AI-Driven Prediction of Treatment Efficacy in Glioblastoma Using Medical Imaging

Noémie N. Moreau[1,2,3][0009−0005−2117−2519], Alexis Desmonts[1,2][0009−0009−4664−8770], Cyril Jaudet[1][0000−0003−2021−1983], Thomas Leleu[4], Alexandre G. Leclercq[1,2,3,5][0009−0004−6862−6613], Carole Brunaud[3][0000−0002−5084−1479], Dinu Stefan[4][0000−0003−4537−2871], Samuel Valable[3][0000−0003−0355−0270], Alexis Lechervy[5][0000−0002−9441−0187], and Aurélien Corroyer-Dulmont[1,2,3][0000−0002−9826−6565]

[1] Medical Physics Department, Centre François Baclesse, 14000 Caen, France
[2] Artificial Intelligence Department, Centre François Baclesse, 14000 Caen, France
[3] Université de Caen Normandie, CNRS, Normandie Université, ISTCT UMR6030, GIP CYCERON, 14000 Caen, France
[4] Radiation Oncology Department, Centre François Baclesse, 14000 Caen, France
[5] Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France
{nmoreau, brunaud, valable}@cyceron.fr
{a.desmonts, c.jaudet, t.leleu, d.stefan, a.corroyer-dulmont}@baclesse.unicancer.fr
{alexandre.leclercq, alexis.lechervy}@unicaen.fr

Abstract. Brain tumors represent a significant proportion of cancers in humans, with an incidence that continues to rise. Glioblastoma, the most aggressive tumor, demonstrates a variable response to treatment. Patients diagnosed with glioblastoma have a median survival of 15 months. A major challenge is that treatment efficacy, evaluated by anatomical MRI, becomes apparent more than two months after initiation. Given the limited survival time, early identification of non-responders before treatment onset is crucial. A binary classification model was performed on a cohort diagnosed with glioblastoma and treated between 2018 and 2023 at our center. Initially, treatment efficacy prediction was assessed using only the surgical criterion. The obtained sensitivity, specificity, and accuracy were 79.78%, 59.30% and 69.71%, respectively. Subsequently, a classifier was pre-trained using transfer learning on the ResNet-51Q model. This model takes as input nine central slices of pre-treatment MRI per patient. The results obtained on the test set were 79.10%, 90.74%, and 81.68% for sensitivity, specificity, and accuracy respectively. Deep hybrid learning (DHL) models were trained to include clinical data, with 84.38%, 94.74% and 90.00% for sensitivity, specificity, and accuracy respectively. Compared with the criterion of surgery alone, the deep learning approach improves the prediction of treatment efficacy prior to its administration. We enhanced performance by incorporating clinical data. Using models to predict treatment efficacy in GBM patients from pre-treatment data has considerable potential for personalising treatment regimens.

Keywords: Deep Learning · medical imaging · outcome prediction.

1 Introduction

Glioblastoma is the most common primary malignant brain tumors in adults. Standard treatment for these tumors follows the Stupp protocol, which includes surgery, radiotherapy and chemotherapy [9]. The median survival for patients with glioblastoma is approximately 15 months [11]. Despite these therapeutic approaches, treatment response remains heterogeneous, with some patients responding well to treatment while others do not. The assessment of treatment efficacy primarily relies on measuring lesion size at various intervals post-treatment, following the Response Assessment in Neuro-Oncology (RANO) criteria [12]. However, in clinical practice, determining the efficacy or the non-efficacy of therapies typically requires a delay of up to two months. This delay is significant given the limited median survival of glioblastoma patients.

Several studies, as highlighted in a literature review [7], have investigated methods to predict treatment response in patients with glioblastomas and brain metastases. Additionally, [2] demonstrated that the quality of surgical intervention plays a crucial role in improving patient survival.

The first approach is based on predicting treatment efficacy using imaging biomarkers. [10, 6] have demonstrated that it is possible to evaluate treatment effectiveness as early as one day after its administration. However, this method has limitations. Patients have already received chemotherapy and radiation therapy, and they may be subject to their side effects.

More recent studies have focused on leveraging radiomics and artificial intelligence (AI) to predict treatment efficacy prior to its administration. These studies indicate that combining imaging data with clinical information significantly improves the performance of predictive models [5]. Nevertheless, these studies are often conducted on small patient cohorts, limiting the robustness and generalizability of the models, which hinders their reliability and applicability in routine clinical routine.

Therefore, developing robust and generalizable predictive tools to assess therapeutic efficacy before treatment administration is essential. These tools must go beyond the limitations of current evaluations based on anatomical magnetic resonance imaging (MRI), providing improved patient management and minimizing the risks associated with ineffective treatments.

However, to our knowledge, no deep learning model with a large cohort is used in routine clinical practice to predict treatment efficacy prior to administration. In this study, we present a binary deep learning model able to predict treatment efficacy before its initiation.

2 Materials and methods

2.1 Patient cohort

The present retrospective study received approval from the local institutional review board and was conducted in compliance with the principles outlined in

the Declaration of Helsinki and the MR-004 guidelines established by the French National Institute for Health Data (INDS) for health research. Informed consent was obtained from all participants for the use of their data. The study population consisted of 175 patients diagnosed with glioblastoma and treated at François Baclesse center between January 2018 and December 2023. For the purposes of this investigation, short-term and long-term survivors were categorized based on the median survival of the entire patient cohort. Patients with a median survival of less than 332 days were classified as short-term survivors, while those with a median survival of 332 days or more were classified as long-term survivors. Detailed information regarding the study population is presented in Table 1.

Table 1: Patient cohort described overall (all patients) and by survival outcome (Short-Term or Long-Term survivors defined as $\leq$ or $>$ 332 days, respectively).

	All Patients (n= 175)	**Short-Term (n= 86)**	**Long-Term (n= 89)**
Survival (days)	mean: 363.4 median: 332.0 min-max: [42.0-1600.0]	mean: 163.6 median: 137.0 min-max: [42.0-329.0]	mean: 556.5 median: 492.0 min-max: [332.0-1600.0]
Sex, n (%)	Female: 74 *(42.3%)* Male: 101 *(57.7%)*	Female: 39 *(45.3%)* Male: 47 *(54.7%)*	Female: 35 *(39.3%)* Male: 54 *(60.7%)*
Age at the diagnosis (years)	mean: 67.7 median: 69.3 min-max: [25.5-88.9]	mean: 70.4 median: 71.3 min-max: [25.5-88.9]	mean: 65.1 median: 65.6 min-max: [42.3-83.7]
Surgery Type, n (%)	Biopsy: 69 *(39.4%)* Incomplete: 62 *(35.4%)* Complete: 44 *(25.1%)*	Biopsy: 51 *(59.3%)* Incomplete: 28 *(32.6%)* Complete: 7 *(8.1%)*	Biopsy: 18 *(20.2%)* Incomplete: 34 *(38.2%)* Complete: 37 *(41.6%)*
WHO performance status, n (%)	0: 27 *(15.4%)* 1: 95 *(54.3%)* 2: 45 *(25.7%)* 3: 8 *(4.6%)* 4: 0 *(0.0%)* 5: 0 *(0.0%)*	0: 7 *(8.1%)* 1: 40 *(46.5%)* 2: 32 *(37.2%)* 3: 7 *(8.1%)* 4: 0 *(0.0%)* 5: 0 *(0.0%)*	0: 20 *(22.5%)* 1: 55 *(61.8%)* 2: 13 *(14.6%)* 3: 1 *(1.1%)* 4: 0 *(0.0%)* 5: 0 *(0.0%)*

2.2 Medical imaging acquisition

MRI was conducted using a 16-channel brain dedicated coil on a 1.5/3 Tesla SIEMENS AREA/VIDA MRI scanner with patients in a supine position. Prior to the examination, each patient received an injection of DOTATEM at a dose of 0.2 mL/kg (500 μmol/ml). The imaging process began with a shimming procedure and scout scan, followed by tumor gadolinium enhancement detection using a post-Gd T1 brain sequence with the following parameters: TR/TEff = 2070/3.15 msec; Angle=15°; NEX = 1; 208 contiguous slices; 3D resolution = 0.5×0.5×1 mm; acquisition matrix = 512×512 pixels. A total of 36,400 2D MR images were

acquired from 175 patients. In order to account for variations in slice thickness and spacing between images, a resampling procedure was performed. For the purposes of this study, nine central slices containing the tumor were selected to ensure inclusion of the largest parts of the tumor.

2.3 Deep learning model

ResNet-51Q architecture: In this study, we employed an AI methodology utilizing a network architecture inspired by ResNet-50 [4]. To enhance efficiency and learning capabilities, the ResNet-51Q [13] network introduced several key modifications to the ResNet-50 framework. While maitaining a bottleneck-like structure similar to ResNet-50, ResNet-51Q incorporates enhanced convolutional operations. It initiates with a quadruple convolutional stem comprising two activation layers, diverging from the traditional max pooling layer found in ResNet-50. The architecture comprises four primary blocks with depths of 2, 4, 6 and 4, constrasting with the ResNet-50 configuration of 3, 4, 6, and 3. A notable distinction lies in the convolutional methodology: ResNet-51Q employs grouped convolutions with a group size of 32 and a bottleneck ratio of 0.25 across the initial three blocks, whereas ResNet-50 utilizes standard convolutions. The final block incorporates depthwise convolutions with a ratio of 1.0 to further optimize efficiency. Additionally, the channel dimensions increase from 256 to 512, reaching a constant value of 1536 at the final stage, in contrast to the ResNet-50 progression to 2048 channels in its final stage. Futhermore, ResNet-51Q enhances gradient propagation and convergence by replacing the standard ReLu activation function of ResNet-50 wuth the SiLU (Sigmoid Linear Unit). The ResNet-51Q system is optimally designed for contemporary deep learning applications that require both high processing capability and effective resource utilisation. This is due to the architectural enhancements that facilitate an optimal balance between depth, computational efficiency and representational power.

Model development: The deep learning model developed for this study employed a fine-tuned ResNet-51Q architecture. The model was developed using a unique set of 1575 T1-Gd brain images obtained from 175 patients. The dataset was split into three subsets: a training set with 936 images (60% of the total), a validation set with 306 images (20% of the total), and a test set with 333 images (20% of the total). Furthermore, it is essential to verify that all slices of a patient are contained within a single dataset. To optimize the model's performance, hyperparameters were adjusted using the Optuna software [1], employing the Bayesian Tree-structured Parzen Estimator (TPE) approach. Hyperparameters tested included a learning rate ranging from 10^{-7} to 10^{-2}, weight decay between 10^{-6} and 10^{-2}, and dropout values of 0.1 to 0.5. The deep learning model was trained using binary cross entropy as the loss function, utilizing one NVIDIA P6000 GPU with 24GB of memory. The code used for developing and training the AI model is publicly available at: https://github.com/AurelienCD/MRI-PRED-GBM.

Deep Hybrid Learning: To leverage the deep learning model using pre-treatment MRI as input and the machine learning model (*DecisionTreeClassifier, LogisticRegression, GaussianNB, SVC, LinearDiscriminantAnalysis, XGBClassifier, KNeighborsClassifier* and *RandomForestClassifier*) using clinical data as input, both models were combined to create a deep hybrid learning (DHL) model. The probabilities predicted by our deep learning model for each patient belonging to a specific class were concatenated with a vector representing the clinical data.

3 Results

3.1 Evaluation of surgery criteria for survival classification without artificial intelligence model

In the literature, the surgical criterion is identified as a predictor of overall survival. Therefore, the first approach involved conducting a one-way analysis of variance (ANOVA) to assess significant differences between the studied groups (biopsy only, incomplete and complete surgery) in relation to overall survival. As illustrated in Figure 1, the results indicate a p-value of 5e-7, a mean value $\pm$ standard deviation of 241.61 $\pm$ 191.24, 404.32 $\pm$ 271.31 and 496.70 $\pm$ 282.83 for biopsy, incomplete surgery and complete surgery, respectively. As shown in study [2] and our own data, the surgical criterion appears to have a significant effect on overall survival. It will be relevant to investigate whether surgery alone could serve as a predictive factor for overall patient survival.

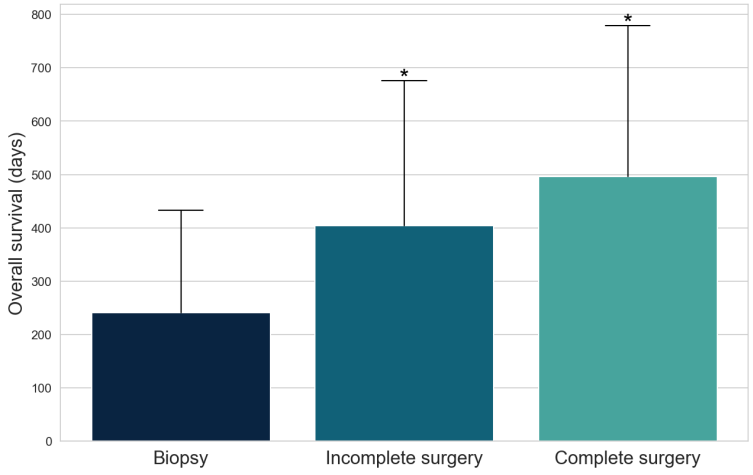


Fig. 1: Impact of surgery quality on overall survival: Mean $\pm$ SD (Standard Deviation) , n=69 ; 62 ; 44 for biopsy, incomplete surgery, complete surgery respectively, * $p < 0.001$ vs biopsy.

The second approach was to analyse the effect of surgery on the classification of patient survival (short-term and long-term survivors). To achieve this, an analysis was conducted without the use of AI. We examined the impact of surgical criteria (biopsy vs. incomplete and complete surgery) on patient survival classification. As shown in the Table 2 this study indicates that surgery alone is not a reliable criterion for classifying patients as responders or non-responders, due to a low accuracy of 69.71%. Moreover, in a medical context, minimizing false positives is crucial. A specificity of 59.30% is therefore not clinically acceptable. The next step is to assess whether AI-based approaches can improve classification performance.

Table 2: Performances based only on surgery criteria.

Accuracy (%)	**Sensitivity (%)**	**Specificity (%)**	**True Positive**	**False Positive**	**False Negative**	**True Negative**
69.71	79.78	59.30	71	35	18	51

3.2 Responding and non-responding patient classification model based only on pre-treatment MRI

We then used an AI model, with only pre-treatment MRI as input data, to classify patients into two categories: short-term or long-term survivors, based on the median survival of all patients (Figure 2). As shown in Figure 2, AI model is able to predict from pre-treatment MRI that a patient will be a long-term survivor (left part of Figure 2) or a short-term survivor (right part of Figure 2). The binary classification model performed well, with an AUC (Area Under the Curve) of 0.89 (Figure 3a) and an accuracy, specificity and sensitivity of 81.68%, 90.74% and 73.10% respectively.

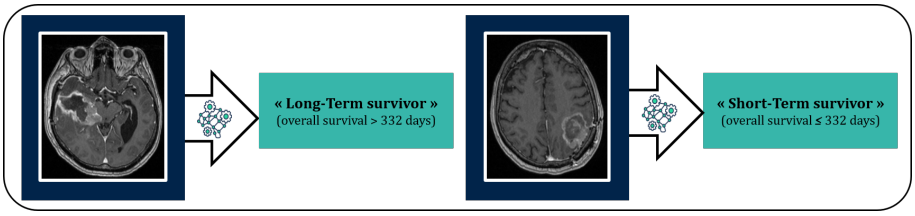
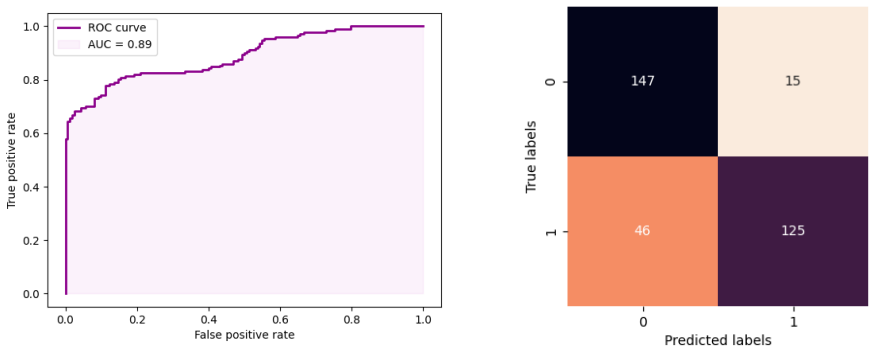


Fig. 2: Long-term survivor (left part) or a short-term survivor (right part).



(a) ROC curve (b) Confusion matrix

Fig. 3: ROC curve and confusion matrix of binary model based on only pre-treatment MRI.

Table 3: Performances based only on pre-treatment MRI.

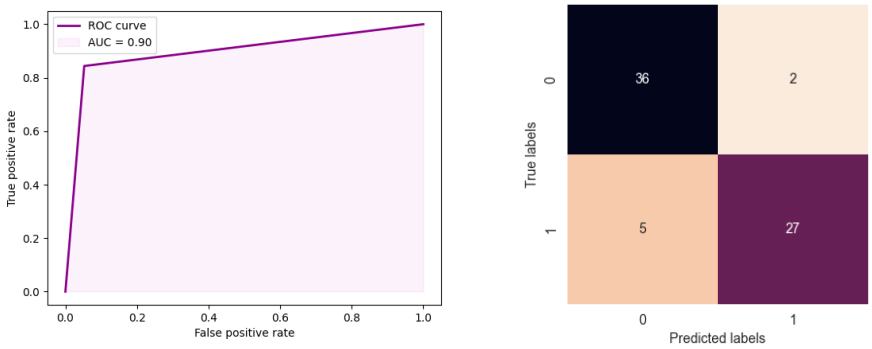
AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	True Positive	False Positive	False Negative	True Negative
0.89	81.68	73.10	90.74	125	15	46	147

3.3 Impact of adding clinical data to the MRI deep learning model in the performance of predicting responding to non-responding patients

We aimed to investigate whether the inclusion of clinical data (WHO performance status, age, sexe and surgery type) improves the performances of the model. For this purpose, we developed a DHL model that combined a pre-treatment MRI deep learning model prediction with a classification machine learning model integrating clinical data. Following performance evaluation and optimization of the hyperparameters, it was determined that *KNeighborsClassifier* demonstrated the optimal performance. The performance evaluation of this model revealed an accuracy 90.00% (+8.32% compared to the model based only on pre-treatment MRI), a specificity of 94.74% (+4.00%) and a sensitivity of 84.38% (+11.28%).

Table 4: Performances based on pre-treatment MRI and clinical data.

AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	True Positive	False Positive	False Negative	True Negative
0.90	90.00	84.38	94.74	27	2	5	36



(a) ROC curve (b) Confusion matrix

Fig. 4: ROC curve and confusion matrix of binary model based on pre-treatment MRI and clinical data.

4 Discussion

In clinical practice, the assessment of glioblastoma treatment efficacy is only possible at least two months after initiation, a relatively long delay considering the median patient survival of 15 months. To anticipate this evaluation, we proposed an AI model using nine central slices of pre-treatment MRI as input data to distinguish short-term survivors to long-term survivors using a threshold of 332 days after MRI acquisition. We experimented with additional slices, but the performance gains were minimal.

A recent study [3] investigated glioblastoma survival prediction eight months after radiotherapy, corresponding to the completion of adjuvant temozolomide treatment. This study, conducted on 206 patients from 11 centers, relied on the first post-radiotherapy brain MRI. The imaging-based model achieved an accuracy of 0.84 and a specificity of 0.77 in a retrospective analysis of 19 patients from two centers, showing performance comparable to ours. Given that this study relied on post-treatment MRI, our results obtained from pre-treatment MRI appear even more promising. Other studies have investigated survival classification for glioblastoma patients based on pre-traitement imaging. To predict survival stratification of 125 patients with GBM, a clinical study [5] was performed. MRI images were used to extract radiomic features. With an AUC of 0.92, the SVM model surpassed the other two machine learning models tested. Another study [6] used a deep convolutional neural network (CNN) with three hidden layers, each with eight neurons, to predict patient survival of 133 patients. The CNN performed well, classifying survival (less than one year, one to two years and more than two years) with an accuracy of 90.6%. The studies are based on small cohorts of data (less than 150 patients). However, for the use of AI models, large data cohorts are required [8]. The performance of the models presented in the

literature is enhanced by the incorporation of clinical data, and the results are consistent with those of our study.
We chose to formulate the problem as a classification task rather than a regression problem, given the limited number of patients within each category. Nevertheless, introducing more classes (e.g., 3 or 4) could approximate a regression framework by enabling the assignment of probability distributions over these classes for each patient.

5 Conclusion

In conclusion, predicting the effectiveness of treatments for patients with glioblastoma is crucial given the median survival of 15 months and high patient response heterogeneity. The results obtained in this study show high accuracy of 90.00% to predict treatment efficacy only using pre-treatment information with MRI and clinical data. However, the addition of clinical data through deep hybrid learning has further improved these results by 8.32%. A potential perspective for this study, is to develop other models for discriminating the model with more class.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019)
2. Brown, T.J., Brennan, M.C., Li, M., Church, E.W., Brandmeir, N.J., Rakszawski, K.L., Patel, A.S., Rizk, E.B., Suki, D., Sawaya, R., Glantz, M.: Association of the Extent of Resection With Survival in Glioblastoma: A Systematic Review and Meta-analysis. JAMA oncology **2**(11), 1460–1469 (Nov 2016). https://doi.org/10.1001/jamaoncol.2016.1373
3. Chelliah, A., Wood, D.A., Canas, L.S., Shuaib, H., Currie, S., Fatania, K., Frood, R., Rowland-Hill, C., Thust, S., Wastling, S.J., et al.: Glioblastoma and radiotherapy: A multicenter ai study for survival predictions from mri (grasp study). Neuro-oncology **26**(6), 1138–1151 (2024)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
5. Jia, X., Zhai, Y., Song, D., Wang, Y., Wei, S., Yang, F., Wei, X.: A Multiparametric MRI-Based Radiomics Nomogram for Preoperative Prediction of Survival Stratification in Glioblastoma Patients With Standard Treatment. Frontiers in Oncology **12**, 758622 (Feb 2022). https://doi.org/10.3389/fonc.2022.758622
6. Luckett, P.H., Olufawo, M., Lamichhane, B., Park, K.Y., Dierker, D., Verastegui, G.T., Yang, P., Kim, A.H., Chheda, M.G., Snyder, A.Z., Shimony, J.S., Leuthardt, E.C.: Predicting survival in glioblastoma with multimodal neuroimaging and machine learning. Journal of Neuro-Oncology **164**(2), 309–320 (Sep 2023). https://doi.org/10.1007/s11060-023-04439-8

7. Moreau, N.N., Valable, S., Jaudet, C., Dessoude, L., Thomas, L., Hérault, R., Modzelewski, R., Stefan, D., Thariat, J., Lechervy, A., Corroyer-Dulmont, A.: Early characterization and prediction of glioblastoma and brain metastasis treatment efficacy using medical imaging-based radiomics and artificial intelligence algorithms. Frontiers in Oncology **15**, 1497195 (Jan 2025). https://doi.org/10.3389/fonc.2025.1497195
8. Sebastian, A.M., Peter, D.: Artificial Intelligence in Cancer Research: Trends, Challenges and Future Directions. Life **12**(12), 1991 (Nov 2022). https://doi.org/10.3390/life12121991, https://www.mdpi.com/2075-1729/12/12/1991
9. Stupp, R., Mason, W.P., Van Den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J., Belanger, K., Brandes, A.A., Marosi, C., Bogdahn, U., Curschmann, J., Janzer, R.C., Ludwin, S.K., Gorlia, T., Allgeier, A., Lacombe, D., Cairncross, J.G., Eisenhauer, E., Mirimanoff, R.O.: Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. New England Journal of Medicine **352**(10), 987–996 (Mar 2005). https://doi.org/10.1056/NEJMoa043330
10. Talati, P., El-Abtah, M., Kim, D., Dietrich, J., Fu, M., Wenke, M., He, J., Natheir, S.N., Vangel, M., Rapalino, O., Vaynrub, A., Arrillaga-Romany, I., Forst, D.A., Yen, Y.F., Andronesi, O., Kalpathy-Cramer, J., Rosen, B., Batchelor, T.T., Gonzalez, R.G., Gerstner, E.R., Ratai, E.M.: MR spectroscopic imaging predicts early response to anti-angiogenic therapy in recurrent glioblastoma. Neuro-Oncology Advances **3**(1), vdab060 (Jan 2021). https://doi.org/10.1093/noajnl/vdab060, https://academic.oup.com/noa/article/doi/10.1093/noajnl/vdab060/6226585
11. Tan, A.C., Ashley, D.M., López, G.Y., Malinzak, M., Friedman, H.S., Khasraw, M.: Management of glioblastoma: State of the art and future directions. CA: A Cancer Journal for Clinicians **70**(4), 299–312 (Jul 2020). https://doi.org/10.3322/caac.21613
12. Wen, P.Y., Bent, M.v.d., Youssef, G., Cloughesy, T.F., Ellingson, B.M., Weller, M., Galanis, E., Barboriak, D.P., Groot, J.d., Gilbert, M.R., Huang, R., Lassman, A.B., Mehta, M., Molinaro, A.M., Preusser, M., Rahman, R., Shankar, L.K., Stupp, R., Villanueva-Meyer, J.E., Wick, W., Macdonald, D.R., Reardon, D.A., Vogelbaum, M.A., Chang, S.M.: RANO 2.0: Update to the Response Assessment in Neuro-Oncology Criteria for High- and Low-Grade Gliomas in Adults. Journal of Clinical Oncology **41**(33), 5187–5199 (2023). https://doi.org/10.1200/JCO.23.01059
13. Wightman, R.: Pytorch image models. https://github.com/huggingface/pytorch-image-models (2019). https://doi.org/10.5281/zenodo.4414861