes to a shared dimension d, we stack them as $\mathbf{H}^{(0)} = [z_V; z_{\mathrm{T1}}; z_{\mathrm{T1c}}; z_{\mathrm{T2}}; z_{\mathrm{FLAIR}}; z_{\mathrm{txt}}] \in \mathbb{R}^{N\times d}$ (N=6) and update them with L standard Transformer encoder layers, where self-attention acts as sample-specific message passing. For brevity, the attention update is:

$$\mathbf{H}^{(l)} = \mathrm{LN}\left(\mathbf{H}^{(l-1)} + \mathrm{softmax}\left(\frac{\mathbf{Q}^{(l)}(\mathbf{K}^{(l)})^\top}{\sqrt{d_k}}\right)\mathbf{V}^{(l)}\right), \quad l = 1, \ldots, L, \tag{1}$$

with $\mathbf{Q}^{(l)} = \mathbf{H}^{(l-1)}\mathbf{W}_Q^{(l)}$, $\mathbf{K}^{(l)} = \mathbf{H}^{(l-1)}\mathbf{W}_K^{(l)}$, and $\mathbf{V}^{(l)} = \mathbf{H}^{(l-1)}\mathbf{W}_V^{(l)}$, where softmax is applied row-wise and d_k denotes the key/query dimension (per head). The final virtual-node embedding $\mathbf{h}_V^{(L)}$ (its row in $\mathbf{H}^{(L)}$) is mapped to a risk score by a lightweight MLP. During training, we apply the same fusion module to both stochastic views $k \in \{B1, B2\}$, producing per-patient risk scores $\hat{r}_i^{(k)}$. During inference, the final risk score is efficiently computed through a single branch of the framework.

2.4 Loss Functions and Joint Optimization

We jointly optimize the diffusion image encoder, multimodal fusion module, and survival head using a composite objective that combines a Cox proportional hazards loss, a DDPM-style noise-prediction loss, and a stop-gradient consistency regularizer across two stochastic views (branches $k \in \{B1, B2\}$). We sample integer diffusion timesteps from a small range for the survival views, $t \sim \mathrm{U}\{0, \ldots, T_s\}$, where $\mathrm{U}\{a, \ldots, b\}$ denotes the discrete uniform distribution over integers $\{a, \ldots, b\}$. Each branch produces risk scores $\{\hat{r}_i^{(k)}\}$, and we compute the negative Cox partial log-likelihood:

$$\mathcal{L}_{\mathrm{surv}}^{(k)} = -\frac{1}{|\mathcal{E}|}\sum_{i\in\mathcal{E}}\left(\hat{r}_i^{(k)} - \log\sum_{j\in\mathcal{R}(i)}\exp(\hat{r}_j^{(k)})\right). \tag{2}$$

where t_i and $\delta_i \in \{0, 1\}$ denote the observed time and event indicator, $\mathcal{E} = \{i \mid \delta_i = 1\}$, and $\mathcal{R}(i) = \{j \mid t_j \geq t_i\}$. For diffusion regularization, we sample $t_k \sim \mathrm{U}\{0, \ldots, T-1\}$ and Gaussian noise ϵ_k, and minimize the MSE between injected and predicted noise. Averaging over branches yields:

$$\mathcal{L}_{\text{diff}} = \frac{1}{2} \sum_{k \in \{B1, B2\}} \mathrm{E}_{x, t_k, \epsilon_k} \left[\|\epsilon_k - \epsilon_\theta(x_{t_k}^{(k)}, t_k)\|_2^2 \right]. \tag{3}$$

To stabilize embeddings under mild perturbations, we apply a stop-gradient (sg) consistency loss:

$$\mathcal{L}_{\text{cons}} = \|z_{\text{img}}^{(B2)} - \text{sg}(z_{\text{img}}^{(B1)})\|_2^2. \tag{4}$$

where $z_{\text{img}}^{(k)}$ denotes the stacked embedding of the four modality-specific image embeddings from branch k. The overall objective is

$$\mathcal{L}_{\text{total}} = \sum_{k \in \{B1, B2\}} \mathcal{L}_{\text{surv}}^{(k)} + \lambda_{\text{diff}} \mathcal{L}_{\text{diff}} + \lambda_{\text{cons}} \mathcal{L}_{\text{cons}}, \tag{5}$$

where λ_{diff} and λ_{cons} are hyperparameters controlling the contribution of the diffusion and consistency tasks, respectively.

3 Experimental Design

3.1 Datasets and Experimental Setup

We evaluated our model on three independent glioma MRI cohorts: UCSF-PDGM [3], UPENN-GBM [2], and RHUH-GBM [4] with 3T mpMRI (T1, T1 contrast (T1c), T2, and FLAIR). UCSF-PDGM includes 494 diffuse glioma patients (501 total; excluding six redundant scans and one case with missing survival labels). The UPENN-GBM cohort consists of 671 de novo glioblastoma cases with mpMRI and overall survival annotations. RHUH-GBM contains 40 GBM patients. For temporal consistency, we used only the earliest available preoperative scan per subject across all datasets. We performed cross-cohort evaluation in two settings: (i) UPENN-GBM was split into training/validation/internal test sets (7:1:2), with UCSF-PDGM and RHUH-GBM held out for external testing; (ii) UCSF-PDGM and UPENN-GBM were swapped under the same protocol. Splits were patient-level with a fixed seed, using identical preprocessing across cohorts; external cohorts were never used for hyperparameter tuning or model selection.

All MRI scans were resized to $120 \times 120 \times 64$ and intensity-normalized to $[0, 1]$. Structured clinical variables were summarized using Gemini 2.0 Flash [26] and encoded into 768-dimensional embeddings using frozen SapBERT [15]. Image embeddings were extracted by a 3D U-Net diffusion encoder trained via a DDPM-style objective (1,000 steps; linear beta schedule, $\beta_{\text{start}} = 10^{-4}$ to $\beta_{\text{end}} = 0.02$). The model was trained for 100 epochs with Adam using decoupled learning rates (10^{-5} for the image encoder and 3×10^{-6} for the survival

module) under a cosine annealing schedule. We set $T_s = 100$ in all experiments. The model used $\lambda_{\text{diff}} = 0.2$, $\lambda_{\text{cons}} = 0.5$, and a 2-layer Transformer (8 heads, $d = 256$), trained with batch size 8, weight decay 10^{-4}, and dropout 0.2.

We report the Concordance Index (C-index) as the primary survival metric. In addition, we compute the hazard ratio (HR) and the log-rank test p-value by stratifying patients into high- and low-risk groups using the median risk score from the training split. We further estimate 95% confidence intervals (CI) using bootstrap resampling with replacement.

3.2 Comparison Methods

To evaluate the efficacy and robustness of the proposed framework, we compared its performance against several representative survival analysis methodologies.

Traditional Statistical Models: We utilized CPH [7] and RSF [9] as foundational baselines. To ensure consistency across multi-institutional cohorts with heterogeneous variables, all models were trained on the same clinical embeddings obtained from patient summaries.

Deep Discriminative Models: DeepSurv [11] and DeepHit [14] used the same 3D U-Net backbone for imaging features, while replacing the diffusion denoising objective with a standard reconstruction loss (MSE) for stable feature learning.

Ablation Study: We evaluated four variants to isolate core components: Clinical-only (clinical summaries only), Imaging-only (diffusion-guided 3D imaging only), Generative-fusion (concatenation of diffusion-guided imaging features and clinical summaries), and DIVER-Surv (proposed model). We optimize $\mathcal{L} = \mathcal{L}_{\text{surv}}^{\text{B1}} + \mathcal{L}_{\text{surv}}^{\text{B2}} + \lambda_{\text{diff}}\mathcal{L}_{\text{diff}} + \lambda_{\text{cons}}\mathcal{L}_{\text{cons}}$. For Clinical-only, we use $\mathcal{L} = \mathcal{L}_{\text{surv}}$ (i.e., $\lambda_{\text{diff}} = \lambda_{\text{cons}} = 0$); for Imaging-only, we remove the clinical branch; Generative-fusion and DIVER-Surv share the same objective and differ only in the fusion operator (concatenation vs. virtual-node graph).

4 Results and Discussion

4.1 Cross-Cohort Generalizability and Clinical Reliability

Tables 1a and 1b summarize cross-cohort survival prediction when training on UCSF-PDGM or UPENN-GBM, with internal testing on the source cohort and external testing on the remaining cohorts. Across six evaluations, our framework is the most robust, achieving the best C-index in most settings and strong worst-case performance under cohort/site shift. When trained on UCSF-PDGM, it attains 0.769 on internal UCSF (95% CI: 0.698–0.823) and generalizes to UPENN and RHUH with 0.589 (95% CI: 0.565–0.613) and 0.638 (95% CI: 0.504–0.753). When trained on UPENN-GBM, it remains stable on internal UPENN (0.644; 95% CI: 0.592–0.697) and on external UCSF and RHUH (0.626; 95% CI: 0.589–0.660, and 0.616; 95% CI: 0.475–0.740). In contrast, classical baselines degrade markedly under shift (e.g., Cox drops to 0.372 on RHUH when trained on UCSF), and RSF cannot reliably report HR on event-limited cohorts.

Table 1: Performance comparison of survival models across multi-institutional datasets. Metrics include Concordance Index (C-idx), Hazard Ratio (HR), and log-rank p-value. Best results are in bold; "–" denotes metrics (HR or p) not computed due to insufficient events ($n \leq 5$).

(a) Trained on **UCSF-PDGM**

Model	Internal UCSF			External UPENN			External RHUH		
	C-idx	HR	p	C-idx	HR	p	C-idx	HR	p
Cox	0.635	2.98	<0.01	0.507	0.95	0.52	0.372	0.58	0.03
RSF	0.715	3.20	<0.01	0.559	–	0.26	0.546	–	0.77
DeepSurv	0.706	**4.15**	<0.01	0.551	1.60	<0.01	0.619	1.46	0.58
DeepHit	0.728	3.58	<0.01	0.555	1.32	0.02	0.619	**2.09**	0.02
DIVER-Surv	**0.769**	2.55	**<0.01**	**0.589**	**1.73**	**0.02**	**0.638**	1.70	**0.01**

(b) Trained on **UPENN-GBM**

Model	Internal UPENN			External UCSF			External RHUH		
	C-idx	HR	p	C-idx	HR	p	C-idx	HR	p
Cox	0.541	1.47	0.10	0.459	0.96	0.65	0.498	1.47	0.45
RSF	0.598	1.78	<0.01	0.497	0.88	0.72	0.589	–	–
DeepSurv	0.639	**1.88**	<0.01	0.600	1.85	<0.01	0.602	**1.82**	0.22
DeepHit	0.605	1.70	0.15	**0.651**	**2.12**	**<0.01**	0.593	1.04	0.97
DIVER-Surv	**0.644**	1.51	**0.01**	0.626	1.87	**<0.01**	**0.616**	1.75	**0.02**

Our method also yields more clinically reliable stratification: median-risk grouping is significant in every evaluation ($p \leq 0.02$), including RHUH. Under UPENN training, DeepHit is higher on UCSF (0.651 vs. 0.626), but Fig. 2 shows similarly clear Kaplan–Meier separation for both methods on UCSF, suggesting a limited practical difference. On RHUH, Fig. 2 shows the key gap: DeepHit fails to stratify ($p = 0.97$, HR ≈ 1), whereas our model remains significant ($p = 0.02$, HR 1.75 with 95% CI: 1.08–2.91), supporting robust cross-cohort reliability.

4.2 Ablation: Effectiveness of Diffusion Regularization and Graph Fusion

Table 2 (UCSF-trained) shows that each added component contributes to improved performance and cross-cohort robustness. Clinical-only achieves strong internal UCSF performance (C-index 0.715, $p < 0.01$) but drops on external RHUH (0.575, $p = 0.04$), whereas adding diffusion-guided imaging (Imaging-only) improves external generalization (RHUH 0.615, $p = 0.03$), supporting acquisition-robust representation learning. Simply concatenating modalities (Generative-fusion) increases internal discrimination (UCSF 0.743, $p < 0.01$) but does not consistently translate to stronger external concordance (RHUH 0.565, $p = 0.03$), suggesting that naive aggregation may miss cross-modal dependencies under shift. In contrast, the Proposed model, which adds interaction-aware

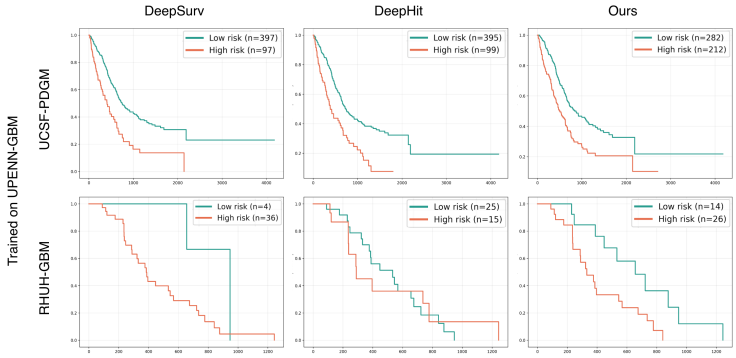

Fig. 2: Kaplan–Meier curves on external cohorts for models trained on UPENN-GBM. Patients are stratified into high/low risk by the median training-split risk score (group sizes shown).

Table 2: Ablation on UCSF-trained setting (Internal UCSF, External RHUH). Checkmarks (✓) indicate the inclusion of Clinical (Clin.), Imaging (Img.), Diffusion (Diff.), and Graph fusion (Graph).

Configuration	Components				Internal UCSF			External RHUH		
	Clin.	Img.	Diff.	Graph	C-idx	HR	p	C-idx	HR	p
(1) Clinical-only	✓				0.715	**3.87**	<0.01	0.575	1.59	0.04
(2) Imaging-only		✓	✓		0.653	2.74	<0.01	0.615	**1.74**	0.03
(3) Generative-fusion	✓	✓	✓		0.743	3.13	<0.01	0.565	1.59	0.03
(4) DIVER-Surv	✓	✓	✓	✓	**0.769**	2.55	**<0.01**	**0.638**	1.70	**0.01**

virtual-node graph fusion on top of diffusion-guided features, achieves the best results on both internal and external evaluation (UCSF 0.769, $p < 0.01$; RHUH 0.638, $p = 0.01$). The same trend is observed on the additional external UPENN test cohort (UCSF-trained; not shown in Table 2): C-index improves from 0.541 (HR 1.44, $p = 0.01$; Clinical-only) to 0.563 (HR 1.15, $p = 0.03$; Imaging-only) and 0.567 (HR 1.49, $p < 0.01$; Generative-fusion), with the largest gain achieved by DIVER-Surv (0.589; HR 1.73; $p = 0.02$).

5 Conclusion

We presented DIVER-Surv, a multimodal survival framework that combines diffusion-guided 3D mpMRI representations, LLM-based clinical text encoding, and virtual-node graph fusion. Across UCSF-PDGM, UPENN-GBM, and RHUH-GBM under cross-cohort evaluation, DIVER-Surv showed robust discrimination and consistent risk-group separation under institutional shift. This study has several limitations. First, the benefit of language-based clinical encoding over fully standardized tabular encoders was not directly isolated. Second, RHUH-GBM was small ($n = 40$), limiting statistical power; therefore, its results

should be interpreted as an external stress test rather than definitive evidence of broad generalizability. Future work will validate DIVER-Surv on larger multi-center cohorts and investigate uncertainty-aware calibration to further improve clinical reliability.

Acknowledgments. This study was supported by the SMC-SKKU Future Convergence Research Program Grant (SMO125123). It was also supported by the AI Graduate School Support Program (Sungkyunkwan University) (RS-2019-II190421), the National Research Foundation of Korea (RS-2024-00408040), the ICT Creative Consilience Program Grant (IITP-2026-RS-2020-II201821), and the AI Computing Infrastructure Enhancement (GPU Rental Support) User Support Program (RQT-25-090077), all funded by the Ministry of Science and ICT (MSIT), Republic of Korea.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abbasi, S., Lan, H., Choupan, J., Sheikh-Bahaei, N., Pandey, G., Varghese, B.: Deep learning for the harmonization of structural mri scans: a survey. BioMedical Engineering OnLine **23**(1), 90 (2024)
2. Bakas, S., Sako, C., Akbari, H., Bilello, M., Sotiras, A., Shukla, G., Rudie, J.D., Santamaría, N.F., Kazerooni, A.F., Pati, S., et al.: The university of pennsylvania glioblastoma (upenn-gbm) cohort: advanced mri, clinical, genomics, & radiomics. Scientific data **9**(1), 453 (2022)
3. Calabrese, E., Villanueva-Meyer, J.E., Rudie, J.D., Rauschecker, A.M., Baid, U., Bakas, S., Cha, S., Mongan, J.T., Hess, C.P.: The university of california san francisco preoperative diffuse glioma mri dataset. Radiology: Artificial Intelligence **4**(6), e220058 (2022)
4. Cepeda, S., García-García, S., Arrese, I., Herrero, F., Escudero, T., Zamora, T., Sarabia, R.: The río hortega university hospital glioblastoma dataset: A comprehensive collection of preoperative, early postoperative and recurrence mri scans (rhuh-gbm). Data in brief **50**, 109617 (2023)
5. Cho, H.h., Lee, S.h., Kim, J., Park, H.: Classification of the glioma grading using radiomics analysis. PeerJ **6**, e5982 (2018)
6. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
7. Cox, D.R.: Regression models and life-tables. Journal of the royal statistical society: Series B (methodological) **34**(2), 187–202 (1972)
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
9. Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S.: Random survival forests (2008)

10. Jacome, M.A., Wu, Q., Piña, Y., Etame, A.B.: Evolution of molecular biomarkers and precision molecular therapeutic strategies in glioblastoma. Cancers **16**(21), 3635 (2024)
11. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC medical research methodology **18**(1), 24 (2018)
12. Królikowska, K., Błaszczak, K., Ławicki, S., Zajkowska, M., Gudowska-Sawczuk, M.: Glioblastoma—a contemporary overview of epidemiology, classification, pathogenesis, diagnosis, and treatment: A review article. International Journal of Molecular Sciences **26**(24), 12162 (2025)
13. Krones, F., Marikkar, U., Parsons, G., Szmul, A., Mahdi, A.: Review of multimodal machine learning approaches in healthcare. Information Fusion **114**, 102690 (2025)
14. Lee, C., Zame, W., Yoon, J., Van Der Schaar, M.: Deephit: A deep learning approach to survival analysis with competing risks. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
15. Liu, F., Shareghi, E., Meng, Z., Basaldella, M., Collier, N.: Self-alignment pretraining for biomedical entity representations. In: Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies. pp. 4228–4238 (2021)
16. Na, I., Noh, J.J., Kim, C.K., Lee, J.W., Park, H.: Combined radiomics-clinical model to predict platinum-sensitivity in advanced high-grade serous ovarian carcinoma using multimodal mri. Frontiers in Oncology **14**, 1341228 (2024)
17. O'donnell, A., Cronin, M., Moghaddam, S., Wolsztynski, E.: A systematic review on machine learning techniques for survival analysis in cancer. Cancer Medicine **14**(22), e71375 (2025)
18. Oh, S., Kim, S., Kim, J.P., Seo, S.W., Park, B.y., Park, H.: Association of apoc1 with cortical atrophy during conversion to alzheimer's disease. GeroScience **47**(6), 6665–6682 (2025)
19. Park, H., Lim, Y., Ko, E.S., Cho, H.h., Lee, J.E., Han, B.K., Ko, E.Y., Choi, J.S., Park, K.W.: Radiomics signature on magnetic resonance imaging: association with disease-free survival in patients with invasive breast cancer. Clinical Cancer Research **24**(19), 4705–4714 (2018)
20. Ra, S., Kim, J., Na, I., Ko, E.S., Park, H.: Enhancing radiomics features via a large language model for classifying benign and malignant breast tumors in mammography. Computer Methods and Programs in Biomedicine **265**, 108765 (2025)
21. Rockenschaub, P., Hilbert, A., Kossen, T., Elbers, P., Von Dincklage, F., Madai, V.I., Frey, D.: The impact of multi-institution datasets on the generalizability of machine learning prediction models in the icu. Critical Care Medicine **52**(11), 1710–1721 (2024)
22. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
23. Sanchez, I., Rahman, R.: Radiogenomics as an integrated approach to glioblastoma precision medicine. Current Oncology Reports **26**(10), 1213–1222 (2024)
24. Singh, G., Manjila, S., Sakla, N., True, A., Wardeh, A.H., Beig, N., Vaysberg, A., Matthews, J., Prasanna, P., Spektor, V.: Radiomics and radiogenomics in gliomas: a contemporary update. British journal of cancer **125**(5), 641–657 (2021)
25. Suleman, M.U., Mursaleen, M., Khalil, U., Saboor, A., Bilal, M., Khan, S.A., Subhani, M.A., Hussnain, M.A., Tabassum, S.N., Tahir, M.: Assessing the generalizability of artificial intelligence in radiology: a systematic review of performance

across different clinical settings. Annals of Medicine and Surgery **87**(12), 8803–8811 (2025)
26. Team, G., Georgiev, P., Lei, V.I., Burnell, R., Bai, L., Gulati, A., Tanzer, G., Vincent, D., Pan, Z., Wang, S., et al.: Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. arXiv preprint arXiv:2403.05530 (2024)
27. Teoh, J.R., Dong, J., Zuo, X., Lai, K.W., Hasikin, K., Wu, X.: Advancing healthcare through multimodal data fusion: a comprehensive review of techniques and applications. PeerJ Computer Science **10**, e2298 (2024)
28. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
29. Wiegrebe, S., Kopper, P., Sonabend, R., Bischl, B., Bender, A.: Deep learning for survival analysis: a review. Artificial Intelligence Review **57**(3), 65 (2024)
30. Yoon, J.S., Oh, K., Shin, Y., Mazurowski, M.A., Suk, H.I.: Domain generalization for medical image analysis: A review. Proceedings of the IEEE **112**(10), 1583–1609 (2024)