Dynamic Modeling of Brain Metastasis Response with Multitask Temporal Learning

Vincent Andrearczyk[1,2] Lorenz Achim Kuhn[1], Jonas Richiardi[3,4], Andreas F. Hottinger[4,5], Luis Schiappacasse[4,6], Vincent Dunet[7], and Adrien Depeursinge[1,2,4]

[1] Institute of Informatics, HES-SO Valais-Wallis
vincent.andrearczyk@hevs.ch
[2] Department of Nuclear Medicine and Molecular Imaging, Lausanne University Hospital (CHUV)
[3] Department of Medical Radiology, Service of Diagnostic and Interventional Radiology, Neuroradiology Unit, CHUV and UNIL
[4] Lundin Family Brain Tumor Research Centre, Departments of Oncology & Clinical Neurosciences, CHUV and UNIL
[5] Department of Oncology, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL)
[6] Department of Oncology, Service of Radiation Oncology, CHUV and UNIL
[7] Department of Medical Radiology, Service of Diagnostic and Interventional Radiology, Neuroradiology Unit, CHUV and UNIL

Abstract. Patients with brain metastases (BMs) treated using stereotactic radiosurgery (SRS) undergo longitudinal MRI to monitor response, yet outcomes vary substantially between lesions across and within patients, reflecting biological heterogeneity and differences in radiosensitivity. Modeling response is challenging due to heterogeneous temporal dynamics and irregular follow-up schedules, with varying numbers of scans per patient. Conventional machine learning relying on fixed, regularly sampled inputs is therefore limited to predict BM response to SRS. We present a large retrospective post-SRS dataset of 4,766 lesions from 568 patients, monitored with contrast-enhanced T1-weighted MPRAGE and T2-weighted MRI over a mean/median duration of 316/214 days. We formulate lesion-level response prediction as (i) future complete response (CR) classification at flexible temporal horizons and (ii) survival analysis with time to complete response (TCR) as the event (53% event rate). We propose a time-aware framework for irregular longitudinal data based on gated recurrent units (GRU). Absolute time since SRS is explicitly encoded and integrated with dynamic lesion-level imaging features and patient-level static covariates. The model processes variable-length input sequences and supports flexible prediction horizons within a unified architecture. Our approach outperforms MLP and XGBoost baselines that require fixed input representations, achieving a C-index of up to 0.77 for TCR prediction and AUCs between 0.74 and 0.90 for CR classification. This work establishes a scalable framework for personalized longitudinal response modeling in SRS-treated BMs. Code will be released upon acceptance.

Keywords: Longitudinal Modeling · Brain Metastases · Survival

1 Introduction

Brain metastases (BM), the most common intracranial tumors in adults, are a major source of morbidity and mortality. Stereotactic radiosurgery (SRS) is a standard treatment for BM, offering high local control while sparing healthy tissue. Following SRS, patients undergo longitudinal MRI to monitor lesion response and detect progression. Despite standardized imaging, treatment adaptation remains largely reactive, relying on visual assessment rather than lesion-level risk modeling. Longitudinal post-SRS imaging enables lesion-level response prediction. Individual metastases can follow heterogeneous trajectories, influenced by tumor biology, treatment, and patient factors. Accurate prediction could enable earlier treatment adaptation, such as intensified monitoring, retreatment, or alternative therapies. Modeling lesion-level outcomes is challenging due to irregular follow-ups, varying observation counts, and the need to integrate lesion- and patient-level information. Clinically relevant prediction horizons also vary from the next follow-up to one year or more.

Machine learning (ML) has increasingly been applied to predict patient outcomes after SRS. Early studies used radiomic features from pre-treatment MRI with clinical variables to predict local control or progression, showing moderate performance (AUC $\approx$ 0.71–0.77) [10]. Later work integrated multimodal MRI and clinical factors, reporting high accuracy (AUC $\approx$ 0.93–0.95) [11], and systematic reviews confirm the promise of ML in predicting treatment response and local failure in SRS-treated BMs, identifying imaging, dosimetric, and clinical biomarkers, while underscoring the need for larger cohorts and consistent outcome definitions [9, 13]. Large-scale challenges such as BRATS-METS [16] have accelerated development and benchmarking for BM segmentation and analysis.

Beyond static baseline prediction, recent work has exploited longitudinal imaging through follow-up curation, lesion trajectory clustering, and graph ML for fixed-horizon response prediction [15]. Cho et al. [6] used a convolutional-GRU architecture to model longitudinal MRI and radiomic features for lesion-level treatment response prediction after SRS, demonstrating the value of temporal information for response modeling. Survival prediction models using clinical variables via hybrid ML have shown potential for estimating overall survival after SRS [17]. However, these approaches typically rely on fixed horizons or static feature aggregation and do not handle irregular follow-ups or dynamic temporal dependencies. More generally, temporal deep learning methods have emerged as a powerful approach for modeling longitudinal clinical and imaging data by capturing disease evolution across multiple follow-up visits while accommodating variable-length sequences and irregular sampling. Recurrent neural networks such as gated recurrent units (GRUs) [5] provide an efficient framework for learning temporal dependencies in longitudinal medical data [7], whereas transformer-based architectures have recently demonstrated strong performance for large-scale sequential modeling by leveraging attention mechanisms [14, 18]. However,

existing approaches typically focus on patient-level outcomes and do not support flexible prediction horizons or jointly model longitudinal lesion response and time-to-event outcomes.

In this work, we introduce a large post-SRS BM MRI cohort of 4,766 lesions from 568 patients. We formulate lesion response prediction as both future outcome classification at flexible temporal horizons and survival analysis for time to complete response (CR), where each lesion is a distinct observation. The contribution lies in the unified lesion-level longitudinal formulation and flexible horizon-conditioned framework, rather than in architectural novelty. We propose a GRU-based temporal framework that encodes irregular time gaps, models variable-length lesion trajectories, and integrates patient covariates. By jointly modeling longitudinal dynamics and time-to-event risk at the lesion level via multitask learning, our approach enables personalized response prediction and data-driven follow-up planning.

2 Methods

2.1 Dataset

This retrospective study includes patients treated for BMs at the Lausanne University Hospital (CHUV) with contrast-enhanced T1-weighted (T1w) MPRAGE and T2-weighted (T2w) MRI acquired prior to SRS (t_0) and approximately every two months thereafter as part of clinical follow-up. The study was approved by the Research Ethics Committee of Vaud Canton, Switzerland (No. 2024-00100).

All data, extended from [15], were derived from a private institutional clinical database and are not publicly available due to privacy and regulatory constraints. Before applying the automated preprocessing pipeline (described below), the cohort comprised 1,134 patients, with an average of 13.6 MRI studies that spanned on average 874 days (min. 0, max. 6358) per patient. About one-third of scans originated from external imaging centers, increasing acquisition diversity.

2.2 Automatic Preprocessing and Feature Extraction

Raw DICOM files were filtered and organized by selecting relevant images and ROIs (RTSTRUCTs from radiotherapy) based on metadata, then converted into a BIDS-like structure. ROIs were automatically matched to the corresponding MRI in the patient timeline, and longitudinal series were cleaned by removing large gaps between follow-ups ($\geq$4 months), redundancies, and pre-treatment planning scans. Spatial alignment was performed by affine registration of baseline and subsequent MRIs using ANTs [3], ensuring consistent lesion localization across timepoints. UNet-based re-segmentation [1] propagated delineations from the initial treatment MRI to all follow-up MRIs, achieving lesion correspondence over time. Patients with brain primary cancer or lesions without follow-up were excluded. After preprocessing and quality control, this pipeline produced a curated dataset of 4,766 lesions from 568 patients with complete longitudinal

data, providing a unique large-scale fully automated resource for longitudinal BM imaging studies. Mean and median lesion follow-up durations were 316 and 214 days, respectively (442 and 280 days at the patient level as lesions may not appear simultaneously). The observed lesion-level CR rate was 47.6%, with a median initial lesion volume of 40 mm^3 (IQR: 134). Each lesion was characterized at every timepoint by 200 dynamic radiomic features extracted from T1w and T2w imaging (93 intensity and texture features from both, and 14 shape features). Approximately 20% of T2w images were unavailable, for which T2w-derived features were set to zero prior to model input. These radiomic features are used by all models including baselines and the proposed temporal GRU. In addition, six static patient-level variables $\mathbf{s}$ were incorporated: primary cancer type, brain region, sex, weight, height, and age at onset. Feature extraction and preprocessing were strictly performed in a time-causal manner, ensuring that for each timepoint only data acquired up to that visit was used, with no access to future scans or outcomes.

For outcome modeling, lesion-level RANO-BM response [2] was automatically derived from longitudinal lesion volumes at each follow-up: complete response (CR), disappearance of the lesion; partial response (PR), >65.7% decrease from baseline; progressive disease (PD), >72.8% increase from nadir; and stable disease (SD), neither PR nor PD.

2.3 Models and Training

Baseline models We compare against multilayer perceptron (MLP) and gradient-boosted decision tree (XGB) [4] baselines. Because these models require fixed-length inputs, they cannot handle variable-length longitudinal sequences. For a given lesion, dynamic radiomic features from $n \in [1, 5]$ timepoints are concatenated and combined with static patient covariates $\mathbf{s}$ to form a fixed-length feature vector. Lesions with fewer than n timepoints are discarded. Separate models are trained for each input length n, for each task-specific prediction setting, and for each horizon (classification task).

For survival modeling formulated as time to complete response (TCR), the MLP and XGB baselines predict lesion-specific risk from the concatenated multi-temporal features. The MLP consists of two fully connected layers with ReLU activations, batch normalization and dropout, followed by a linear layer producing a scalar risk score, and is trained using the Cox partial likelihood (learning rate, $lr = 1e-3$, early stop on validation C-index with patience = 5). The XGB model (1000 trees, $lr = 0.05$) is trained with a Cox proportional hazards loss, directly optimizing the Cox negative log-likelihood and outputting a scalar risk.

For response prediction, the same MLP and XGB architectures are used to predict RANO-BM response categories at a fixed future horizon ($\Delta t \in [6, 9, 12, 18]$ months). In addition to the input length n, separate models are trained for each prediction horizon Δt, and lesions with insufficient follow-up are excluded. The MLP outputs logits over four response classes (complete response, CR, partial response, PR, stable disease, SD and progressive disease, PD) and is trained

using focal loss, while XGB performs multiclass classification trained with cross-entropy loss.

Temporal GRU The architecture is illustrated in Fig. 1. Full implementation details are available on GitHub. Given dynamic radiomic features extracted from B lesions $\mathbf{X} \in \mathbb{R}^{B\times T\times D_{\text{dyn}}}$ observed at irregular follow-up times $\mathbf{t} \in \mathbb{R}^{B\times T}$ ($T = 12$), padding mask $\mathbf{M}$, and static patient covariates $\mathbf{s} \in \mathbb{R}^{B\times D_{\text{stat}}}$, follow-up times are first log-transformed:

$$\tilde{\mathbf{t}} = \log(1 + \mathbf{t}),$$

and concatenated with the dynamic lesion features before projection:

$$\mathbf{Z}_0 = [\mathbf{X}, \tilde{\mathbf{t}}]\mathbf{W}_{\text{dyn}} + \mathbf{b}_{\text{dyn}} \in \mathbb{R}^{B\times T\times d},$$

where $d = 64$.

Longitudinal lesion evolution is modeled using a GRU encoder operating on variable-length sequences. The temporal encoder is interchangeable; a transformer implementation was also evaluated but performed worse. Padded observations are ignored via packed sequence processing, and the final hidden state is used as the temporal representation:

$$\mathbf{h} = \text{GRU}(\mathbf{Z}_0, \mathbf{M}).$$

Static patient covariates are concatenated with the temporal representation:

$$\mathbf{z} = [\mathbf{h}; \mathbf{s}],$$

and passed to two prediction heads.

A survival head produces a lesion-specific risk score:

$$r = f_{\text{risk}}(\mathbf{z}),$$

optimized using the Cox proportional hazards loss [8].

For future response prediction at horizon Δt (in months), the representation is conditioned on the target time:

$$\mathbf{d}_{\Delta t} = \text{MLP}_t(\log(1 + \Delta t)),$$

which is projected and added to the fused representation:

$$\mathbf{z}_{\Delta t} = \mathbf{z} + \mathbf{W}_{\Delta t}\mathbf{d}_{\Delta t}.$$

The final representation is used to predict lesion response (CR/non-CR) at horizon Δt, optimized with a focal loss.

While the primary clinical endpoint of interest is binary CR/non-CR prediction, we supervise training using four-class RANO response categories to leverage finer-grained intermediate response information.

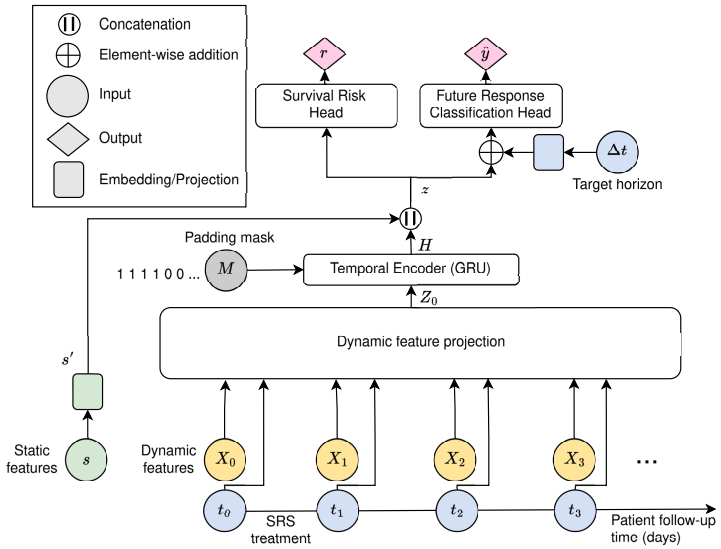
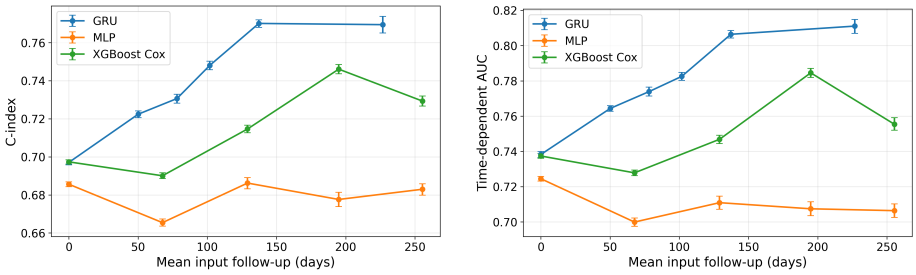


Fig. 1: Overview of the proposed temporal GRU model for variable-length longitudinal lesion follow-up. The GRU encoder models lesion evolution from irregular follow-up observations and jointly predicts time-to-complete response (TCR) risk and future lesion response at flexible prediction horizons.

Training Protocol Models are trained using mini-batch ($B = 128$) gradient descent with the AdamW optimizer. Categorical encodings, feature normalization and selection are fitted on the training split. Low-variance (< 0.01) and highly correlated (> 0.90) features are removed and the resulting feature subset is used in the validation and test sets. Regularization is achieved through dropout on dynamic and static features, as well as temporal truncation.

Training is performed using patient-level 5-fold cross-validation, ensuring that lesions from the same patient do not appear across splits. Early stopping (patience $= 5$, $lr = 1e-4$) is applied using 80/20 internal train–validation split within each training fold of the cross-validation. Average training runtime per fold on a 32GB V100 is 682s for the temporal model (45k trainable parameters), 181s for an MLP, and 58s for XGBoost (both single-timepoint survival models).

Evaluation Survival performance (TCR) is evaluated using Harrell's concordance index (C-index) and time-dependent AUC. Future response prediction is assessed using binary CR/non-CR ROC AUC. Similar trends were observed with Precision-Recall AUC and Matthews correlation coefficient, which are not reported for brevity. Predictions at a given horizon Δt are evaluated only for lesions still under observation at that time. Statistical significance is evaluated using a nested bootstrap, resampling patients within each fold and aggregating fold-level metrics to estimate the mean and its 95% confidence interval (CI).



(a) C-index (b) Mean Cumulative/Dynamic AUC

Fig. 2: TCR survival performance with CIs.

Statistical tests, namely two-sided paired non-parametric bootstrap hypothesis tests, are performed only between models sharing the same architecture (temporal models in ablation studies). Other statistical comparisons between temporal and baseline models are performed with CIs, where hypothesis testing with p-values is not conducted due to differences in input sequence lengths.

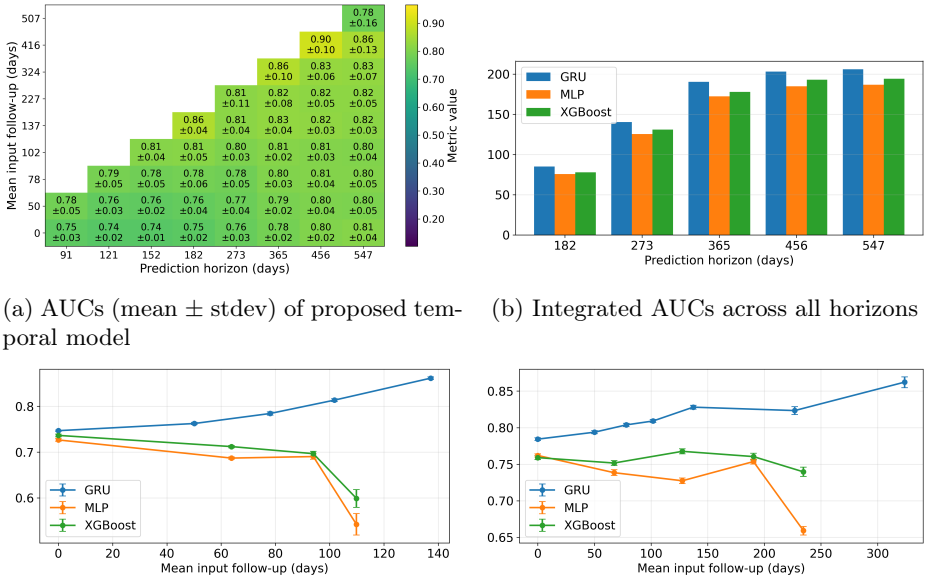
To summarize global classification performance, we integrate the AUC across different numbers of input follow-ups for a given prediction horizon (Figure 3b). For fair comparison, the integration range is identical across models, from 0 to the minimum of their maximum input lengths.

Evaluation is challenging due to comparisons across models, horizons, and input lengths. For classification, longer horizons reduce the number of eligible lesions, while longer inputs exclude lesions that already achieved CR. While longer inputs provide more context, they also reduce the proportion of future CR events and require the model to distinguish lesions that persist over time. Survival analysis is simpler in that C-index does not require multiple horizon evaluation, yet input lengths also vary together with the number of lesions available. Comparing baseline and temporal models adds further complexity, as baselines use a fixed number of inputs, whereas the temporal model can leverage all inputs up to a given follow-up (e.g. 3–9 months). In both cases, we report the mean number of input follow-ups for comparison.

3 Results

3.1 Complete Response Risk Prediction

TCR risk prediction (survival task) results are shown in Fig. 2. The baselines and the temporal model have different mean input follow-ups because the former must be evaluated at fixed number of timepoints, which are of varying days, while the latter is evaluated at different numbers of timepoints, using all follow-ups $\leq m$, $m \in [0, 3, 4, 5, 6, 9]$ months.



(a) AUCs (mean ± stdev) of proposed temporal model

(b) Integrated AUCs across all horizons

(c) AUCs horizon 182 days

(d) AUCs horizon 365 days

Fig. 3: Classification results including (a) CR AUCs of the temporal model at multiple horizons and with varying follow-up lengths; (b) Integrated CR AUCs for all horizons for the 3 models; and CR AUCs at horizons (c) 182 and (d) 365 days.

3.2 Horizon Future Response Classification

CR classification results are reported in Fig. 3. All predictions of the proposed temporal model are obtained with a single model, using all available types of dynamic and static features. Individual MLP and XGB models are trained, with the same features, for each horizon/input-length combination.

3.3 Ablation Studies

Effect of multitask learning. We evaluated the benefit of jointly modeling survival and future response. Multitask learning improved the average C-index by +1.98 percentage points (pp) as compared to using the survival head only ($p < .001$). Conversely, using the classification head showed that multitask learning also boosts CR classification performance by +3.17pp ($p < .001$) on average. These results confirm that combining both tasks benefits each objective.

Impact of input features. For the temporal model, the best performance was achieved using all dynamic and static features. Compared to using only lesion volume and only radiomic features, the C-index improved by +8.99pp ($p < .001$) and +0.48pp ($p < .001$). CR classification AUCs also increased, with +8.98pp ($p < .001$) and +0.43pp ($p < .001$) on average across all horizons versus

volume-only and radiomics-only, respectively. This indicates that the proposed model effectively leverages the full set of dynamic and static covariates, while simpler baseline models benefit less from additional features (best MLP C-index with volume only). Most of the informative dynamic signal is captured by T1w features. Using only T1w features resulted in a marginal improvement of +0.10pp in C-index and +0.15pp in CR AUC, whereas using only T2w features led to a decrease of -1.59pp in C-index and -1.45pp in CR AUC.

4 Discussions and Conclusion

We presented a large-scale, longitudinal dataset of 4,766 brain metastases from 568 patients, curated via an automated pipeline and featuring long-term follow-up. This cohort is unique in both size and temporal coverage, enabling robust evaluation of lesion-level response prediction. A limitation of this study is the retrospective, single-center nature of the dataset, without external validation, although the cohort includes MRI examinations acquired at multiple referring imaging centers, providing substantial imaging heterogeneity.

Our time-aware framework handles variable-length inputs, flexible prediction horizons, and joint TCR survival and response classification. In contrast, baseline MLP and XGBoost models are limited to fixed input lengths, fixed prediction horizons (e.g. 1 year) and single tasks. The proposed model consistently outperformed baselines, achieving the highest C-index (up to 0.77) and improving CR classification across all horizons. Although instantiated here with a GRU, the framework can accommodate other temporal encoders. A transformer variant was also evaluated but achieved lower performance.

The proposed model performed consistently across follow-up durations and horizons (C-indices > 0.69, AUCs > 0.74). Predictions improve with more timepoints, though this is partly obscured by evaluations limited to lesions with long follow-up and no CR events within the input period. Small lesions at baseline were associated with more CR events (Spearman $\rho = -0.34$, $p < .001$) and longer time to CR ($\rho = 0.28$, $p < 0.001$). However, volume had minimal impact on predictive performance (ΔC-index Q4–Q1=+0.009, average across input lengths). Primary cancer type showed stronger outcome heterogeneity (Chi-square $p < 0.001$; Kruskal–Wallis $p < 0.001$), with subtype-specific C-indices ranging from 0.768 (kidney) to 0.804 (melanoma) (ΔC-index=0.036).

These results demonstrate the potential clinical value of lesion-level, horizon-flexible prediction. By identifying lesions at higher risk of persistent disease or delayed response, the model could inform more proactive monitoring, retreatment planning, or individualized follow-up schedules. Future work will explore additional features including imaging features from MRI foundation-models [12] and additional treatment information. This analysis isolates imaging-derived and temporal signal from clinical and treatment context; the incorporation of systemic therapy, dosimetric, and histological variables is left to future clinically-oriented work. Pretraining strategies such as predicting features at random timepoints, as well as regressing lesion volume may further improve performance.

Overall, our study establishes a scalable, multitask framework for personalized, lesion-level longitudinal response modeling in brain metastases, with both methodological and clinical relevance.

Acknowledgments. This work was partially funded by the Swiss Cancer Research foundation with the project TARGET (KFS-5549-02-2022-R), a donation from the Lundin Cancer Fund, and the Swiss National Science Foundation (SNSF) with projects 205320_219430 and CR00-5-242825_10.007.857.

Disclosure of Interests. No competing interests to declare.

References

1. Andrearczyk, V., Schiappacasse, L., Abler, D., Wodzinski, M., Hottinger, A., Raccaud, M., Bourhis, J., Prior, J.O., Dunet, V., Depeursinge, A.: Automatic detection and multi-component segmentation of brain metastases in longitudinal MRI. Scientific reports **14**(1), 31603 (2024)
2. Andrearczyk, V., Schiappacasse, L., Raccaud, M., Bourhis, J., Prior, J.O., Cuendet, M.A., Hottinger, A.F., Dunet, V., Depeursinge, A.: The value of ai for assessing longitudinal brain metastases treatment response. Neuro-oncology advances **7**(1), vdae216 (2025)
3. Avants, B.B., Tustison, N., Song, G., et al.: Advanced normalization tools (ANTS). Insight j **2**(365), 1–35 (2009)
4. Chen, T.: XGBoost: A scalable tree boosting system. Cornell University (2016)
5. Cho, K., Van Merriënboer, B., Gulçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder–decoder for statistical machine translation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1724–1734 (2014)
6. Cho, S.J., Cho, W., Choi, D., Sim, G., Jeong, S.Y., Baik, S.H., Bae, Y.J., Choi, B.S., Kim, J.H., Yoo, S., et al.: Prediction of treatment response after stereotactic radiosurgery of brain metastasis using deep learning and radiomics on longitudinal mri data. Scientific Reports **14**(1), 11085 (2024)
7. Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., Sun, J.: Doctor ai: Predicting clinical events via recurrent neural networks. In: Machine learning for healthcare conference. pp. 301–318. PMLR (2016)
8. Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) **34**(2), 187–202 (1972)
9. Desai, S.P., Hori, Y.S., Kattaa, A.H., Izhar, M., Lam, F.C., Reesh, D.A., Tayag, A., Ustrzynski, L., Emrich, S.C., Gu, X., et al.: Artificial intelligence (AI) uses in stereotactic radiosurgery (SRS): outcome prediction with brain metastasis (BM)-A systematic review. Journal of Clinical Neuroscience **145**, 111854 (2026)
10. DeVries, D.A., Lagerwaard, F., Zindler, J., Yeung, T.P.C., Rodrigues, G., Hajdok, G., Ward, A.D.: Performance sensitivity analysis of brain metastasis stereotactic radiosurgery outcome prediction using MRI radiomics. Scientific Reports **12**(1), 20975 (2022)
11. Du, P., Liu, X., Shen, L., Wu, X., Chen, J., Chen, L., Cao, A., Geng, D.: Prediction of treatment response in patients with brain metastasis receiving stereotactic radiosurgery based on pre-treatment multimodal MRI radiomics and clinical risk factors: A machine learning model. Frontiers in Oncology **13**, 1114194 (2023)

12. Gordaliza, P.M., Banus, J., Gérin, B., Wynen, M., Molchanova, N., Richiardi, J., Cuadra, M.B.: From 100,000+ images to winning the first brain MRI foundation model challenges: Sharing lessons and models. arXiv preprint arXiv:2601.13166 (2026)
13. Habibi, M.A., Rashidi, F., Habibzadeh, A., Mehrtabar, E., Arshadi, M.R., Mirjani, M.S.: Prediction of the treatment response and local failure of patients with brain metastasis treated with stereotactic radiosurgery using machine learning: A systematic review and meta-analysis. Neurosurgical Review **47**(1), 199 (2024)
14. Holste, G., Lin, M., Zhou, R., Wang, F., Liu, L., Yan, Q., Van Tassel, S.H., Kovacs, K., Chew, E.Y., Lu, Z., et al.: Harnessing the power of longitudinal medical imaging for eye disease prognosis using transformer-based sequence modeling. NPJ digital medicine **7**(1), 216 (2024)
15. Kuhn, L., Abler, D., Richiardi, J., Hottinger, A.F., Schiappacasse, L., Dunet, V., Depeursinge, A., Andrearczyk, V.: AI-based response assessment and prediction in longitudinal imaging for brain metastases treated with stereotactic radiosurgery. In: Learning with Longitudinal Medical Images and Data (LMID at MICCAI) (2025)
16. Moawad, A.W., Janas, A., Baid, U., Ramakrishnan, D., Saluja, R., Ashraf, N., Maleki, N., Jekel, L., Yordanov, N., Fehringer, P., et al.: The brain tumor segmentation-metastases (brats-mets) challenge 2023: Brain metastasis segmentation on pre-treatment mri. arxiv pp. arXiv–2306 (2024)
17. Öznacar, T., Aral, İ.P., Zengin, H.Y., Tezcan, Y.: Survival prediction in brain metastasis patients treated with stereotactic radiosurgery: A hybrid machine learning approach. Brain Sciences **15**(3), 266 (2025)
18. Siebra, C.A., Kurpicz-Briki, M., Wac, K.: Transformers in health: a systematic review on architectures for longitudinal data analysis. Artificial Intelligence Review **57**(2), 32 (2024)