Evaluation of Monte Carlo Dropout for Uncertainty Quantification in Multi-task Deep Learning-Based Glioma Subtyping

Gonzalo Esteban Mosquera Rojas[1 [0009−0008−9901−4141]], Sebastian van der Voort[2 [0000−0002−6526−8126]], Carolin M. Pirkl[3], Sandeep Kaushik[4 [0000−0003−0654−0799]], Marion Smits[1,5,6 [0000−0001−5563−2871]], and Stefan Klein[1 [0000−0003−4449−6784]]

[1] Department of Radiology and Nuclear Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands g.mosquerarojas@erasmusmc.nl
[2] Department of Medical Informatics, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands
[3] GE HealthCare, Munich, Germany
[4] GE HealthCare, USA
[5] Brain Tumor Center, Erasmus MC Cancer Institute, Rotterdam, the Netherlands
[6] Medical Delta, Delft, the Netherlands

Abstract. Uncertainty Quantification (UQ) is essential for enhancing the trustworthiness of Deep Learning (DL) models in high-stakes medical imaging applications. Monte Carlo Dropout (MCD) remains one of the most widely used and foundational approaches for UQ, often serving as a baseline in comparative studies. In this work, we systematically evaluate MCD in the context of DL-assisted glioma diagnosis, focusing on a less-explored yet clinically relevant multi-task setting that combines glioma subtyping and segmentation. We investigate how key parameters of MCD, namely the number of MC samples and the dropout rate, may affect the quality of uncertainty estimates. Additionally, we disentangle epistemic and aleatoric uncertainty components to gain deeper understanding of model confidence. The results demonstrate that, when appropriately tuned, MCD produces well-calibrated uncertainty estimates. The segmentation task was primarily influenced by epistemic uncertainty, whereas aleatoric uncertainty constituted the main source of uncertainty in all classification tasks.

Keywords: Uncertainty quantification · Monte Carlo Dropout · Magnetic Resonance Imaging · Glioma

1 Introduction

Gliomas are among the most aggressive brain tumors, with high morbidity and mortality worldwide [14]. The World's Health Organization (WHO) classification standards incorporate molecular markers such as isocitrate dehydrogenase (IDH) mutation status and 1p19q co-deletion [11]. These features, together with tumor grade, provide crucial information on tumor biology and patient prognosis.

Accurate tumor segmentation in brain magnetic resonance imaging (MRI) is also essential for treatment planning and patient monitoring [12]. Concurrently, DL methods have enabled automatic prediction of IDH mutation, 1p19q co-deletion and tumor grade from imaging data with promising performance [3, 9]. Multi-task learning frameworks have shown strong potential in clinical applications, often outperforming single-task models. For instance, authors in [24] improved breast tumor classification using a multi-task approach combining segmentation and classification from ultrasound images, while the study in [10] achieved better results for brain tumor diagnosis through joint segmentation and classification of gliomas, meningiomas, and pituitary tumors. Similarly, in [23], a multi-task model was proposed to segment gliomas and classify their molecular features and grade, demonstrating promising performance but struggling with 1p19q co-deletion and grade 3 prediction.

Nonetheless, the analysis of uncertainty in multi-task applications remains limited. The study in [23] noted the absence of uncertainty estimates and emphasized the importance of including and rigorously evaluating them. Addressing this gap is crucial since multi-task learning introduces challenges such as shared representations and heterogeneous uncertainty sources, making reliable uncertainty estimation essential for trustworthy clinical decision-making.

In this work, we build upon the multi-task framework from [23], which jointly performs glioma segmentation and subtyping. We extend it by including UQ using Monte Carlo Dropout (MCD) [6]. Our study aims to systematically evaluate how key MCD parameters, namely the number of MC samples and the dropout rate, affect the quality of uncertainty estimates in this medical application. Additionally, we disentangle epistemic and aleatoric uncertainty to (i) determine which source dominates in each task and (ii) examine how each is influenced by these parameters.

2 Materials and Methods

2.1 Uncertainty Quantification with Monte Carlo Dropout

Let N be a neural network with weights $\mathbf{W}$. In MCD, Bayesian inference is approximated by applying dropout at rate p at train time, and then performing T stochastic forward passes with the same dropout rate enabled at test time. Given some training data $\mathcal{D}$ and a new input x^*, each forward pass samples a different set of network weights $\mathbf{W}_t$, yielding a collection of predictions $\{p(y^*|x^*, \mathbf{W}_t)\}_{t=1}^{T}$. The predictive distribution is then approximated by [6]:

$$p(y^*|x^*, \mathcal{D}) \approx \frac{1}{T} \sum_{t=1}^{T} p(y^*|x^*, \mathbf{W}_t). \tag{1}$$

The total predictive uncertainty can be quantified by computing the entropy of this predictive distribution [5, 21]:

$$U_{\text{predictive}} = -\sum_{c} \left(\frac{1}{T} \sum_{t=1}^{T} p(y^* = c \mid x^*, \mathbf{W}_t) \right) \log \left(\frac{1}{T} \sum_{t=1}^{T} p(y^* = c \mid x^*, \mathbf{W}_t) \right), \tag{2}$$

where c corresponds to the number of classes for the classification task.
The predictive uncertainty can be decomposed into two components. The aleatoric uncertainty, representing the inherent noise in the data, is estimated by averaging the entropy of the predictive distribution from all t samples [5, 21]:

$$U_{\text{aleatoric}} = \frac{1}{T} \sum_{t=1}^{T} \left(-\sum_{c} p(y^* = c \mid x^*, \mathbf{W}_t) \log p(y^* = c \mid x^*, \mathbf{W}_t) \right). \tag{3}$$

The epistemic uncertainty describes the uncertainty coming from the model parameters. It is computed as the difference between the predictive and aleatoric components [5, 21]:

$$U_{\text{epistemic}} = \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{c} p_{t,c} \log p_{t,c} \right) - \left(\sum_{c} \bar{p}_c \log \bar{p}_c \right), \tag{4}$$

where $\bar{p}_c$ represents the mean predicted probability for class c across T stochastic forward passes, and $p_{t,c}$ describes the class c probability at step t.

2.2 Monte Carlo Dropout Evaluation

Multi-task subtyping framework: As shown in figure 1, the pipeline operates on four pre-operative structural MRI modalities: T1-weighted (T1w), contrast-enhanced T1-weighted (T1wCE), T2-weighted (T2w), and T2-weighted fluid-attenuated inversion recovery (FLAIR). It includes a preprocessing module, an encoder-decoder network [23] inspired by the U-net [17] for tumor segmentation, and a classification branch composed of fully connected layers for the final tumor suptying. The latter uses imaging features extracted from several resolution levels of the segmentation pathway, and predicts the IDH status (wildtype, mutated), 1p19q co-deletion status (intact, co-deleted), and tumor grade (2, 3, and 4).

The model was trained and tested in a collection of both in-house data and six publicly available datasets of adult-type glioma patients, including BraTS [1, 2, 13], REMBRANDT [18], IvyGAP [20], CPTAC-GBM [4], TCGA-GBM [19] and TCGA-LGG [16]. Further information about the data and architecture implementation are provided as supplementary material.

Evaluation Experiments: We performed a systematic evaluation of MCD, varying the dropout rate p and the number of MC samples t. Specifically, we explored values of p between 0.2 and 0.6 in increments of 0.05, and values of t from 10 to 100 in steps of 10. These ranges were selected to analyze the effect of

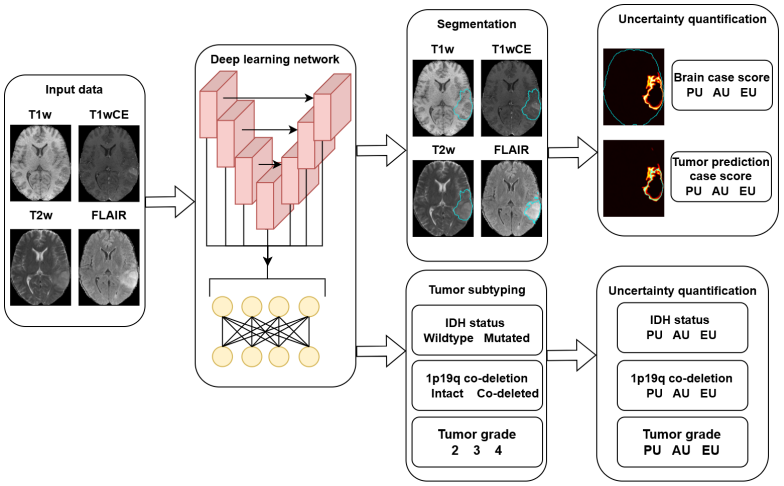

Fig. 1. Overview of the multi-task glioma subtyping framework with UQ. The input scans are preprocessed by means of registration, bias field correction and skull-stripping, followed by a DL model that performs both segmentation and tumor subtyping. For each of the tasks, the predictive (PU), aleatoric (AU) and epistemic uncertainty (EU) are computed. In the segmentation task, a case-level uncertainty score is computed by averaging the values of the uncertainty map comprised by either the brain mask or the tumor mask predicted by the MCD model.

both parameters in a wide spectrum, covering both moderate and higher values, and considering theoretical and empirical insights from the literature [6, 8, 22].

Importantly, we interpret the dropout rate as the probability of dropping out units, meaning that higher rates reduce the network predictive capacity more aggresively. Too low a dropout rate may lead to underestimation of uncertainty, while excessively high values can harm predictive stability or lead to underfitting. The upper bound of 0.6 is chosen such that even with aggressive dropout, sufficient model capacity is retained without requiring network width scaling. Likewise, the number of MC samples influences the fidelity of uncertainty estimates: too few may yield unstable approximations, but there is an optimal value where the predictive distribution is approximate enough to the theoretical Bayesian posterior.

Uncertainty quality in classification tasks was assessed using Expected Calibration Error (ECE) and Negative Log-Likelihood (NLL). ECE quantifies how well predicted probabilities align with actual outcome frequencies, while NLL penalizes confident but incorrect predictions, capturing both accuracy and calibration aspects. In addition, we computed the Receiver Operating Characteristic Area Under the Curve (ROC-AUC) to contrast uncertainty estimates with model accuracy, providing an alternative view of how uncertainty correlates with predictive performance. In the segmentation task, the quality of the uncertainty estimates was evaluated using the Pearson correlation coefficient ρ between case-

level uncertainty scores and the Dice Similarity Coefficient (DSC). This analysis was based on the assumption that higher uncertainty should generally be associated with lower segmentation accuracy, resulting in a negative ρ.

To derive the case-level uncertainty, we averaged voxel-wise uncertainty values with two aggregation methods: the average value within the whole brain (generated with HD-BET [7]) and the tumor region predicted by the MCD model. While focusing on the tumor localizes uncertainty to clinically relevant areas, it may neglect broader anatomical or contextual details. In contrast, averaging across the entire brain incorporates global information but may dilute the focus on the target structure. Therefore, this setting allowed us to investigate quantitatively which region yielded more informative uncertainty estimates.

3 Results

3.1 Effect of the number of Monte Carlo samples and dropout rate

As shown in Figure 2, uncertainty values stabilized consistently across all tasks once the number of MC samples exceeds 20-30. Increasing dropout rates generally led to proportional increases in predictive and aleatoric uncertainties. Epistemic uncertainty followed this trend up to a dropout rate of 0.5, after which it peaked at 0.55 and slightly declined at 0.6.

For the IDH and 1p19q prediction tasks, both predictive and aleatoric uncertainties rose with dropout rate, with some decrease at higher rates (between 0.55-0.6). In the tumor grade prediction, the same trend was observed, although epistemic uncertainty reached its highest value at 0.55. In the tumor segmentation, all types of uncertainty increased consistently with dropout rate and the epistemic component was the main contributor to the total uncertainty. Conversely, aleatoric uncertainty dominated in all the tumor subtyping tasks.

3.2 Calibration of uncertainty estimates

Based on the findings from figure 2, we chose 30 as the optimal t value, since it was the point where stabilization was most consistently achieved throughout the tasks. Figure 3 summarizes calibration of the uncertainty estimates and predictive accuracy across classification tasks. For the IDH prediction, both ECE and NLL reached their lowest values at dropout 0.25, with subtle fluctuations between 0.2 and 0.4 and then a steady increase thereafter. ROC-AUC was highest at 0.2 and generally declined as dropout increased. In the 1p19q prediction, ECE and NLL were minimized at 0.2 and generally rose with higher dropout rates. ROC-AUC followed a similar decreasing pattern. For the tumor grade prediction, ECE was lowest at 0.25 and gradually increased with dropout, while NLL rose consistently. ROC-AUC also showed a decreasing trend for dropout rates higher than 0.35. In the IDH and 1p19q predictions, the confidence intervals for the ECE were the narrowest at the dropout rate where it was minimized.

Figure 4 presents the uncertainty quality analysis for the segmentation case. In the brain mask-based approach, correlation values ranged from -0.1 to slighly

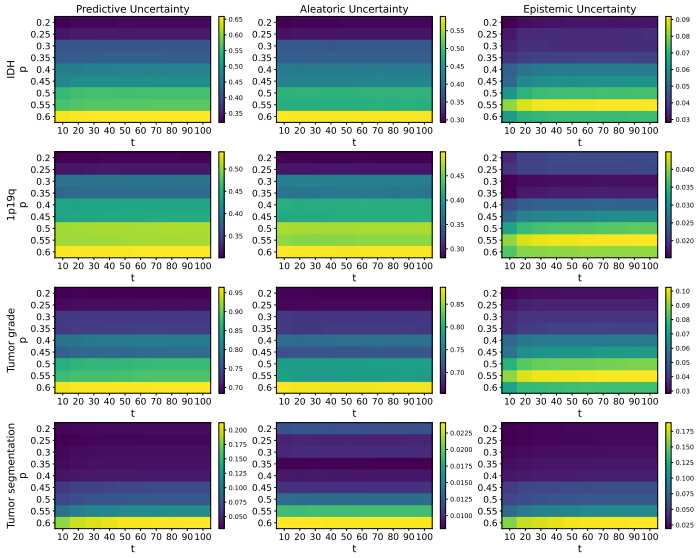

Fig. 2. Predictive, aleatoric and epistemic uncertainty values accross all tasks as a function of the number of Monte Carlo samples t and the dropout rate p.

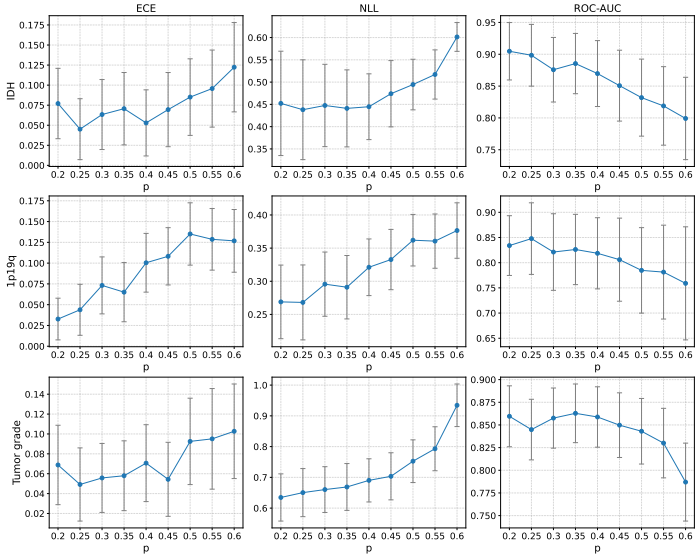

Fig. 3. ECE, NLL and ROC-AUC values as a function of dropout rate p. For each point, a 95% confidence interval is presented. This interval was obtained by 1000× bootstrap resampling of the test set.

above 0.2, with stronger negative values observed at dropout rates of 0.2, 0.25,

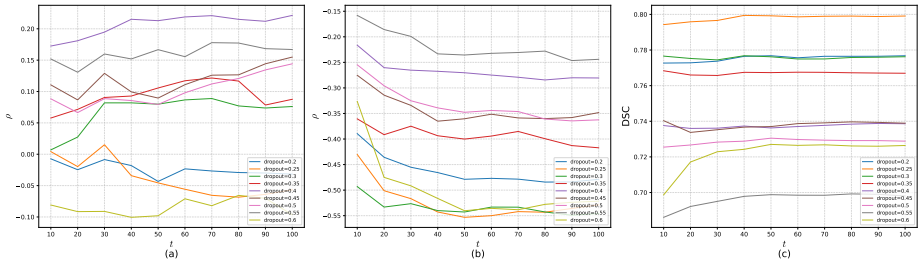

Fig. 4. Pearson correlation coefficient ρ between the segmentation case-level total uncertainty score and the Dice Similarity Coefficient DSC. The correlation is presented as a function of the number of MC samples t and dropout rates p with two different voxel-wise uncertainty score aggregation methods: (a) using a brain mask and (b) using the tumor mask predicted by the MCD model. (c) DSC between the prediction of different MCD models on the test set with varying number of MC samples t.

and 0.6. Conversely, in the tumor mask-based approach, the correlation scores spanned from -0.55 to -0.15, which indicated stronger inverse relationship between the DSC and the uncertainty estimates. The most negative values were found at dropout rates of 0.25, 0.3, and 0.6. Despite the high negative correlation at dropout rate of 0.6 in both settings, it was noted that the average DSC of the MCD model on the test set was lower than those at moderate dropout rates such as 0.2, 0.25 and 0.3.

Our evaluation of the uncertainty estimates across tasks and metrics demonstrated that the MCD model with a dropout rate of 0.25 offered the best trade-off between predictive performance and uncertainty calibration, and was therefore selected as the optimal configuration. Figure 5 further highlights the model's ability to produce calibrated uncertainty scores, where low quality predictions consistenly exhibit higher uncertainty in all tasks.

4 Discussion

In this work, we systematically evaluated the effect of Monte Carlo sample size and dropout rate on uncertainty estimates in a multi-task DL framework for glioma segmentation and subtyping. Uncertainty values consistently stabilized after 20-30 samples across all tasks, with no apparent variability between classification and segmentation. Increasing the dropout rate led to higher uncertainty estimates. This is in line with the MCD theory, where higher dropout injects greater variability into the network during test-time sampling, effectively broadening the variational posterior used to approximate the true weight distribution, and thus giving place to higher uncertainty values.

Minor deviations, such as some cases with higher epistemic uncertainty at dropout 0.55 compared to 0.6, reflect instability of the model in the high dropout regime, where it may begin to underfit or behave erratically due to excessive regularization. This is potentially exacerbated by limited supervision, especially in

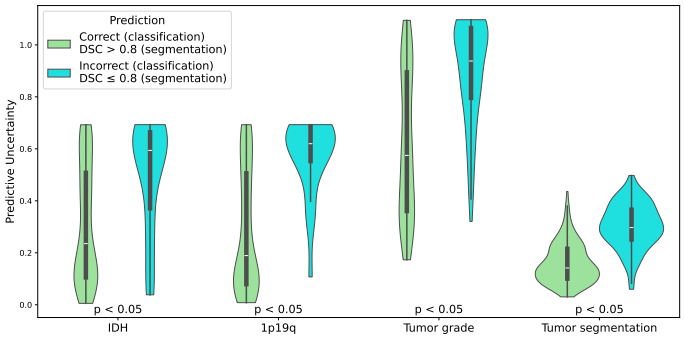

Fig. 5. Violin plots of the total uncertainty across different tasks, comparing the distributions of high quality (light green) and low quality (light blue) predictions. For the classification tasks (IDH, 1p19q and tumor grade), high and low quality predictions correspond to correctly and incorrectly classified cases, respectively. For the segmentation task, high and low quality predictions represent cases with $DSC > 0.8$ and $DSC \leq 0.8$, respectively. The label $p < 0.05$ indicates that the distributions were found to be significantly different under Mann-Whitney U test.

the 1p19q and grade prediction tasks. Furthermore, aleatoric uncertainty dominated in classification tasks, indicating data-driven ambiguity, which may be amplified by label noise due to human error. These findings agree with current discussions in neuro-oncology, where it has been noted that some tumor features, such as the grade, can be very flexible given that there are no strict thresholds to define it [15]. Epistemic uncertainty was more prominent in the segmentation task, which may be attributed to the variability in tumor presentation across patients, including differences in size, shape, and intensity profiles, making generalization more difficult.

The calibration results revealed distinct patterns across tasks that can be interpreted in light of both the nature of each metric and the available training data. ECE was lowest at dropout 0.25 for IDH and grade, and at 0.2 for 1p19q, indicating better calibration at moderate dropout levels. These dropout rates likely strike a balance between model stochasticity and stability, especially relevant in low-data regimes. In contrast, NLL increased consistently across dropout rates in all tasks, which relates to how it penalizes incorrect and overconfident predictions. Higher dropout rates introduce greater variability and reduce prediction confidence, thereby lowering the average predictive likelihood and resulting in elevated NLL values.

The decrease in ROC-AUC with increasing dropout rates corresponds with rising NLL and ECE values, indicating that higher uncertainty and poorer calibration negatively impact the model's discriminative performance. This suggests that while moderate dropout improves both calibration and accuracy, excessive dropout leads to less confident predictions and reduced predictive reliability.

In the segmentation task, the tumor mask-based uncertainty showed a clear negative correlation with the DSC, which aligns with theoretical expectations: higher uncertainty should correspond to poorer segmentation performance. Such an inverse relationship was stronger than in the brain mask-based approach, where correlations fluctuated around zero. This could be due to the fact that in most of the brain area the model is generally certain about it not being a tumor, thus resulting in a very low case-level score. Future work could be directed towards exploring alternative voxel-wise uncertainty score aggregation methods that consider false positive and false negative cases.

In general, higher dropout corresponded to lower quality in the uncertainty estimates for the tumor segmentation. Nonetheless, instability of the model was also observed in the high dropout regime. Although the uncertainty estimates had a negative correlation with the DSC at dropout of 0.6, the average DSC on the test set was lower than those at moderate dropout, indicating lower practical value of uncertainty estimates when derived from an erratic model.

In this work, we extended a multi-task DL-based framework for glioma subtyping by integrating and systematically evaluating Monte Carlo Dropout for Uncertainty Quantification. Our analysis revealed that the dropout rate plays a more critical role than the number of MC samples in determining the quality of uncertainty estimates across all tasks. Notably, we demonstrated that, when properly tuned, Monte Carlo Dropout yields well-calibrated predictive uncertainty, thereby effectively enhancing model interpretability and trustworthiness.

Acknowledgments. This work is part of the "Trustworthy AI for MRI" ICAI lab within the project ROBUST: Trustworthy AI-based Systems for Sustainable Growth with project number KICH3.LTP.20.006, financed by the Dutch Research Council (NWO), GE HealthCare, and the Dutch Ministry of Economic Affairs and Climate Policy (EZK) under the program LTP KIC 2020-2023.

Disclosure of Interests. Authors GEMR, MS, and SKlein are part of the Erasmus MC ICAI lab "Trustworty AI for MRI", a public-private research program partially funded by GE HealthCare (payment to institution). MS received consultancy honoraria (paid to institution) from Bracco. CP and SKaushik are employees of GE HealthCare. SV declares no competing interests.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific data **4**(1), 1–13 (2017)
2. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 (2018)

3. Chang, P., Grinband, J., Weinberg, B.D., Bardis, M., Khy, V., Cadena, G., Su, M., Cha, S., Filippi, C.G., Bota, D., et al.: Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. AJNR. American journal of neuroradiology **39**(7), 1201–1207 (2018)
4. CPTAC: The clinical proteomic tumor analysis consortium glioblastoma multiforme collection (cptac-gbm) (version 16) [dataset]. The Cancer Imaging Archive (2018)
5. Depeweg, S., Hernandez-Lobato, J.M., Doshi-Velez, F., Udluft, S.: Decomposition of uncertainty in bayesian deep learning for efficient and risk-sensitive learning. In: International conference on machine learning. pp. 1184–1193. PMLR (2018)
6. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: international conference on machine learning. pp. 1050–1059. PMLR (2016)
7. Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.P., Heiland, S., Wick, W., et al.: Automated brain extraction of multisequence mri using artificial neural networks. Human brain mapping **40**(17), 4952–4964 (2019)
8. Kendall, A., Badrinarayanan, V., Cipolla, R.: Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. arXiv preprint arXiv:1511.02680 (2015)
9. Kickingereder, P., Bonekamp, D., Nowosielski, M., Kratz, A., Sill, M., Burth, S., Wick, W., Schlemmer, H.P., Radbruch, A., Bendszus, M., et al.: Radiogenomics of glioblastoma: machine learning–based classification of molecular characteristics by using multiparametric and multiregional mr imaging features. Radiology **290**(3), 674–682 (2019)
10. Kordnoori, S., Sabeti, M., Shakoor, M.H., Moradi, E.: Deep multi-task learning structure for segmentation and classification of supratentorial brain tumors in mr images. Interdisciplinary Neurosurgery **36**, 101931 (2024)
11. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al.: The 2021 who classification of tumors of the central nervous system: a summary. Acta Neuropathologica **143**(1), 11–28 (2021)
12. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694
13. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
14. Ostrom, Q.T., Cioffi, G., Gittleman, H., Patil, N., Waite, K., Kruchko, C., Barnholtz-Sloan, J.S.: Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2012–2016. Neuro-oncology **21**(Supplement_5), v1–v100 (2019)
15. Park, Y.W., Vollmuth, P., Foltyn-Dumitru, M., Sahm, F., Ahn, S.S., Chang, J.H., Kim, S.H.: The 2021 who classification for gliomas and implications on imaging diagnosis: part 1—key points of the fifth edition and summary of imaging findings on adult-type diffuse gliomas. Journal of Magnetic Resonance Imaging **58**(3), 677–689 (2023)

16. Pedano, N., Flanders, A.E., Scarpace, L., Mikkelsen, T., Eschbacher, J.M., Hermes, B., Sisneros, V., Barnholtz-Sloan, J., Ostrom, Q.: The cancer genome atlas low grade glioma collection (tcga-lgg). (No Title) (2016)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
18. Scarpace, L., Flanders, A., Jain, R., Mikkelsen, T., Andrews, D.: Data from rembrandt [data set]. the cancer imaging archive (2019)
19. Scarpace, L., Mikkelsen, T., Cha, S., Rao, S., Tekchandani, S., Gutman, D., Saltz, J.H., Erickson, B.J., Pedano, N., Flanders, A.E., et al.: The cancer genome atlas glioblastoma multiforme collection (tcga-gbm). The Cancer Imaging Archive (2016)
20. Shah, N., Feng, X., Lankerovich, M., Puchalski, R., Keogh, B.: Data from ivy glioblastoma atlas project (ivygap)[data set]. The Cancer Imaging Archive (2016)
21. Smith, L., Gal, Y.: Understanding measures of uncertainty for adversarial example detection. arXiv preprint arXiv:1803.08533 (2018)
22. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)
23. van der Voort, S.R., Incekara, F., Wijnenga, M.M., Kapsas, G., Gahrmann, R., Schouten, J.W., Nandoe Tewarie, R., Lycklama, G.J., De Witt Hamer, P.C., Eijgelaar, R.S., et al.: Combined molecular subtyping, grading, and segmentation of glioma using multi-task deep learning. Neuro-oncology **25**(2), 279–289 (2023)
24. Zhou, T., Ruan, S., Canu, S.: Multi-task learning for segmentation and classification of tumors in 3d medical images. In: Machine Learning in Medical Imaging. pp. 339–347. Springer (2019)