

Standardization of imaging methods for machine learning in neuro-oncology

Xiao Tian Li and Raymond Y. Huang^o

Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts, USA (X.T.L., R.Y.H.)

Corresponding Author: Raymond Y. Huang, MD, PhD, Department of Radiology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA (ryhuang@bwh.harvard.edu).

Abstract

Radiomics is a novel technique in which quantitative phenotypes or features are extracted from medical images. Machine learning enables analysis of large quantities of medical imaging data generated by radiomic feature extraction. A growing number of studies based on these methods have developed tools for neuro-oncology applications. Despite the initial promises, many of these imaging tools remain far from clinical implementation. One major limitation hindering the use of these models is their lack of reproducibility when applied across different institutions and clinical settings. In this article, we discuss the importance of standardization of methodology and reporting in our effort to improve reproducibility. Ongoing efforts of standardization for neuro-oncological imaging are reviewed. Challenges related to standardization and potential disadvantages in over-standardization are also described. Ultimately, greater multi-institutional collaborative effort is needed to provide and implement standards for data acquisition and analysis methods to facilitate research results to be interoperable and reliable for integration into different practice environments.

Keywords

- machine learning | neuro-oncology | radiomics | reproducibility | standardization.

The field of artificial intelligence (AI), particularly machine learning, enables analysis of medical imaging data and helps augment current diagnostic imaging practice in the era of increasing clinical demand.¹ AI algorithms are currently being researched to solve a number of clinical problems.² In neuro-oncology, machine learning tools have been utilized for the development of radiomics models for tumor diagnosis and therapeutic response using non-invasive imaging techniques such as MRI and positron emission tomography (PET). Furthermore, radiomics can sometimes be categorized into predefined feature-based techniques versus deep learning-based techniques that do not require predefinition or the intermediate feature extraction step.^{3,4} Despite the promises of this developing field, most radiomic tools remain far from clinical implementation due to a number of challenges such as technical complexity, poor study design, overfitting of data, and lack of standards for validating results.^{5,6} The purpose of this article is to review current efforts of standardization for neuroimaging, more specifically to discuss the advantages of standardized

methodology and reporting within radiomic and neuro-oncology research. Challenges related to standardization and potential disadvantages in over-standardization are also discussed.

Need for Imaging Standardization

Advances in radiomic and machine learning methods have led to rapidly increasing efforts to generate automated tools and to help solve clinical problems in medicine. The field of radiomics is a swiftly developing novel technique in which image-based phenotypes of lesions including shape, size, and texture are extracted from medical images.⁷ A growing number of studies based on radiomic and machine learning methods have aimed at several neuro-oncology applications, including preoperative classification of central nervous system neoplasms including their histology, grade, and molecular signatures, as well as

evaluating treatment response for better determination of treatment efficacy and prediction of subsequent clinical outcome. For example, machine learning models based on radiomic features of preoperative MRI have been developed to predict IDH1/2 mutations in both high- and low-grade gliomas.^{8,9} Classification models have also been developed to distinguish “pseudoprogression” from true tumor progression in glioblastomas receiving chemoradiation treatment.¹⁰ Other studies have shown that machine learning classifier can predict the tumor type of brain metastases.^{11,12} Studies in radiomics with additional modalities such as PET have gained increasing attention and shown to differentiate brain tumor or metastasis versus radiation injury,^{13,14} lymphoma versus glioblastoma,¹⁵ or predict underlying mutational status.¹⁶ Advanced MRI techniques have also incorporated radiomics such as MR perfusion for the identification of pseudoprogression in glioblastoma.¹⁷

Many researchers are increasingly automating the pipelines to generate radiomic models in order for easier integration into clinical use. A majority of these pipelines consist of numerous steps such as imaging acquisition and postprocessing. When examining the methods by which most of these models are constructed, one can readily discover that there is wide variability in every step leading to the final model. Variations in methodology among studies using radiomic approach often include patient demographics, patient selection criteria, patient cohort design, molecular or pathology data generation, imaging acquisition, image preprocessing, lesion segmentation, radiomic feature extraction, and machine learning computation. Even a small change in one or more of these key steps can result in significant differences in model accuracies and reproducibility. Due to these variations in the methodology, it is often difficult to show that one set of results are better or worse compared to the others by only directly comparing accuracies. Similarly, it is also not clear whether the results from one study can be applied to data obtained at another institution. Finally, a significant part of the challenge in determining the reproducibility of machine learning models is the frequent lack of sufficient methodology details in the published studies.

Due to the rapid development and application of radiomics methods, there is a lack of a standardized approach to analysis and reporting. To date, only a small number of studies that developed radiomic models for cancer care investigated their repeatability and reproducibility and provided details related to imaging processing and feature extraction.¹⁸ While researchers often employ cross-validation or include independent validation datasets to show reproducibility of trained radiomic models,¹⁹ the generalizability of these models can still be limited if they are applied to datasets that are sufficiently different from those that are used in training and validation. Thus, the development and implementation of standard imaging data acquisition methods can reduce variations that limit generalizability.

Sources of Variability in Radiomic Methodology

Radiomics is a complex multistep process that requires a close examination of each of these steps in order to

investigate sources of variability and to improve the reproducibility of the end results.

Patient Cohort and Study Design

Within a different patient population, the type and prevalence of certain imaging findings and underlying pathology may be variable. One study with a homogenous patient cohort may get completely different results compared to another patient cohort with a more diverse underlying population. The inclusion and exclusion criteria for patient selection should therefore be clearly defined and potential biases related to these selection criteria should be discussed. In neuro-oncology, recent updates to the World Health Organization CNS tumor classification include molecular data in defining tumor subtypes.²⁰ It is important to include details related to these molecular markers for the study population so that the resultant radiomic/AI models can be correctly applied clinically according to the updated criteria. Due to these changes in disease definitions, one should also be cautious in comparing radiomic models that were constructed using different disease criteria.

Imaging Acquisition

Imaging acquisition methods including scanner types and sequence parameters can affect radiomic feature calculations. Studies show that computed tomography (CT) features may be nonreproducible and redundant by comparing 5 different scanners with the same CT parameters.²¹ Unlike CT where pixel intensities are directly related to tissue attenuation, the intensity values derived from MRI typically do not have physical meaning and widely vary depending on acquisition parameters and this results in significant variations among radiomic features derived from MRI.²² Based on MRI phantom analyses of commonly used MRI sequences, a study showed that FLAIR sequence is the most robust, although only 15 out of 45 radiomic features showed excellent robustness and reproducibility across all sequences.²³ Evaluation of imaging protocol variability using a PET phantom demonstrates at least good reliability when parameters are in clinical range though there should be caution combining patients from different scanners into the same radiomics dataset.²⁴ Another study using a PET phantom concluded that the repeatability of PET radiomic features depends on a number of factors and recommends high-level image acquisition and preprocessing standardization.²⁵

Imaging Preprocessing

Several imaging preprocessing steps are often performed prior to radiomic feature extraction in order to minimize intensity-related variations. These include noise filtering, inhomogeneity correction, and intensity normalization. When preprocessing steps are applied to MRI dataset, the robustness of radiomic features has improved based on phantom analyses²⁶ as well as data from patients with glioblastoma.²⁷

Intensity standardization or normalization is a preprocessing technique that changes the values of pixel intensity using a predefined algorithm in order to reduce variations from data acquisition. There are several normalization methods that can affect the reproducibility of radiomic features.²⁸ Histogram matching is a commonly used normalization method and has been shown to contribute more than other preprocessing steps in reducing feature variability of conventional MRI sequences.²⁹

Tumor Segmentation

Segmentation of lesions is an important step to focus feature extraction on imaging pixels that are most relevant to the pathological process of interest and excluding other tissue types such as normal brain or treatment-related changes. Different lesion segmentation methods can introduce variability in subsequent feature calculations. For manual segmentation, it has been shown that interobserver variability in segmentation does impact radiomic analysis.³⁰ Newer techniques of segmentation are developed to shorten the processing time. Semiautomatic techniques such as using software like 3D Slicer and TumorPrism3D have shown to reduce variability between observers.^{31,32} One study claims that interactive methods decrease variability compared to semiautomatic methods in segmenting glioblastomas.³³ The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) introduced a large public dataset for the purpose of evaluating brain tumor segmentation algorithms with annual challenges.³⁴ This benchmarking resource allowed the testing of a number of algorithms including fully automated ones. The development of a fully automated segmentation technique would tremendously benefit clinical care and treatment planning. Initial results showed that different algorithms, including fully automated ones, showed promise for different subregions of segmentation, but no single algorithm was top tier in all of the subregions³⁴ in comparison to expert manually annotated results. Fully automated techniques are technically challenging partly due to variations in tumor structures and the displacement of normal structures due to mass effect.³⁴ Since 2015, deep learning-based models, specifically convolutional neural networks (CNN), have been the top-performing techniques based on BRATS benchmarking.³⁵ Many different deep learning techniques have been used including ensembles and a variety of neural networks.³⁶ Recently, CNN models trained using sequential MRI data provide automated segmentation of patients with glioblastoma in clinical trials and provide objective response assessment comparable to expert evaluation.^{37,38}

Radiomic Feature Extraction and Selection

Given a data type, some radiomic features show greater robustness compared to other features. Therefore, it is often advantageous to use the features with greater robustness to improve model reproducibility. In the textural analysis of MRI from patients with a brain tumor, many features derived from a standard radiomic library lack robustness.³⁹ Many radiomic features are shown to not be robust in the setting of different matrix size and dynamic

range configurations.⁴⁰ Without standardization of matrix size and dynamic range configurations, it would be challenging to obtain reliable and comparative results across different centers.³⁹ Initial evaluation of feature variability in FDG-PET showed that many of the radiomic features do have high retest and interobserver stability with increased robustness seen in features more stable in repeated PET imaging.⁴¹ Gray-level co-occurrence matrix and shape features have been shown to be least sensitive to PET imaging system variations.⁴²

Techniques have been developed to increase the robustness of radiomic features. One commonly used method is to introduce perturbations to images to help select robust features that are not susceptible to these perturbations.⁴³ Automated machine learning systems have also been developed to scan for novel features and optimize parameters.⁴⁴ There are methods to assess the stability and discriminatory capacity of radiomic features on apparent diffusion coefficient images.⁴⁵

Machine Learning Algorithms

As different machine learning approaches are used for neuroimaging, the types of models used will impact the variability. While it is possible that with sufficient optimization of machine learning hyperparameters, the performance of different algorithms can be equivalent. In practice, however, the fine-tuning process of these parameters can vary due to differences in expertise. The difference in the choice of classification method can be a dominant source of performance variation.⁴⁶ The recent development in automatic machine learning algorithms such as Tree-based Pipeline Optimization Tool can select the best performing machine learning algorithms or hybrid algorithms.⁴⁷ This automatic machine learning approach helps to remove the variation from machine learning algorithm selection and the differences in expertise between groups.

In recent years, deep learning-based radiomics have been rapidly developed. Whereas other techniques are generally based on predefined features, deep learning methods allow for automatic extraction of high-level features by using multiple layers of neural networks that imitate the functions of the human visual network.^{3,48} While this technique has the advantage of not requiring image segmentation or predefined features, it also increases the difficulty in performing step-wise quality control due to the intrinsic nature of complex, high-integrated neural network layers. In addition, deep learning techniques often need significantly larger datasets compared to traditional machine learning techniques to reduce the chance of overfitting due to a much greater number of features generated within the neural networks. The technique of transfer learning can help reduce the data requirement by using an artificial neural network model generated for another similar task as a basis for a new task.⁴⁹

Radiomic Model Validation

Since many commonly used machine learning algorithms can train a large number of features in proportion to the size of data, it is important to minimize the risk of overfitting.

At the minimum, the trained models should be evaluated in the internal dataset by cross-validation methods. Ideally, further validation using independent external cohorts can further strengthen the credibility of the radiomic models.^{50,51} In order for this validation to be meaningful, it is necessary to ensure that the study populations of the internal and external datasets are comparable. The details of both populations should be clearly described. Similarly, the reference standard used to supervise the training of a machine learning model should also be comparable to that used in the validation population.

Pros and Cons of Standardization

There are several benefits to the standardization of radiomic methodology. When imaging acquisition, preprocessing, and postprocessing steps are standardized, the variations of calculated radiomic features will be reduced, and the performance of radiomic models will be improved. Besides improving interoperability and reproducibility, standardization can help define the processes of pooling results from multiple institutions and reduce the chances of incompatible datasets that would lower the accuracy of radiomic models. By defining the methodology and reporting guidelines, a growing body of literature with standardized nomenclature can facilitate comparison of results.⁵²

However, is standardization always good? The fields of AI and radiomics are rapidly developing and consistently changing. New emerging techniques and algorithms are being published and implemented rapidly. By forcing researchers and organizations to follow a strict guideline, this may only be effective if the guidelines are updated to reflect the current development. If the guidelines are outdated, such efforts for standardization can potentially hinder innovation. Since technology is constantly improving, today's "standard" is likely going to be different from tomorrow. For these concerns, collaborative efforts should be dedicated to constantly reviewing and updating standard protocols.

Another potential downside for standardization is a more narrow range of imaging parameters or postprocessing methods that can limit the discovery of clinically useful biomarkers from "nonstandard" protocols. In other words, standardization facilitates the validation side of efforts but may impede the exploration of novel features. If the size of data collection is sufficiently large, a likely scenario due to ongoing efforts in open data sharing, this may obviate the need for restricting imaging parameters and allow a certain degree of data heterogeneity.

Current Efforts to Improve Standardization

Standardization efforts are being investigated throughout the field of AI. Leading organizations such as ISO and IEEE are involved in a broader sense as governing bodies.⁵³ Within neuroimaging, the American Society of Neuroradiology has initiated a task force to develop

training and educational programs related to AI in neuroimaging.⁵⁴ For example, one branch is addressing quality assurance issues in regard to standardization, reliability, and reproducibility and also working in conjunction with the National Institute of Standards and Technology.

Within the scope of cancer biomarker development, the most comprehensive effort to standardize the extraction of image biomarkers comes from Image Biomarker Standardization Initiative (IBSI). The IBSI helps define some of the reporting guidelines.⁵² The main challenge according to these initiatives is to help solidify the process of translating acquired imaging into high-throughput image markers. Before this initiative, there is felt to be a lack of available consensus-based guidelines. This effort provides guidelines into standardized image biomarker nomenclature and definitions, general image processing workflow, tools for software implementation, and reporting guidelines. Another article by the Cancer Research UK and the European Organization for Research and Treatment of Cancer described the creation of an imaging biomarker roadmap for cancer studies by assembling a group of experts to discuss the challenges of imaging biomarker validation and standardization.⁵⁵

Due to existing variations in imaging hardware, standardization of imaging acquisition is often challenging. Most ongoing efforts are to minimize acquisition variations by defining a standard range of acquisition parameters for major equipment vendors or platforms. For example, brain tumor imaging protocol (BTIP) was developed as a standard brain MRI protocol for imaging of primary brain gliomas⁵⁶ that are increasingly adopted in recent clinical trials. This standard protocol includes a 3DT1-weighted sequence that was adopted from a preexisting standard imaging sequence used for imaging of Alzheimer's disease.⁵⁷ More recently, BTIP-BM has been proposed as standard protocol for assessment of brain metastases.⁵⁸ Similar imaging acquisition criteria are created for PET imaging for glioma.⁵⁹ Here the challenge is in making a standard protocol that clinical groups of different practice settings are willing to adopt without clear proven clinical benefit. Thus, if a new standard protocol can be modified from another existing, commonly used standard protocol it will allow easier implementation.

The IBSI also helps to make sure that the feature values from different groups follow the same feature definition. There are open-source tools that conform to the IBSI such as pyradiomics.⁶⁰ Calculators, such as RadCaT, help provide a calculation for a large number of radiomic features that are in compliance with IBSI standards.⁷ Imaging biomarker explorer is a software platform that is designed to implement common radiomic tasks.⁶¹ This software has also been shown to be highly compliant to the IBSI standards.⁶² LIFEx is a radiomic feature calculator designed for multiple modalities including PET, SPECT, MRI, CT, and ultrasound which is at least partly compliant with IBSI standards.⁶³ Brain Cancer Imaging Phenomics Toolkit (brain-CaPTk) is developed as a type of cancer imaging phenomics toolkit for quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome.^{64,65} Even with the development of these software tools, there may still be significant variations in features among software packages and indicates the need for standardization.⁶⁶ Recently,

the IBSI standardized a set of 169 radiomic features, which could verify and calibrate different radiomics software.⁶⁷

The creation of open-access high-quality imaging datasets has been promoted as a way to improve the model training and validation process.⁵⁵ A popular repository that encourages the sharing of data is The Cancer Imaging Archive.⁶⁸ One paper recommends at least 10–15 patients for each feature evaluated.⁶⁹ Having sufficient high-quality public data can make this task much easier.

Despite the best effort for researchers to adopt these strategies to reduce variability, it is often not possible to completely control the numerous variables within each step in developing a machine learning-based imaging biomarker. Investigators should still provide details of all essential steps from acquisition to pre- and postprocessing in accordance with a standard reporting guideline (ie, IBSI) to maximize the ability for other researchers to reproduce research results.

Conclusions

Today, despite rapid advances in the development of AI technologies, major research organizations that perform AI research are often absent from the efforts in the standardization of imaging biomarker methodology.⁵³ Whether the research is performed in the industry or within academic institutions, successful AI and radiomics applications in clinical medicine require a clear demonstration of their reproducibility and generalizability. With greater collaborative efforts among different institutions for collecting data such as from oncology patients, standardization of data acquisition and analysis methods may facilitate research results to be interoperable and reliable for integration into different practice environments.

Funding

None.

Conflict of interest statement. None.

Authorship Statement. Writing and revision: X.T.L. and R.Y.H.

References

- Hainc N, Federau C, Stieltjes B, Blatow M, Bink A, Stippich C. The bright, artificial intelligence-augmented future of neuroimaging reading. *Front Neurol.* 2017;8(1):489.
- Davatzikos C. Machine learning in neuroimaging: progress and challenges. *Neuroimage.* 2019;197(1):652–656.
- Lohmann P, Kocher M, Ruge M, et al. PET/MRI radiomics in patients with brain metastases. *Front Neurol.* 2020;11(1):1.
- Parekh VS, Jacobs MA. Deep learning and radiomics in precision medicine. *Expert Rev Precis Med Drug Dev.* 2019;4(2):59–72.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* 2016;278(2):563–577.
- Sollini M, Antunovic L, Chiti A, Kirienco M. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. *Eur J Nucl Med Mol Imaging.* 2019;46(13):2656–2672.
- Pfaehler E, Zwanenburg A, de Jong JR, Boellaard R. RaCaT: an open source and easy to use radiomics calculator tool. *PLoS One.* 2019;14(2):e0212223.
- Zhang B, Chang K, Ramkissoon S, et al. Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro Oncol.* 2017;19(1):109–117.
- Zhou H, Chang K, Bai HX, et al. Machine learning reveals multimodal MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low- and high-grade gliomas. *J Neurooncol.* 2019;142(2):299–307.
- Jang BS, Jeon SH, Kim IH, Kim IA. Prediction of pseudoprogression versus progression using machine learning algorithm in glioblastoma. *Sci Rep.* 2018;8(1):12516.
- Kniep HC, Madesta F, Schneider T, et al. Radiomics of brain MRI: utility in prediction of metastatic tumor type. *Radiology.* 2019;290(2):479–487.
- Ortiz-Ramón R, Larroza A, Ruiz-España S, Arana E, Moratal D. Classifying brain metastases by their primary site of origin using a radiomics approach based on texture analysis: a feasibility study. *Eur Radiol.* 2018;28(11):4514–4523.
- Lohmann P, Kocher M, Ceccon G, et al. Combined FET PET/MRI radiomics differentiates radiation injury from recurrent brain metastasis. *Neuroimage Clin.* 2018;20(1):537–542.
- Hotta M, Minamimoto R, Miwa K. 11C-methionine-PET for differentiating recurrent brain tumor from radiation necrosis: radiomics approach with random forest classifier. *Sci Rep.* 2019;9(1):15666.
- Kong Z, Jiang C, Zhu R, et al. 18F-FDG-PET-based radiomics features to distinguish primary central nervous system lymphoma from glioblastoma. *Neuroimage Clin.* 2019;23(1):101912.
- Haubold J, Demircioglu A, Gratz M, et al. Non-invasive tumor decoding and phenotyping of cerebral gliomas utilizing multiparametric (18) F-FET PET-MRI and MR fingerprinting. *Eur J Nucl Med Mol Imaging.* 2019;47(1):1435–1445.
- Elshafeey N, Kotrotsou A, Hassan A, et al. Multicenter study demonstrates radiomic features derived from magnetic resonance perfusion images identify pseudoprogression in glioblastoma. *Nat Commun.* 2019;10(1):3170.
- Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys.* 2018;102(4):1143–1158.
- Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur Radiol Exp.* 2018;2(1):35.
- Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 2016;131(6):803–820.
- Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, et al. Radiomics of ct features may be nonreproducible and redundant: influence of CT acquisition parameters. *Radiology.* 2018;288(2):407–415.
- Ford J, Dogan N, Young L, Yang F. Quantitative radiomics: impact of pulse sequence parameter selection on MRI-based textural features of the Brain. *Contrast Media Mol Imaging.* 2018;2018(1):1729071.
- Baeßler B, Weiss K, Pinto Dos Santos D. Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. *Invest Radiol.* 2019;54(4):221–228.

24. Ger RB, Meier JG, Pahlka RB, et al. Effects of alterations in positron emission tomography imaging parameters on radiomics features. *PLoS One*. 2019;14(9):e0221877.
25. Pfaehler E, Beukinga RJ, de Jong JR, et al. Repeatability of 18 F-FDG PET radiomic features: a phantom study to explore sensitivity to image reconstruction settings, noise, and delineation method. *Med Phys*. 2019;46(2):665–678.
26. Bologna M, Corino V, Mainardi L. Technical Note: virtual phantom analyses for preprocessing evaluation and detection of a robust feature set for MRI-radiomics of the brain. *Med Phys*. 2019;46(11):5116–5123.
27. Moradmamand H, Aghamiri SMR, Ghaderi R. Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma. *J Appl Clin Med Phys*. 2020;21(1):179–190.
28. Scalco E, Belfatto A, Mastropietro A, et al. T2w-MRI signal normalization affects radiomics features reproducibility. *Med Phys*. 2020;47(4):1680–1691.
29. Um H, Tixier F, Bermudez D, Deasy JO, Young RJ, Veeraraghavan H. Impact of image preprocessing on the scanner dependence of multi-parametric MRI radiomic features and covariate shift in multi-institutional glioblastoma datasets. *Phys Med Biol*. 2019;64(16):165011.
30. Pavic M, Bogowicz M, Würms X, et al. Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncol*. 2018;57(8):1070–1074.
31. Parmar C, Rios Velazquez E, Leijenaar R, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One*. 2014;9(7):e102107.
32. Lee M, Woo B, Kuo MD, Jamshidi N, Kim JH. Quality of radiomic features in glioblastoma multiforme: impact of semi-automated tumor segmentation software. *Korean J Radiol*. 2017;18(3):498–509.
33. Tixier F, Um H, Young RJ, Veeraraghavan H. Reliability of tumor segmentation in glioblastoma: impact on the robustness of MRI-radiomic features. *Med Phys*. 2019;46(8):3582–3591.
34. Menze BH, Jakab A, Bauer S, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans Med Imaging*. 2015;34(10):1993–2024.
35. Ghaffari M, Sowmya A, Oliver R. Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the BRATS 2012–2018 challenges. *IEEE Rev Biomed Eng*. 2020;13(1):156–168.
36. Bakas S, Reyes M, Jakab A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv e-prints*. 2018;arXiv:1811.02629. <https://ui.adsabs.harvard.edu/abs/2018arXiv181102629B>. Accessed November 01, 2018.
37. Chang K, Beers AL, Bai HX, et al. Automatic assessment of glioma burden: a deep learning algorithm for fully automated volumetric and bidimensional measurement. *Neuro Oncol*. 2019;21(11):1412–1422.
38. Kickingereder P, Isensee F, Tursunova I, et al. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *Lancet Oncol*. 2019;20(5):728–740.
39. Molina D, Pérez-Beteta J, Martínez-González A, et al. Lack of robustness of textural measures obtained from 3D brain tumor MRIs impose a need for standardization. *PLoS One*. 2017;12(6):e0178843.
40. Molina D, Pérez-Beteta J, Martínez-González A, et al. Influence of gray level and space discretization on brain tumor heterogeneity measures obtained from magnetic resonance images. *Comput Biol Med*. 2016;78(1):49–57.
41. Leijenaar RT, Carvalho S, Velazquez ER, et al. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol*. 2013;52(7):1391–1397.
42. Papp L, Rausch I, Grahovac M, Hacker M, Beyer T. Optimized feature extraction for radiomics analysis of 18F-FDG PET imaging. *J Nucl Med*. 2019;60(6):864–872.
43. Zwanenburg A, Leger S, Agolli L, et al. Assessing robustness of radiomic features by image perturbation. *Sci Rep*. 2019;9(1):614.
44. Le TT, Fu W, Moore JH. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics*. 2019;36(1):250–256.
45. Bologna M, Corino VDA, Montin E, et al. Assessment of stability and discrimination capacity of radiomic features on apparent diffusion coefficient images. *J Digit Imaging*. 2018;31(6):879–894.
46. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep*. 2015;5(1):13087.
47. Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Creis Kidd L, Moore JH. Automating biomedical data science through tree-based pipeline optimization. *arXiv e-prints*. 2016;arXiv:1601.07925. <https://ui.adsabs.harvard.edu/abs/2016arXiv160107925O>. Accessed January 01, 2016.
48. Cha YJ, Jang WI, Kim MS, et al. Prediction of response to stereotactic radiosurgery for brain metastases using convolutional neural networks. *Anticancer Res*. 2018;38(9):5437–5445.
49. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. Paper presented at: Proceedings, Part III Artificial Neural Networks and Machine Learning—ICANN 2018; October 4–7, 2018; Cham.
50. Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp*. 2018;2(1):36.
51. Wang H, Zheng H. Model validation, machine learning. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, eds. *Encyclopedia of Systems Biology*; 10.1007/978-1-4419-9863-7_233. New York, NY: Springer New York; 2013:1406–1407.
52. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *arXiv e-prints*. 2016;arXiv:1612.07003. <https://ui.adsabs.harvard.edu/abs/2016arXiv161207003Z>. Accessed December 01, 2016.
53. Cihon P. *Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development*. Oxford, UK: Oxford University, Future Humanity Institute; 2019.
54. ASNR Artificial Intelligence Task Force. <https://www.asnr.org/research/asnr-study-groups/artificial-intelligence/artificial-intelligence-task-force/>. Accessed February 15, 2020.
55. O'Connor JP, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol*. 2017;14(3):169–186.
56. Ellingson BM, Bendszus M, Boxerman J, et al.; Jumpstarting Brain Tumor Drug Development Coalition Imaging Standardization Steering Committee. Consensus recommendations for a standardized brain tumor imaging protocol in clinical trials. *Neuro Oncol*. 2015;17(9):1188–1198.
57. Jack CR Jr, Bernstein MA, Fox NC, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging*. 2008;27(4):685–691.
58. Kaufmann TJ, Smits M, Boxerman J, et al. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases (BTIP-BM). *Neuro Oncol*. 2020.
59. Law I, Albert NL, Arbizu J, et al. Joint EANM/EANO/RANO practice guidelines/SNMMI procedure standards for imaging of gliomas using PET with radiolabelled amino acids and [18F]FDG: version 1.0. *Eur J Nucl Med Mol Imaging*. 2019;46(3):540–557.
60. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104–e107.

61. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys*. 2015;42(3):1341–1353.
62. Bettinelli A, Branchini M, De Monte F, Scaggion A, Paiusco M. Technical note: an IBEX adaption toward image biomarker standardization. *Med Phys*. 2019;47(3):1167–1173.
63. Nioche C, Orhac F, Boughdad S, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res*. 2018;78(16):4786–4789.
64. Davatzikos C, Rathore S, Bakas S, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. *J Med Imaging (Bellingham)*. 2018;5(1):011018.
65. Rathore S, Bakas S, Pati S, et al. Brain cancer imaging phenomics toolkit (brain-CaPTk): an interactive platform for quantitative analysis of glioblastoma. *Brainlesion*. 2018;10670(1):133–145.
66. Foy JJ, Robinson KR, Li H, Giger ML, Al-Hallaq H, Armato SG 3rd. Variation in algorithm implementation across radiomics software. *J Med Imaging (Bellingham)*. 2018;5(4):044505.
67. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020.
68. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26(6):1045–1057.
69. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. *Phys Med Biol*. 2016;61(13):R150–R166.