Artificial intelligence in neuro-oncology

Vihang Nakhate[1,2] and L. Nicolas Gonzalez Castro[1,3,4]*

[1]Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States, [2]Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States, [3]Harvard Medical School, Boston, MA, United States, [4]The Center for Neuro-Oncology, Dana–Farber Cancer Institute, Boston, MA, United States

Artificial intelligence (AI) describes the application of computer algorithms to the solution of problems that have traditionally required human intelligence. Although formal work in AI has been slowly advancing for almost 70 years, developments in the last decade, and particularly in the last year, have led to an explosion of AI applications in multiple fields. Neuro-oncology has not escaped this trend. Given the expected integration of AI-based methods to neuro-oncology practice over the coming years, we set to provide an overview of existing technologies as they are applied to the neuropathology and neuroradiology of brain tumors. We highlight current benefits and limitations of these technologies and offer recommendations on how to appraise novel AI-tools as they undergo consideration for integration into clinical workflows.

Introduction

As Artificial intelligence (AI) continues to shape and reshape various aspects of our physical and virtual lives, its growing impact on and promise in medicine are hard to ignore. One of the first definitions of "artificial intelligence" was formulated in 1956 by Prof. John McCarthy at Dartmouth University, to refer to "making a machine behave in ways that would be called intelligent if a human were so behaving." (Nillson, 2010). In a broad sense, AI signifies machines that can simulate human intelligence with tasks like learning, visual processing, problem-solving, decision-making, and that increasingly can extend the reaches of human intelligence with enhanced classification and prediction. While Artificial General Intelligence (AGI), or "strong AI," refers to systems that can perform a wide range of tasks comparably to humans, most existing systems are considered Artificial Narrow Intelligence (ANI), or "weak AI," signifying systems capable of performing a defined task (Russel and Norvig, 2020). Narrow AI systems can be further classified based on physical (robotic/automation systems) and cognitive applications (machine learning, computer vision, natural language processing). Most AI applications in medicine are comprised of machine learning (ML) applications. ML refers to the ability of algorithms (see Table 1) to derive patterns and rules ("learn") from large sets of data to recognize patterns, perform tasks or make predictions without being explicitly programmed to do so (Kann et al., 2021). Within ML, learning algorithms can be characterized as supervised (using data with labeled input–output pairs), unsupervised (using data without labeled inputs) or reinforcement (using a reinforcement feedback signal for learning). While conventional ML requires manual engineering of raw data to create representations suitable for ML algorithms to learn, deep learning (DL) refers to a subset of ML techniques that can extract and learn features from raw, unstructured and multimodal data (e.g., raw imaging, text, audio-visual data) using

TABLE 1 Glossary of commonly used artificial intelligence terms.

Artificial intelligence (AI)	Computer algorithms that can solve problems, inform decision-making and perform complex tasks that have traditionally required human intelligence
Machine learning (ML)	Discipline in AI involving computers or "machines" as agents that can learn patterns and rules from large sets of data to build predictive models and solve problems, without being explicitly programmed to do so
Algorithm	Set of rules that an agent, in this case a "machine," can follow to complete a set of tasks
Supervised learning	Machine learning in which the agent observes data consisting of input–output pairs with manually assigned labels, and learns a function that predicts output from input
Unsupervised learning	Machine learning in which the agent learns patterns in the input without any explicit manual feedback
Random forest algorithm	Type of supervised learning algorithm used for classification and regression tasks, in which a large number of decision trees operate together as an ensemble to reach a common output result
Support vector machine	Type of supervised learning algorithm used for classification and regression tasks, in which a subset of training points (support vectors) from the decision function are used to complete a given task
Deep learning (DL)	Machine learning techniques in which the computational path between input and output consists of multiple layers of simple, adjustable computing elements. The resulting computational circuit, termed a **neural network**, allows for a large number of input variables to interact in complex ways. Widely used in visual object recognition, speech recognition, image and speech synthesis.
Convolutional neural networks	Type of DL neural network architecture commonly used in analysis of images
Transfer learning	Machine learning method in which knowledge gained by an agent from one domain can be transferred and applied to a new domain, so that learning can proceed faster with less data
Large language model	A neural network-based model trained on large text datasets using self-supervised learning.

Adapted from *Artificial Intelligence: A Modern Approach*, Russel and Norvig (2020).

layered neural networks (LeCun et al., 2015). DL algorithms can be supervised or unsupervised (see Figure 1).

With the advent of increasing computing power in recent decades, DL has achieved remarkable results in areas including image classification, speech recognition, and game playing, among others (LeCun et al., 2015; Silver et al., 2016). Remarkable flexibility of input and output structures coupled with modern computing power have positioned ML and DL well to analyze large data sets that are increasingly being generated in modern medicine and oncology, and to aid in using such data to guide decision making.

The applications of ML to medicine, oncology, and neuro-oncology are myriad, spanning enhanced screening, diagnosis, prognosis, classification, drug discovery, precision medicine, and more (di Nunno et al., 2022). In neuro-oncology, work utilizing ML algorithms has so far predominantly been focused on neuropathology and neuroradiology applications, including tumor diagnosis and grading, prediction of molecular features, and automated assessment of tumor volume (Figure 2).

Neuropathology applications of AI for neuro-oncology

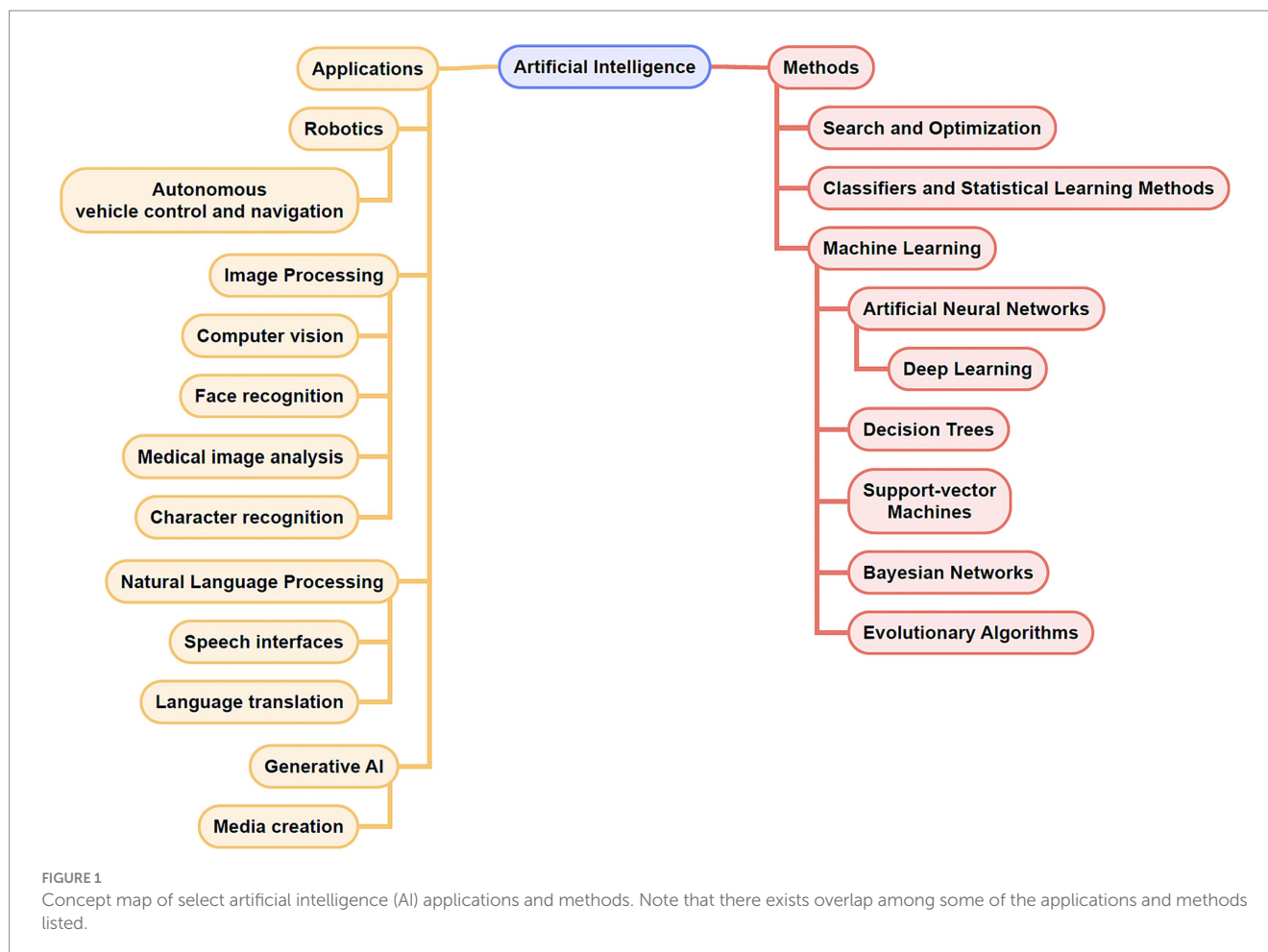## Histopathologic and genomic features

Histopathologic analysis has long been at the heart of diagnosis in oncology. However, it is susceptible to interobserver variability that can impede accurate diagnosis and optimized management (van den Bent, 2010). In neuro-oncology, grading of gliomas based on atypia, mitosis, microvascular proliferation and necrosis entails some degree of subjectivity that is contributory. The introduction of molecular features such as isocitrate dehydrogenase (IDH) mutation and 1p/19q co-deletion status into WHO grading of gliomas, as well as the

burgeoning availability of individualized tumor genetic data, leaves AI well-positioned to assist pathologists in interpreting large and multiparametric data to establish diagnoses (see Table 2).

The advent of high quality digitized whole slide images (WSIs) has allowed for the application of DL in histopathologic diagnosis. Broadly in oncology, DL algorithms have been used to detect metastatic breast cancer in lymph node biopsies (Litjens et al., 2016; Ehteshami Bejnordi et al., 2017), assign Gleason scores in prostate cancer biopsies (Litjens et al., 2016; Nagpal et al., 2020), and distinguish lung adenocarcinoma and squamous cell carcinoma from normal lung tissue (Coudray et al., 2018), among others, with high accuracy.

In neuro-oncology, convolutional neural networks (CNNs) trained on WSIs of gliomas have been used to render nonbiased neuropathologic diagnoses of gliomas. Ertosun et al. trained two CNNs on publicly available hematoxylin and eosin (H&E) stained histopathology images of gliomas from The Cancer Genome Atlas (TCGA). One CNN aimed to distinguish glioblastoma (GBM) vs. low-grade glioma (LGG), the other to distinguish grade 2 from grade 3 LGGs. When tested on an independent data set of glioma WSIs, the CNNs determined histopathologic grade with 96% accuracy for GBM vs. LGG and 71% accuracy for grade 2 vs. grade 3 (Ertosun and Rubin, 2015). A similar study by Truong et al. trained multiple CNNs using TCGA WSIs, with the best models achieving 73% mean accuracy in distinguishing GBM from LGG, and 53% accuracy in distinguishing grade 2 from grade 3 LGGs (Truong et al., 2020). Limitations in both included absence of IDH mutant/1p19q codeletion status of the tumors.

Jin et al. developed a platform named "AI Neuropathologist," whereby a CNN was trained on over 79,000 H&E-stained histologic patch WSIs from 267 patients from an institutional biobank to distinguish GBM, anaplastic astrocytoma (AA), anaplastic oligodendroglioma (AO), astrocytoma (A), oligodendroglioma (O), and background glia. The CNN derived histopathologic features and

FIGURE 1
Concept map of select artificial intelligence (AI) applications and methods. Note that there exists overlap among some of the applications and methods listed.

classified gliomas from 56 independent patients with over seventeen thousand images into the above categories with an average patch-level accuracy of 86.5%, and patient-level accuracy of 87.5% (Jin et al., 2021). However when the tumors' IDH/1p19q status was assessed, the numbers of patients with each genetically classified tumor subtype in the training sample were in some cases found to be relatively low (e.g., 16 "GBM with IDH mutant" and 39 "GBM with IDH-wild type") (Komori, 2021). Im et al. used deep transfer learning to classify subtypes of gliomas from histopathologic images generated in routine clinical practice from a single institution cohort of 468 patients. Their model distinguished oligodendroglial tumors from non-oligodendroglial tumors with an accuracy of 87.3%, whereas in distinguishing glioma grade 2 vs. 3 vs. 4 the accuracy was 58% (Im et al., 2021). Pie et al. developed a deep learning-based model that fused molecular and histopathologic features to predict glioma grade. They used digital WSIs from 549 patients in the TCGA with molecular information on IDH, 1p/19q, *ATRX*, and O⁶-methylguanine-DNA methyltransferase (*MGMT*) promoter alterations. Their model achieved an accuracy of 93.8% in distinguishing high grade glioma (HGG) from LGG, and 74% distinguishing grade 2 vs. grade 3 gliomas, the latter outperforming state-of-the-art methods (Pei et al., 2021). Finally, Hollon et al. developed a DL-based method of rapid automated molecular classification of diffuse glioma from intraoperative tissue samples (Hollon et al., 2023). They trained a CNN using histologic images from 373 diffuse glioma patients,

acquired by Stimulated Raman Histology (SRH) imaging. They also trained a genetic embedding model using TCGA and other public glioma genomic databases to learn labels that define molecular subgroups of diffuse gliomas. The SRH and genetic encoders were integrated to predict IDH, 1p19q, and ATRX mutations and thereby achieve molecular classification of gliomas by WHO criteria. When prospectively tested on 153 patients, the model predicted WHO glioma classification with a mean 93.3% accuracy, including IDH mutation (94.7%), 1p19q co-deletion (94.1%), and ATRX mutation (91.0%).

Tumor classification based on DNA methylome profiling

In addition to histopathology and direct genomic alterations, DNA methylome profiling has emerged as a valuable method for classifying CNS tumors. Cancer cells undergo substantial alterations in DNA methylation patterns, which when profiled by epigenome-wide methylation assays may be used to classify tumor types with high specificity (Moran et al., 2016). Seminal work in harnessing the methylome was conducted by Capper et al., who developed a ML algorithm to classify CNS tumors based on DNA methylation profiles (Capper et al., 2018). The authors trained the algorithm with methylation data for 2,801 pre-classified samples of almost every CNS
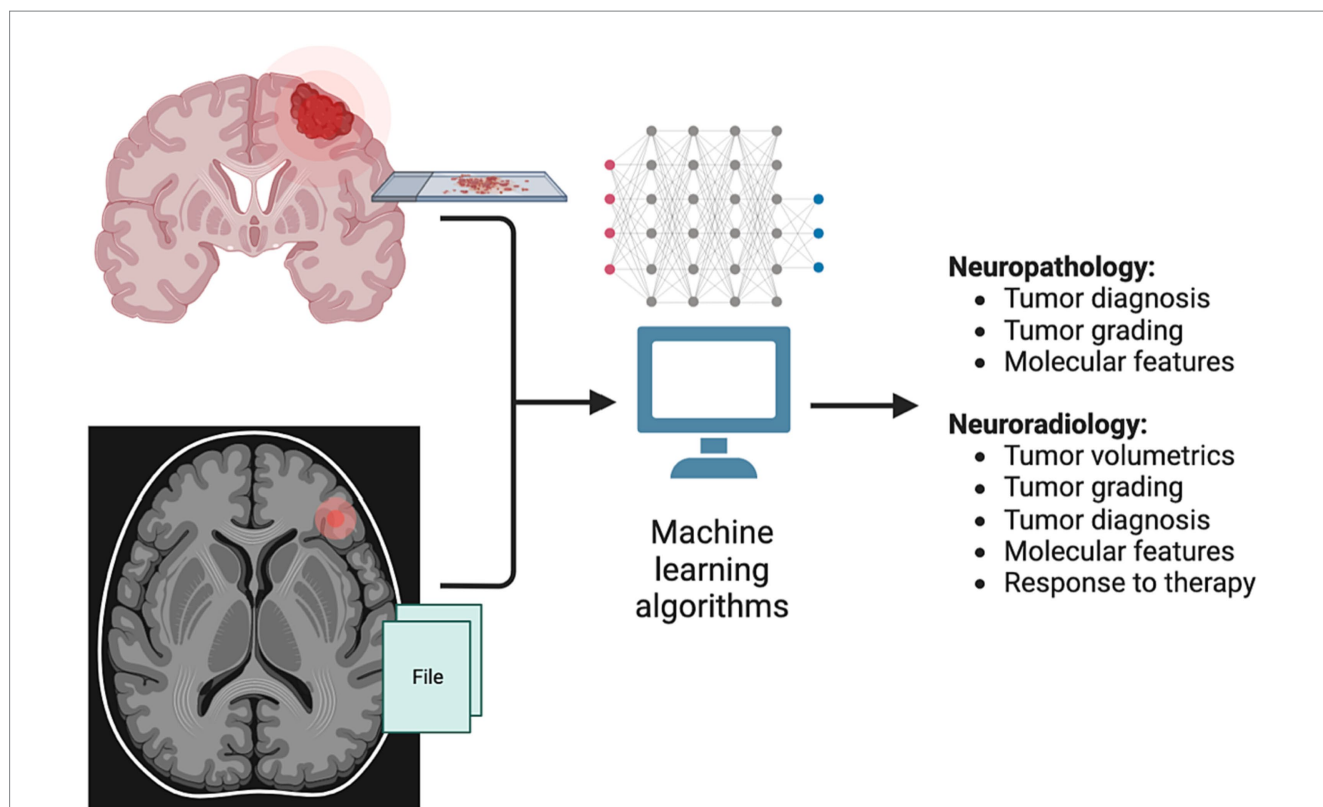
FIGURE 2
Applications of machine learning (ML) in neuro-oncology. Current research applications of ML methods in neuro-oncology have mainly introduced in the fields of neuropathology and neuro-radiology. In neuropathology, ML methods have been developed to establish a tumor diagnosis based on histopathologic and genomic features, as well as to classify tumors based on DNA methylation status. In neuroradiology, ML has been used to generate automated 3D tumor measurements (volumetrics), as well as to help predict tumor diagnosis, grading, molecular features and response to therapy.

TABLE 2 Select studies on the application of AI/ML to neuropathology in neuro-oncology.

Authors and year	Study sample (total n)	Task	ML algorithm	Performance
Ertosun and Rubin (2015)	Gliomas grade 2–4 (44 whole tissue slides)	Glioma grade	CNN	Accuracy 96% GBM vs. LGG; 71% grade 2 vs. grade 3
Jin et al. (2021)	Glioma grade 2–4 (323 patients)	Glioma classification	CNN	Accuracy 87.5%
Pei et al. (2021)	Glioma grade 2–4 (549 patients)	Glioma grade	DNN	Accuracy 93.8% HGG vs. LGG; 74% grade 2 vs. grade 3
Hollon et al. (2023)	Diffuse glioma (373 patients)	Glioma WHO classification	CNN	93.3% accuracy
Capper et al. (2018)	Most WHO-classified CNS tumors (2801)	CNS Tumor WHO classification	Supervised ML (random forest classifier), unsupervised ML	60.4% agree with pathologist; 15.5% better subclass; 12.6% did not match pathologist but most eventually proved accurate; 11.5% unclassified

LGG, low-grade glioma; HGG, high-grade glioma; GBM, glioblastoma; WHO, world health organization; ML, machine learning; CNN, convolutional neural network; DNN, deep neural network.

tumor type. The algorithm used supervised machine learning to recognize methylation patterns based on the known classifications, as well as unsupervised learning to search for patterns to independently assign samples into computer generated categories. In so doing the algorithm assigned the tumors to 82 distinct classes – around one-third matched known WHO tumor types; one-third represented sub-classes of WHO tumor types; notably, the remainder were classifications that did not match WHO groupings, including previously unrecognized tumor types, and those with histologic overlap but distinct methylation profiles. When prospectively tested on 1,104 new samples, the algorithm's classification matched the pathologist's diagnosis in 60.4% of cases; in 15.5%, the two classifications matched but the algorithm classified the tumor into a subgroup that could not be assigned by histopathology alone. In 12.6% of the cases, the algorithm's diagnosis did not match the pathologist's and, remarkably, further analysis (including by gene sequencing)

revealed that 92.8% of these unmatched tumors were reclassified from the pathologists' diagnosis to the algorithm's diagnosis, including a majority that were assigned a new tumor grade. Finally, 11.5% could not be classified by the algorithm (Capper et al., 2018; Machine Learning Improves Diagnosis of CNS Cancers, 2018; Wong and Yip, 2018). Since then, multiple studies have corroborated the algorithm's fidelity, and it has been incorporated into clinical pipelines at centers across the world (Capper et al., 2018; Jaunmuktane et al., 2019; Karimi et al., 2019; Priesterbach-Ackley et al., 2020). It has been especially useful in the classification of tumors with morphology that is heterogeneous or otherwise challenging to distinguish, including ependymomas, medulloblastomas, and diffuse glioneuronal tumors (Capper et al., 2018; Pickles et al., 2020). Its utility in guiding diagnoses for these tumors has been incorporated into the 2021 WHO guidelines for CNS tumor classification (Louis et al., 2021).

Neuroradiology applications of AI for neuro-oncology

MRI imaging is the mainstay of diagnosis, radiographic surveillance, and assessment of treatment response in neuro-oncology. However, MRI interpretation in brain tumor patients can sometimes be challenging – treatment related changes may resemble tumor progression; histologic and molecular features that drive prognosis and guide treatment often lack readily apparent imaging correlates; and determining tumor size can pose a challenge in tumors with heterogeneous and infiltrative components. AI methods including ML, DL, and radiomics have been employed to extract from images clinically relevant information that may not be apparent visually (see Table 3). Radiomics is the process of extracting quantitative and mineable data or "features" (e.g., shape, intensity, texture) from clinical imaging. ML methods are often used to build models using these features that can predict various clinical variables. In neuro-oncology, ML/DL have been used to quantify tumor size and type, predict tumor grade, molecular features, and survival. Typically, MRI data are pre-processed and standardized, labeled/annotated by radiologists to establish ground truth for training of ML algorithms, and then may undergo augmentation, transformation and further pre-processing before being used for the training of ML/DL algorithms (Zhu et al., 2022). Often the performance of these trained algorithms is assessed on a "test" cohort of patient images not encountered in training. Technical aspects of AI in brain tumor imaging have also been reviewed elsewhere (Afridi et al., 2022; Aftab et al., 2022).

Tumor volumetrics

Reliably delineating tumor size and burden on structural MRI brain is necessary to longitudinally assess tumor progression and response to treatment, and is thus critical to both effective clinical care and the assessment of response in clinical trials. However, identifying tumor boundaries manually on MRI per the RANO (Response Assessment in Neuro-Oncology) criteria, which involve quantitative 2D measurements of contrast-enhancing and FLAIR hyperintense lesions, can be challenging for infiltrative tumors like high grade gliomas, and is associated with high interrater variability (Vos et al., 2003; Pope and Hessel, 2011; Ford et al., 2016).

Chang et al. developed a deep learning algorithm to automatically segment T2/FLAIR and T1-post contrast MRI images of adult gliomas to quantify both 2D RANO measurements as well as 3D tumor volumes (Chang et al., 2019). They used MRIs from 800 patients with newly diagnosed LGG and HGGs, and over 700 post-op longitudinal MRIs from 50 patients with newly diagnosed GBMs. Their automated tumor quantification was reproducible in double baseline MRIs (interclass correlation coefficients, or ICCs, > 0.97), with high agreement between manual and automated tumor volumes (ICCs >0.91), and between manually and automatically derived longitudinal changes in tumor burden (ICCs >0.85). Though their automated RANO measurements were reproducible and internally consistent, they were often larger than manual RANO measurements. Taken together with inconsistency of RANO measurements found between the two human raters, these findings suggested that the automated measurements may be more accurate (detecting longer diameters than can be visualized by eye), and more precise. Peng et al. used a similar deep-learning approach to develop an algorithm that determines two dimensional measurements and three-dimensional volume in pediatric high grade gliomas, medulloblastomas, and other leptomeningeal-seeding tumors, with high repeatability and agreement with human raters (Peng et al., 2022).

3D volumetric measurements are not routinely used in response assessment for gliomas, in part due to the labor intensive, time-consuming, and variable nature of segmentation approaches, though they are likely more reliable and accurate than 2D measurements (Sorensen et al., 2001, 2008). Kickingereder et al. trained a DL-based CNN to carry out automated tumor segmentation on MRI data from 455 patients with brain tumors (mostly gliomas), and tested it on over two thousand MRIs from over 500 patients. The algorithm demonstrated high precision with Dice coefficients of 0.91 (T1 post-contrast) and 0.93 (T2/FLAIR) (Barash and Klang, 2019; Kickingereder et al., 2019).

Although a recent evaluation of the RANO criteria suggests that analysis of FLAIR data (when performed by humans) does not add additional information in terms of predicting survival (Youssef et al., 2023), analysis of FLAIR sequences with automated algorithms such as developed by Chang et al. and Kickingereder et al. may enable the incorporation of volumetric assessment of gliomas into research assessments and potentially clinical practice.

Prediction of molecular features

As signified by their growing prominence in the WHO classification of central nervous system tumors in 2016 and 2021, molecular features of CNS tumors are increasingly informing diagnosis, prognosis, and management (Louis et al., 2021; Gritsch et al., 2022). For instance, the presence of an IDH mutation in adult gliomas precludes a diagnosis of glioblastoma, WHO grade 4, regardless of histologic grade given its favorable prognosis compared to IDH wildtype tumors. IDH mutant gliomas with 1p/19q co-deletion are classified as oligodendrogliomas (WHO grade 2–3), and those without 1p/19q co-deletion are classified as astrocytomas (WHO grade 2–4). This highlights the importance of molecular testing, which can be time consuming and requires adequate surgical tissue for histopathologic and genetic analysis, in routine clinical practice. Noninvasive determination of a tumor's molecular features via

TABLE 3 Select studies on the application of AI/ML to neuroradiology in neuro-oncology.

Authors and Year	Study sample (total n)	Task	ML algorithm	Performance
Chang et al. (2019)	Newly diagnosed LGG/ HGG (800), longitudinal newly diagnosed GBM (50)	tumor volumes and RANO measurements	DL	Double baseline MRIs ICCs >0.97; manual vs. automated ICC > 0.85
Kickingereder et al. (2019)	Mostly gliomas (1027)	Tumor volumes	ANN	DICE coefficients 0.91 (T1-post), 0.93 (T2/FLAIR)
Zhang et al. (2017)	HGGs (120)	IDH mutation status	Random forest	Accuracy 89%, AUC 0.92
Chang et al. (2018)	LGGs and HGGs (259)	Automated segmentation; IDH, 1p/19q, *MGMT* promoter status	CNN	Accuracy 94% (IDH), 92% (1p/19q), 83% (MGMT promoter)
Akkus et al. (2017)	LGGs (159)	1p/19q status	CNN	Accuracy 87.7%
van der Voort et al. (2019)	Presumed LGG on pre-op MRI (413)	1p/19q status	Support vector machine	AUC 0.72
Yogananda et al. (2020)	LGGs and HGGs (368)	1p/19q status	3D CNN	Accuracy 93.46%
Zhou et al. (2019)	Grade 2–4 glioma (744)	IDH and 1p/19q status	Random forest	AUC 0.92 (IDH status); accuracy 78.2% (3-group classification)
Cluceru et al. (2022)	Grade 2–4 glioma (531)	IDH and 1p/19q status	CNN	Accuracy 85.7% (overall 3-group), 95.2% (IDHwt), 88.9% (IDHmut-intact), 60.0% IDHmut-codel
Korfiatis et al. (2017)	GBM (155)	*MGMT* promoter status	CNN	Accuracy 94.9%
Chen et al. (2022)	Diffuse glioma (111)	*MGMT* promoter status	CNN	Accuracy 91%; AUC 0.90
Sun et al. (2019)	Midline glioma (100)	H3 K27 mutation status	DL	AUC 0.85
Prasanna et al. (2017)	GBM (65)	Survival	Random forest	CI = 0.70 (short term vs. long term survival)
Park et al. (2020)	GBM (216)	Survival	LASSO cox regression	CI = 0.70 (overall survival)
Beig et al. (2018)	GBM (115)	Survival	Random forest	CI = 0.83 (short term vs. long term survival)
Kim et al. (2019)	GBM (95)	Progression vs. Pseudoprogression	Hybrid ML-DL	AUC = 0.85

ICC, interclass correlation coefficient; AUC, area under the curve (referring to receiver operating characteristics curve); ANN, artificial neural network; CNN, convolutional neural network; LGG, low-grade glioma; HGG, high-grade glioma; GBM, glioblastoma; IDH, isocitrate dehydrogenase; MGMT, O⁶-methylguanine-DNA methyltransferase; RANO, response assessment in neuro-oncology; CI, concordance index. *n* reflects training and validation cohorts where applicable.

imaging would be valuable not only in cases of inoperable tumors or insufficient surgical samples, but also in all cases to guide early diagnosis or early pre-surgical enrollment into clinical trials (Gonzalez Castro et al., 2023).

Certain qualitative radiographic correlates of molecular features have previously been recognized. For instance, IDH wildtype LGGs have poorer definition of non-enhancing margin and more multifocal distributions than IDH mutant LGGs; IDH wildtype tumors have larger percentage core enhancing component; and the "T2-FLAIR mismatch" sign can identify IDH-mutant 1p19q-intact gliomas with good specificity (Ellingson, 2015; Patel et al., 2017; Park et al., 2018; Lasocki et al., 2021; Miller et al., 2023). IDH mutant gliomas have been shown to have higher ADC and lower relative cerebral blood volume (rCBV) values, though these findings have wide ranging sensitivities (56 to 100%) and specificities (51–100%) depending on the study (Xing et al., 2017; Suh et al., 2019). 2-HG MR spectroscopy has better sensitivity in detecting IDH mutant status, in one analysis sensitivity 96% and specificity 85% (Suh et al., 2019). Additionally, 1p/19q co-deleted tumors are associated with indistinct tumor borders, frontal tumor location, heterogeneous T2 signal intensity,

and cortical/subcortical tumor infiltration (Smits and van den Bent, 2017).

ML algorithms developed to predict molecular features based on imaging data offer the promise of automated recognition of these and other features, and have the advantage of being independent of operator experience, more accessible, and more amenable to training on improved data sets. Here we highlight some salient studies among the numerous studies that have employed ML approaches to predict molecular features including IDH mutation, 1p/19q codeletion, *MGMT* promoter methylation status, and other relevant features.

Zhang et al. developed a ML-based model using a random forest classifier to predict IDH mutation status based on patient age and pre-operative MRIs of 90 patients with HGGs. Using T1, T2/FLAIR, and ADC sequences, the model achieved accuracy of 89% (AUC 0.9231) in the validation cohort of 30 HGGs (Zhang et al., 2017). Imaging features contributing the most to IDH genotyping were patient age and MRI parametric intensity, texture and shape features. In a similar study, Chang et al. trained a CNN to predict IDH mutation status from pre-operative MRIs of patients with grade II-IV gliomas, and accuracies improved from 85.7 to 89.1% with incorporation of

patient age into the predictive model (Chang et al., 2018). These models required manual tumor segmentation which limits clinical viability, but automated segmentation approaches (as described in the previous section) may help overcome this. A meta-analysis of 9 studies by Zhao et al. employing ML to radiographically predict IDH mutations in gliomas found pooled sensitivity and specificity of 87 and 88%, respectively, in the training set, and 87 and 90%, respectively, in the validation set (Zhao et al., 2020). Finally, Chang, P et al. used MRI data from 259 patients with low or high grade gliomas from The Cancer Imaging Archive (TCIA) to train a CNN to predict IDH mutation, 1p/19q codeletion, and *MGMT* promoter methylation status simultaneously and using an automated segmentation tool (Chang et al., 2018). They achieved a high accuracy of 94% in IDH mutation status, while accuracies for 1p19q codeletion and *MGMT* promoter methylation were 92 and 83%, respectively.

Codeletion of the 1p/19q chromosome arms in IDH mutant gliomas is characteristic of oligodendroglioma and associated with increased survival and better response to treatment, and is another important part of glioma classification by WHO criteria (Taal et al., 2015; Louis et al., 2021). Fellah et al. used multivariate random forest models to retrospectively predict 1p/19q codeletion status from conventional MRI (cMRI) sequences (T1- and T2-weighted sequences) and from diffusion-weighted imaging (DWI), perfusion-weighted imaging (PWI), and MRI spectroscopy (MRS). Their model had misclassification rate of 48% and established that inclusion of DWI, PWI, and MRS did not help improve the prediction of 1p/19q codeletion relative to cMRI sequences alone (Fellah et al., 2013). Akkus et al. used 159 preoperative cMRIs of LGGs to train and test a CNN to predict 1p/19q codeletion status, and achieved an accuracy of 87.7% (Akkus et al., 2017). Van der woort et al. trained a support vector machine (SVM) algorithm on cMRI images of 284 patients who had undergone biopsy or resection for presumed LGG (rather than histologically confirmed LGG, so as to reflect a more clinically relevant population). Their model, which also incorporated age and sex data, predicted 1p/19q codeletion in 129 patients from an external test cohort from the TCIA, with AUC of 0.72. The authors compared this to predictions by clinical experts, who achieved AUCs of 0.52 (two neurosurgeons) and 0.81 (two neuroradiologists) albeit with wide variability among the clinical experts (AUC 0.45–0.83) (van der Voort et al., 2019). Finally, Yogananda et al. used only T2-weighted MRI sequences from a cohort of 368 patients from the TCIA/TCGA with low and high-grade gliomas, divided into training, validation and testing sets, to predict 1p/19q-codeletion. Their 3D CNN achieved an accuracy of 93.46% (Yogananda et al., 2020). Their exclusive use of T2-weighted images, as well as of automated tumor segmentation, signified a step forward in terms of potential implementation in a clinical setting.

Given the need to identify multiple molecular alterations simultaneously (e.g., IDH and 1p/19q codeletion status) for accurate classification of gliomas, some investigators have worked to develop models for simultaneous classification into one of 3 groups: IDH wild type (IDHwt), IDH mutant and 1p/19q-codeleted (IDHmut-codel), and IDH mutant and 1p/19q non-codeleted (IDHmut-non-codel). Matsui et al. used multi-modal MRI ^{11}C,-methionine PET, and CT images as well as age/gender data from 217 LGG patients to develop a DL model to predict glioma classification, achieving 68.7% accuracy in the test dataset. They noted lower accuracies with only MRI, MRI and PET, and MRI and CT, and reasoned that ^{11}C-methionine-PET

increased yield for oligodendrogliomas and IDH wild type astrocytomas, while CT increased yield for oligodendrogliomas by detecting calcification (Matsui et al., 2020). Zhou et al. trained a random forest algorithm on preoperative cMRI in 538 patients with grade 2–4 gliomas from three different institutions. Integrating patient age, they developed two models to sequentially detect IDH mutation status, then 1p/19q status among the IDH mutants. When tested on an external validation cohort from the TCIA of 206 patients with glioma, their model achieved AUC of 0.919 for IDH mutation, and an overall accuracy for glioma classification of 78.2% (Zhou et al., 2019). The authors suggest that a larger sample size may enhance 1p/19q codeletion status prediction in this model. Finally, Cluceru et al. trained a CNN to identify IDH mutation and 1p/19q co-deletion in pre-operative MRIs of newly diagnosed grade 2–4 gliomas, using a cohort of 384 patients from a single institution and 147 patients from the TCGA dataset (Cluceru et al., 2022). They trained multiple CNN classifiers, including using a sequential model (predicting IDH mutation first, then 1p19q codeletion) and a simultaneous 3-group model; they also trained CNNs with or without DWI sequences in addition to cMRI sequences. They found that their best classifier was a 3-group CNN that included DWI as input, predicted molecular features with an overall test accuracy of 85.7%, and correctly classified 95.2% IDHwt, 88.9% IDHmut-intact, and 60.0% IDHmut-codel gliomas. The authors suggested that incorporating susceptibility-weighted imaging (SWI) and rCBV sequences into future algorithms may improve diagnostic accuracy in IDHmut-codel gliomas.

Methylation of the *MGMT* promoter in gliomas predicts longer survival and better response to alkylating chemotherapy agents such as temozolomide, and is thus a clinically vital molecular feature to determine (Stupp et al., 2009). Radiographically, gliomas with *MGMT* promoter methylation have been associated with less vasogenic edema, higher ADC values, and lower cerebral blood flow and blood volume on MR PWI, relative to unmethylated tumors according to a meta-analysis of relevant studies (Suh et al., 2019). Several studies have endeavored to noninvasively assess *MGMT* promoter methylation status via MRI using ML and DL methods.

Li et al. used a cohort of 193 patients with newly diagnosed GBM to build a ML-based random forest classifier for prediction of *MGMT* promoter methylation status in pre-operative cMRIs. Their model selected six features including location, geometry, intensity and texture features; it predicted *MGMT* promoter methylation status with 80% accuracy (AUC 0.88), and the addition of clinical features did not lead to an improvement of this result (Li et al., 2018). Crisi et al. used MR PWI in a cohort of 59 patients with GBM to identify 14 quantitative radiomic features that were used to build a DL model to classify *MGMT* promoter methylation status into three groups: unmethylated (<10% methylated), intermediate-methylated (10–30% methylated), and methylated (>29% methylated). Their model classified *MGMT* promoter methylation status into these three groups with AUC 0.84, sensitivity 75% and specificity 85% (Crisi and Filice, 2020). This lends support to MR PWI as a potential biomarker for *MGMT* promoter methylation status using ML/DL classifiers. Korfiatis et al. used T2 MRI images from 155 patients with newly diagnosed GBM to train and test three different residual CNNs to predict *MGMT* promoter methylation status in each image slice. Their best performing CNN had 50 layers, and predicted MGMT status (methylated, unmethylated, or no tumor) with 94.90% accuracy in the test set. Notably their model eliminated the need for a manual tumor

segmentation step (Korfiatis et al., 2017). Chen et al. built a DL model to assess the predictive value of cMRI and ADC sequences in 111 patients using two regions of interest (ROIs), tumor core and tumor whole (the latter including tumor edema). They found highest predictive value in the tumor core ROI using T1-post contrast combined with ADC sequences, with 91% accuracy and AUC 0.90 (Chen et al., 2022).

A review and meta-analysis of ML-based prediction of molecular features in glioma using MRI by Jian et al. examined 44 studies and found a pooled sensitivity and specificity for IDH mutation of 0.83 and 0.85, respectively. Pooled sensitivities and specificities for 1p/19q codeletion and *MGMT* promoter methylation ranged between 0.76 and 0.83. Of the 44, 7 studies utilized DL, while most used ML-based random forest or SVM classifiers (Jian et al., 2021). Another review and meta-analysis by Bhandari et al. on using MRI radiomics to predict IDH and 1p/19q status in LGGs examined 14 studies. They found that for IDH mutation status prediction, conventional radiomics combined with DL based CNN derived features was the most accurate approach, with 94.4% sensitivity and 86.7% specificity. In contrast, conventional texture-based radiomics performed best in predicting 1p/19q codeletion status, with 90% sensitivity and 96% specificity (Bhandari et al., 2021). These results should be interpreted cautiously, as there was a high degree of heterogeneity among the studies reviewed, with varying radiomic pipelines many of which required manual tumor segmentation, making direct comparisons challenging.

In diffuse midline gliomas, H3 K27 mutation is commonly observed in both pediatric and adult patients, and in pediatric patients portends decreased overall survival regardless of tumor location or histopathological grade (Karremann et al., 2018; Kleinschmidt-DeMasters and Mulcahy Levy, 2018; Ebrahimi et al., 2019). As many of these tumors are located in the brainstem, surgical intervention, including biopsy, can be morbid and is sometimes foregone, increasing the utility of accurate non-invasive H3 K27 mutation status prediction. Su et al. developed deep learning models to predict H3 K27 mutation using only T2 weighted MRI sequences in a cohort of 100 patients with midline gliomas, including 40 mutant and 60 wild type tumors, with three quarters of the cohort reserved for a training set and one quarter for testing. Of ten generated prediction models, accuracies ranged 60 to 84% in the testing cohort, and the best model had a AUC of 0.85 in the test cohort. Larger sample sizes, may help further refine the accuracy of this approach.

Prognostic models

Discussing prognosis is of major importance at the time of brain tumor diagnosis, especially for GBM where the median survival is approximately 16–18 months despite completion of standard-of-care therapy (Wen et al., 2020). Risk factors for poor survival in GBMs include older age and lower Karnofsky Performance Scale (KPS) scores at time of diagnosis, surgery without adjuvant chemoradiation, and absence of *MGMT* promoter methylation (Krex et al., 2007; Thumma et al., 2012). Radiographic MRI features have also been associated with worse overall survival including degree of necrosis and contrast enhancement, multifocality, peritumor edema and higher

rCBV (Hammoud et al., 1996; Lacroix et al., 2001; Pope et al., 2005; Jain et al., 2014).

ML and DL-based algorithms have been developed and evaluated to predict survival using a combination of radiographic and clinical features. Sun et al. used a 3D CNN for automated segmentation of cMRI images from 210 HGG and 75 LGG patients, and then used a ML-based random forest classifier to extract radiomics features and predict overall survival. They classified 66 gliomas in a validation cohort into short-term (<10 months), mid-term (10–15 months) and long-term (>15 months) survivors with a modest (61%) accuracy (Sun et al., 2019). Prasanna et al. used cMRI sequences from 65 patients with GBM from the TCIA, manually segmented into enhancing, peritumoral brain zone, and tumor necrosis regions; they extracted 402 radiomics features and used a random forest classifier to isolate features most predictive of short-term (< 7 months) vs. long term (>18 months) survival. They found that peritumoral radiomic features combined with multiparametric MRI sequences performed best at predicting long- vs. short-term survival with a concordance index (CI) of 0.70 (as opposed to combining tumor necrosis features with specific T1 or T2 sequences). When combined with clinical features the model's highest predictive accuracy was achieved at a CI of 0.735 (Prasanna et al., 2017). Lao et al. developed a DL-based model using cMRI combined with clinical data (age and KPS) from 112 patients with GBM from TCIA and institutional cohorts, to predict overall survival with a similar CI of 0.710 (Lao et al., 2017).

Nie et al. used T1 MRI, resting state functional MRI (rs-fMRI), and diffusion tensor imaging (DTI) from 68 HGG patients, and develop a 3D CNN to extract predictive radiomics features. These were combined with clinical features including age, gender, tumor location/size, and WHO grade, and incorporated into a SVM model to predict short vs. long overall survival time, defined as less than, or greater than 650 days, respectively, with 88% accuracy on a 25 patient validation cohort (Nie et al., 2019). Limitations of this study include the its small sample sizes, as well as a binary cutoff of 650 days defining short- vs. long-term survival. Park et al. extracted radiomics features from MRI DWI and PWI in addition to cMRI from 158 patients with newly diagnosed GBM, and combined these with clinical features including age, gender, KPS, *MGMT* promoter methylation status, and extent of surgical resection to develop a ML-based predictive model for survival. On a test set of 58 patients the model predicted OS with a CI of 0.70, performing better than the authors' models that used radiomics features or clinical predictors alone (Park et al., 2020).

As tumor hypoxia is considered an important molecular mechanism driving treatment resistance and poor prognosis, Beig et al. aimed to study radiomics features that predict tumor hypoxia, and utilized these to develop a predictive model for survival in GBM. Radiomics features extracted from cMRI of 115 subjects from the TCIA, coupled with RNA seq data from 21 genes implicated in GBM hypoxia, were used to generate a hypoxia enrichment score (HES). A random forest classifier was then used to stratify patients into short-term (OS <7 months), mid-term (OS 7–16 months) and long-term (OS >16 months) survival based on radiomic markers of hypoxia and clinical features (age, gender, KPS). On a validation subset, the model was able to predict a statistically significant separation between the Kaplan–Meier curves of short-term and long-term survivors, with a CI of 0.83 (Beig et al., 2018). In addition to predicting survival, non-invasive assessment of tumor hypoxia may

guide selection of patients for clinical trials or management with anti-angiogenic therapy (Rahman et al., 2010). Future studies on prognostication may benefit from greater incorporation of molecular features, including IDH mutation, *MGMT* promoter methylation and 1p/19q codeletion status into predictive models.

Differentiating progression from treatment-related radiographic changes (PseudoProgression)

Assessment of true progression (TP) of brain tumors (particularly HGG) on surveillance MRI often presents a significant clinical and radiologic challenge as true progression can appear radiographically similar to pseudoprogression (PsP), i.e., radiation treatment-related inflammatory changes most common 3–6 months after completing radiotherapy (Ellingson et al., 2017). Distinguishing TP from PsP is vital in guiding management and enrollment in (or withdrawal from) clinical trials. In practice, while pathological diagnosis is often considered gold standard to distinguish the two, serial MRI is often used for practical reasons as treatement-related changes regress over time (Youssef et al., 2023). However, this approach can lead to diagnostic delay. Moreover, TP and PsP may co-exist. Studies have suggested that recurrent tumors have lower ADC values than radiation necrosis on DWI sequences, and higher rCBV on PWI can predict PsP with 81.5% sensitivity and 77.8% specificity (Kong et al., 2011; Chu et al., 2013). Approaches utilizing radiomics, ML and DL have ventured to make this distinction noninvasively in the hopes of improving diagnostic fidelity.

Kim et al. studied cMRI, ADC and CBV sequences in 61 patients with GBMs who had undergone resection and standard concurrent chemoradiation therapy (CCRT), and had developed new contrast enhancing lesions within 12 weeks of completion of the latter. Ground truth of TP vs. PsP was based mostly on subsequent serial MRIs, though 8 cases were confirmed with pathology. They extracted radiomics features from the contrast-enhancing portion of the MRIs and used a ML-based classifier to develop a model to distinguish TP vs. PsP. Their multiparametric model (incorporating cMRI, ADC, CBV) performed the best with AUC 0.85 on an external validation cohort of 34 patients (Kim et al., 2019). Jang et al. used a similar cohort of 59 GBM patients to train a hybrid ML-DL model with CNN-LSTM (long short-term memory) on T1 pre- and post-contrast MRI, as well as clinical and molecular features, and were able to distinguish TP from PsP with AUC of 0.83 on an external validation set of 19 patients (Jang et al., 2018). Pathologic confirmation was available for 20 TP and 3 PsP cases. In a similar study employing data from 124 GBM patients with new enhancing lesion after resection and CCRT, Moassefi et al. trained a CNN that achieved AUC 0.75 in distinguishing TP from PsP, with all ground truth determination of TP vs. PSP based on serial imaging (Moassefi et al., 2022).

Discussion

The practice of neuro-oncology is developing at an ever-faster pace, propelled by advances in our understanding of brain tumor biology and technical innovations in allied fields such as neuropathology and neuroradiology. In parallel, advances in AI methods hold increasing promise to optimize workflows in many aspects of neuro-oncology care, as well as to generate new insights regarding tumor biology and therapeutic mechanisms. In neuropathology, to date, AI algorithms have been applied to WSI data to resolve histopathologic features, aiding brain tumor diagnosis and grading. In addition, ML is increasingly being applied to tumor classification on the basis of DNA methylome profiling. In neuroradiology, AI algorithms have been applied to the problem of tumor measurement (volumetrics), to the prediction of grade, molecular features and diagnosis, as well as to the discrimination between progression and treatment-related changes, and the determination of prognosis. As highlighted above, the accuracy of the output of many of these analyses depends on the complexity and diversity of the training datasets, and the AI methods applied to tackle each problem. There is clearly room for improvement, and this is expected through collaboration across centers (leading to more extensive and diverse datasets) and improvements in computational methods and hardware.

Witnessing the current progress, a natural question is if these algorithms will one day come to replace the work of neuropathologists and neuroradiologists in neuro-oncology practice. Although the roles of neuropathologists and neuroradiologists will evolve, we do not expect these specialists to come "out of the loop," as their expertise is irreplaceable, particularly when it comes to diagnosing and evaluating difficult cases. AI will not replace neuropathology or neuroradiology but rather expedite and enhance their workflows. With advances in large language models (LLM; e.g., ChatGPT), which are able to address complex queries with increasing accuracy (Haupt and Marks, 2023), the relevance of the clinical neuro-oncologist also comes into question. Here again, we think that AI will not replace but rather support the role of neuro-oncologists, putting the latest clinical evidence and treatment algorithms at their fingertips, systematizing part of their role but unable to replace the physical touch that enables patient assessment and the development of a relationship that helps guide patients through difficult decisions.

As the field of AI continues to develop and progressively integrate into research and clinical practice, we need to remain aware of the limitations of each method/algorithm, particularly since their underpinnings are often not clearly explained and, more importantly, are difficult to assess by end users. Guidelines for evaluating, validating and approving AI systems for their use in medicine in general, and neuro-oncology specifically, will be fundamental to the safe introduction of these methods into practice. Elements to consider when evaluating novel AI tools include, (1) the characteristics of training datasets (data types and standards, diversity of dataset elements, size of dataset, accuracy of data annotations, if relevant), (2) the specifics of the algorithms involved, (3) the characteristics of the validation dataset (including metrics that are consistent with those of the training dataset), and (4), the performance of the system at the moment of its release as well as over time, including specific warnings regarding blindspots of classification or systematic errors regarding output for specific inputs. Related to this last point, it is important to note that AI systems have the potential to perpetuate clinical and social biases (Larrazabal et al., 2020; Seyyed-Kalantari et al., 2021).

In the end, how AI will continue to integrate into the practice of neuro-oncology remains to be determined. We hope to have updated neuro-oncology clinicians and researchers on current advances in the field of AI to help them inform how to incorporate AI tools into their practice. In the words of the Nobel Prize winning physicist, Dennis Gabor, "the future cannot be predicted, but futures can be invented (Gabor, 1964)."

Author contributions

VN performed the literature review and drafted the manuscript. LG conceived the project, performed the literature review, and drafted the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Afridi, M., Jain, A., Aboian, M., and Payabvash, S. (2022). Brain tumor imaging: applications of artificial intelligence. *Semin. Ultrasound CT MR* 43, 153–169. doi: 10.1053/j.sult.2022.02.005

Aftab, K., Aamir, F. B., Mallick, S., Mubarak, F., Pope, W. B., Mikkelsen, T., et al. (2022). Radiomics for precision medicine in glioblastoma. *J. Neuro-Oncol.* 156, 217–231. doi: 10.1007/s11060-021-03933-1

Akkus, Z., Ali, I., Sedlář, J., Agrawal, J. P., Parney, I. F., Giannini, C., et al. (2017). Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence. *J. Digit. Imaging* 30, 469–476. doi: 10.1007/s10278-017-9984-3

Barash, Y., and Klang, E. (2019). Automated quantitative assessment of oncological disease progression using deep learning. *Ann. Trans. Med.* 7:S379. doi: 10.21037/atm.2019.12.101

Beig, N., Patel, J., Prasanna, P., Hill, V., Gupta, A., Correa, R., et al. (2018). Radiogenomic analysis of hypoxia pathway is predictive of overall survival in glioblastoma. *Sci. Rep.* 8:7. doi: 10.1038/s41598-017-18310-0

Bhandari, A. P., Liong, R., Koppen, J., Murthy, S. V., and Lasocki, A. (2021). Noninvasive determination of IDH and 1p19q status of lower-grade gliomas using MRI Radiomics: a systematic review. *AJNR Am. J. Neuroradiol.* 42, 94–101. doi: 10.3174/ajnr.A6875

Capper, D., Jones, D. T. W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., et al. (2018). DNA methylation-based classification of central nervous system tumours. *Nature* 555, 469–474. doi: 10.1038/nature26000

Capper, D., Stichel, D., Sahm, F., Jones, D. T. W., Schrimpf, D., Sill, M., et al. (2018). Practical implementation of DNA methylation and copy-number-based CNS tumor diagnostics: the Heidelberg experience. *Acta Neuropathol.* 136, 181–210. doi: 10.1007/s00401-018-1879-y

Chang, K., Bai, H. X., Zhou, H., Su, C., Bi, W. L., Agbodza, E., et al. (2018). Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging. *Clin. Cancer Res.* 24, 1073–1081. doi: 10.1158/1078-0432.CCR-17-2236

Chang, K., Beers, A. L., Bai, H. X., Brown, J. M., Ly, K. I., Li, X., et al. (2019). Automatic assessment of glioma burden: a deep learning algorithm for fully automated volumetric and bidimensional measurement. *Neuro-Oncology* 21, 1412–1422. doi: 10.1093/neuonc/noz106

Chang, P., Grinband, J., Weinberg, B. D., Bardis, M., Khy, M., Cadena, G., et al. (2018). Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *AJNR Am. J. Neuroradiol.* 39, 1201–1207. doi: 10.3174/ajnr.A5667

Chen, S., Xu, Y., Ye, M., Li, Y., Sun, Y., Liang, J., et al. (2022). Predicting MGMT promoter methylation in diffuse gliomas using deep learning with Radiomics. *J. Clin. Med.* 11:3445. doi: 10.3390/jcm11123445

Chu, H. H., Choi, S. H., Ryoo, I., Kim, S. C., Yeom, J. A., Shin, H., et al. (2013). Differentiation of true progression from pseudoprogression in glioblastoma treated with radiation therapy and concomitant temozolomide: comparison study of standard and high-b-value diffusion-weighted imaging. *Radiology* 269, 831–840. doi: 10.1148/radiol.13122024

Cluceru, J., Interian, Y., Phillips, J. J., Molinaro, A. M., Luks, T. L., Alcaide-Leon, P., et al. (2022). Improving the noninvasive classification of glioma genetic subtype with deep learning and diffusion-weighted imaging. *Neuro-Oncology* 24, 639–652. doi: 10.1093/neuonc/noab238

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., et al. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* 24, 1559–1567. doi: 10.1038/s41591-018-0177-5

Crisi, G., and Filice, S. (2020). Predicting MGMT promoter methylation of glioblastoma from dynamic susceptibility contrast perfusion: a Radiomic approach. *J. Neuroimaging* 30, 458–462. doi: 10.1111/jon.12724

di Nunno, V., Fordellone, M., Minniti, G., Asioli, S., Conti, A., Mazzatenta, D., et al. (2022). Machine learning in neuro-oncology: toward novel development fields. *J. Neuro-Oncol.* 159, 333–346. doi: 10.1007/s11060-022-04068-7

Ebrahimi, A., Skardelly, M., Schuhmann, M. U., Ebinger, M., Reuss, D., Neumann, M., et al. (2019). High frequency of H3 K27M mutations in adult midline gliomas. *J. Cancer Res. Clin. Oncol.* 145, 839–850. doi: 10.1007/s00432-018-02836-5

Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast Cancer. *JAMA* 318, 2199–2210. doi: 10.1001/jama.2017.14585

Ellingson, B. M. (2015). Radiogenomics and imaging phenotypes in glioblastoma: novel observations and correlation with molecular characteristics. *Curr. Neurol. Neurosci. Rep.* 15:506. doi: 10.1007/s11910-014-0506-0

Ellingson, B. M., Chung, C., Pope, W. B., Boxerman, J. L., and Kaufmann, T. J. (2017). Pseudoprogression, radionecrosis, inflammation or true tumor progression? Challenges associated with glioblastoma response assessment in an evolving therapeutic landscape. *J. Neuro-Oncol.* 134, 495–504. doi: 10.1007/s11060-017-2375-2

Ertosun, M. G., and Rubin, D. L. (2015). Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. *AMIA Annu. Symp. Proc.* 2015, 1899–1908.

Fellah, S., Caudal, D., de Paula, A. M., Dory-Lautrec, P., Figarella-Branger, D., Chinot, O., et al. (2013). Multimodal MR imaging (diffusion, perfusion, and spectroscopy): is it possible to distinguish oligodendroglial tumor grade and 1p/19q codeletion in the pretherapeutic diagnosis? *AJNR Am. J. Neuroradiol.* 34, 1326–1333. doi: 10.3174/ajnr.A3352

Ford, R. R., O'Neal, M., Moskowitz, S. C., and Fraunberger, J. (2016). Adjudication rates between readers in blinded independent central review of oncology studies. *J. Clin. Trials* 6, 1–9. doi: 10.4172/2167-0870.1000289

Gabor, D. (1964) *Inventing the future.* (New York: Alfred A Knopf).

Gonzalez Castro, L. N., Arrillaga-Romany, I. C., and Batchelor, T. T. (2023). Challenges and opportunities for clinical trials in patients with glioma. *JAMA Neurol.* 80, 227–228. doi: 10.1001/jamaneurol.2022.4924

Gritsch, S., Batchelor, T. T., and Gonzalez Castro, L. N. (2022). Diagnostic, therapeutic, and prognostic implications of the 2021 World Health Organization classification of tumors of the central nervous system. *Cancer* 128, 47–58. doi: 10.1002/cncr.33918

Hammoud, M. A., Sawaya, R., Shi, W., Thall, P. F., and Leeds, N. E. (1996). Prognostic significance of preoperative MRI scans in glioblastoma multiforme. *J. Neuro-Oncol.* 27, 65–73. doi: 10.1007/BF00146086

Haupt, C. E., and Marks, M. (2023). AI-generated medical advice-GPT and beyond. *JAMA* 329, 1349–1350. doi: 10.1001/jama.2023.5321

Hollon, T., Jiang, C., Chowdury, A., Nasir-Moin, M., Kondepudi, A., Aabedi, A., et al. (2023). Artificial-intelligence-based molecular classification of diffuse gliomas using rapid, label-free optical imaging. *Nat. Med.* 29, 828–832. doi: 10.1038/s41591-023-02252-4

Im, S., Hyeon, J., Rha, E., Lee, J., Choi, H. J., Jung, Y., et al. (2021). Classification of diffuse glioma subtype from clinical-grade pathological images using deep transfer learning. *Sensors (Basel)* 21:3500. doi: 10.3390/s21103500

Jain, R., Poisson, L. M., Gutman, D., Scarpace, L., Hwang, S. N., Holder, C. A., et al. (2014). Outcome prediction in patients with glioblastoma by using imaging, clinical,

and genomic biomarkers: focus on the nonenhancing component of the tumor. *Radiology* 272, 484–493. doi: 10.1148/radiol.14131691

Jang, B.-S., Jeon, S. H., Kim, I. H., and Kim, I. A. (2018). Prediction of Pseudoprogression versus progression using machine learning algorithm in glioblastoma. *Sci. Rep.* 8:12516. doi: 10.1038/s41598-018-31007-2

Jaunmuktane, Z., Capper, D., Jones, D. T. W., Schrimpf, D., Sill, M., Dutt, M., et al. (2019). Methylation array profiling of adult brain tumours: diagnostic outcomes in a large, single Centre. *Acta Neuropathol. Commun.* 7:24. doi: 10.1186/s40478-019-0668-8

Jian, A., Jang, K., Manuguerra, M., Liu, S., Magnussen, J., and di Ieva, A. (2021). Machine learning for the prediction of molecular markers in glioma on magnetic resonance imaging: a systematic review and Meta-analysis. *Neurosurgery* 89, 31–44. doi: 10.1093/neuros/nyab103

Jin, L., Shi, F., Chun, Q., Chen, H., Ma, Y., Wu, S., et al. (2021). Artificial intelligence neuropathologist for glioma classification using deep learning on hematoxylin and eosin stained slide images and molecular markers. *Neuro-Oncology* 23, 44–52. doi: 10.1093/neuonc/noaa163

Kann, B. H., Hosny, A., and Aerts, H. J. W. L. (2021). Artificial intelligence for clinical oncology. *Cancer Cell* 39, 916–927. doi: 10.1016/j.ccell.2021.04.002

Karimi, S., Zuccato, J. A., Mamatjan, Y., Mansouri, S., Suppiah, S., Nassiri, F., et al. (2019). The central nervous system tumor methylation classifier changes neuro-oncology practice for challenging brain tumor diagnoses and directly impacts patient care. *Clin. Epigenetics* 11:185. doi: 10.1186/s13148-019-0766-2

Karremann, M., Gielen, G. H., Hoffmann, M., Wiese, M., Colditz, N., Warmuth-Metz, M., et al. (2018). Diffuse high-grade gliomas with H3 K27M mutations carry a dismal prognosis independent of tumor location. *Neuro-Oncology* 20, 123–131. doi: 10.1093/neuonc/nox149

Kickingereder, P., Isensee, F., Tursunova, I., Petersen, J., Neuberger, U., Bonekamp, D., et al. (2019). Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *Lancet Oncol.* 20, 728–740. doi: 10.1016/S1470-2045(19)30098-1

Kim, J. Y., Park, J. E., Jo, Y., Shim, W. H., Nam, S. J., Kim, J. H., et al. (2019). Incorporating diffusion- and perfusion-weighted MRI into a radiomics model improves diagnostic performance for pseudoprogression in glioblastoma patients. *Neuro-Oncology* 21, 404–414. doi: 10.1093/neuonc/noy133

Kleinschmidt-DeMasters, B. K., and Mulcahy Levy, J. M. (2018). H3 K27M-mutant gliomas in adults vs. children share similar histological features and adverse prognosis. *Clin. Neuropathol.* 37, 53–63.

Komori, T. (2021). AI neuropathologist: an innovative technology enabling a faultless pathological diagnosis? *Neuro-Oncology* 23, 1–2. doi: 10.1093/neuonc/noaa229

Kong, D.-S., Kim, S. T., Kim, E. H., Lim, D. H., Kim, W. S., Suh, Y. L., et al. (2011). Diagnostic dilemma of pseudoprogression in the treatment of newly diagnosed glioblastomas: the role of assessing relative cerebral blood flow volume and oxygen-6-methylguanine-DNA methyltransferase promoter methylation status. *AJNR Am. J. Neuroradiol.* 32, 382–387. doi: 10.3174/ajnr.A2286

Korfiatis, P., Kline, T. L., Lachance, D. H., Parney, I. F., Buckner, J. C., and Erickson, B. J. (2017). Residual deep convolutional neural network predicts MGMT methylation status. *J. Digit. Imaging* 30, 622–628. doi: 10.1007/s10278-017-0009-z

Krex, D., Klink, B., Hartmann, C., von Deimling, A., Pietsch, T., Simon, M., et al. (2007). Long-term survival with glioblastoma multiforme. *Brain* 130, 2596–2606. doi: 10.1093/brain/awm204

Lacroix, M., Abi-Said, D., Fourney, D. R., Gokaslan, Z. L., Shi, W., DeMonte, F., et al. (2001). A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *J. Neurosurg.* 95, 190–198. doi: 10.3171/jns.2001.95.2.0190

Lao, J., Chen, Y., Li, Z. C., Li, Q., Zhang, J., Liu, J., et al. (2017). A deep learning-based Radiomics model for prediction of survival in glioblastoma Multiforme. *Sci. Rep.* 7:10353. doi: 10.1038/s41598-017-10649-8

Larrazabal, A. J., Nieto, N., Peterson, V., Milone, D. H., and Ferrante, E. (2020). Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc. Natl. Acad. Sci. U. S. A.* 117, 12592–12594. doi: 10.1073/pnas.1919012117

Lasocki, A., Anjari, M., Örs Kokurcan, S., and Thust, S. C. (2021). Conventional MRI features of adult diffuse glioma molecular subtypes: a systematic review. *Neuroradiology* 63, 353–362. doi: 10.1007/s00234-020-02532-7

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Li, Z.-C., Bai, H., Sun, Q., Li, Q., Liu, L., Zou, Y., et al. (2018). Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: a multicentre study. *Eur. Radiol.* 28, 3640–3650. doi: 10.1007/s00330-017-5302-1

Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., et al. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep.* 6:26286. doi: 10.1038/srep26286

Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., et al. (2021). The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-Oncology* 23, 1231–1251. doi: 10.1093/neuonc/noab106

Machine Learning Improves Diagnosis of CNS Cancers (2018). Machine learning improves diagnosis of CNS cancers. *Cancer Discov.* 8, 523–524. doi: 10.1158/2159-8290.CD-NB2018-040

Matsui, Y., Maruyama, T., Nitta, M., Saito, T., Tsuzuki, S., Tamura, M., et al. (2020). Prediction of lower-grade glioma molecular subtypes using deep learning. *J. Neuro-Oncol.* 146, 321–327. doi: 10.1007/s11060-019-03376-9

Miller, J. J., Gonzalez Castro, L. N., McBrayer, S., Weller, M., Cloughesy, T., Portnow, J., et al. (2023). Isocitrate dehydrogenase (IDH) mutant gliomas: a Society for Neuro-Oncology (SNO) consensus review on diagnosis, management, and future directions. *Neuro-Oncology* 25, 4–25. doi: 10.1093/neuonc/noac207

Moassefi, M., Faghani, S., Conte, G. M., Kowalchuk, R. O., Vahdati, S., Crompton, D. J., et al. (2022). A deep learning model for discriminating true progression from pseudoprogression in glioblastoma patients. *J. Neuro-Oncol.* 159, 447–455. doi: 10.1007/s11060-022-04080-x

Moran, S., Martínez-Cardús, A., Sayols, S., Musulén, E., Balañá, C., Estival-Gonzalez, A., et al. (2016). Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol.* 17, 1386–1395. doi: 10.1016/S1470-2045(16)30297-2

Nagpal, K., Foote, D., Tan, F., Liu, Y., Chen, P. H. C., Steiner, D. F., et al. (2020). Development and validation of a deep learning algorithm for Gleason grading of prostate Cancer from biopsy specimens. *JAMA Oncol.* 6, 1372–1380. doi: 10.1001/jamaoncol.2020.2485

Nie, D., Lu, J., Zhang, H., Adeli, E., Wang, J., Yu, Z., et al. (2019). Multi-Channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal Neuroimages. *Sci. Rep.* 9:1103. doi: 10.1038/s41598-018-37387-9

Nillson, N. J. *The quest for artificial intelligence*. Cambridge: Cambridge University Press, (2010).

Park, Y. W., Han, K., Ahn, S. S., Bae, S., Choi, Y. S., Chang, J. H., et al. (2018). Prediction of IDH1-mutation and 1p/19q-Codeletion status using preoperative MR imaging phenotypes in lower grade gliomas. *AJNR Am. J. Neuroradiol.* 39, 37–42. doi: 10.3174/ajnr.A5421

Park, J. E., Kim, H. S., Jo, Y., Yoo, R. E., Choi, S. H., Nam, S. J., et al. (2020). Radiomics prognostication model in glioblastoma using diffusion- and perfusion-weighted MRI. *Sci. Rep.* 10:4250. doi: 10.1038/s41598-020-61178-w

Patel, S. H., Poisson, L. M., Brat, D. J., Zhou, Y., Cooper, L., Snuderl, M., et al. (2017). T2-FLAIR mismatch, an imaging biomarker for IDH and 1p/19q status in lower-grade gliomas: a TCGA/TCIA project. *Clin. Cancer Res.* 23, 6078–6085. doi: 10.1158/1078-0432.CCR-17-0560

Pei, L., Jones, K. A., Shboul, Z. A., Chen, J. Y., and Iftekharuddin, K. M. (2021). Deep neural network analysis of pathology images with integrated molecular data for enhanced glioma classification and grading. *Front. Oncol.* 11:668694. doi: 10.3389/fonc.2021.668694

Peng, J., Kim, D. D., Patel, J. B., Zeng, X., Huang, J., Chang, K., et al. (2022). Deep learning-based automatic tumor burden assessment of pediatric high-grade gliomas, medulloblastomas, and other leptomeningeal seeding tumors. *Neuro-Oncology* 24, 289–299. doi: 10.1093/neuonc/noab151

Pickles, J. C., Stone, T. J., and Jacques, T. S. (2020). Methylation-based algorithms for diagnosis: experience from neuro-oncology. *J. Pathol.* 250, 510–517. doi: 10.1002/path.5397

Pope, W. B., and Hessel, C. (2011). Response assessment in neuro-oncology criteria: implementation challenges in multicenter neuro-oncology trials. *AJNR Am. J. Neuroradiol.* 32, 794–797. doi: 10.3174/ajnr.A2582

Pope, W. B., Sayre, J., Perlina, A., Villablanca, J. P., Mischel, P. S., and Cloughesy, T. F. (2005). MR imaging correlates of survival in patients with high-grade gliomas. *AJNR Am. J. Neuroradiol.* 26, 2466–2474.

Prasanna, P., Patel, J., Partovi, S., Madabhushi, A., and Tiwari, P. (2017). Radiomic features from the peritumoral brain parenchyma on treatment-naïve multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. *Eur. Radiol.* 27, 4188–4197. doi: 10.1007/s00330-016-4637-3

Priesterbach-Ackley, L. P., Boldt, H. B., Petersen, J. K., Bervoets, N., Scheie, D., Ulhøi, B. P., et al. (2020). Brain tumour diagnostics using a DNA methylation-based classifier as a diagnostic support tool. *Neuropathol. Appl. Neurobiol.* 46, 478–492. doi: 10.1111/nan.12610

Rahman, R., Smith, S., Rahman, C., and Grundy, R. (2010). Antiangiogenic therapy and mechanisms of tumor resistance in malignant glioma. *J. Oncol.* 2010:251231, 1–16. doi: 10.1155/2010/251231

Russel, S., and Norvig, P. *Artificial intelligence: a modern approach*. London: Pearson (2020).

Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y., and Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat. Med.* 27, 2176–2182. doi: 10.1038/s41591-021-01595-0

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature* 529, 484–489. doi: 10.1038/nature16961

Smits, M., and van den Bent, M. J. (2017). Imaging correlates of adult glioma genotypes. *Radiology* 284, 316–331. doi: 10.1148/radiol.2017151930

Sorensen, A. G., Batchelor, T. T., Wen, P. Y., Zhang, W.-T., and Jain, R. K. (2008). Response criteria for glioma. *Nat. Clin. Pract. Oncol.* 5, 634–644. doi: 10.1038/ncponc1204

Sorensen, A. G., Patel, S., Harmath, C., Bridges, S., Synnott, J., Sievers, A., et al. (2001). Comparison of diameter and perimeter methods for tumor volume calculation. *J. Clin. Oncol.* 19, 551–557. doi: 10.1200/JCO.2001.19.2.551

Stupp, R., Hegi, M. E., Mason, W. P., van den Bent, M., Taphoorn, M. J., Janzer, R. C., et al. (2009). Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol.* 10, 459–466. doi: 10.1016/S1470-2045(09)70025-7

Suh, C. H., Kim, H. S., Jung, S. C., Choi, C. G., and Kim, S. J. (2019). Imaging prediction of isocitrate dehydrogenase (IDH) mutation in patients with glioma: a systemic review and meta-analysis. *Eur. Radiol.* 29, 745–758. doi: 10.1007/s00330-018-5608-7

Sun, L., Zhang, S., Chen, H., and Luo, L. (2019). Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning. *Front. Neurosci.* 13:810. doi: 10.3389/fnins.2019.00810

Taal, W., Bromberg, J. E. C., and van den Bent, M. J. (2015). Chemotherapy in glioma. *CNS Oncologia* 4, 179–192. doi: 10.2217/cns.15.2

Thumma, S. R., Fairbanks, R. K., Lamoreaux, W. T., Mackay, A. R., Demakas, J. J., Cooke, B. S., et al. (2012). Effect of pretreatment clinical factors on overall survival in glioblastoma multiforme: a surveillance epidemiology and end results (SEER) population analysis. *World J. Surg. Oncol.* 10:75. doi: 10.1186/1477-7819-10-75

Truong, A. H., Sharmanska, V., Limbäck-Stanic, C., and Grech-Sollars, M. (2020). Optimization of deep learning methods for visualization of tumor heterogeneity and brain tumor grading through digital pathology. *Neurooncol Adv.* 2, 1–13. doi: 10.1093/noajnl/vdaa110

van den Bent, M. J. (2010). Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective. *Acta Neuropathol.* 120, 297–304. doi: 10.1007/s00401-010-0725-7

van der Voort, S. R., Incekara, F., Wijnenga, M. M. J., Kapas, G., Gardeniers, M., Schouten, J. W., et al. (2019). Predicting the 1p/19q Codeletion status of presumed low-grade glioma with an externally validated machine learning algorithm. *Clin. Cancer Res.* 25, 7455–7462. doi: 10.1158/1078-0432.CCR-19-1127

Vos, M. J., Uitdehaag, B. M. J., Barkhof, F., Heimans, J. J., Baayen, H. C., Boogerd, W., et al. (2003). Interobserver variability in the radiological assessment of response to chemotherapy in glioma. *Neurology* 60, 826–830. doi: 10.1212/01.WNL.0000049467.54667.92

Wen, P. Y., Weller, M., Lee, E. Q., Alexander, B. M., Barnholtz-Sloan, J. S., Barthel, F. P., et al. (2020). Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro-Oncology* 22, 1073–1113. doi: 10.1093/neuonc/noaa106

Wong, D., and Yip, S. (2018). Machine learning classifies cancer. *Nature* 555, 446–447. doi: 10.1038/d41586-018-02881-7

Xing, Z., Yang, X., She, D., Lin, Y., Zhang, Y., and Cao, D. (2017). Noninvasive assessment of IDH mutational status in World Health Organization grade II and III Astrocytomas using DWI and DSC-PWI combined with conventional MR imaging. *AJNR Am. J. Neuroradiol.* 38, 1138–1144. doi: 10.3174/ajnr.A5171

Yogananda, C. G. B., et al. (2020). A novel fully automated MRI-based deep-learning method for classification of 1p/19q co-deletion status in brain gliomas. *Neuro Oncol. Adv.* 2, vdaa066. iv42–iv48.

Youssef, G., Rahman, R., Bay, C., Wang, W., Lim-Fat, M. J., Arnaout, O., et al. (2023). Evaluation of standard response assessment in neuro-oncology, modified response assessment in neuro-oncology, and immunotherapy response assessment in neuro-oncology in newly diagnosed and recurrent glioblastoma. *J. Clin. Oncol.* 41, 3160–3171. doi: 10.1200/JCO.22.01579

Zhang, B., Chang, K., Ramkissoon, S., Tanguturi, S., Bi, W. L., Reardon, D. A., et al. (2017). Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro-Oncology* 19, 109–117. doi: 10.1093/neuonc/now121

Zhao, J., Huang, Y., Song, Y., Xie, D., Hu, M., Qiu, H., et al. (2020). Diagnostic accuracy and potential covariates for machine learning to identify IDH mutations in glioma patients: evidence from a meta-analysis. *Eur. Radiol.* 30, 4664–4674. doi: 10.1007/s00330-020-06717-9

Zhou, H., Chang, K., Bai, H. X., Xiao, B., Su, C., Bi, W. L., et al. (2019). Machine learning reveals multimodal MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low- and high-grade gliomas. *J. Neuro-Oncol.* 142, 299–307. doi: 10.1007/s11060-019-03096-0

Zhu, M., Li, S., Kuang, Y., Hill, V. B., Heimberger, A. B., Zhai, L., et al. (2022). Artificial intelligence in the radiomic analysis of glioblastomas: a review, taxonomy, and perspective. *Front. Oncol.* 12:924245. doi: 10.3389/fonc.2022.924245