Opportunities and Challenges for Deep Learning in Brain Lesions

Jay Patel[1,2], Ken Chang[1,2], Syed Rakin Ahmed[1,3,4], Ikbeom Jang[1,5], and Jayashree Kalpathy-Cramer[1,5(✉)]

[1] Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Boston, MA, USA
jkalpathy-cramer@mgh.harvard.edu
[2] Massachusetts Institute of Technology, Cambridge, MA, USA
[3] Harvard Graduate Program in Biophysics, Harvard Medical School, Harvard University, Cambridge, MA, USA
[4] Geisel School of Medicine at Dartmouth, Dartmouth College, Hanover, NH, USA
[5] Department of Radiology, Harvard Medical School, Boston, MA, USA
https://qtim-lab.github.io/

Abstract. In recent years, deep learning techniques have shown potential for incorporation in many facets of the medical imaging pipeline, from image acquisition/reconstruction to segmentation/classification to outcome prediction. Specifically, these models can help improve the efficiency and accuracy of image interpretation and quantification. However, it is important to note the challenges of working with medical imaging data, and how this can affect the effectiveness of the algorithms when deployed. In this review, we first present an overview of the medical imaging pipeline and some of the areas where deep learning has been used to improve upon the current standard of care for brain lesions. We conclude with a section on some of the current challenges and hurdles facing neuroimaging researchers.

Keywords: Deep learning · Imaging · Neuro-oncology

1 Introduction

The advent of noninvasive imaging technologies such as magnetic resonance imaging (MRI) and computed tomography (CT) has revolutionized medicine, enabling clinicians to make informed decisions for diagnosis, surgical planning, and treatment response assessment. In recent years, access to larger and more comprehensive repositories of patient imaging data along with advances in computational resources has closed the gap between machine and human. Specifically, artificial intelligence (AI) based algorithms can now interpret imaging scans at the level of expert clinicians.

While the majority of current research is focused on the interpretation of medical imaging, upstream aspects of the imaging pipeline are primed to be

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
A. Crimi and S. Bakas (Eds.): BrainLes 2021, LNCS 12962, pp. 25–36, 2022.
https://doi.org/10.1007/978-3-031-08999-2_2

improved via AI as well. Briefly, the imaging pipeline can be broken into three steps: 1) acquisition/reconstruction, 2) analysis, and 3) interpretation (Fig. 1). The first step in the pipeline is image acquisition, wherein raw data that is not visually interpretable by a human is gathered. This raw data must then be reconstructed into an anatomical image. For example, when performing an MRI, data is acquired at specific frequency bands in the Fourier domain and is then reconstructed into the spatial domain for human interpretation. The next step is image analysis, wherein both qualitative and quantitative information regarding the pathology of interest is gleaned. Finally, the last step is image interpretation, wherein a trained clinician makes judgments regarding tasks such as diagnosis or treatment planning. For instance, given a tumor's volume and location in the brain, a clinician may decide to utilize radiation in lieu of surgery. This general workflow is shown in Fig. 1.

Even though imaging has been used in clinical practice for many decades, problems still persist that hamper its efficacy. For example, patient motion during image acquisition may render a scan unreadable since most reconstruction algorithms are incapable of correcting for motion blur. Even when a scan is perfectly acquired, the complete manual analysis may be too time-consuming to be feasible, resulting in metrics such as the response assessment in neuro-oncology (RANO) criteria [49] to be used as a proxy measure for full volumetric tumor burden. In the following sections, we will discuss some of the problems that arise in the standard imaging pipeline and the opportunities that exist to utilize advanced deep learning techniques to improve the efficiency of each of these steps.

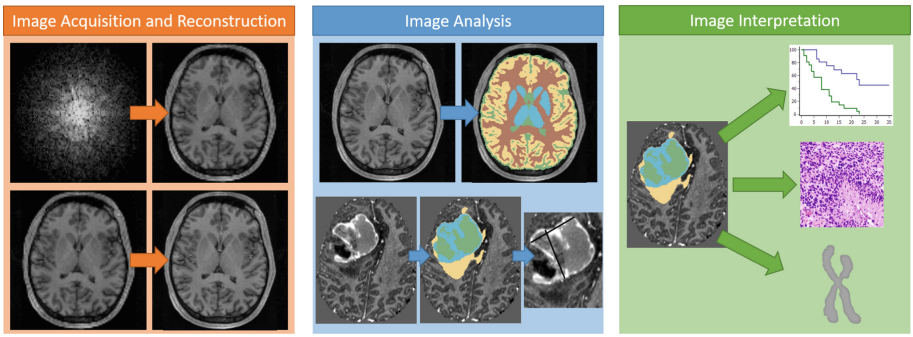

Fig. 1. The imaging pipeline is made up of three main components: 1) acquisition/reconstruction, 2) analysis, and 3) interpretation. Image acquisition and reconstruction entails converting sensor domain data into the spatial domain. Image enhancement/super-resolution can either be done in parallel with reconstruction, or as a separate step. Image analysis for brain lesions includes anatomical and tumor segmentations, along with automatic RANO measures. Finally, image interpretation includes survival prediction, tumor histopathologic grading, and radiogenomic correlations, among other applications.

2 Opportunities in Image Acquisition and Reconstruction

The first step in the imaging pipeline is acquisition and reconstruction. When an image is acquired, it is encoded into an intermediate representation of the image target known as the sensor domain. For this intermediate representation to lead to an image, the function or encoding method used to encode the image into the sensor domain must be inverted in a process known as reconstruction. Image reconstruction is required for many kinds of medical imaging, including MRI, CT, and positron emission tomography. Existing approaches for reconstruction are incomplete since noisy, real-world data often precludes knowledge of an exact inverse transform. To overcome the problems with conventional image reconstruction methods, researchers have in recent years begun testing deep learning-based approaches.

One example of a unified framework for deep learning-based image reconstruction is Automated Transform by Manifold Approximation (AUTOMAP) [54]. AUTOMAP is implemented with a deep neural network architecture composed of fully connected layers followed by convolutional layers. Zhu et al. generated training data by taking a large set of images from a natural scene and inverse encoding them into the sensor domain with the desired encoding function to create a paired dataset. The network was then trained in a supervised learning manner, enabling the network to learn the optimal strategies for image reconstruction. The trained neural network was then applied to MRI images of the human brain. Surprisingly, they found that training on images of objects such as animals and plants (rather than MRI of the brain) still allowed for accurate reconstruction of brain MRI images for three of the four commonly used encoding schemes they tested, which implies the robustness of their approach. Moreover, AUTOMAP implicitly learned how to denoise imaging, removing common artifacts such as zipper artifacts that would have persisted if the image had been reconstructed by conventional methods. When tested against simulated data using known ground truth, AUTOMAP reconstructed images were thus more accurate and had a higher SNR. The study opened opportunities for adopting deep learning approaches for image reconstruction of a wide range of different imaging modalities without having to learn complex, modality-specific physics.

Another groundbreaking reconstruction model for accelerated MRI is the Variational Network (VN) [21]. One of the biggest concerns about using learning-based reconstruction methods in the clinical workflow was that they may not preserve pathology-related features that are rare or specific to certain patients. For efficient and accurate reconstruction of MRI data, they proposed a trainable formulation for accelerated parallel imaging-based MRI reconstruction inspired by variational methods and deep learning. VN incorporates key concepts from compressed sensing, formulated as a variational model within a deep learning approach. This approach is designed to learn a complete reconstruction procedure for complex multi-channel MRI data, including all free parameters that need to be established empirically. Hammernik et al. train the model on a complete clinical protocol for musculoskeletal imaging, evaluate its performance on various accelerating factors, and train on both normal and pseudo-random Cartesian

2D sampling. Using clinical patient data, they investigated the ability of the VN approach to preserve unique pathologies not included in the training dataset. Surprisingly, it was able to preserve important features not present in the training data, outperforming conventional reconstructions for a range of pathologies while providing unprecedented reconstruction speeds.

3 Opportunities in Image Analysis

The second step in the imaging pipeline is analysis. Here, information necessary for downstream tasks is either manually or automatically extracted. Medical image analysis covers a wide span of topics, including but not limited to anatomical segmentation and volumetric quantification, extraction of parameter maps from diffusion/perfusion imaging, and groupwise population analyses. In this section, we will specifically look at examples involving brain tumor segmentation.

Primary and metastatic brain tumors account for nearly 200,000 new cases in the US every year, and imaging plays a crucial role in optimizing patient care [43,48]. Segmentation of tumor boundaries is a necessary component for successful surgical and radiotherapy treatment planning [14]. Unfortunately, tumor segmentation is a challenging task requiring substantial domain expertise. Furthermore, as many studies have shown, motion artifacts, field inhomogeneities, and differences in imaging protocols both within and across medical institutions lead to non-negligible amounts of human error as well as significant amounts of intra- and inter-rater variability [31].

To combat these issues, researchers have turned to deep learning as it has the potential to produce accurate and reproducible results many orders of magnitude faster than can be accomplished manually. The shift to trainable AI is being further encouraged by the release of open-source datasets with high-quality annotations such as that from the Multimodal Brain Tumor Segmentation Challenge (BraTS) [4,6–8,33].

Variations of 3D U-Nets [46] have provided state-of-the-art results for segmentation of primary brain tumors. For example, Myronenko won the 2018 BRATS challenge utilizing an asymmetrical residual U-Net, where most of the trainable parameters of the model resided in the encoder. Furthermore, in contrast to the standard U-Net framework which uses four or five downsampling operations in the encoder, he applied only three in order to preserve spatial context [36]. Other modifications to the U-Net structure have also been used with success. Jiang et al. won the 2019 challenge using a two-stage cascaded asymmetrical residual U-Net, where the second stage of their cascade was used to refine the coarse segmentation maps generated by the first stage [27]. The second place that year was awarded to Zhao et al., who utilized dense blocks along with various optimization strategies such as variable patch/batch size training, heuristic sampling, and semi-supervised learning [52]. It is important to note that while architectural modifications to the U-Net can provide performance boosts, they are not always necessary. Indeed, Isensee et al. won the 2020 challenge

with their architecture coined "No New-Net", highlighting that a vanilla U-Net coupled with excellent training and optimization strategies can still achieve state-of-the-art results. Moreover, they achieved an average testing set dice score of 88.95% for whole tumor segmentation, achieving segmentation performance indistinguishable from human experts [25].

Similar strategies have been shown to work for metastatic brain tumors, which present additional hurdles compared to primary brain tumors. Patients with metastases often present with more than one target lesion along with micro-metastases spread systemically across the brain parenchyma. Micro-metastases are particularly challenging to segment due to their size and limited contrast enhancement. Various approaches have been proposed, from two-stage detection/segmentation pipelines to modifications of the loss function. While these approaches have yielded some improvement in performance, much work is still needed. For example, Zhou et al. developed a two-stage pipeline consisting of a detection stage followed by a segmentation stage. While they reported an excellent dice score of 87% on large metastases ($\geq$6 mm), their results dropped to just 17% for micro-metastases ($<$3 mm) [53]. This trend is seen in other studies as well [11,47], indicating the strong need for better segmentation algorithms for brain metastases cases.

Longitudinal measurement of lesion burden is the basis for treatment response assessment. While volumetric measurement would be the ideal metric for lesion burden, the aforementioned issues with manual tumor segmentation necessitate the use of proxy measures such as RANO. RANO for gliomas is defined as the product of the maximum bidimensional diameters of the largest axial cross-section of the tumor on MRI [49]. Even this metric is subject to inter-rater variability, since different raters may choose differing slices based on their subjective assessment of which axial slice has the largest tumor area. To automate this process, Chang et al. developed a tool called AutoRANO which used the outputs of a segmentation model capable of running on post-operative imaging to derive RANO measurements. He noted that AutoRANO had a higher correlation with manual contrast-enhancing volume than did manual RANO measures performed by expert radiologists, suggesting that AutoRANO may be a more accurate measure of tumor burden than manual RANO [14]. Similar work has been done to automate bi-directional measurements for other tumor types, with equally promising results [40].

4 Opportunities in Image Interpretation

The final step in the imaging pipeline is interpretation. From a machine learning standpoint, this is often framed as a classification problem. For example, with regards to brain tumors, image classification tasks include but are not limited to identifying subtypes, predicting pseudo-progression versus true progression, ascertaining tumor malignancy status, and identifying treatment responders. Indeed, two key facets in which the rise of AI has been particularly exciting include radiogenomics and survival prediction.

Radiogenomics refers to the correlation between imaging features and specific gene expression patterns/molecular profiles of tumors. Such approaches have mainly been studied for primary gliomas, but interest is accruing to replicate such studies for brain metastases and spinal cord tumors as well. The ability to predict molecular marker status noninvasively is important since a priori knowledge of the mutational status of key genes together with radiographic suspicion of a neoplasm might favor early intervention and/or mutation-specific therapeutic interventions. In the case of gliomas, the MGMT gene, which codes for an enzyme responsible for DNA repair following alkylating agent chemotherapy, may be silenced by methylation of its promoter during tumor development, thereby preventing repair of DNA damage. This increases the potential effectiveness of alkylating agent chemotherapy for these patients [23]. In order to demonstrate that a deep learning model could predict MGMT methylation status from imaging without the need for explicitly providing a tumor segmentation, Korfiatis et al. [30] trained three deep residual neural networks of varying sizes on a training dataset of 110 patients with T2-weighted MRI, artificially increasing the size of this dataset by splitting all 3D imaging into 2D axial slices. Here, the authors found that deeper, more parametrized networks produce better results, with their ResNet50 model achieving an accuracy of 94.9% on the test set (45 patients with 2612 slices). Another key gene conferring longer survival in glioma patients is IDH, which in its wild-type form codes for an enzyme responsible for the conversion of isocitrate to a-ketoglutarate in the Krebs cycle. Gliomas harboring the IDH1/2 mutation carry a significantly increased overall survival than the corresponding wild type [12]. Chang et al. [12] used a similar methodology as Korfiatis et al. [30] for the prediction of IDH status, utilizing a residual neural network with 2D inputs. In this case, the network required a predefined tumor segmentation, since it was trained on cropped tumor images only. The authors performed exceptional multi-institutional evaluation, acquiring data from three different sites, and reporting a final accuracy and AUC on a testing set of 147 patients of 87.6% and 0.95, respectively. Similarly, Akkus et al. [2] focused on the prediction of 1p19q co-deletion, a highly prognostic molecular marker associated with longer survival in low-grade glioma (LGG) patients. With only 387 slices in the training data, the authors noted extreme overfitting, initially seeing perfect training sensitivity, specificity, and accuracy. To mitigate this, they made use of data augmentation techniques such as random translations, rotations, and flips, resulting in an increased final test set accuracy from 63.3% to 87.7%. Additionally, Chang et al. [16] aimed to integrate prediction of MGMT methylation status, IDH mutation status, and 1p19q codeletion into a single residual network. After five-fold cross-validation on their dataset of 259 patients (5259 slices), they achieved mean accuracy of 83%, 94%, and 92%, respectively, on the three tasks. Finally, MGMT methylation status prediction from MRI was a key component of the BraTs 2021 challenge, in which many teams utilized machine learning techniques for non-invasive assessment.

Survival analysis is a technique employed in cohort and other longitudinal studies to predict the time it takes for a particular event to occur. In these stud-

ies, individuals are followed from an initial observation (e.g. study enrollment, time of diagnosis/treatment) until the occurrence of a subsequent event (e.g., death, disease, relapse) or until follow-up is no longer possible. Depending on what event is used, the time between the two is denoted as progression-free survival or overall survival (OS) [39]. Survival analyses of brain tumors have utilized both radiomics based approaches and deep learning, as well as an integration of the two. Ujjwal et al. [5] proposed a three-step framework for OS prediction which involved segmentation, radiomic feature extraction, and a survival prediction model to stratify patients into three survival groups (short-, mid-, and long-term survivors) and to predict OS. This approach achieved accuracy scores of 0.571 and 0.558 on validation and testing cohorts of 53 and 130 cases respectively. Finally, Han et al. [22] incorporated both hand-crafted radiomics features and deep features generated by a pretrained CNN on a dataset of 178 high-grade glioma patients (50 local, and 128 from TCGA), applying feature selection and Elastic Net-Cox modeling to classify patients into short- and long-term survivors. This combined feature analysis framework resulted in a log-rank test p-value of <0.001 for the 50 patient local cohort, and a corresponding value of 0.014 for the 128 patient TCGA cohort.

5 Challenges

As mentioned in the previous sections, there are significant opportunities to improve clinical decision-making and patient management using AI. However, it is important to keep in mind certain caveats and challenges to developing effective deep learning models for healthcare applications. First, it is important to acknowledge the brittleness of deep learning models, or in other words, the lack of generalizability across different acquisition settings and patient populations [15,18]. For example, different hospitals may have MRI scanners with different field strengths or use different scanning protocols. Different hospitals may also admit patients of different age groups or racial backgrounds. These institutional differences are further exacerbated by the fact that many medical datasets are small, either due to rare pathology, costly human annotations, or simply due to difficulty in extracting data from antiquated electronic medical record systems. Indeed, empirical studies have shown that there is a drop in the performance of deep learning models for brain lesions when evaluated at institutions different from the ones in which they were trained [3,44]. One approach to handle the issue of generalizability is to accumulate large quantities of diverse, multi-institutional patient data. However, logistical issues, as well as patient privacy concerns may render this impractical. Another approach involves fine-tuning the existing model on a small quantity of new data when there is dataset shift [44]. More generally, continuous learning methods allow models to be "living" and to be refined as the data changes [42]. Other approaches include methods to adapt either the data or the model itself to be able to handle new domains with approaches under the umbrella of domain adaptation [28,51]. If large quantities of data are available, but not shareable between institutions, distributed learning approaches can be

used to train models without the need to share patient data, overcoming patient privacy barriers [13,45].

Another major challenge facing trainable AI models is the lack of definitive ground truth. For example, for the segmentation of brain lesions, there is often subjectivity involved in determining tumor boundaries, especially for lesions that are diffusely edematous. Similarly, the boundaries of contrast enhancement may be ambiguous as well due to the presence of necrotic regions. This subjectivity is primarily due to the spatial resolution limitations of MRI, which makes categorizing tumor components into discrete bins of necrosis, enhancing, or edema difficult. Thus, it is unsurprising that there is significant intra- and inter-rater variability for neuroimaging related segmentation [10,17,34]. In the case of radiogenomic prediction using ground truth from a single biopsy site, there is also uncertainty stemming from regional intra-lesional genetic heterogeneity of tumors [37,38,41,50]. This is further compounded by multi-focal lesions, which can also display genetic heterogeneity across lesions from the same patient [1]. For other prediction tasks, such as prognostic assessment, there may be significant confounders that are not incorporated into the inputs, such as degree of resection and chemotherapeutic regimen. Taken together, the clinical utility and efficacy of machine learning models may be limited if there is no way to handle uncertainty within the data. One way to potentially mitigate this problem is to utilize deep learning methods that can estimate uncertainty to provide multiple possible outputs, mimicking variability by different clinicians [29]. Another viable approach is to train networks to directly report a measure of uncertainty, thus allowing clinicians to stratify network outputs by the degree of confidence [24,32]. This would enable flagging of highly uncertain cases for further manual expert review.

A final challenge that should be mentioned is the reproducibility of deep learning studies for neuroimaging. With the rapid pace of advances within the field, new research often builds upon previous work to yield improvements in performance. However, without the release of code, much effort would need to be devoted to reproducing previously published results for further evaluation and development [20]. As such, there has been a growing trend towards the release of open-source frameworks for medical AI to allow for greater collaboration within the research and clinical communities [9,19,26]. On a similar front, the public release of code is increasingly becoming the expectation for publication [35]. However, this is not without potential concerns of its own, since it may result in the accidental leaking of protected patient health information or may deter the commercialization of research.

6 Conclusion

Significant progress has been made in the last few years to automate and increase the efficiency of all steps in the imaging pipeline via the use of deep learning. Specifically, greater accessibility to large-scale multi-institutional datasets and better computational resources together have led to advances in image reconstruction, analysis, and interpretation. Our review has highlighted some of the

exciting AI research being performed at each of these steps in the imaging pipeline, and some challenges and pitfalls that all researchers working with neuroimaging data must acknowledge.

References

1. Abou-El-Ardat, K., et al.: Comprehensive molecular characterization of multifocal glioblastoma proves its monoclonal origin and reveals novel insights into clonal evolution and heterogeneity of glioblastomas. Neuro Oncol. **19**(4), 546–557 (2017)
2. Akkus, Z., et al.: Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence. J. Digit. Imaging **30**(4), 469–476 (2017). https://doi.org/10.1007/s10278-017-9984-3, https://pubmed.ncbi.nlm.nih.gov/28600641, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5537096/
3. AlBadawy, E.A., Saha, A., Mazurowski, M.A.: Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. Med. Phys. **45**(3), 1150–1158 (2018)
4. Baid, U., et al.: The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021)
5. Baid, U., et al.: Overall survival prediction in glioblastoma with radiomic features using machine learning. Front. Comput. Neurosci. **14**, 61 (2020). https://doi.org/10.3389/FNCOM.2020.00061/BIBTEX
6. Bakas, S., et al.: Segmentation labels for the pre-operative scans of the TCGA-GBM collection (2017)
7. Bakas, S., Akbari, H., Sotiras, A., et al.: Segmentation labels for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017)
8. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data **4**(1), 1–13 (2017)
9. Beers, A., et al.: DeepNeuro: an open-source deep learning toolbox for neuroimaging. Neuroinformatics **19**(1), 127–140 (2021)
10. Bø, H.K., Solheim, O., Jakola, A.S., Kvistad, K.A., Reinertsen, I., Berntsen, E.M.: Intra-rater variability in low-grade glioma segmentation. J. Neuro-oncol. **131**(2), 393–402 (2017)
11. Bousabarah, K., et al.: Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data. Radiat. Oncol. **15**(1) (2020). https://doi.org/10.1186/s13014-020-01514-6
12. Chang, K., et al.: Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging. Clin. Cancer Res. **24**(5), 1073–1081 (2018). https://doi.org/10.1158/1078-0432.CCR-17-2236, https://clincancerres.aacrjournals.org/content/24/5/1073
13. Chang, K., et al.: Distributed deep learning networks among institutions for medical imaging. J. Am. Med. Inform. Assoc. **25**(8), 945–954 (2018)
14. Chang, K., et al.: Automatic assessment of glioma burden: a deep learning algorithm for fully automated volumetric and bidimensional measurement. Neuro-oncology **21**(11), 1412–1422 (2019). https://doi.org/10.1093/neuonc/noz106
15. Chang, K., et al.: Multi-institutional assessment and crowdsourcing evaluation of deep learning for automated classification of breast density. J. Am. Coll. Radiol. **17**(12), 1653–1662 (2020)

16. Chang, P., et al.: Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. Am. J. Neuroradiol. **39**, 1201–1207 (2018). https://doi.org/10.3174/ajnr.A5667, http://www.ajnr.org/content/early/2018/05/10/ajnr.A5667
17. Deeley, M., et al.: Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. Phys. Med. Biol. **56**(14), 4557 (2011)
18. Finlayson, S.G., et al.: The clinician and dataset shift in artificial intelligence. N. Engl. J. Med. **385**(3), 283 (2021)
19. Gibson, E., et al.: NiftyNet: a deep-learning platform for medical imaging. Comput. Methods Programs Biomed. **158**, 113–122 (2018)
20. Haibe-Kains, B., et al.: Transparency and reproducibility in artificial intelligence. Nature **586**(7829), E14–E16 (2020)
21. Hammernik, K., et al.: Learning a variational network for reconstruction of accelerated MRI data. Magn. Reson. Med. **79**(6), 3055–3071 (2018). https://doi.org/10.1002/mrm.26977
22. Han, W., et al.: Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas. Am. J. Neuroradiol. **41**(1), 40–48 (2020). https://doi.org/10.3174/AJNR.A6365, http://www.ajnr.org/content/41/1/40, http://www.ajnr.org/content/41/1/40.abstract
23. Hegi, M.E., et al.: MGMT gene silencing and benefit from temozolomide in glioblastoma. N. Engl. J. Med. **352**(10), 997–1003 (2005). https://doi.org/10.1056/NEJMOA043331/SUPPL_FILE/997SA1.PDF, https://www.nejm.org/doi/full/10.1056/nejmoa043331
24. Hoebel, K., et al.: An exploration of uncertainty information for segmentation quality assessment. In: Medical Imaging 2020: Image Processing, vol. 11313, p. 113131K. International Society for Optics and Photonics (2020)
25. Isensee, F., Jaeger, P.F., Full, P.M., Vollmuth, P., Maier-Hein, K.H.: nnU-Net for brain tumor segmentation (2020)
26. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
27. Jiang, Z., Ding, C., Liu, M., Tao, D.: Two-stage cascaded U-Net: 1st place solution to BraTS challenge 2019 segmentation task. In: Crimi, A., Bakas, S. (eds.) BrainLes 2019. LNCS, vol. 11992, pp. 231–241. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-46640-4_22
28. Kamnitsas, K., et al.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.-T., Shen, D. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 597–609. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_47
29. Kohl, S., et al.: A probabilistic U-Net for segmentation of ambiguous images. In: Advances in Neural Information Processing Systems, vol. 31 (2018)
30. Korfiatis, P., Kline, T.L., Lachance, D.H., Parney, I.F., Buckner, J.C., Erickson, B.J.: Residual deep convolutional neural network predicts MGMT methylation status. J. Digit. Imaging **30**(5), 622–628 (2017). https://doi.org/10.1007/s10278-017-0009-z
31. Mazzara, G.P., Velthuizen, R.P., Pearlman, J.L., Greenberg, H.M., Wagner, H.: Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. Int. J. Radiat. Oncol. Biol. Phys. **59**(1), 300–312 (2004). https://doi.org/10.1016/j.ijrobp.2004.01.026

32. Mehta, R., et al.: Qu-BraTS: MICCAI BraTS 2020 challenge on quantifying uncertainty in brain tumor segmentation-analysis of ranking metrics and benchmarking results. arXiv preprint arXiv:2112.10074 (2021)
33. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694
34. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2014)
35. Mongan, J., Moy, L., Kahn, C.E., Jr.: Checklist for artificial intelligence in medical imaging (claim): a guide for authors and reviewers. Radiol. Artif. Intell. **2**(2), e20029 (2020)
36. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization (2018)
37. Parker, N.R., et al.: Intratumoral heterogeneity identified at the epigenetic, genetic and transcriptional level in glioblastoma. Sci. Rep. **6**(1), 1–10 (2016)
38. Parker, N.R., Khong, P., Parkinson, J.F., Howell, V.M., Wheeler, H.R.: Molecular heterogeneity in glioblastoma: potential clinical implications. Front. Oncol. **5**, 55 (2015)
39. Peacock, J., Peacock, P.: Oxford Handbook of Medical Statistics. Oxford University Press, New York (2010) https://doi.org/10.1093/MED/9780199551286.001.0001
40. Peng, J., et al.: Deep learning-based automatic tumor burden assessment of pediatric high-grade gliomas, medulloblastomas, and other leptomeningeal seeding tumors. Neuro Oncol. **24**(2), 289–299 (2022)
41. Pfisterer, W.K., et al.: Diagnostic and prognostic significance of genetic regional heterogeneity in meningiomas. Neuro Oncol. **6**(4), 290–299 (2004)
42. Pianykh, O.S., et al.: Continuous learning AI in radiology: implementation principles and early applications. Radiology **297**(1), 6–14 (2020)
43. Porter, K.R., McCarthy, B.J., Freels, S., Kim, Y., Davis, F.G.: Prevalence estimates for primary brain tumors in the United States by age, gender, behavior, and histology. Neuro-oncology **12**(6), 520–527 (2010). https://doi.org/10.1093/neuonc/nop066
44. Rauschecker, A.M., et al.: Interinstitutional portability of a deep learning brain MRI lesion segmentation algorithm. Radiol. Artif. Intell. **4**(1), e200152 (2021)
45. Rieke, N., et al.: The future of digital health with federated learning. NPJ Digit. Med. **3**(1), 1–7 (2020)
46. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. CoRR abs/1505.0 (2015) http://arxiv.org/abs/1505.04597
47. Rudie, J.D., et al.: Three-dimensional U-Net convolutional neural network for detection and segmentation of intracranial metastases. Radiol. Artif. Intell. **3**(3), e200204 (2021). https://doi.org/10.1148/ryai.2021200204
48. Tabouret, E., Chinot, O., Metellus, P., Tallet, A., Viens, P., Goncalves, A.: Recent trends in epidemiology of brain metastases: an overview. Anticancer Res. **32**(11), 4655–4662 (2012). https://ar.iiarjournals.org/content/32/11/4655
49. Wen, P.Y., et al.: Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. J. Clin. Oncol. **28**(11), 1963–1972 (2010)
50. Wenger, A., Ferreyra Vega, S., Kling, T., Bontell, T.O., Jakola, A.S., Carén, H.: Intratumor DNA methylation heterogeneity in glioblastoma: implications for DNA methylation-based classification. Neuro Oncol. **21**(5), 616–627 (2019)

51. Yan, W., et al.: MRI manufacturer shift and adaptation: increasing the generalizability of deep learning segmentation for MR images acquired with different scanners. Radiol. Artif. Intell. **2**(4), e190195 (2020)
52. Zhao, Y.-X., Zhang, Y.-M., Liu, C.-L.: Bag of tricks for 3D MRI brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) BrainLes 2019. LNCS, vol. 11992, pp. 210–220. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-46640-4_20
53. Zhou, Z., et al.: MetNet: computer-aided segmentation of brain metastases in post-contrast T1-weighted magnetic resonance imaging. Radiother. Oncol. **153**, 189–196 (2020). Physics Special Issue: ESTRO Physics Research Workshops on Science in Development, https://doi.org/10.1016/j.radonc.2020.09.016, https://www.sciencedirect.com/science/article/pii/S016781402030788X
54. Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., Rosen, M.S.: Image reconstruction by domain-transform manifold learning. Nature **555**(7697), 487–492 (2018). https://doi.org/10.1038/nature25988