Radiomics in neuro-oncological clinical trials

*Philipp Lohmann, *Enrico Franceschi, Philipp Vollmuth, Frédéric Dhermain, Michael Weller, Matthias Preusser, Marion Smits, Norbert Galldiks

The development of clinical trials has led to substantial improvements in the prevention and treatment of many diseases, including brain cancer. Advances in medicine, such as improved surgical techniques, the development of new drugs and devices, the use of statistical methods in research, and the development of codes of ethics, have considerably influenced the way clinical trials are conducted today. In addition, methods from the broad field of artificial intelligence, such as radiomics, have the potential to considerably affect clinical trials and clinical practice in the future. Radiomics is a method to extract undiscovered features from routinely acquired imaging data that can neither be captured by means of human perception nor conventional image analysis. In patients with brain cancer, radiomics has shown its potential for the non-invasive identification of prognostic biomarkers, automated response assessment, and differentiation between treatment-related changes from tumour progression. Despite promising results, radiomics is not yet established in routine clinical practice nor in clinical trials. In this Viewpoint, the European Organization for Research and Treatment of Cancer Brain Tumour Group summarises the current status of radiomics, discusses its potential and limitations, envisions its future role in clinical trials in neuro-oncology, and provides guidance on how to address the challenges in radiomics.

Lancet Digit Health 2022; 4: e841–49

Published Online September 28, 2022 https://doi.org/10.1016/S2589-7500(22)00144-3

*Contributed equally

Institute of Neuroscience and Medicine (INM-3, INM-4), Research Center Juelich (FZJ), Juelich, Germany (P Lohmann PhD, Prof N Galldiks MD); **Department of Stereotactic and Functional Neurosurgery** (P Lohmann), **and Department of Neurology** (Prof N Galldiks), **Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany; Brain Tumour Group, European Organization for Research and Treatment of Cancer, Brussels, Belgium** (P Lohmann, E Franceschi MD, P Vollmuth MD, F Dhermain MD, Prof M Weller MD, Prof M Preusser MD, Prof M Smits MD, Prof N Galldiks); **IRCCS Istituto Scienze Neurologiche di Bologna, Nervous System Medical Oncology Department, Bologna, Italy** (E Franceschi); **Department of Neuroradiology, Heidelberg University Hospital, Heidelberg, Germany** (P Vollmuth); **Radiation Oncology Department, Gustave Roussy University Hospital, Cancer Campus Grand Paris, Villejuif, France** (F Dhermain); **Department of Neurology, University Hospital and University of Zurich, Zurich, Switzerland** (Prof M Weller); **Division of Oncology, Department of Internal Medicine I, Medical University of Vienna, Vienna, Austria** (Prof M Preusser); **Department of Radiology and Nuclear Medicine and Brain Tumour Center, Erasmus Medical Center, Rotterdam, Netherlands** (Prof M Smits); **Center for Integrated Oncology, Universities of Aachen, Bonn, Cologne, and Duesseldorf, Cologne, Germany** (Prof N Galldiks)

Introduction

Advances in medicine, such as improved surgical techniques, the development of new drugs and devices, the use of statistical methods in research, the recognition of the need for regulation, and the development of codes of ethics have influenced the way patients with diseases such as cancer can be treated nowadays. Artificial intelligence (AI) and machine learning have become an integral part of our daily lives and might have the potential to similarly change the way we provide care to patients with brain cancer.

In 2019, an editorial in *Nature* referred to machine learning as the breakthrough of the decade: "few fields are untouched by the machine learning revolution, from materials science to drug exploration; quantum physics to medicine".[1] In general, medicine is moving towards the incorporation of AI technologies in health care.[2] Since AI-based methods, such as deep learning and radiomics, have been developed and achieved remarkable success especially in image classification, it is not surprising that AI technologies have developed strongly in image-based disciplines, such as dermatology, gastroenterology, ophthalmology, neuro-pathology, and neuroradiology.[2,3]

Neuropathological analyses and the acquisition of neuroimages by MRI and PET are of utmost importance for the diagnosis and follow-up of patients with brain cancer; therefore, these are ideal fields for the application of AI technologies. Moreover, limitations and risks in obtaining tumour tissue from the brain increase the value of non-invasive, next generation assessments. In addition, efforts to expand the standard criteria for response assessment to include advanced image analysis techniques, such as radiomics, clearly show the widespread recognition, desire, and hope associated with these technologies.[4,5]

Structural MRI remains the method of choice for the diagnosis, follow-up, and treatment monitoring of patients with brain tumours. In addition, advanced imaging techniques based on either MRI, such as perfusion or diffusion-weighted imaging, or amino acid PET are increasingly applied in neuro-oncology. Consequently, the amount and complexity of the available data is steadily increasing, but at most only a small proportion of the available information is used routinely in the clininc or in clinical trials. With sufficient computer support, allowing a timely evaluation of these complex multiparametric data, neuro-oncology could benefit from the application of methods from the field of AI to automate and optimise time-consuming processes, such as lesion detection, tumour segmentation, and response assessment. Moreover, new information can be derived from existing data (eg, non-invasive identification of prognostic molecular alterations).

After the introduction of the concept by Lambin and colleagues,[6] radiomics has become particularly relevant to medical subdisciplines with a strong connection to imaging, such as neuro-oncology. Subsequently, Gillies and colleagues[7] summarised the whole idea of radiomics in one sentence: "images are more than pictures, they are data".

Radiomics is a method to extract undiscovered imaging features by converting routinely acquired medical images into higher dimensional data, which are not accessible by conventional visual image analysis.[6–8] Especially in combination with other established clinical parameters or molecular markers, predictive or prognostic mathematical models can be established to support clinical decision making. Radiomics might be the answer to the demands of modern medicine.[9]

In brief, the radiomics workflow (figure) typically includes several preprocessing steps, such as image coregistration and intensity normalisation, before the identification and segmentation of the tumour.[10] After tumour segmentation, radiomics features are extracted that can be either mathematically predefined (feature-based radiomics) or automatically extracted (learned) from the input images (deep learning-based radiomics).

Correspondence to:
PD Dr Philipp Lohmann, Institute of Neuroscience and Medicine (INM-4), Research Center Juelich (FZJ), 52425 Juelich, Germany
p.lohmann@fz-juelich.de

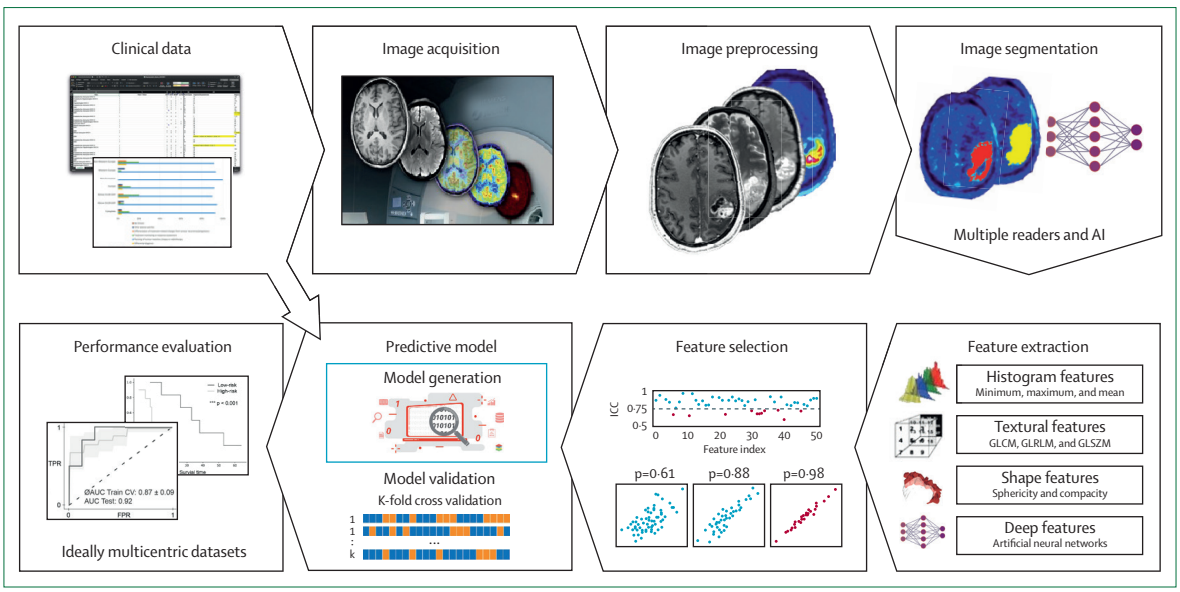

Figure: **Schematic workflow of a radiomics analysis**
Clinical data from the patients is collected and reviewed followed by image acquisition and preprocessing. Afterwards, the tumour or region of interest is segmented either manually or using deep learning-based segmentation tools. From the segmented tumours, different mathematically predefined features, such as histogram, textural, shape features, or deep features, which are learned from the input data, can be extracted. After extracting numerous features, the most relevant ones are identified by a process called feature selection. A predictive machine-learning model is generated using several different classifiers and validated using techniques like cross validation. Finally, the performance and generalisability is evaluated, ideally using several large independent multicentric datasets. AI=artificial intelligence. GLCM=grey level co-occurrence matrix. GLRLM=grey level run length matrix. GLSZM=grey level size zone matrix. ICC=intraclass correlation coefficient.

After the most relevant and informative features have been identified by feature selection, a machine-learning model using different classifiers can be generated and validated. The developed model is then applied to external multicentric datasets to evaluate its generalisability.

Several studies have shown the potential of radiomics in neuro-oncology (eg, for survival prediction; response assessment; the identification of important biomarkers, such as isocitrate dehydrogenase (IDH) mutation status or O-6-methylguanine-DNA-methyltransferase promoter methylation status; and differentiation between treatment-induced changes from local brain tumour relapse).[11–20] However, most studies do not aim to improve mechanism-based understanding of the developed models, which still restricts their trustworthiness and acceptance for a successful clinical translation due to a missing link between the radiomics signature and the underlying pathology, including specific biological pathways.[21,22]

A tissue-based pathological validation is crucial for a deeper understanding of the relationship between structural and metabolic neuroimaging, and pathological features of the tumour. Most commonly, tissue samples collected from contrast-enhanced areas on preoperative MRI during open brain surgery or stereotactic biopsies are used in studies comparing neuroimaging with pathology. Even though initial studies have shown biological links between neuroimaging, radiomics, and the underlying pathology in various types of cancers,[23–26] restrictions in the availability, number, and size of brain tissue samples result in a loss of the ability to fully characterise the heterogeneous tumours. These restrictions with respect to the histopathological correlation of radiomics features might only be overcome by large-scale tissue samples or even whole-brain specimens.[27] To collect a substantial number of whole-brain specimens from autopsies of patients with brain tumours is extremely challenging, not least due to ethical reasons. Nevertheless, a stronger emphasis on the biological understanding of radiomics features could support the clinical translation and use of the technique in neuro-oncological clinical trials.[28] Although correlation with the underlying biology is desirable and certainly beneficial to increase the understanding and acceptance of AI technologies, it might not be mandatory in the long term.[21] In medicine, there are many examples of established diagnostic and therapeutic procedures for which the underlying biological mechanisms are not fully understood. Therefore, similar standards could be applied to all technologies, even if the initial scepticism against AI technologies in medicine is understandable, justified, and important.

Despite the rapidly increasing number of studies investigating the potential of radiomics in neuro-oncology with encouraging results, this has led only sparsely to the translation of radiomics in clinical trials. In this Viewpoint, we summarise the current status, discuss the potential and limitations, and envision the future role of radiomics in neuro-oncological clinical trials.

Current status, potential, and limitations

Despite a large number of studies suggesting an added value of radiomics for diagnosis and disease monitoring in patients with brain tumours,[12,14,16,17,19,20,29–37] this technique is far from being routinely used in neuro-oncological clinical trials, and the number of radiomics studies based on at least evaluating data from clinical trials is low.[20,29–37] The main reasons for the low number of radiomics studies evaluating data from or being part of clinical trials are probably a missing standardisation of both imaging protocols and reporting of the results, a frequent absence of validation of the developed machine-learning model in large multicentre trials, a missing emphasis on the interpretability and biological meaning of identified radiomics features, and poorly defined evaluation criteria for editors and reviewers to identify high quality radiomics studies. Conversely, there are several initiatives attempting to offer mitigations to overcome these limitations. The Image Biomarker Standardization Initiative (IBSI) aims to improve standardisation of imaging protocols and reporting of results,[38] and the radiomics quality score, introduced by Lambin and colleagues in 2017,[8] provides evaluation criteria for researchers and reviewers to identify high quality radiomics studies. In addition, the transparent reporting of multivariable prediction model of individual prognosis or diagnosis (TRIPOD) statement and the prediction model risk of bias assessment tool (PROBAST) have been extended to also consider studies that applied machine learning techniques. These extensions are referred to as TRIPOD-AI and PROBAST-AI.[39] Furthermore, the US Food and Drug Administration (FDA) provided guiding principles to enable an improved standardisation and use of machine learning in medicine.[40] Despite such efforts, these existing guidelines are rarely implemented.

There are only a few clinical trials originally designed to evaluate radiomics for the characterisation of patients with newly diagnosed brain tumours. Hollon and colleagues[32] developed a deep convolutional neural network that generates radiomics features from the data without previous mathematical definition, resulting in a machine-learning model for the intraoperative diagnosis of brain tumours. The model was trained on a large dataset of more than 2·5 million images derived from stimulated Raman histology. For diagnosis, the neuronal convolutional network learned a hierarchy of recognisable histological features, similar to the morphological features of tissue samples that a neuropathologist uses for visual characterisation. Subsequently, the developed machine-learning model was tested in a prospective multicentre trial of 278 patients, and it showed a similar diagnostic accuracy (95%) for the diagnosis of brain tumours compared with the neuropathological interpretation of conventional histological images (94%). Of note, the radiomics model obtained the diagnosis around ten times faster than the conventional neuropathological workflow (2–3 min for the radiomics model *vs* 20–30 min for the conventional workflow). Thus, this study suggested the potential use of radiomics for the intraoperative diagnosis of brain cancer.

The prospective SPORT (German clinical trials register number DRKS00019855) trial aimed to characterise and predict molecular signatures of brain lesions using radiomics signatures obtained from proton magnetic resonance spectroscopy.[30,41] In this prospective single-centre study, 120 patients with newly diagnosed brain lesions were included. By combining deep autoencoder and linear discriminant models, the authors developed a classification algorithm that allows for the prediction of the origin of a brain lesion (ie, glial or metastatic) and the IDH genotype from magnetic resonance spectroscopy. In that study, tumour characteristics were predicted by an overall accuracy of more than 90% using a classifier score.

To date, no prospective clinical trials have used radiomics to evaluate the effect of treatment interventions (eg, response assessment). However, at least one clinical trial is planned to investigate the predictive value of radiomics following a radio-oncological intervention. The planned multicentre, phase 2 clinical trial, by Takami and colleagues,[42] aims to determine whether the rate of symptomatic radiation toxicity at 12 months after neoadjuvant stereotactic radiosurgery in patients with brain metastases differs from toxicity rates in patients given postoperative stereotactic radiosurgery of the resection cavity. The secondary endpoints of the trial are the 1-year local control of the treated lesion, 1-year rates of leptomeningeal dissemination, and 2-year rates of progression-free survival and overall survival. Radiomics features derived from MRI will also be evaluated with respect to the prediction of primary and secondary endpoints. The results of this clinical trial are expected to be published no earlier than 2023.

Other studies retrospectively evaluated the value of radiomics by reanalysing data from prospective interventional trials that did not originally include radiomics as part of their analysis.[31,33,34] George and colleagues[31] did a post-hoc analysis of structural MRI data from a multicentre trial on the efficacy of durvalumab, a programmed cell death ligand 1 (PD-L1) inhibitor, in 113 patients with glioblastoma and investigated the potential of radiomics for prognosis estimation. The developed machine-learning model, which used radiomics features extracted from the first MRI scan after therapy, showed promising results. Future studies are necessary to assess the generalisability of the model and to integrate additional clinical and advanced imaging features to improve the diagnostic performance.

Kickingereder and colleagues[33] showed that automated volumetric quantification of tumour burden using artificial neural networks for the assessment of response to bevacizumab plus lomustine is highly accurate and outperformed the response assessment in neuro-oncology (RANO) criteria for predicting overall survival.

	Modality	Centres	Patients	Timepoints	Lesion type	Segmentation	Evaluation of clinical features	Classifier and statistics	Data split (training:test)	Endpoints	Performance	Main results
Grossmann et al (2017)[20]	Structural MRI	2	291	Pretreatment and first follow-up (6 weeks after treatment)	Recurrent GBM	Semiautomatic using thresholding	Yes, maximum axial diameter, total tumour volume, age, and sex, KPS	Cox proportional hazard models	126:165	PFS and OS	PFS HR, 4·5; OS HR, 2·5	MRI radiomics provided prognostic value for survival and progression in recurrent GBM treated with BEV
Carles et al (2021)[29]	FET PET	1	32	At least 6 months between first and second RT	Recurrent GBM	Semiautomatic using thresholding	Yes, tumour volume, SUVmax, SUVmean, and SUVmin	Logistic regression, Cox proportional hazards models, and log-rank tests	32:0 (cross validation)	TTP, OS, and RL	AUC TTP 0·68; statistically significant in log-rank test for prediction of OS and RL (p<0·05)	FET PET radiomics signatures distinguished re-irradiation responders from non-responders
Franco et al (2021)[30]	MR spectroscopy	1	120	Pretreatment	Mainly gliomas (73%) and non-glial lesions	Semiautomatic	No	Autoencoder and linear discriminant analysis	80:40	Prediction of molecular subtypes	Accuracy 91·2%	Developed MRS classifier allowed a prediction of molecular subtypes
George et al (2022)[31]	Structural MRI	5	113	Pretreatment and first follow-up	Newly diagnosed and recurrent GBM	Semiautomatic using thresholding	Yes, volume and maximal axial diameter	Random survival forest	70:30 (approximate)	PFS and OS	Median PFS 106 days; PFS CI 0·68–0·72; Median OS 208 days; CI 069-0·75	MR radiomics predicted survival in patients with GBM on PD-L1 inhibition immunotherapy
Hollon et al (2020)[32]	SRH images	2	693	Intraoperative during tumour resection	Various brain lesions, including gliomas, metastases, meningiomas, TRC, and healthy tissue	Automatic	No	Linear discriminant analysis	415:278	Histological brain tumour diagnosis	Accuracy 94·6%	Automated diagnosis of SRH images was non-inferior to pathologist-based interpretation of conventional histological images
Kickingereder et al (2019)[33]	Structural MRI	35	1493	Preoperative, early postoperative, or follow-up	Gliomas and various other brain tumours	Automatic	Yes, RANO criteria	Dice similarity coefficient, Cox proportional hazard models, and log-rank tests	921:572	Accuracy of automated versus manual tumour segmentation, response assessment and prediction of TTP	Dice coefficient 0·89–0·93; agreement in volumetrically defined time to progression 87–90%; HR time to progression, 2·59	Automated assessment of tumour response and imaging biomarker discovery in neuro-oncology at high-throughput
Muzi et al (2020)[34]	FMISO PET	2	72	Preoperative and postoperative	GBM and anaplastic astrocytoma	Manual and semiautomatic using thresholding	Yes, age, volume, TBRmax, SUVmean, SUVmax, SUVpeak, and HV	Cox proportional hazard models	72:0	OS	OS concordance index 0·77; statistically significant differentiation between high-risk and low-risk patients (p<0·05)	FMISO PET radiomics combined with clinical features improved survival prediction in glioma patients
Shen et al (2021)[35]	NIR fluorescence imaging	1	23	Intraoperative during tumour resection	Gliomas	NA	No	Logistic regression	16:7	Determination of glioma and prediction of grade and Ki-67 level	AUC determination of glioma 0·95; AUC grade 0·87; AUC Ki-67 level 0·63	Promising classifier for intraoperative pathological diagnosis and supporting neurosurgeons during tumour resection

(Table continues on next page)

	Modality	Centres	Patients	Timepoints	Lesion type	Segmentation	Evaluation of clinical features	Classifier and statistics	Data split (training:test)	Endpoints	Performance	Main results
(Continued from previous page)												
Vils et al (2021)[36]	Structural MRI	2	118	After treatment	Recurrent GBM	Manual	No	Cox regression and logistic regression	69:49	PFS, OS, and MGMT promoter methylation status	PFS and OS models not confirmed in validation data; AUC MGMT 0·67	Model predicted MGMT promoter methylation status in patients with recurrent GBM
Yan et al (2021)[37]	DTI	3	2086	After treatment	Gliomas	Manual	Yes, gender, grade, KPS, EoR, IDH status, and previous therapies	Cox proportional hazard models, and log-rank test	688:1398	OS and identification of biological pathways associated with radiomics signature	Signature associated with survival (p<0·001); concordance index (accuracy) 0·84–0·90	Prediction of clinical outcomes using DTI radiomics and link of radiomics signature with signalling pathways

AUC=area under the receiver operating characteristic curve. BEV=bevacizumab. DTI=diffusion tensor imaging. EoR=extent of resection. FET=O-(2-[^{18}F]fluoroethyl)-L-tyrosine. FMISO=^{18}F-fluoromisonidazole. GBM=glioblastoma. HR=hazard ratio. HV=hypoxic volume. IDH=isocitrate dehydrogenase. KPS=Karnofsky performance status. MGMT=O-6-methylguanine-DNA methyltransferase. NA=not applicable. NIR=near-infrared. OS=overall survival. PD-L1=programmed death-ligand 1. PFS=progression-free survival. RANO=response assessment in neuro-oncology. RI=recurrence location. RT=radiotherapy. SRH=stimulated Raman histology. SUVmax=maximum standardised uptake value. SUVmin=minimum standardised uptake value. SUVmean=mean standardised uptake value. SUVpeak=peak standardised uptake value. TBRmax=maximum tumour-to-brain-ratio. TRC=treatment-related changes. TTP=time-to-progression.

Table: **Summary of essential elements and main findings of the discussed studies**

Because this method has been integrated into a ready-to-use software infrastructure and the response assessment is fully automated on routinely acquired conventional MRI, this method has the potential to improve the assessment of response in future clinical trials. The main findings and essential elements of the ten studies discussed in this Viewpoint are summarised in table.

Outlook

For radiomics to be incorporated into clinical trials, there are several obstacles to overcome.[43]

Different approaches and methods can be followed in all steps of the radiomics workflow, which might result in varying results. Frequently, approaches differ by reducing the involvement of a human expert to a minimum to obtain results that are as standardised and reproducible as possible. However, these approaches have disadvantages that need to be considered. For example, fully automated tumour segmentations often give good results, but visual inspection of the segmentations by an experienced neuroradiologist is required. In addition, either all lesions or only the target lesion can be included in the radiomics analysis, and whether the entire lesion, individual compartments, or areas outside the lesions are segmented will affect the outcome and should be reported in a consistent manner.

The role of input data should also not be underestimated. If highly homogeneous and standardised imaging data are used (eg, from clinical trials), there is a risk that the resulting radiomics signatures do not reflect the underlying pathology but are instead linked to the imaging process. Consequently, such models will not provide reliable results in clinical practice. Therefore, radiomics models should be developed on datasets with a diversity of imaging protocols and images of disease and normal findings.[21]

The development and optimisation of the machine-learning models involve additional trade-offs that directly affect the results of the radiomics models. For example, mathematically predefined radiomics features can be calculated that are not correlated with the underlying data and can therefore be used in small datasets. Alternatively, deep learning-based radiomics can be used to identify or learn features from the data, which are then highly correlated with the underlying data and are therefore more suitable for very large datasets. Techniques, such as transfer learning, that use models that have already been pretrained for a different purpose and then optimised on the existing data, can circumvent the drawbacks of deep-learning-based approaches.[44]

Feature selection could be integrated into radiomics-specific trials or represent exploratory endpoints within an established trial. Radiomics signatures that have already been extensively validated could serve as primary or secondary endpoints, especially if specific biological processes or processes targeted within a clinical trial can be linked with the radiomics signature.[21]

***Panel 1*: Guidelines and recommendations for radiomics study and future clinical trial designs**

IBSI
The Image Biomarker Standardization Initiative (IBSI) provides recommendations and definitions for standardisation of image biomarker extraction, imaging protocols, and reporting of results.[38]

TRIPOD-AI
Extension of the Transparent Reporting of multivariable prediction model of Individual Prognosis or Diagnosis (TRIPOD) statement also takes into consideration prediction model studies that applied artificial intelligence (AI) machine-learning techniques.[39]

PROBAST-AI
Extension of Prediction model risk of Bias Assessment Tool (PROBAST) also takes into consideration prediction model studies that applied machine-learning techniques.[39]

RQS
The Radiomics Quality Score (RQS) provides evaluation criteria for researchers, reviewers, and editors to identify high-quality radiomics studies.[8]

CONSORT-AI
Reporting guidelines for clinical trial reports for interventions involving AI.[55]

SPIRIT-AI
Guidelines for clinical trial protocols for interventions involving AI.[56]

FAIR guiding principles
The findability, accessibility, interoperability, and reusability guiding principles for scientific data management.[47]

Software as medical device regulatory framework
Proposed regulatory framework by the US Food and Drug Administration for modifications to AI and machine learning-based software as medical device.[51]

Another important trade-off is to balance between moderate performing explainable radiomics models and high-performance inexplainable models. Finding the correct combination of all these approaches for specific tasks will be another major challenge in the successful translation of radiomics into clinical practice.

In addition to the obvious need for standardisation of methods and reporting, rigorous repeatability, reproducibility, and efficacy analyses of these approaches should be done at multiple sites. Of note, large multicentre studies sharing medical data between institutions are not only an organisational challenge, but they also often fail because of data and privacy protection reasons. Without access to sufficient data, radiomics and AI methods cannot reach their full potential and ultimately fail to make the transition from the experimental stage to clinical practice.

Federated learning is a concept adapted to medical imaging by Sheller and colleagues[45] and Rieke and colleagues[46] that potentially overcomes the issue of restricted access to multicentric data. The idea behind the approach is that a consensus radiomics or machine-learning model is developed without exchanging patient data between institutions. Instead, the radiomics workflow is done locally at each participating institution and only model parameters and settings are transferred before being aggregated to a final consensus model. It was shown that a machine-learning model developed by federated learning achieved 99% of the quality of a model that used centralised data.[45] Thus, federated learning might be key for a successful translation of many promising radiomics studies to clinical routine and trials.

Despite these promising developments, all institutions should pay attention to the consistent use of enabling technologies such as the findable, accessible, interoperable, and reusable (FAIR) data principles.[47] Federated learning also has disadvantages, such as single point failure and the way it is affected by malicious data.[48] Single point failure describes the issue that the whole federated learning system will fail if the central server that integrates the results from the local training to a global model is compromised by accidental network connection failures, unexpected external attacks, or malicious data. Malicious data is caused by dishonest clients that train local models in a way that is not in agreement with the predefined federated learning protocols or submit false data about their training results and thereby contaminate the global federated learning model. Blockchain is a technology that could be used to overcome these shortcomings and the combination of federated learning and blockchain efficiently addresses privacy and security issues of distributed machine learning.[48,49]

Another important driver for the fast and efficient integration of radiomics into clinical routine and clinical trials is the availability of Conformité Européene (CE)-marked and FDA-approved radiomics or machine-learning software integrated suitably into the clinical workflow.[50] The ability of radiomics or machine learning software to continuously learn from real-world data and improve its performance makes these technologies uniquely situated among software as a medical device (SaMD).[51,52] To date, the CE and FDA have cleared or approved several machine-learning-based SaMD that have only included algorithms that are locked (ie, provide the same result each time the same input is applied to it and does not change with use, such as decision trees or look-up tables). One example is an imaging system that uses algorithms to give diagnostic information for skin cancer.[50,51] However, the power of AI technologies lies in their ability to continuously learn and improve over time, which might provide a different output compared with the output that was initially cleared for a given set of inputs.[51] To address this, the FDA has proposed an innovative approach to embrace the iterative improvement power of AI-based SaMD while assuring that patient safety is maintained.[51]

A publicly accessible database with device details for CE-marked medical devices in Europe and FDA-approved devices in the USA needs to be established to increase transparency of the regulatory pathways and approval processes to build public confidence in AI technologies in medicine.[50]

***Panel 2*: Recommendations to address sources of bias in the radiomics workflow**

Patient selection
- Selected patients should be fully representative of the population of interest and not only include extreme cases

Reference labels
- Reference labels (eg, molecular marker, treatment response, and survival) need to be clinically relevant, standardised, and their assessment reproducible
- Reference labels should not be defined on the imaging data that is later used for feature extraction
- The distribution of reference labels should be balanced between the groups
- In case of unbalanced data, try resampling methods, generate synthetic data, or use performance metrics that consider class imbalances (eg, Cohen's kappa or F1 score)

Image acquisition
- Design standardised image acquisition protocols, including software version control
- Phantom or test-retest experiments are recommended to assess intrascanner and interscanner or intrasequence and intersequence variability
- Quantitative imaging techniques (eg, PET and quantitative MRI) are less strongly affected by scanner and sequence or protocol variabilities
- For the discovery of a robust radiomics signature, a heterogeneous dataset comprising different scanners and imaging protocols is advantageous
- Within clinical trials, optimised and highly standardised image protocols that ensure image quality and reproducibility should be preferred
- Establish multi-institutional collaborations to increase the amount of available data
- Use distributed learning approaches to reduce the burden of direct data sharing (eg, federated learning)

Image (pre)processing
- Motion correction, normalisation, and harmonisation of imaging data is essential
- Use standardised (pre)processing routines from established (open source) software or release source code publicly

Segmentation
- Automatic segmentation methods should be preferred, which improve the reproducibility of the results and can be easily integrated in an automated workflow
- Manual or semiautomated segmentation should, if necessary, be done by more than one observer to increase robustness and reproducibility
- Use established (open source) software, if available, or clinically applied and certified software tools
- Careful visual inspection of the final segmentations is essential

Feature extraction
- Adhere to standardised recommendations regarding feature extraction and reporting of the results to ensure a common definition of all features for fair comparison, reproducibility, and quantitative analysis (eg, Image Biomarker Standardization Initiative)[38]
- Use established standardised (open source) software tools or release source code publicly

Feature selection
- Select features on the basis of model performance by association with reference labels (eg, treatment response or survival)
- Balance between moderate-performance explainable radiomics models using simple and low numbers of features and high-performance inexplainable models using abstract and high numbers of features
- Report the statistical criteria that were used during the process of feature selection

Model generation and validation
- Resampling methods, such as cross validation and penalised regression methods, should be used to generate the models
- Model evaluation should be done on an independent validation dataset
- Cross validation can be used in the absence of an independent validation dataset

Model testing
- Apply the best-performing model to an ideally large, independent, multi-institutional dataset

Reporting of the results
- Report the study results in a standardised way (eg, based on the radiomics quality score),[8] to improve reproducibility of the results but also supports reviewers to objectively assess the quality of the radiomics study

Clinical translation
- Try to link the developed radiomics signature with a biological meaning
- Aim towards a full integration of the radiomics workflows in clinical routine
- Emphasise economic benefits of integrating radiomics into the routine clinical workflow—if applicable
- Monitor and evaluate the use and regularly solicit feedback from users to continuously improve the developed workflow

To support this process and improve trust and transparency with these technologies, Wu and colleagues[53] created an annotated database of FDA-approved medical AI devices and analysed how they were evaluated before approval. Of note, 37 of the 130 FDA-approved medical AI devices used multisite data and prospective studies were done in only four cases. Consequently, the authors recommend that the performance of AI devices must be

For the **annotated database** see https://ericwu09.github.io/medical-ai-evaluation

Search strategy and selection criteria

We searched the PubMed database, according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, to identify relevant studies published before May 20, 2022, using the following search terms "Artificial-Intelligence", "AI", "Machine-Learning", "Radiomics", "Deep-Learning", "Neural-Networks", "Trial", "Clinical-Trial", "Clinical-Study", "Phase-0", "Phase-I", "Phase-II", "Phase-III", "Brain-Tumour", "Brain-Tumour", "Glioma", "Brain-Metastases", "Brain-Metastasis", "Glioblastoma", "GBM", "Astrocytoma", "Meningioma", "Neuro-Oncology", and "Neurooncology". Following the initial search, articles listed in the references of the identified studies were also evaluated. The literature search returned 122 articles, of which 87 were excluded after review of the title or abstract. The remaining 35 articles were reviewed and studies that met any one of the following criteria were excluded: review article or meta-analysis, absence of neuroimaging data, focus on tumour detection, segmentation or data augmentation and synthesis, no data from clinical trials, study protocols, no extraction of radiomics or textural features, and studies on extracranial tumours. Ultimately, ten studies were included in the final analysis.

evaluated in multiple clinical sites to ensure that the algorithms and models perform well across representative populations. Prospective studies of AI devices in comparison with standard of care are needed to reduce the risk of harmful model overfitting and capture true clinical outcomes more realistically. Furthermore, the authors state that postmarket surveillance is also needed for a deeper understanding and monitoring of unintended outcomes and biases that are not assessable in prospective multicentre trials.[53,54]

Furthermore, future radiomics studies need to put stronger emphasis on the biological interpretation and validation of radiomics signatures, beyond using an independent test cohort, to improve our understanding of its biological significance and cement the role of radiomics in clinical decision making.[22,23] Such studies, which provide deeper insights into the biological significance of radiomics features, will ultimately reduce scepticism about AI technologies and pave the way for successful incorporation of radiomics into neuro-oncology clinical trials.

Conclusions

In summary, radiomics is a promising approach that has a great potential to considerably affect the design of neuro-oncology trials in the future. A list of recommended guidelines and recommendations that should be considered in every future study of radiomics and in the design of future clinical trials aiming to integrate radiomics into clinical practice is provided in panel 1. Furthermore, recommendations on how to address and avoid sources of bias in the radiomics workflow are provided in panel 2.

Nevertheless, to achieve this, several methodological hurdles still need to be overcome to finally prove the value of radiomics in clinical trials and to make it a helpful next-generation tool for clinical practice.

Contributors

PL, EF, and NG conceptualised and wrote, reviewed, and edited the original draft. All authors contributed equally to writing, review, and editing to the final manuscript. The corresponding author takes full responsibility for the decision to submit for publication.

Declaration of interests

MW reports research grants from Apogenix and Quercis, and honoraria for lectures, advisory board participation, or consulting from Bayer, Medac, Merck EMD, Nerviano Medical Sciences, Novartis, Orbus, Philogen, and y-mAbs. MP reports honoraria for lectures, consultation, or advisory board participation from Bayer, Bristol-Myers Squibb, Novartis, Gerson Lehrman Group, CMC Contrast, GlaxoSmithKline, Mundipharma, Roche, British Medical Journals, MedMedia, AstraZeneca, AbbVie, Lilly, Medahead, Daiichi Sankyo, Sanofi, Merck Sharp & Dohme, Tocagen, Adastra, and Gan & Lee Pharmaceuticals. MS reports speaker fees from AuntMinnie paid to their institution. All other authors declare no competing interests.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation; project number 428090865/SPP 2177 [PL and NG], and project number 491111487).

References

1 Nature. The scientific events that shaped the decade. *Nature* 2019; **576:** 337–38.
2 Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med* 2022; **28:** 31–38.
3 Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. *Nat Rev Cancer* 2018; **18:** 500–10.
4 Gerwing M, Herrmann K, Helfen A, et al. The beginning of the end for conventional RECIST — novel therapies require novel imaging approaches. *Nat Rev Clin Oncol* 2019; **16:** 442–58.
5 Ko CC, Yeh LR, Kuo YT, Chen JH. Imaging biomarkers for evaluating tumour response: RECIST and beyond. *Biomark Res* 2021; **9:** 52.
6 Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; **48:** 441–46.
7 Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; **278:** 563–77.
8 Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14:** 749–62.
9 Sinigaglia M, Assi T, Besson FL, et al. Imaging-guided precision medicine in glioblastoma patients treated with immune checkpoint modulators: research trend and future directions in the field of imaging biomarkers and artificial intelligence. *EJNMMI Res* 2019; **9:** 78.
10 Lohmann P, Galldiks N, Kocher M, et al. Radiomics in neuro-oncology: basics, workflow, and applications. *Methods* 2021; **188:** 112–21.
11 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:** 299–307.
12 Lohmann P, Lerche C, Bauer EK, et al. Predicting IDH genotype in gliomas using FET PET radiomics. *Sci Rep* 2018; **8:** 13328.
13 Peng L, Parekh V, Huang P, et al. Distinguishing true progression from radionecrosis after stereotactic radiation therapy for brain metastases with machine learning and radiomics. *Int J Radiat Oncol Biol Phys* 2018; **102:** 1236–43.
14 Lohmann P, Kocher M, Ceccon G, et al. Combined FET PET/MRI radiomics differentiates radiation injury from recurrent brain metastasis. *Neuroimage Clin* 2018; **20:** 537–42.
15 Zhang Z, Yang J, Ho A, et al. A predictive model for distinguishing radiation necrosis from tumour progression after gamma knife radiosurgery based on radiomic features from MR images. *Eur Radiol* 2018; **28:** 2255–63.
16 Kickingereder P, Burth S, Wick A, et al. Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology* 2016; **280:** 880–89.

17 Kickingereder P, Gotz M, Muschelli J, et al. Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. *Clin Cancer Res* 2016; **22:** 5765–71.

18 Kickingereder P, Bonekamp D, Nowosielski M, et al. Radiogenomics of glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology* 2016; **281:** 907–18.

19 Choi YS, Bae S, Chang JH, et al. Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics. *Neuro Oncol* 2021; **23:** 304–13.

20 Grossmann P, Narayan V, Chang K, et al. Quantitative imaging biomarkers for risk stratification of patients with recurrent glioblastoma treated with bevacizumab. *Neuro Oncol* 2017; **19:** 1688–97.

21 Fournier L, Costaridou L, Bidaut L, et al. Incorporating radiomics into clinical trials: expert consensus endorsed by the European Society of Radiology on considerations for data-driven compared to biologically driven quantitative biomarkers. *Eur Radiol* 2021; **31:** 6001–12.

22 Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. *Radiology* 2021; **299:** E256.

23 Geady C, Keller H, Siddiqui I, Bilkey J, Dhani NC, Jaffray DA. Bridging the gap between micro- and macro-scales in medical imaging with textural analysis - a biological basis for CT radiomics classifiers? *Phys Med* 2020; **72:** 142–51.

24 McGarry SD, Bukowy JD, Iczkowski KA, et al. Gleason probability maps: a radiomics tool for mapping prostate cancer likelihood in MRI space. *Tomography* 2019; **5:** 127–34.

25 Saltz J, Almeida J, Gao Y, et al. Towards generation, management, and exploration of combined radiomics and pathomics datasets for cancer research. *AMIA Jt Summits Transl Sci Proc* 2017; **2017:** 85–94.

26 Grossmann P, Stringfield O, El-Hachem N, et al. Defining the biological basis of radiomic phenotypes in lung cancer. *Elife* 2017; **6:** e23421.

27 Bobholz SA, Lowman AK, Brehler M, et al. Radio-pathomic maps of cell density identify glioma invasion beyond traditional MR imaging defined margins. *AJNR AM J Neuroradiol* 2022; **43:** 682–88.

28 Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. *Radiology* 2021; **298:** 505–16.

29 Franco P, Huebschle I, Simon-Gabriel CP, et al. Mapping of metabolic heterogeneity of glioma using MR-spectroscopy. *Cancers* 2021; **13:** 2417.

30 Franco P, Würtemberger U, Dacca K, et al. SPectroscOpic prediction of bRain Tumours (SPORT): study protocol of a prospective imaging trial. *BMC Med Imaging* 2020; **20:** 123.

31 George E, Flagg E, Chang K, et al. Radiomics-based machine learning for outcome prediction in a multicenter phase II study of programmed death-ligand 1 inhibition immunotherapy for glioblastoma. *AJNR Am J Neuroradiol* 2022; **43:** 675–81.

32 Hollon TC, Pandian B, Adapa AR, et al. Near real-time intraoperative brain tumour diagnosis using stimulated Raman histology and deep neural networks. *Nat Med* 2020; **26:** 52–58.

33 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:** 728–40.

34 Muzi M, Wolsztynski E, Fink JR, et al. Assessment of the prognostic value of radiomic features in ^{18}F-FMISO PET imaging of hypoxia in postsurgery brain cancer patients: secondary analysis of imaging data from a single-center study and the multicenter ACRIN 6684 Trial. *Tomography* 2020; **6:** 14–22.

35 Shen B, Zhang Z, Shi X, et al. Real-time intraoperative glioma diagnosis using fluorescence imaging and deep convolutional neural networks. *Eur J Nucl Med Mol Imaging* 2021; **48:** 3482–92.

36 Vils A, Bogowicz M, Tanadini-Lang S, et al. Radiomic analysis to predict outcome in recurrent glioblastoma based on multi-center MR imaging from the prospective DIRECTOR Trial. *Front Oncol* 2021; **11:** 636672.

37 Yan J, Zhao Y, Chen Y, et al. Deep learning features from diffusion tensor imaging improve glioma stratification and identify risk groups with distinct molecular pathway activities. *EBioMedicine* 2021; **72:** 103583.

38 Zwanenburg A, Vallieres M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020; **295:** 328–38.

39 Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 2021; **11:** e048008.

40 US Food and Drug Administration. Good machine learning practice for medical device development: guiding principles: U.S. Food and Drug Administration. 2021. https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles (accessed May 15, 2022).

41 Franco P, Würtemberger U, Dacca K, et al. SPectroscOpic prediction of bRain Tumours (SPORT): study protocol of a prospective imaging trial. *BMC Med Imaging* 2020; **20:** 123.

42 Takami H, Nassiri F, Moraes FY, et al. A phase II study of neoadjuvant stereotactic radiosurgery for large brain metastases: clinical trial protocol. *Neurosurgery* 2020; **87:** 403–07.

43 O'Connor JP, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* 2017; **14:** 169–86.

44 Raghu M, Zhang C, Kleinberg J, Bengio S. Transfusion: understanding transfer learning for medical imaging. 33rd International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA. 2019; Article 301. 3347–3357.

45 Sheller MJ, Edwards B, Reina GA, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* 2020; **10:** 12598.

46 Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digit Med* 2020; **3:** 119.

47 Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016; **3:** 160018.

48 Wang Z, Hu Q. Blockchain-based federated learning: a comprehensive survey. *arXiv* 2021; published online Oct 5, 2021. https://arxiv.org/abs/2110.02182 (preprint).

49 Zerka F, Barakat S, Walsh S, et al. Systematic review of privacy-preserving distributed machine learning from federated databases in health care. *JCO Clin Cancer Inform* 2020; **4:** 184–200.

50 Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Health* 2021; **3:** e195–203.

51 US Food and Drug Administration. Proposed Regulatory Framework for Modifications to Arfificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD). https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf (accessed May 15, 2022).

52 Walsh S, de Jong EEC, van Timmeren JE, et al. Decision support systems in oncology. *JCO Clin Cancer Inform* 2019; **3:** 1–9.

53 Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med* 2021; **27:** 582–84.

54 Ferryman K. Addressing health disparities in the Food and Drug Administration's artificial intelligence and machine learning regulatory framework. *J Am Med Inform Assoc* 2020; **27:** 2016–19.

55 Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health* 2020; **2:** e537–48.

56 Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med* 2020; **26:** 1351–63.