

Review

Artificial intelligence in neuro-oncology imaging: Advancing brain tumor detection, grading, and treatment response evaluation



Mohamed Mustaf Ahmed^a, Uthman Okikiola Adebayo^{b,c,*}, Zhinya Kawa Othman^d, Majd Oweidat^e, Olalekan John Okesanya^f, Omar Kasimieh^g, Shuaibu Saidu Musa^a, Christian Joseph N. Ong^h, Francesco Brandaⁱ, Don Eliseo Lucero-Prisno III^j

^a School of Global Health, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand

^b Department of Medical Laboratory Services, Federal University of Health Sciences Teaching Hospital, Ila-Orangun 234001, Nigeria

^c Department of Public Health, Faculty of Medicine and Health Sciences, Hormuud University, Mogadishu BN 03010, Somalia

^d Department of Pharmacy, Kurdistan Technical Institute, Sulaymaniyah 46001, Iraq

^e College of Medicine, Hebron University, Hebron P727, Palestine

^f Department of Public Health and Maritime Transport, University of Thessaly, Volos 3822, Greece

^g Department of Medicine, University of the East Ramon Magsaysay Memorial Medical Center, Quezon City 1113, Philippines

^h Department of Biology, College of Science, De La Salle University, Manila 0922, Philippines

ⁱ Unit of Medical Statistics and Molecular Epidemiology, Campus Bio-Medico University of Rome, Rome 00128, Italy

^j Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

ARTICLE INFO

Keywords:

Brain tumor
Neuroimaging
Deep learning
Radiomics
Pseudoprogression
Interoperability

ABSTRACT

This narrative review synthesizes how artificial intelligence (AI) is reshaping neuro-oncology imaging through automated detection, segmentation, grading, and longitudinal monitoring of brain tumors using multiparametric magnetic resonance imaging (MRI) and hybrid positron emission tomography (PET) combined with MRI. Conventional interpretation remains constrained by interobserver variability, subjective estimation of tumor burden, and persistent difficulty in separating treatment-related changes, particularly pseudoprogression and radiation effects, from true recurrence, creating uncertainty in clinical decision-making. Deep learning segmentation pipelines now support fast, reproducible delineation of enhancing tumors, non-enhancing tumors, edema, and postoperative cavities, generating quantitative two-dimensional and three-dimensional metrics that better reflect complex, infiltrative diseases than manual measurements. Radiomics and multimodal machine learning models leveraging diffusion, perfusion, spectroscopy, and amino-acid PET features provide biologically informed characterization, support radiogenomic associations aligned with the 2021 World Health Organization Classification of Central Nervous System Tumors, fifth edition molecular taxonomy, and improve post-therapy response evaluation when morphology alone is equivocal. However, real-world translation remains limited by domain shifts across scanners and protocols, missing or inconsistent sequences, label noise, and uncertain ground truth in the post-treatment setting. Clinical adoption further depends on standards-based interoperability, human-in-the-loop quality assurance, usability and workflow fit, rigorous external validation including prospective and multi-reader studies, and clear lifecycle governance. Regulatory readiness requires alignment with United States Food and Drug Administration Software as a Medical Device guidance and the evolving European Union Medical Device Regulation and AI Act landscape, alongside transparent reporting and risk-of-bias appraisal using emerging guidelines that target prediction models incorporating AI. This review synthesizes end-to-end evidence and practical pathways for the safe and scalable deployment of AI in brain tumor neuroimaging, with an emphasis on segmentation-driven quantification, distinguishing pseudoprogression from true progression, probability calibration, and decision-curve evaluation of clinical utility.

* Corresponding author at: Department of Medical Laboratory Services, Federal University of Health Sciences Teaching Hospital, Ila-Orangun 234001, Nigeria.
E-mail address: uthmanadebayo85@gmail.com (U.O. Adebayo).

<https://doi.org/10.1016/j.intonc.2026.100071>

Received 16 April 2026; Received in revised form 22 June 2026; Accepted 25 June 2026

Available online 29 June 2026

2950-2616/© 2026 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Brain tumors represent a major clinical and public health challenge, with high morbidity and mortality driven by diagnostic complexity, heterogeneous biological behavior, and variability in treatment response.^{1,2} Neuroimaging, anchored by magnetic resonance imaging (MRI), remains central to the diagnosis and management of brain tumors, guiding initial detection, surgical planning, radiotherapy targeting, and monitoring after treatment.^{3–5} However, conventional neuroimaging interpretation is limited by interobserver variability, subjective assessment of disease burden, and difficulties in distinguishing post-treatment effects, such as pseudoprogression, from true tumor recurrence. These challenges create uncertainty in clinical decision-making and delay optimal care delivery. Recent advances in artificial intelligence (AI), especially machine learning (ML) and deep learning (DL), offer substantive opportunities to enhance neuroimaging by enabling automated tumor segmentation, objective volumetric quantification, predictive grading, and improved post-therapy evaluation.^{6,7}

In recent years, AI model development has produced numerous research articles and proof-of-concept systems demonstrating improved accuracy for tumor detection and classification.^{7–9} Despite this progress, the translation of these findings into real-world clinical practice remains limited. Most studies have focused on model accuracy without addressing key questions such as the harmonization of input imaging protocols, robustness across institutions, integration into radiology workflows, and compliance with regulatory frameworks such as the U.S. Food and Drug Administration (FDA) or the European Union Medical Device Regulation (EU MDR).^{10,11} As a result, clinicians, radiologists, and researchers increasingly need a consolidated, clinically oriented synthesis that connects algorithm development to operational deployment.¹²

This review responds to this need by providing an end-to-end evaluation of the field, covering technical methods, the supporting evidence base, post-treatment disease assessment, and practical pathways for safe, reliable, and scalable adoption. Compared with existing AI in neuro-oncology reviews, which typically concentrate on segmentation benchmarks or model accuracy in isolation, the present synthesis is distinguished by (i) coupling technical performance to interoperability, lifecycle governance, and AI-specific reporting standards, (ii) providing an evidence-graded comparison of MRI radiomics, perfusion MRI, amino-acid positron emission tomography (PET), and longitudinal AI models for the pseudoprogression versus true-progression problem, and (iii) embedding probability calibration, decision curve analysis, and clinical-readiness tiers into the assessment of clinical utility. Particular attention is given to AI use after therapy, where detecting subtle treatment-related effects is among the most clinically consequential challenges. Accordingly, we examined AI approaches in brain tumor neuroimaging, with an emphasis on tumor segmentation, three-dimensional volumetric assessment, and differentiation between pseudoprogression and true tumor recurrence, alongside the requirements for embedding AI in practice, including interoperability, governance, validation, and regulatory compliance.

2. Literature search and selection

This narrative review was conducted to examine the integration of AI in neuroimaging for brain tumor detection, grading, volumetric assessment, and post-treatment response evaluation, with particular attention to its clinical applicability, workflow integration, and regulatory readiness. Searches were conducted in PubMed, Scopus, Web of Science, and Google Scholar for articles published in English. Recent articles were prioritized, but the search applied no lower date limit; it was last updated in May 2026, with the earliest and most recent included sources published in 2006 and April 2026, respectively. Beyond recency, sources were prioritized according to peer-reviewed status,

methodological rigor and validation strategy (for example, external or multi-institutional validation), dataset size and quality, and direct relevance to the clinical translation of AI in brain-tumor neuroimaging; foundational works were retained where needed to provide conceptual context. The search strategy incorporated both MeSH (Medical Subject Headings) and free-text keywords, which were adapted to each database's indexing system using the Boolean operators AND and OR. Core search terms included “artificial intelligence,” “machine learning,” “deep learning,” “MRI,” “neuroimaging,” “brain tumor segmentation,” “tumor grading,” “pseudoprogression,” “treatment response evaluation,” “clinical integration,” and “medical device regulation.” The screening was done by two independent reviewers (O. J. O. and U. O. A.) to assess relevance. Any disagreements were resolved by consensus with a third reviewer (M.M.A). The inclusion criteria comprised primary research articles on AI applications in brain tumor neuroimaging; narrative and systematic reviews and meta-analyses on these topics; and regulatory or professional guidelines and reporting standards relevant to the clinical deployment of imaging AI (for example, those addressing the FDA and EU MDR pathways). Studies were excluded if they focused solely on non-AI imaging approaches, applied AI only to non-tumor neurological conditions, were unrelated to clinical decision-making in brain tumor care, or were not accessible in English. Eligible publications were read and summarized, with attention to AI model type, imaging modality, dataset source and size, reported performance metrics, validation approach, clinical or post-treatment context, workflow considerations, regulatory aspects, and known methodological limitations. Findings were synthesized narratively across the sections of this review. In total, 125 sources (primary research studies, reviews and meta-analyses, and regulatory or professional guidelines and standards) met these criteria and were synthesized in this review.

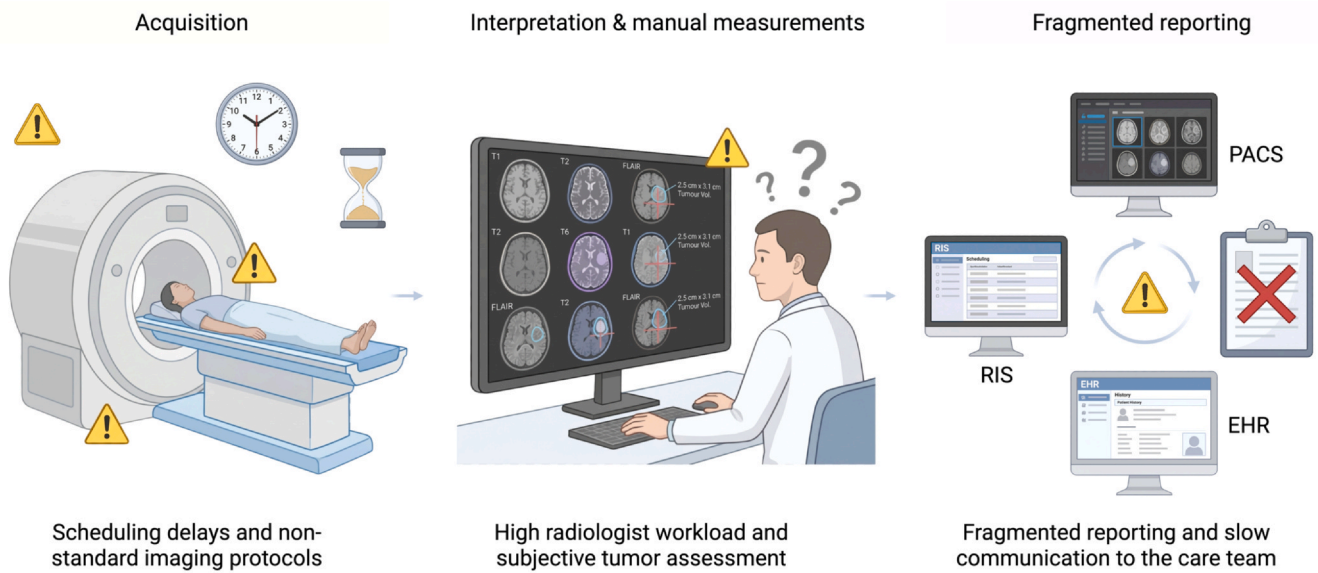
3. Clinical and imaging context

To frame the baseline workflow and its key interpretive bottlenecks, we summarize the conventional neuroimaging pathway in [Figure 1](#).

3.1. Molecular classification and imaging implications

The 2021 World Health Organization Classification of Central Nervous System Tumors, fifth edition (WHO CNS5) represents a paradigm shift from a primarily histology-driven system to one in which molecular alterations define tumor type. Core diagnostic biomarkers, including isocitrate dehydrogenase (IDH) mutations, 1p/19q codeletion, telomerase reverse transcriptase promoter mutations, α -thalassemia/mental retardation X-linked loss, and CDKN2A/CDKN2B homozygous deletion, now define tumor type and grade, while O⁶-methylguanine-DNA methyltransferase (MGMT) promoter methylation functions as a complementary predictive biomarker for alkylating-chemotherapy response; copy number variations and epigenetic signatures are increasingly used to resolve ambiguous cases.² Methylation profiling is increasingly used as an effective ancillary method for precise classification, whereas molecular integration has reshaped imaging priorities. For instance, IDH-mutant diffuse gliomas often exhibit extensive non-enhancing regions, necessitating imaging strategies that are aligned with their biology.¹³

Radiogenomic studies have revealed that multiparametric MRI, including anatomical, diffusion, perfusion, and susceptibility sequences, captures subtle correlates of tumor genetics. Quantitative features such as normalized relative cerebral blood volume (rCBV) and relative cerebral blood flow (rCBF) have been associated with epidermal growth factor receptor (EGFR) amplification, CDKN2A loss, and receptor tyrosine kinase subgroup status, although ML models currently achieve only moderate predictive accuracy.¹⁴ Emerging MRI techniques, including magnetic resonance (MR) fingerprinting, diffusion kurtosis imaging, neurite orientation dispersion and density imaging, chemical exchange saturation transfer, and arterial spin labeling, alongside



From scan to treatment decision: cumulative delays and failure points

Figure 1. Traditional neuroimaging workflow for brain tumors and key bottlenecks, from image acquisition and manual interpretation to fragmented reporting. Abbreviations: 1. PACS: picture archiving and communication system 2. RIS: radiology information system 3. EHR: electronic health record. This Figure was created in BioRender. <https://BioRender.com/a0xpcni>.

hybrid PET-MRI with novel tracers, for example, O-(2-[18F]fluoroethyl)-L-tyrosine and 3,4-dihydroxy-6-[18F]fluoro-L-phenylalanine PET, have enhanced the mapping of molecular, metabolic, and microstructural tumor features.^{15–17}

Tumor heterogeneity, both spatial and temporal, necessitates biologically informed segmentation and longitudinal imaging to capture clonal evolution, treatment-induced molecular shifts, and intratumoral gradients, which are critical for therapy planning and trial stratification purposes. AI models should increasingly adopt molecularly defined labels rather than histology alone, integrating multimodal data (imaging, genomics, pathology, and clinical) through approaches such as vision transformers, weakly supervised learning, and federated networks to accommodate data limitations while preserving privacy.^{18,19} Clinically, molecularly informed imaging supports non-invasive tumor characterization, aids in predicting treatment response, including radiotherapy, targeted therapies, and immunotherapy, and helps distinguish true progression from pseudoprogression or treatment-related changes in the tumor. Standardization initiatives, such as the Quantitative Imaging Biomarkers Alliance guidelines, ensure the reproducibility of quantitative imaging biomarkers across scanners and centers, facilitating reliable radiogenomic correlations for precision neuro-oncology.²⁰

3.2. Response assessment frameworks

The updated Response Assessment in Neuro-Oncology (RANO) 2.0 criteria refine long-standing response assessment practices by unifying the guidelines for both high- and low-grade gliomas. The framework preserves the strengths of earlier versions by emphasizing reproducible measures of enhancing versus non-enhancing disease, clear definitions of measurable lesions, standardized MRI protocols, and consideration of steroid dosing and clinical status, which influence radiologic interpretation.²¹ Two-dimensional (2D) measurements remain the primary metric, supplemented by standardized baseline imaging and careful confirmation of early post-radiotherapy changes to account for pseudoprogression.

Despite its widespread use, RANO 2.0 is constrained by labor-intensive, shape-dependent 2D metrics that are sensitive to head positioning, inter-reader variability, and scanner differences, often underestimating

infiltrative or non-enhancing tumor components.²² Manual lesion counting and differentiation between measurable and non-measurable diseases remain inconsistent, particularly near resection cavities. Emerging automated segmentation and volumetry tools provide rapid and reproducible quantification, supporting 3D volumetric growth rates, functional tumor volume assessments, and longitudinal time-series analyses that more accurately reflect the true tumor burden.

The integration of multiparametric MRI (diffusion, perfusion, and spectroscopy), advanced PET tracers (e.g., fluoroethyltyrosine [FET]-PET and fluorodopa [FDOPA]-PET), and radiomic features (texture, heterogeneity, and shape complexity) allows for a biologically informed, molecularly correlated response assessment. These approaches are increasingly aligned with the WHO CNS5 molecular categories, facilitating noninvasive predictions of IDH, EGFR, MGMT, and other alterations, and enabling treatment-stratified endpoints. AI-driven multiparametric radiomics and multimodal fusion models enhance reproducibility and predictive power, reducing subjectivity for molecular-genotype prediction and for distinguishing tumor progression from treatment-related changes.^{23,24} Standardization initiatives, such as a consensus brain-tumor imaging protocol, ensure the harmonization of image acquisition across scanners and institutions.²⁰ Collectively, automated, volumetric, and biologically informed tools represent the next evolution of RANO-style frameworks, providing objective, reproducible, and clinically actionable metrics for precision neuro-oncology.²²

3.3. Modalities and sequences

Advanced neuroimaging modalities provide physiological and metabolic contexts that complement conventional MRI for glioma characterization. Diffusion-weighted imaging and apparent diffusion coefficient (ADC) maps remain core markers of cellularity, with restricted diffusion indicating dense tumor tissue.²⁵ Emerging diffusion techniques, including Diffusion Kurtosis Imaging, Neurite Orientation Dispersion and Density Imaging, free-water imaging, and Restriction Spectrum Imaging, offer improved microstructural specificity for infiltrative tumor margins and correlate with molecular features such as IDH and MGMT status.²⁶ Perfusion MRI using dynamic susceptibility

Table 1
Clinical framework and imaging inputs relevant to AI neuro-oncology.

Clinical framework element	What it changes clinically	Imaging inputs most relevant to AI	Reference(s)
WHO CNS5 molecular taxonomy	Tumor entities increasingly defined by molecular alterations, pushing imaging toward biology-aligned characterization and endpoints	Multisequence MRI patterns linked to molecular subtype (enhancing vs. non-enhancing burden; microstructural/vascular signatures)	2,14
Radiogenomics in gliomas	Enables non-invasive prediction of select molecular traits (often moderate accuracy), supporting stratification and hypothesis generation	Multiparametric MRI radiomics (anatomic + diffusion + perfusion + susceptibility) and multiregional feature extraction	14
RANO 2.0	Standardizes response interpretation but remains limited by 2D, labor-intensive, reader-variable measurements	Baseline + follow-up MRI with consistent acquisition; measurable enhancing vs. non-enhancing components	21
Need for automated longitudinal quantification	Creates clinical demand to move beyond 2D measurements toward reproducible volumetric and longitudinal tracking, especially for complex/non-spherical disease and post-operative cavities	Serial post-treatment MRI capturing non-spherical residual disease and cavity geometry suitable for volumetric assessment	22
Core physiologic/metabolic imaging inputs	Adds cellularity, perfusion/vascularity, and metabolic specificity when morphology is equivocal	DWI/ADC; perfusion; MRS (incl. 2-HG for IDH); amino-acid PET (e.g., FET) and PET/MRI combinations	24,25,29,32,33
Standardized trial-style imaging protocols	Reduces variability across scanners/sites and supports more reliable longitudinal comparison	Protocol-driven MRI acquisition schedules and sequence selection for multicenter consistency	21,40

Abbreviations: 1. 2D: two-dimensional 2. ADC: apparent diffusion coefficient 3. FET: fluoroethyltyrosine 4. IDH: isocitrate dehydrogenase 5. MRI: magnetic resonance imaging 6. MRS: magnetic resonance spectroscopy 7. PET: positron emission tomography 8. RANO: Response Assessment in Neuro-Oncology 9. WHO CNS5: World Health Organization Classification of Central Nervous System Tumors, fifth edition 10. DWI: diffusion-weighted imaging 11. 2-HG: 2-hydroxyglutarate 12. AI: artificial intelligence.

contrast and dynamic contrast-enhanced imaging quantifies rCBV, rCBF, and vascular permeability and reliably identifies angiogenesis and high-grade transformation.^{27,28} Arterial spin labeling is a non-contrast alternative with predictive value for glioma grading and IDH mutation.^{29,30}

MR spectroscopy extends the physiological assessment by detecting metabolic signatures, such as elevated choline and lactate and reduced N-acetylaspartate levels.³¹ Advanced spectroscopy targeting 2-hydroxyglutarate enables the noninvasive identification of IDH-mutant gliomas, an essential WHO CNS5 biomarker.^{32,33} Additional metabolites, including glutamate/glutamine and lipids, further differentiate viable tumors from treatment-related changes.³¹ Complementary physiologic MRI techniques, such as susceptibility-weighted imaging, quantitative susceptibility mapping, chemical exchange saturation transfer, and MR fingerprinting, provide insights into vascular microarchitecture, iron load, tissue pH, and quantitative T1/T2 mapping.^{15,16,34,35} PET using amino-acid tracers (FET, FDOPA, methionine) refines the detection of recurrence versus pseudoprogession and adds metabolic specificity beyond MRI.³⁶ Emerging hybrid PET/MR workflows are also exploring additional molecular targets beyond amino-acid uptake, including translocator protein for neuroinflammation and prostate-specific membrane antigen for neovascularity, although clinical evidence in neuro-oncology remains preliminary.^{37,38}

DL-based segmentation frameworks increasingly rely on multi-modal MRI inputs (T1, T2, fluid-attenuated inversion recovery [FLAIR], and contrast-enhanced T1-weighted [T1CE]), which improve tumor compartment delineation; however, their performance deteriorates when sequences are missing or inconsistently acquired, and models remain vulnerable to scanner variability and domain shift.³⁹ Standardizing heterogeneous datasets remains a major challenge, and translation from benchmark datasets to routine clinical imaging is limited. Recent developments, including sequence synthesis models, domain adaptation, harmonization techniques, and self-supervised learning, aim to improve robustness across clinical environments, while ongoing studies are expanding segmentation beyond gliomas to other intracranial tumor types. Collectively, these structural, physiological, metabolic, and AI-enhanced imaging modalities provide a multimodal foundation for radiogenomic profiling, treatment response assessment, and WHO CNS5-aligned precision neuro-oncology. Ultimately, the evolving molecular taxonomy, response frameworks, and advanced imaging modalities define a complex clinical landscape in which

accurate and reproducible quantification of tumor burden is increasingly essential. Key clinical frameworks and imaging inputs that shape AI-enabled neuro-oncology are summarized in Table 1.

4. AI for tumor segmentation, 3D volumetry, and biomarker characterization

The end-to-end AI-enabled neuroimaging pipeline across segmentation, quantification, and monitoring is outlined in Figure 2.

4.1. Benchmarks and datasets

Recent Brain Tumor Segmentation Challenge (BraTS) editions have evolved from a single pre-treatment glioma benchmark into a cluster of clinically realistic tasks. The 2024 BraTS cluster now includes post-treatment adult glioma segmentation (BraTS–2024 post-treatment), pre- and post-treatment brain metastases (BraTS-METS, originally launched in 2023 and extended in 2024), meningioma radiotherapy planning, pediatric brain tumors, and the Sub-Saharan Africa glioma cohort, alongside generalizability, augmentation, MRI synthesis, and MRI inpainting subtasks. Across these tasks, BraTS 2024 comprises adult post-treatment glioma (≈2,200 cases), pediatric tumors (≈2,490 train+validation), pre- and post-treatment brain metastases (≈460 cases), meningioma radiotherapy planning (≈650 train+validation), BraTS-Africa (≈750 cases), and ≈280,000 histology samples in the pathology track, substantially expanding demographic, scanner, and disease diversity compared with earlier editions.⁴¹

Quantitatively, post-treatment glioma segmentation pipelines built around the BraTS 2024 task reported whole-lesion Dice in the high 0.8 s (≈0.87–0.92) and tumor-core Dice across a wider range (≈0.61–0.91), depending on the subregion and the presence of resection cavities and treatment-related inflammatory or radiation effects.^{42,43} The brain-metastases track reported systematically lower lesion-wise performance (driven by small, miliary lesions where false positives and false negatives are heavily penalized), while the Sub-Saharan Africa and pediatric tracks expose persistent domain-shift challenges with lower out-of-distribution performance than adult Western cohorts.^{44,45} These developments provide richer training distributions, support research on longitudinal segmentation and treatment monitoring, and help address long-standing generalizability gaps in the deployment of clinical AI.

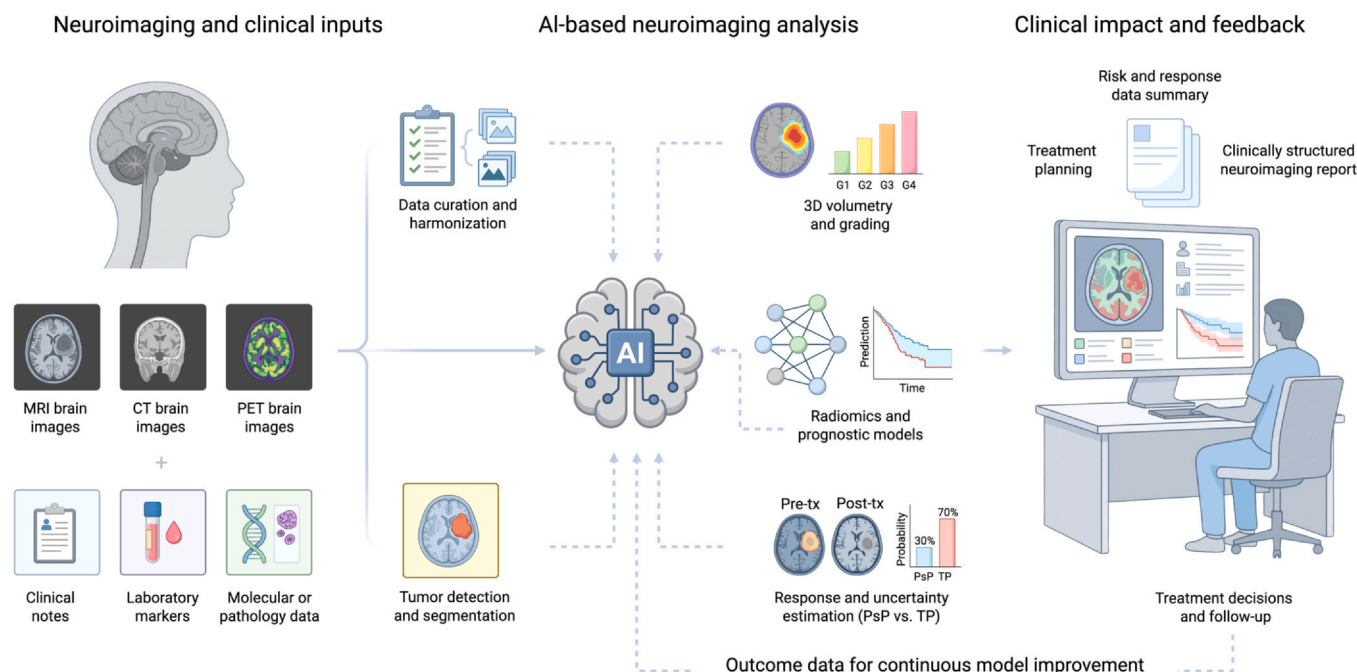


Figure 2. AI-enabled neuroimaging pipeline for brain tumors. Abbreviations: 1. MRI, magnetic resonance imaging 2. CT: computed tomography 3. PET: positron emission tomography 4. 3D: three-dimensional 5. PsP: pseudoprogression 6. TP: true progression 7. AI: artificial intelligence. This Figure was created in BioRender. <https://BioRender.com/2z1c486>.

4.2. Model families and performance

Convolutional neural network (CNN)-based architectures, particularly nnU-Net, remain foundational, demonstrating that systematic auto-configuration of preprocessing, network design, and post-processing can outperform manually engineered pipelines across biomedical domains. On BraTS 2020, the BraTS-specific nnU-Net ensemble reported Dice scores of approximately 0.89 for whole tumor, 0.85 for tumor core, and 0.82 for enhancing tumor,⁴⁶ with subsequent transformer-enhanced and hybrid variants on BraTS 2021 approaching or modestly exceeding these baselines (whole-tumor Dice approximately 0.92–0.93 in winning entries).⁴⁷ The dominance of nnU-Net established a performance baseline that catalyzed the rise of transformer-enhanced and hybrid architectures (e.g., Swin-UNETR, SegResNet, MedNeXt) leveraging long-range spatial dependencies and multi-sequence integration to push enhancing-tumor Dice modestly higher (approximately 0.86–0.89 on BraTS 2021) at the cost of substantially greater compute.⁴⁸

However, even high-capacity DL models exhibit sharp performance drops under domain shift, with Dice degradations of approximately 0.04–0.17 reported between in-domain and external test sets, driven by variations in scanner vendors, pulse sequences, reconstruction parameters, and patient populations.⁴⁹ Missing sequences, inconsistent acquisitions, and imbalanced data distributions further degrade robustness. This underscores the need for harmonization strategies, physics-based augmentation, domain adaptation, self-supervised pretraining, and federated learning to reduce cross-site variability and improve out-of-distribution performance. From a clinical-readiness perspective, nnU-Net-based models with rigorous local acceptance testing are closest to deployable status for pre-treatment glioma volumetry, whereas transformer and foundation model variants remain at the late-stage validation tier pending prospective, multi-reader evidence.

4.3. Clinical validation and impact

Fully automated DL pipelines now provide end-to-end assessments of enhancing tumors, non-enhancing tumors, edema, and postoperative cavities on both pre- and post-treatment MRI, generating volumetric

and bidimensional RANO-aligned metrics directly from segmentation outputs.⁵⁰ Built on large expert-annotated multicenter datasets, these systems produce results up to $20\times$ faster than manual contouring, with intraclass correlation coefficients of approximately 0.96, matching experienced radiologists for key subregions. They also improve reproducibility, reduce inter-reader variability, and provide structured quantitative endpoints that are suitable for clinical trials. Emerging evidence suggests that combining radiomics, deep features, and temporal volumetry may further improve monitoring after radiotherapy and immunotherapy, although differentiating pseudoprogression, radiation necrosis, and immune-related responses remains a major challenge.^{24,51}

4.4. Standardization and interoperability

The clinical deployment of segmentation-derived biomarkers requires standardized, interoperable output formats. The Digital Imaging and Communications in Medicine (DICOM) Template Identifier – 1500 structured reporting template provides a machine-readable standard for encoding volumetric measurements, contours, and quantitative maps into picture archiving and communication system (PACS) and radiology workstations.⁵² In parallel, the Brain Imaging Data Structure (BIDS) organizes heterogeneous MRI and derivative segmentation outputs in a reproducible, metadata-rich format to support research scalability.⁵³ Integrating the Healthcare Enterprise (IHE) AI Results (AIR) profile further specifies how AI-generated contours, probability maps, and measurements should be packaged and transmitted within clinical systems, enabling plug-and-play interoperability and reducing vendor-specific fragmentation.⁵² These standards are essential for embedding AI-driven volumetry into routine workflows, thereby enabling quality assurance, audit trails, and large-scale quantitative imaging. Despite these advances, distinguishing true progression from treatment-related changes remains a critical unresolved bottleneck that requires tighter integration of longitudinal AI models, multimodal imaging, and molecularly informed biomarkers.

Recent advances are reshaping each layer of the segmentation-to-decision pipeline. In medical image delineation, foundation models such as the Segment Anything Model (SAM) and its medical adaptations

(MedSAM, SAM-Med3D) now provide promptable, weakly supervised baselines that approach task-specific performance with a fraction of the labeled data,^{54–56} while diffusion-based and self-supervised pretraining improve robustness when annotations are scarce. Radiomic pipelines have matured beyond hand-crafted feature extraction toward stability-aware feature selection, harmonization (ComBat, modality-adaptive schemes), and reproducibility audits compatible with the Image Biomarker Standardization Initiative; explainable, calibrated radiomics pipelines for postoperative glioblastoma surveillance and MGMT methylation prediction illustrate this shift toward clinically interpretable outputs.^{57,58}

Patient-profiling models increasingly fuse imaging with molecular, clinical, and emerging hematologic biomarkers to enable individualized risk stratification and treatment monitoring rather than population-level classification.⁵⁹ In parallel, explainability methods have moved from generic post-hoc saliency toward neuroimaging-specific tools, including counterfactual visualizations and emerging concept-based explanations, with growing attention to faithfulness, stability, and clinician usability.⁶⁰ Together, these advances define a research agenda in which delineation, radiomic feature engineering, patient profiling, and explainability are co-designed and evaluated against prospective, calibration-aware clinical endpoints rather than retrospective accuracy alone.

4.5. Molecular and radiogenomic characterization

Beyond delineating tumor boundaries, AI is increasingly applied to infer the molecular alterations that now define brain tumor diagnosis under the WHO CNS5, an approach termed radiogenomics. ML applied to preoperative MRI could classify glioblastoma molecular characteristics directly from imaging phenotypes.¹⁴ Radiomic and DL models have since been developed to predict key drivers non-invasively, including IDH mutation, 1p/19q codeletion, and tumor grade, frequently within automated pipelines that couple segmentation with downstream classification.^{23,61} Conventional sequences also carry molecular signal; for example, the T2-FLAIR mismatch sign is a specific imaging biomarker of IDH-mutant, 1p/19q-non-codeleted astrocytoma.¹³ Advanced acquisitions, including dynamic contrast-enhanced and arterial spin labeling perfusion, further support IDH and grade estimation.³⁰ More recent research emphasizes MGMT promoter methylation prediction and explainable, modality-adaptive models that improve transparency and clinical trust.⁵⁸ Although reported accuracies are encouraging, external validation across scanners, institutions, and molecular assays remains limited, and radiogenomic predictions are presently best positioned as complements to, rather than replacements for, tissue-based molecular testing.

5. AI for treatment-response assessment and pseudoprogression

5.1. Clinical problem and criteria

Differentiating pseudoprogression from true tumor progression remains one of the most persistent and clinically consequential challenges in post-treatment neuro-oncologic imaging. Immune-related therapies, including checkpoint inhibitors and vaccine-based regimens, can induce transient inflammatory enlargement and new enhancement that radiographically mimic worsening disease, particularly within the first 3–6 months following standard chemoradiation, and within the first 6 months of immunotherapy.⁶²

Even outside immunotherapy, therapy-induced injury, such as radiation necrosis, blood–brain barrier disruption, and delayed treatment response, can coexist with viable tumors, generating highly heterogeneous imaging patterns whose temporal evolution may extend for months. Conventional MRI offers limited discriminatory value in this context; enhancement, mass effect, and FLAIR changes cannot reliably distinguish biologically active tumors from inflammation or necrosis,

and inter-reader variability remains high. As a result, diagnostic uncertainty is frequent, motivating the use of standardized criteria such as RANO and immunotherapy RANO, which emphasize delayed confirmation of progression, integration of clinical stability, and temporal constraints around early post-treatment intervals.^{21,63}

5.2. MRI radiomics and DL

Radiomics applied to standard post-chemoradiation MRI provides a fine-grained quantitative characterization of lesion heterogeneity that surpasses visual assessment. Features derived from T1CE, T2, and ADC quantify intensity distributions, higher-order texture, shape, and microstructural disruption, capturing subtle differences in cellularity, necrosis, and edema that help separate pseudoprogression from true progression.⁵¹

ADC-based texture metrics, such as kurtosis, contrast, and correlation, have been shown to reflect the underlying tumor cellular density versus treatment injury.

Multiparametric radiomic models that integrate clinical factors (MGMT methylation, treatment timelines), molecular biomarkers, and imaging features can outperform imaging-only baselines for early treatment-response prediction, an interval where conventional visual assessment is often unreliable. However, performance variability across folds underscores persistent challenges, including small sample sizes, heterogeneous acquisition protocols, and limited availability of histologically confirmed ground truth, which complicate model generalizability. These limitations affirm the need for large multi-institutional cohorts, harmonization pipelines, and rigorous external validation.⁵¹

5.3. PET-augmented approaches

PET using radiolabelled amino acids has become an important adjunct for disentangling treatment-related changes from active glioma. Tracers such as FET and FDOPA exploit increased amino acid transport in tumor cells and typically show higher tumor-to-background ratios (TBRs) in true progression than in pseudoprogression.⁶⁴ Static metrics such as TBR_mean and TBR_max provide robust quantitative readouts; threshold values of approximately 1.9–2.3 for TBR_max have been associated with high sensitivity and specificity for detecting progression, including within the early post-chemoradiation interval defined by response criteria.

Dynamic protocols further analyze time–activity curves, where earlier time-to-peak uptake and characteristic curve shapes strengthen the discrimination between recurrent tumors and transient therapy effects. Emerging ML models that integrate static and dynamic PET parameters can refine this separation when conventional PET interpretation is equivocal, offering a pathway to a more standardized and reproducible PET-augmented assessment of suspected progression in glioblastoma.⁶⁴ These well-established PET metrics provide structured quantitative inputs that can be readily integrated into multimodal AI classifiers, complementing MRI radiomics when conventional morphologic assessment is equivocal.

5.4. Frontier methods

State-of-the-art multimodal DL architectures increasingly incorporate not only multi-sequence MRI but also radiotherapy dose maps, spatial context, and pre-treatment biology to predict recurrence patterns with greater fidelity. Transformer-based models, capable of capturing long-range spatial dependencies and integrating heterogeneous inputs, achieve robust cross-validated performance and maintain predictive independence from demographic and clinical confounders. These approaches address a key limitation of conventional models by leveraging spatiotemporal information on where recurrence is biologically and dosimetrically likely to emerge. However, significant barriers persist, including limited histopathological confirmation,

Table 2
Core AI applications across segmentation/volumetry and post-treatment evaluation.

AI use case	Typical inputs	Key outputs/metrics	Reference(s)
Benchmark training/evaluation (incl. post-treatment)	Multisequence MRI datasets; growing inclusion of post-treatment anatomy (cavities, treatment effects)	Comparable model evaluation across cohorts; clinically realistic segmentation tasks	41–43
Tumor segmentation (glioma subregions)	T1, T1CE, T2, FLAIR (multimodal MRI)	Compartment masks (enhancing tumor, core, edema/cavity); Dice/HD and related segmentation metrics	39,46
Robustness to domain shift	Cross-site heterogeneity (scanner/vendor/protocol differences; missing sequences)	Domain adaptation/harmonization to preserve performance out-of-distribution	39,65,66
Automated 2D + 3D tumor burden measurement	Pre-/post-operative MRI; longitudinal series	Bidimensional + volumetric burden measures generated from segmentation; faster, more reproducible quantification	50
Conventional-MRI radiomic ML for PsP vs. TP	Post-CRT MRI (T1CE, T2, ADC ± FLAIR/perfusion) + clinical/molecular covariates (e.g., MGMT)	Classifier outputs distinguishing pseudoprogression vs. true progression; cross-validated performance (variable by cohort)	51,67
Adjunct amino-acid PET for PsP vs. TP	Amino-acid PET (e.g., FET/FDOPA) ± MRI	Quantitative PET metrics (e.g., TBR) strengthening recurrence vs. treatment effect separation	24,64
Frontier multimodal DL	Multisequence MRI + treatment metadata (incl. radiotherapy dose); emerging transformer/self-supervised designs	Improved recurrence/pseudoprogression prediction with multimodal integration; external cohort validation as a generalization target	68,69
Explainability for decision support	CNN/DL classifiers on MRI	Saliency/attribution outputs (e.g., Grad-CAM/SHAP-style explanations) to support clinician review	70,71

Abbreviations: 1. 2D: two-dimensional 2. 3D: three-dimensional 3. ADC: apparent diffusion coefficient 4. AI: artificial intelligence 5. CRT: chemoradiotherapy 6. DL: deep learning 7. FDOPA: fluorodopa 8. FET: fluoroethyltyrosine 9. FLAIR: fluid-attenuated inversion recovery 10. Grad-CAM: Gradient-weighted Class Activation Mapping 11. HD: Hausdorff distance 12. MGMT: O⁶-methylguanine-DNA methyltransferase 13. ML: machine learning 14. MRI: magnetic resonance imaging 15. PET: positron emission tomography 16. PsP: pseudoprogression 17. SHAP: SHapley Additive exPlanations 18. T1CE: contrast-enhanced T1-weighted 19. TP: true progression 20. TBR: tumor-to-background ratio 21. CNN: convolutional neural network.

variability in follow-up intervals, and the inherently uncertain ground truth for labeling pseudoprogression, which continue to constrain reliability. Emerging research directions consist of the integration of longitudinal imaging trajectories, uncertainty quantification, model calibration, and self-supervised or weakly supervised learning to reduce dependence on scarce labels. Federated learning and harmonization frameworks are also gaining traction for mitigating institutional variability and expanding access to diverse training data.¹⁹ Major AI use cases for segmentation/volumetry and post-treatment evaluation are summarized in Table 2.

Taken together, the four discriminative modalities differ in clinical utility and limitations and should be viewed as complementary rather than competing. Standard-of-care MRI radiomics is the most widely accessible approach, requires no additional acquisition cost, and achieves reported areas under the receiver operating characteristic curve (AUC) of approximately 0.75–0.90 for separating pseudoprogression from true progression in single- and multi-institutional cohorts,⁷² but its performance is highly sensitive to scanner and protocol variation, segmentation choices, and the small, single-center cohorts that dominate the literature. Dynamic susceptibility contrast and dynamic contrast-enhanced perfusion MRI add hemodynamic information (rCBV, K^{trans}) and improve discrimination when standard sequences are equivocal, but they depend on standardized acquisition and post-processing and remain under-harmonized across centers. Amino-acid PET (FET, FDOPA) provides metabolic specificity and well-validated cut-offs (TBR_{max} approximately 1.9–2.3) for progression, with multicenter evidence supporting reproducibility, yet limited tracer availability and access to PET/MRI restrict deployment.

Longitudinal multimodal AI models that ingest serial MRI, dose maps, and clinical variables offer the most biologically informed predictions and the strongest spatiotemporal performance, but they are still largely experimental, depend on uncertain ground truth, and have not yet undergone prospective multi-reader validation. A practical synthesis is therefore to treat advanced MRI plus amino-acid PET as the current clinically deployable tier and to position longitudinal AI as a late-stage validation tier whose introduction should be accompanied by probability calibration, decision curve evaluation of clinical net benefit, and prospective endpoints. Recent calibrated radiomics pipelines for

postoperative glioblastoma surveillance illustrate this trajectory, reporting an AUC of approximately 0.80 with stability-aware feature selection and explainability outputs designed to support clinical decision support.⁵⁷ These applications are also summarized in Table 2.

6. Clinical deployment: validation, workflow, interoperability, and human oversight

Realizing the clinical value of the segmentation, characterization, and response-assessment methods reviewed above depends on reliable validation, workflow integration, interoperability, and human oversight. Interpretability is integral to this transition, with model outputs increasingly paired with explainable AI (XAI) methods, such as Gradient-weighted Class Activation Mapping (Grad-CAM), SHapley Additive exPlanations (SHAP), and Local Interpretable Model-agnostic Explanations, that expose the features driving a prediction and support clinician review.^{6,71} Clinical success further depends on integration with radiological workflows and standardized data protocols.⁷³

6.1. Interoperability standards

Current research underscores that interoperability is essential for the real-time deployment of AI in neuroimaging, particularly for integrating AI-generated results into clinical workflows. While the literature highlights the substantial potential of AI in brain tumor imaging and diagnosis, it also points to ongoing challenges in data standardization, quality, and integration with existing radiological systems, such as PACS and radiology information system.⁷⁴ Standardized frameworks and protocols, such as DICOM Structured Reporting (with emerging HL7 Fast Healthcare Interoperability Resources support), enable the storage and exchange of AI-derived imaging results within clinical systems and are a prerequisite for the longer-term goal of multimodal radiomic-genomic-clinical data integration.^{52,74}

However, studies have noted that inconsistent DICOM conformance across AI vendors and uneven adoption of broader interoperability standards (including HL7 Fast Healthcare Interoperability Resources and IHE profiles) remain significant barriers to multi-institutional scalability and seamless workflow.⁵² Addressing these interoperability

challenges is critical for ensuring reproducibility, auditability, and broad clinical adoption of AI tools in neuroimaging, ultimately supporting more personalized and efficient patient care.

6.2. PACS and radiology information system integration and structured reporting

AI-driven decision support is increasingly streamlining radiology workflows by automating the extraction of imaging biomarkers and pre-populating structured report templates, allowing radiologists to focus on verification and editing rather than manual entry.⁷⁴ This type of AI-to-report pipeline can shorten reporting time and support more consistent longitudinal follow-up, especially when outputs are carried forward in a standardized format across serial scans. Structured reporting systems that combine quantitative imaging biomarkers with qualitative scoring have been designed for seamless PACS and radiology information system integration and template-based reporting; however, their real-world benefits depend on how well they fit radiologists' reading behavior.⁷⁵

If the interface interrupts the usual "eyes-on-images" pattern (e.g., forcing attention back and forth between form fields and images), it can increase effort and time, potentially offsetting efficiency gains. At the infrastructure level, vendor-neutral and scalable deployments are increasingly supported by modular orchestration approaches that emphasize interoperability and integration across heterogeneous systems, including architectures compatible with cloud-based and container-friendly deployments.⁷⁶ However, sustained adoption remains strongly shaped by sociotechnical factors: radiologists report barriers such as added steps, workflow delays, and unstable tool behavior, while highlighting usability and fit with local routines as key facilitators.⁷⁷

6.3. Human-in-the-loop quality assurance

Human-in-the-loop frameworks are widely recognized as essential in AI-assisted diagnostics to ensure diagnostic safety, maintain clinician oversight, and foster trust in clinical decision making. These frameworks leverage active learning, real-time feedback, and collaborative workflows, allowing clinicians to guide model development, interpret outputs, and intervene when uncertainties or errors arise.⁷⁸ Collaborative reader study designs and human-AI partnership frameworks have demonstrated that combining human expertise with AI can outperform either alone, reducing interobserver variability and amplifying diagnostic accuracy in real-world settings.⁷⁹

Key fail-safe mechanisms, such as confidence scoring, uncertainty visualization, and human review of low-confidence outputs, are highlighted as critical safeguards against erroneous outputs, supporting regulatory compliance and clinical reliability.⁸⁰ Audit trails and local acceptance testing provide traceability for algorithm updates and performance, aligning with regulatory expectations such as the EU MDR and FDA's Good Machine Learning Practice (GMLP).^{11,81} Furthermore, studies show that multi-institutional training improves generalization compared with single-site training, underscoring the importance of multi-center cohorts and domain adaptation in hospital-specific contexts.⁸²

6.4. Implementation checklist

Successful real-time integration of AI in neuroimaging requires a systematic, multi-step implementation strategy that begins with a thorough site readiness assessment, including validation of DICOM conformance, computational resources, network bandwidth, and cybersecurity measures to support high-throughput image processing.^{81,83} Routine performance monitoring, version tracking, and robust incident handling protocols, such as rollback procedures and user feedback loops, are essential for maintaining system reliability and meeting regulatory requirements, especially as clinical environments and imaging protocols evolve.⁸⁴

Research emphasizes that the primary goal of real-time AI decision support is to augment, not replace, human expertise by delivering timely, reproducible, and context-aware insights.⁸⁵ Achieving this vision depends on harmonized standards, rigorous validation, and adaptive quality assurance frameworks that ensure AI tools are trusted partners within the neuroimaging workflow continuum.^{83,85} Additionally, strategies, including continual model calibration, domain adaptation, and local acceptance testing, are critical for maintaining diagnostic accuracy and regulatory compliance as AI systems are deployed across diverse clinical settings.⁸⁶ Key requirements for clinical translation are summarized in Table 3.

7. Regulatory, reporting, and ethics

AI now supports brain tumor detection, grading, segmentation, and treatment monitoring across the applications reviewed above.^{70,71} However, sustainable clinical translation demands rigorous regulatory oversight, transparent reporting, robust data governance, and ethical safeguards to ensure reliability, reproducibility, fairness, and clinician trust.^{3,6} Regulators in the United States, the European Union, and other jurisdictions are emphasizing lifecycle management, post-market monitoring, risk mitigation, and demonstrable clinical effectiveness to an increasing extent, particularly for AI systems that shift from static to adaptive models.

7.1. United States: FDA SaMD, GMLP, and AI/ML regulation

In the United States, neuroimaging AI systems fall under the FDA's Software as a Medical Device (SaMD) framework, which mandates transparency, clinically relevant validation, and adherence to GMLP principles, emphasizing dataset representativeness, traceability, continuous learning oversight, and clear documentation of algorithmic behavior.¹¹ These requirements are crucial in brain tumor imaging because of substantial variations in scanners, acquisition parameters, tumor heterogeneity, and demographic differences, all of which heighten the risk of domain shift and unpredictable degradation in performance.^{49,88}

The FDA's predetermined change control plan is especially relevant for adaptive AI/ML systems, allowing manufacturers to pre-specify the types of permissible model updates, validation protocols, and performance expectations necessary to ensure safety without requiring full re-approval for each modification.¹¹ However, many developers provide insufficient detail on retraining triggers, validation matrices, and update boundaries, raising concerns about algorithmic drift and the disproportionate impact on underrepresented groups.⁸⁷ Compounding this are persistent gaps in public validation datasets and inconsistent external validation practices, rendering it difficult to evaluate real-world generalizability and safety.^{11,88} To foster clinician trust, experts advocate for standardized disclosure formats analogous to pharmaceutical labeling, including intended use statements, performance stratified by demographic and technical subgroups, explanation of methods used, update histories, and known failure modes.^{87,102}

7.2. European Union: MDR and AI Act interplay and high-risk classification

In the EU, AI-enabled neuroimaging systems must comply with both the MDR and the newly enacted EU AI Act, which collectively classify these tools as high-risk, requiring stringent conformity assessments, rigorous clinical evaluation, detailed technical documentation, and comprehensive post-market surveillance.^{10,91} The AI Act works in parallel with the EU MDR and the In Vitro Diagnostic Regulation, imposing standards on data governance, algorithmic transparency, explainability, risk and bias management, and mandatory human oversight.^{10,91}

While the dual-compliance structure aims to enhance algorithmic accountability, critics highlight regulatory redundancies and added certification complexity, which may slow innovation and increase

Table 3
Clinical translation requirements: Interoperability, quality assurance, and regulation/reporting.

Translation domain	Practical requirement	What “good” looks like in practice	Reference(s)
Standards-based interoperability	AI outputs must be ingestible by clinical systems	Structured AI results integrated into reporting/workflow; standards-first approach via IHE profiles and structured outputs	52,74
Structured reporting integration	Reduce manual transcription and enable consistent capture of quantitative results over time	Auto-populated structured report sections with AI-derived quantitative results and traceability	74,75
Deployment architecture	Scale across sites without locking into one vendor	Orchestration and modular deployment patterns supporting multiple tools and updates	12,76
Human factors & adoption	Minimize workflow disruption and cognitive load	UI/fit-to-routine emphasized; barriers captured and mitigated during rollout	12,77,83
Human-in-the-loop quality assurance	Maintain clinician oversight and safe failover	Confidence/uncertainty-aware review, feedback loops, and audit trails for updates	78,80,83
Local acceptance testing & calibration	Verify site performance before and after updates	Local validation/retraining or fine-tuning when appropriate; monitor drift over time	83,86
US regulatory pathway	SaMD expectations for validation and lifecycle control	Clear intended use, representative validation, and updated governance aligned with FDA review practices	11,87,88
EU regulatory pathway	Dual compliance for high-risk systems	MDR clinical evaluation + AI Act requirements for risk management, transparency, and human oversight	10,89–91
Reporting & risk-of-bias appraisal	Reproducible, comparable evidence	TRIPOD-AI/PROBAST-AI direction plus consolidated reporting checklists to reduce avoidable bias	92–94
Privacy-preserving multi-site learning	Enable scale while protecting data	Federated learning with attention to privacy/security risks and governance constraints	95–98
Ethics, fairness, sustainability	Prevent harm and inequity	Subgroup analyses (demographic/scanner/geography), transparency about limits, and attention to broader ethical impacts	99–101

Abbreviations: 1. AI: artificial intelligence 2. EU: European Union 3. FDA: Food and Drug Administration 4. IHE: Integrating the Healthcare Enterprise 5. MDR: Medical Device Regulation 6. PROBAST: Prediction model Risk Of Bias Assessment Tool 7. SaMD: Software as a Medical Device 8. TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis 9. US: United States 10. UI: user interface.

development costs.⁹⁰ Nevertheless, the phased implementation in 2026–2027 is expected to reshape the European AI landscape and strengthen trust and safety across high-stakes clinical domains, such as neuro-oncologic imaging.^{10,89} Developers must also consider compliance with medical software standards (e.g., IEC 62304, ISO 14971, and IEC 62366), cybersecurity requirements, and documentation obligations for adaptive models.⁸⁹

7.3. Reporting, appraisal, and reproducibility standards

Transparent reporting is essential for reproducibility, fair comparison, and ethical clinical adoption. The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) + AI extension provides a harmonized checklist for reporting prediction models integrating regression and ML, emphasizing dataset descriptions, preprocessing, model architectures, calibration, subgroup performance, and explainability methods.⁹³ Complementary tools, such as Prediction model Risk Of Bias Assessment Tool (PROBAST) + AI, enable standardized risk-of-bias appraisal and assessment of applicability across clinical contexts.¹⁰³ Despite these frameworks, adherence is still inconsistent; many studies omit key methodological details, such as cross-site demographics, calibration curves, handling of missing data, retraining strategies, or external validation procedures, limiting reproducibility and meta-analytic synthesis.⁹² Overfitting, selective reporting, and inadequate disclosure of model updates remain persistent concerns, prompting calls for preregistered study protocols, open-source code releases, and the adoption of Model Cards, Data Cards, and algorithmic change logs to standardize transparency practices.¹⁰⁴ Emerging consolidated guidelines for AI in healthcare further strengthen the expectations for rigor, fairness, and reproducibility.^{92,94}

Beyond discrimination metrics such as the AUC, reporting should foreground probability calibration and clinical utility evaluation. Probability calibration, assessed with calibration plots, the Brier score, and integrated calibration index, ensures that predicted probabilities map faithfully to observed event rates, which is essential when AI outputs inform thresholded treatment decisions in glioma surveillance. Decision

curve analysis quantifies the net benefit of a model across plausible risk thresholds compared with default strategies (treat-all, treat-none), making it the most direct way to demonstrate clinically meaningful improvement over standard care.¹⁰⁵ The TRIPOD + AI extension explicitly recommends reporting calibration metrics and decision curve evaluations alongside discrimination. Recent calibrated radiomics studies in postoperative glioblastoma surveillance and MGMT methylation prediction have provided concrete templates for this practice.^{57,58,93,103}

7.4. Data governance, privacy, and security

Neuroimaging AI development relies heavily on large-scale, multi-institutional datasets, raising critical challenges regarding privacy, data rights, security, and equity. Regulatory frameworks, such as the General Data Protection Regulation and the Health Insurance Portability and Accountability Act, tightly restrict data sharing, necessitating compliance-preserving alternatives.⁹⁸ Federated learning enables collaborative training without transferring raw data and has demonstrated performance comparable to that of centrally trained models while preserving confidentiality.^{96,97} However, federated learning introduces technical risks and operational burdens, including the need for consistent cross-site normalization, secure parameter aggregation (e.g., encrypted or multiparty computation), and safeguards against privacy leakage through gradient inversion attacks. Additional concerns include computational overhead, heterogeneous data quality, and the lack of harmonized global governance frameworks capable of addressing the ethical, legal, and societal implications of distributed neuroimaging AI. Interoperability standards, such as DICOM-SEG, DICOM-SR, BIDS, and IHE AIR, are increasingly essential for encoding, auditing, and securely exchanging AI-derived outputs across clinical systems.

7.5. Ethical, fairness, and human-factors considerations

Beyond regulation and reporting, the ethical deployment of neuroimaging AI requires attention to bias, fairness, clinician oversight, and patient autonomy. Model evaluation should include demographic,

scanner-based, and geographic subgroup analyses to expose inequities and mitigate systematic performance gaps. This is key for under-represented populations and low-field MRI systems, which are prevalent in low- and middle-income countries.^{100,106} Ethical governance also includes transparency regarding AI use during clinical care, informed consent for secondary data use, appropriate human-in-the-loop oversight, and a clear delineation of liability between clinicians and vendors.^{85,101} Explainability tools, Grad-CAM and SHAP for example, facilitate interpretability but remain limited; effective clinical deployment warrants user-centered design, training for clinicians, and validation that the explanations faithfully represent model reasoning. Environmental sustainability (the energy demands of AI model training and deployment and the carbon footprint of healthcare data centers) is an emerging ethical consideration.⁹⁹

8. Challenges and future directions

8.1. Cross-cutting challenges

AI and ML have the potential to substantially improve neuroimaging; however, their reliable and reproducible clinical applications are still in their early stages. Although DL and other AI methods offer improved disease prediction and diagnostic accuracy, their application in clinical practice is still constrained by several persistent and pervasive challenges.^{107,108} One of the most significant challenges in neuroimaging AI is the lack of generalizability across diverse datasets, populations, and imaging conditions (Figure 3). Models are often trained on limited, site-specific datasets that vary in scanner type, acquisition protocols, and demographic composition, resulting in poor transferability to new domains.¹⁰⁷

Even subtle differences in image resolution, hardware versions, or preprocessing pipelines can introduce substantial distributional shifts, an issue that is particularly pronounced when combining data from multiple institutions or integrating multimodal neuroimaging sources such as structural MRI and functional MRI.¹⁰⁹ Such variability amplifies

the risk of overfitting, where models achieve high accuracy on local datasets but fail to validate externally.⁴⁹ Pediatric versus adult imaging further complicates this, as developmental differences alter brain morphology and signal characteristics, making domain adaptation essential for accurate analysis. Statistical harmonization methods, transfer learning, and federated learning offer partial solutions by correcting scanner effects or enabling decentralized model training without data exchange; however, the standardization of acquisition protocols and quality assurance remains foundational.^{19,65,110}

AI models in neuroimaging also depend critically on accurate and consistent ground truth labels; however, establishing such labels remains a formidable challenge. In contrast to natural image datasets such as ImageNet, where labeling is straightforward, medical image annotation demands expert radiological interpretation and often suffers from inter- and intra-observer variability.^{5,111} Subtle pathologic findings may be inconsistently identified, and the absence of standardized annotation guidelines further compounds label noise.¹¹² Moreover, defining “ground truth” in neuroscience is complex, as diagnostic labels derived from clinical assessments or follow-ups may not perfectly align with underlying biological or histopathological realities. This disconnect limits the ultimate threshold of AI model performance. The use of multiple expert raters, consensus labeling, and probabilistic modeling of uncertainty can partially mitigate these issues; nonetheless, scalable and high-quality labeling remains labor-intensive and costly.¹¹³ Advanced approaches, such as weakly supervised and semi-supervised learning, which leverage partially labeled data or infer labels from multimodal information, show promise in this regard. Even so, their interpretability and reliability in clinical applications require further validation. Without more robust labeling frameworks, even the most sophisticated algorithms risk learning spurious correlations rather than clinically meaningful patterns.¹¹²

AI systems in neuroimaging often operate as black boxes, generating highly accurate predictions without providing clear explanations for their decisions. This opacity poses significant safety and ethical concerns, particularly in clinical settings, where false negatives or

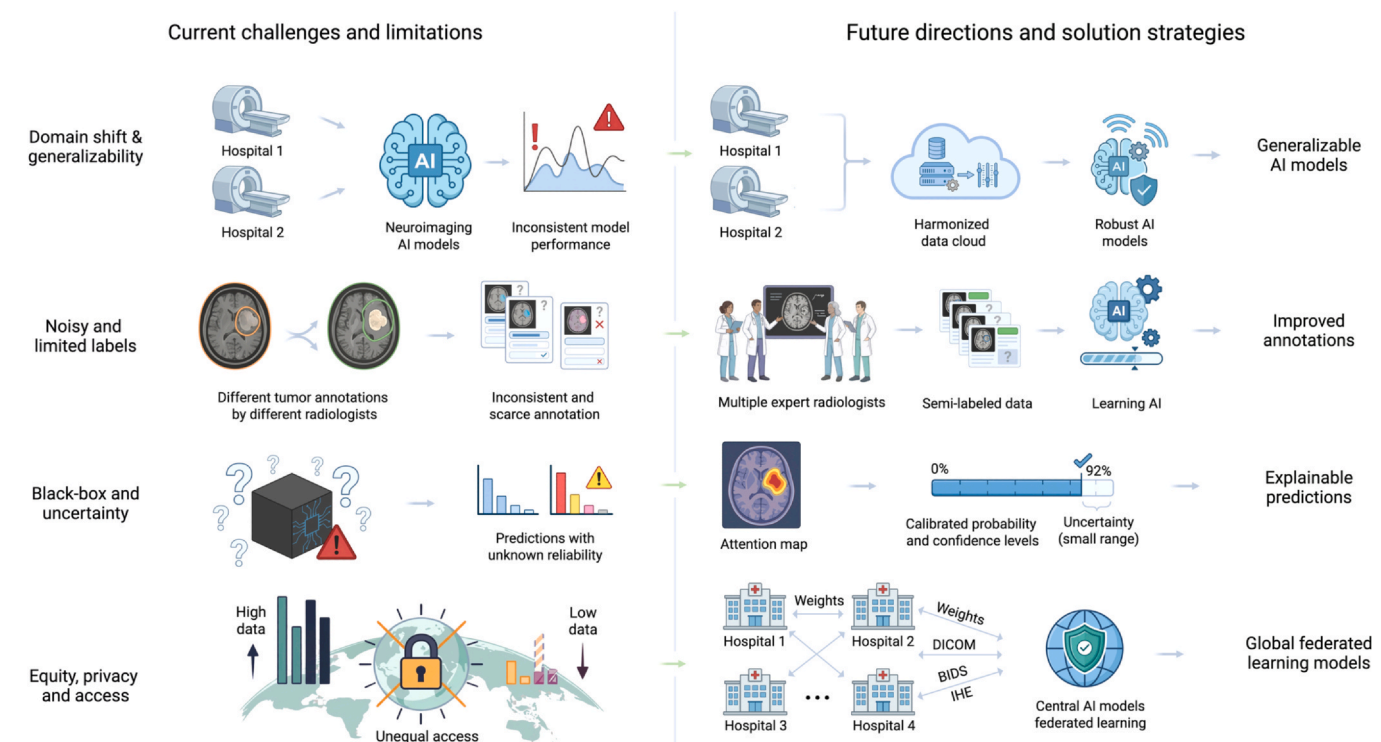


Figure 3. Cross-cutting challenges and future solution pathways for AI in brain tumor neuroimaging. **Abbreviations:** 1. DICOM: Digital Imaging and Communications in Medicine 2. BIDS: Brain Imaging Data Structure 3. IHE: Integrating the Healthcare Enterprise 4. AI: artificial intelligence. This Figure was created in BioRender. <https://BioRender.com/uu9nhvz>.

overconfident misclassifications can have severe consequences.¹¹⁴ Models are rarely calibrated to account for uncertainty or to define explicit thresholds for clinical actions. Calibration is a crucial probabilistic output that must correspond accurately to real-world likelihoods if AI is to inform decisions, such as disease progression prediction or treatment response assessment.¹¹⁵ DL models are especially prone to overconfidence and misestimating uncertainty when confronted with out-of-distribution data. Techniques such as Bayesian DL, ensemble methods, and conformal prediction have been proposed to better quantify uncertainty, whereas XAI frameworks, such as saliency maps, SHAP, and attention mechanisms, aim to elucidate the features driving predictions.¹¹⁶ However, post-hoc interpretability tools, often designed for natural images, may be ill-suited for neuroimaging, producing misleading or unstable explanations. A shift toward ante hoc, intrinsically interpretable models could enhance both reliability and scientific insight, aligning AI outputs with neurobiological understanding.¹¹⁴

Equity and access remain pressing concerns regarding the deployment of AI-based neuroimaging tools. The concentration of high-quality datasets in resource-rich regions, predominantly Western institutions, creates biases that undermine the performance of algorithms when applied to underrepresented populations. This algorithmic bias can exacerbate existing healthcare disparities, as models trained on homogeneous data may misclassify or underperform in diverse ethnic, socioeconomic, and geographic groups.¹¹⁷ Furthermore, the high computational and financial costs of developing and deploying AI systems limit accessibility in low-resource settings, perpetuating the digital divide in clinical care. Identifiable features in MRI and 3D reconstructions can enable patient re-identification, motivating privacy-preserving techniques such as differential privacy that allow aggregated learning while protecting sensitive information.^{95,118} Nevertheless, these privacy-preserving approaches can reduce model accuracy or obscure rare disease signals, necessitating careful calibration of privacy and utility. Ensuring equitable access to AI tools also demands investments in infrastructure, workforce training, and open-access data initiatives, such as the Human Connectome Project and OpenNeuro, which democratize data availability.^{119,120}

A critical appraisal of the current evidence highlights four recurring shortfalls. First, external validation is uneven: most published models are evaluated on a single internal cohort, and performance often declines materially on independent multicenter or international data, with absolute Dice or AUC drops of 0.05–0.20 commonly reported.¹²¹ Second, prospective, workflow-embedded testing remains rare; only a small minority of brain-tumor AI studies report multi-reader multi-case designs or impact on radiologist decisions, time to report, or patient outcomes. Third, probability calibration is seldom assessed: calibration plots, Brier scores, and integrated calibration indices are absent from many AI imaging studies, despite their direct relevance to thresholded clinical decisions in glioma surveillance. Fourth, transparent reporting of cohort flow, handling of missing sequences, retraining triggers, and subgroup performance remains uncommon in published AI imaging studies, leaving substantial residual bias risk.^{92,103} Addressing these gaps is a prerequisite for moving AI tools from late-stage validation to deployment, and should be encoded in study design from the outset rather than added as an afterthought.

8.2. Future directions and research agenda

AI continues to advance neuroimaging, with the prospect of measurable gains in disease diagnosis, treatment monitoring, and personalized therapy. However, to move beyond proof-of-concept models toward clinical integration, future research must emphasize methodological robustness, data diversity, interpretability, and regulatory alignment.⁶⁸ Traditional segmentation algorithms perform well in pre-treatment imaging but often fail to capture the complexity of post-surgical or post-therapeutic changes. For instance, after tumor resection

or radiation therapy, residual cavities, necrotic tissue, and treatment effects create ambiguous imaging signatures that can mislead conventional AI models. Future work must incorporate post-treatment-aware segmentation using datasets such as the BRaTS Challenge 2024 post-treatment cohort, which includes detailed annotations for resection cavities and treatment-related tissue compartments (enhancing tissue, non-enhancing tumor core, and surrounding non-enhancing FLAIR hyperintensity).^{41–43} Advanced architectures, including transformer-based and hybrid CNN–transformer models, should be trained to reliably differentiate post-treatment pseudoprogression from true tumor progression, thereby improving prognostic accuracy and supporting more informed and adaptive treatment decisions for patients with glioblastoma. Integrating clinical metadata, multimodal imaging, and treatment records into AI pipelines can support more biologically informed assessment of tumor status and therapeutic response, moving diagnostic imaging closer to longitudinal patient management.⁶⁹

AI-driven multimodal fusion holds the potential to meaningfully improve the early detection and discrimination of pseudoprogression from true progression in neuro-oncology. DL models capable of jointly processing multimodal datasets can detect subtle cross-modality correlations that elude human interpretations. Recent progress on transformer-based architectures and attention-based multimodal fusion supports this direction, allowing the effective integration of heterogeneous modalities.¹⁸ Future research should explore adaptive fusion frameworks in which multimodal data dynamically inform treatment decisions, for example, by adjusting chemotherapy or radiotherapy based on AI-inferred progression likelihood. Such models must also be clinically feasible, leveraging imaging sequences routinely available in hospital workflows rather than experimental protocols.⁶⁹

To ensure clinical credibility, AI systems should move from retrospective accuracy metrics (e.g., Dice scores) to prospective real-world evidence and multi-reader multi-case trials. These evaluations should measure practical outcomes, such as diagnostic time savings, error reduction, safety impact, and influence on clinical decision-making.¹²² Prospective studies embedded in radiology departments can assess the effects of AI integration on workflow efficiency, inter-observer consistency, and patient outcomes. Moreover, standardized benchmarks comparing algorithm performance with that of human experts are essential to establish clinical validity. The inclusion of post-market surveillance and tracking of AI system behavior across institutions will further strengthen confidence in reliability and reproducibility. Such trials should align with regulatory frameworks such as the FDA's SaMD guideline to support global clinical adoption.^{11,123}

As neuroimaging datasets become larger and more distributed, federated learning will play a pivotal role in enabling collaborative model training without compromising patient privacy. Under this paradigm, models are trained locally at participating hospitals, and only algorithmic updates, not raw data, are shared, thereby maintaining compliance with data protection laws.¹⁹ Future research should incorporate privacy-preserving mechanisms, such as differential privacy, homomorphic encryption, and secure multiparty computation, to safeguard sensitive brain data, especially in rare neurological conditions.¹²⁴ However, ensuring that these models remain generalizable requires continuous uncertainty quantification and drift monitoring during production.⁸⁶

Translating AI models into clinical practice merits strict adherence to established medical data standards and to interoperable workflows. Future AI systems should support standards-first deployment pipelines, transforming research data from BIDS formats into DICOM structured reports for seamless integration into PACS.^{52,53} Conformance with IHE profiles, such as AIR and AI Workflow for Imaging, will facilitate cross-platform compatibility and secure communication between imaging modalities, hospital information systems, and AI applications.⁵² Structured reporting templates embedding AI-generated metrics, segmentation maps, and uncertainty scores directly into radiology reports enhance interpretability and traceability. Furthermore, modular software

architectures and containerized AI deployments ensure scalability and regulatory compliance, enabling efficient updates and version control within clinical environments.¹²⁵

9. Conclusions

AI is positioned to move brain tumor neuroimaging from subjective, reader-dependent interpretation toward more reproducible and quantitative decision support. Across detection, segmentation, grading, and longitudinal monitoring, DL and radiomics can generate fast, standardized measures of tumor burden and improve assessment in complex post-treatment settings, where pseudoprogression, radiation effects, and true recurrence overlap. However, consistent clinical impact still depends on solving domain shift, protocol heterogeneity, missing-sequence resilience, and ground-truth uncertainty while demonstrating benefits through robust external validation and prospective, workflow-embedded studies. To translate this potential into routine practice, the field should prioritize five concrete actions: (i) prospective, multi-reader multi-case validation in real-world neuro-oncology workflows; (ii) routine reporting of probability calibration and decision curve analysis alongside discrimination metrics; (iii) explicit uncertainty quantification and drift monitoring post-deployment; (iv) adherence to AI-specific reporting and risk-of-bias standards (TRIPOD + AI, PROBAST + AI) with standards-first interoperability via DICOM structured reporting, BIDS, and IHE AI Results profiles; and (v) evaluation against clinically meaningful endpoints, including time to report, change in management, and patient-relevant outcomes, rather than retrospective accuracy alone. Aligned with FDA SaMD expectations and the EU MDR and AI Act, these foundations can move AI-enabled neuroimaging from proof-of-concept performance to reliable, scalable tools that support precision neuro-oncology and more confident treatment decisions.

CRedit authorship contribution statement

Mohamed Mustaf Ahmed: Writing – review & editing, Conceptualization. **Don Eliseo Lucero-Prisno III:** Supervision. **Francesco Branda:** Writing – review & editing. **Christian Joseph N. Ong:** Writing – original draft, Methodology. **Shuaibu Saidu Musa:** Validation. **Omar Kasimieh:** Conceptualization. **Olalekan John Okesanya:** Writing – original draft, Methodology. **Majd Oweidat:** Writing – original draft, Methodology. **Zhinya Kawa Othman:** Investigation, Data curation. **Uthman Okikiola Adebayo:** Investigation, Data curation.

Declaration of Competing Interest

All authors of this review declare that they have no conflicts of interest or competing interests.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Paperpal AI (<https://paperpal.com/>) for its “Language Edit” and “Make Academic” features to improve clarity and readability of the manuscript text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

The authors thank the editorial team for constructive feedback on earlier versions of this manuscript.

References

- Chang H, Yan X, Zhong Q, Ning P, Weng M, Liu Y. Global burden of brain and central nervous system cancer among people aged 20–64 years, 1992–2021 and projections to 2050: a population-based study. *BMJ Open*. 2025;15(9):e094462.
- Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol*. 2021;23(8):1231–1251.
- Pan I, Huang RY. Artificial intelligence in neuroimaging of brain tumors: reality or still promise? *Curr Opin Neurol*. 2023;36(6):549–556.
- Pope WB, Brandal G. Conventional and advanced magnetic resonance imaging in patients with high-grade glioma. *Q J Nucl Med Mol Imaging*. 2018;62(3):239–253.
- Visser M, Müller DMJ, van Duijn RJM, et al. Inter-rater agreement in glioma segmentations on longitudinal MRI. *Neuroimage Clin*. 2019;22:101727.
- Chukwujindu E, Faiz H, Ai-Douri S, Faiz K, De Sequeira A. Role of artificial intelligence in brain tumour imaging. *Eur J Radiol*. 2024;176:111509.
- Ranjbarzadeh R, Caputo A, Tirkolaee EB, Jafarzadeh Ghouschi S, Bendechache M. Brain tumor segmentation of MRI images: a comprehensive review on the application of artificial intelligence tools. *Comput Biol Med*. 2023;152:106405.
- Mathivanan SK, Sonaimuthu S, Murugesan S, Rajadurai H, Shivahare BD, Shah MA. Employing deep learning and transfer learning for accurate brain tumor detection. *Sci Rep*. 2024;14(1):7232.
- Zubair Rahman AMJ, Gupta M, Aarathi S, et al. Advanced AI-driven approach for enhanced brain tumor detection from MRI images utilizing EfficientNetB2 with equalization and homomorphic filtering. *BMC Med Inform Decis Mak*. 2024;24(1):113.
- Kotter E, D'Antonoli TA, Cuocolo R, et al. Guiding AI in radiology: ESR's recommendations for effective implementation of the European AI Act. *Insights Imaging*. 2025;16(1):33.
- Zhang K, Khosravi B, Vahdati S, Erickson BJ. FDA review of radiologic AI algorithms: process and challenges. *Radiology*. 2024;310(1):e230242.
- Korfatis P, Kline TL, Meyer HM, et al. Implementing artificial intelligence algorithms in the radiology workflow: challenges and considerations. *Mayo Clin Proc Digit Health*. 2024;3(1):100188.
- Patel SH, Poisson LM, Brat DJ, et al. T2-FLAIR mismatch, an imaging biomarker for IDH and 1p/19q status in lower-grade gliomas: a TCGA/TCIA project. *Clin Cancer Res*. 2017;23(20):6078–6085.
- 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(3):907–918.
- Hirschler L, Sollmann N, Schmitz-Abecassis B, et al. Advanced MR techniques for preoperative glioma characterization: part 1. *J Magn Reson Imaging*. 2023;57(6):1655–1675.
- Hangel G, Schmitz-Abecassis B, Sollmann N, et al. Advanced MR techniques for preoperative glioma characterization: part 2. *J Magn Reson Imaging*. 2023;57(6):1676–1695.
- Albert NL, Weller M, Suchorska B, et al. Response assessment in neuro-oncology working group and european association for neuro-oncology recommendations for the clinical use of PET imaging in gliomas. *Neuro Oncol*. 2016;18(9):1199–1208.
- Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med*. 2022;28(9):1773–1784.
- Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digit Med*. 2020;3(1):119.
- Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology standards for quantitative imaging biomarkers. *Radiology*. 2015;277(3):813–825.
- Wen PY, van den Bent M, Youssef G, et al. RANO 2.0: update to the response assessment in neuro-oncology criteria for high- and low-grade gliomas in adults. *J Clin Oncol*. 2023;41(33):5187–5199.
- Suter Y, Notter M, Meier R, et al. Evaluating automated longitudinal tumor measurements for glioblastoma response assessment. *Front Radiol*. 2023;3:1211859.
- Bai J, Cui B, Li F, Han X, Yang H, Lu J. Multiparametric radiomics signature for predicting molecular genotypes in adult-type diffuse gliomas utilizing 18F-FET PET/MRI. *BMC Med Imaging*. 2025;25(1):187.
- Paprottka KJ, Kleiner S, Preibisch C, et al. Fully automated analysis combining [18F]-FET-PET and multiparametric MRI including DSC perfusion and APTw imaging: a promising tool for objective evaluation of glioma progression. *Eur J Nucl Med Mol Imaging*. 2021;48(13):4445–4455.
- Padhani AR, Liu G, Mu-Koh D, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. *Neoplasia*. 2009;11(2):102–125.
- Romano A, Palizzi S, Romano A, et al. Diffusion weighted imaging in neuro-oncology: diagnosis, post-treatment changes, and advanced sequences—an updated review. *Cancers*. 2023;15(3):618.
- Boxerman JL, Quarles CC, Hu LS, et al. Consensus recommendations for a dynamic susceptibility contrast MRI protocol for use in high-grade gliomas. *Neuro Oncol*. 2020;22(9):1262–1275.
- Shiroishi MS, Castellazzi G, Boxerman JL, et al. Principles of T2*-weighted dynamic susceptibility contrast MRI technique in brain tumor imaging. *J Magn Reson Imaging*. 2015;41(2):296–313.
- Falk Delgado A, De Luca F, van Westen D, Falk Delgado A. Arterial spin labeling MR imaging for differentiation between high- and low-grade glioma—a meta-analysis. *Neuro Oncol*. 2018;20(11):1450–1461.
- Brendle C, Hempel JM, Schittenhelm J, et al. Glioma grading and determination of IDH mutation status and ATRX loss by DCE and ASL perfusion. *Clin Neuroradiol*. 2018;28(3):421–428.
- Horská A, Barker PB. Imaging of brain tumors: MR spectroscopy and metabolic imaging. *Neuroimaging Clin N Am*. 2010;20(3):293–310.

32. Verma G, Mohan S, Nasrallah MP, et al. Non-invasive detection of 2-hydroxyglutarate in IDH-mutated gliomas using two-dimensional localized correlation spectroscopy (2D L-COSY) at 7 Tesla. *J Transl Med*. 2016;14(1):274.
33. Choi C, Ganji SK, DeBerardinis RJ, et al. 2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas. *Nat Med*. 2012;18(4):624–629.
34. Liu C, Li W, Tong KA, Yeom KW, Kuzminski S. Susceptibility-weighted imaging and quantitative susceptibility mapping in the brain: Brain Susceptibility Imaging and Mapping. *J Magn Reson Imaging*. 2015;42(1):23–41.
35. Deistung A, Schweser F, Reichenbach JR. Overview of quantitative susceptibility mapping. *NMR Biomed*. 2017;30(4):e3569.
36. Galldiks N, Langen KJ. Amino Acid PET - an imaging option to identify treatment response, posttherapeutic effects, and tumor recurrence? *Front Neurol*. 2016;7:120.
37. Filippi L, Frantellizzi V, Vincentis GD, Schillaci O, Evangelista L. Clinical applications of TSPO PET for glioma imaging: current evidence and future perspective—a systematic review. *Diagnostics*. 2023;13(10):1813.
38. Muoio B, Albano D, Dondi F, et al. Diagnostic accuracy of PET/CT or PET/MRI using PSMA-targeting radiopharmaceuticals in high-grade gliomas: a systematic review and a bivariate meta-analysis. *Diagnostics*. 2022;12(7):1665.
39. Diana-Albelda C, García-Martín Á, Bescos J. A review on deep learning methods for glioma segmentation, limitations, and future perspectives. *J Imaging*. 2025;11(8):269.
40. Kaufmann TJ, Smits M, Boxerman J, et al. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases. *Neuro Oncol*. 2020;22(6):757–772.
41. Bonato B, Nanni L, Bertoldo A. Advancing precision: a comprehensive review of MRI segmentation datasets from BraTS challenges (2012–2025). *Sensors*. 2025;25(6):1838.
42. Baig S, Vidic I, Mastorakos GM, et al. Segmentation of pre- and posttreatment diffuse glioma tissue subregions including resection cavities. *Neurooncol Adv*. 2024;6(1):vd4e140.
43. de Verdier MC, Saluja R, Gagnon L, et al. The 2024 brain tumor segmentation (BraTS) challenge: Glioma segmentation on post-treatment MRI. *arXiv preprint arXiv:2405.18368*. Published online 2024.
44. Moawad AW, Janas A, Baid U, et al. The Brain Tumor Segmentation (BraTS-METS) Challenge 2023: Brain Metastasis Segmentation on Pre-treatment MRI. *arXiv preprint arXiv:2306.00838*. Published online 2023.
45. Adewole M, Rudie JD, Gbadamosi A, et al. The Brain Tumor Segmentation (BraTS) Challenge 2023: Glioma Segmentation in Sub-Saharan Africa Patient Population (BraTS-Africa). *arXiv preprint arXiv:2305.19369*. Published online 2023.
46. Isensee F, Jaeger PF, Full PM, Vollmuth P, Maier-Hein KH. nnU-Net for Brain Tumor Segmentation. *arXiv preprint arXiv:2011.00848*. Published online 2020.
47. Luu HM, Park SH. Extending nn-UNet for brain tumor segmentation. *arXiv preprint arXiv:2112.04653*. Published online 2021.
48. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth H, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. *arXiv preprint arXiv:2201.01266*. Published online 2022.
49. AlBadawy EA, Saha A, Mazurowski MA. Deep learning for segmentation of brain tumors: impact of cross-institutional training and testing. *Med Phys*. 2018;45(3):1150–1158.
50. Nalepa J, Kotowski K, Machura B, et al. Deep learning automates bidimensional and volumetric tumor burden measurement from MRI in pre- and post-operative glioblastoma patients. *Comput Biol Med*. 2023;154:106603.
51. Patel M, Zhan J, Natarajan K, et al. Machine learning-based radiomic evaluation of treatment response prediction in glioblastoma. *Clin Radiol*. 2021;76(8):628.e17–628.e27.
52. Tejjani AS, Cook TS, Hussain M, Sippel Schmidt T, O'Donnell KP. Integrating and adopting AI in the radiology workflow: a primer for standards and integrating the healthcare enterprise (IHE) profiles. *Radiology*. 2024;311(3):e232653.
53. Gorgolewski KJ, Auer T, Calhoun VD, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data*. 2016;3(1):160044.
54. Kirillov A, Mintun E, Ravi N, et al. Segment anything. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023:4015–4026.
55. Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. *Nat Commun*. 2024;15(1):654.
56. Wang H, Guo S, Ye J, et al. SAM-Med3D: Towards General-purpose Segmentation Models for Volumetric Medical Images. *arXiv preprint arXiv:2310.15161*. Published online 2023.
57. Christodoulou RC, Vamvouras G, Pitsillos R, Solomou EE, Georgiou MF. Explainable radiomics with probability calibration for postoperative glioblastoma surveillance. *Eur J Radiol Artif Intell*. 2026;5:100074.
58. Christodoulou RC, Vamvouras G, Pitsillos R, Solomou EE, Georgiou MF. Explainable, modality-adaptive radiomics for MGMT methylation prediction in high-grade glioma: a decision-curve analysis study. *Front Oncol*. 2026;15:1731258.
59. Christodoulou RC, Papageorgiou PS, Eller D, et al. Multimodal artificial intelligence in glioma management: integrating neuroimaging and hematologic biomarkers for precision oncology. *Front Oncol*. 2026;16:1812518.
60. Abraham LA, Palanisamy G, Veerappu G, Nisha JS. Exploring the potential of explainable AI in brain tumor detection and classification: a systematic review. *Artif Intell Rev*. 2025;9(1):14.
61. Decuyper M, Bonte S, Deblaere K, Van Hoven R. Automated MRI based pipeline for segmentation and prediction of grade, IDH mutation and 1p19q co-deletion in glioma. *Comput Med Imaging Graph*. 2021;88:101831.
62. Brandsma D, Stalpers L, Taal W, Smnina P, van den Bent MJ. Clinical features, mechanisms, and management of pseudoprogression in malignant gliomas. *Lancet Oncol*. 2008;9(5):453–461.
63. Okada H, Weller M, Huang R, et al. Immunotherapy response assessment in neuro-oncology: a report of the RANO working group. *Lancet Oncol*. 2015;16(15):e534–e542.
64. Young JS, Al-Adli N, Scotford K, Cha S, Berger MS. Pseudoprogression versus true progression in glioblastoma: what neurosurgeons need to know. *J Neurosurg*. 2023;139(3):748–759.
65. Seoni S, Shahini A, Meiburger KM, et al. All you need is data preparation: A systematic review of image harmonization techniques in Multi-center/device studies for medical support systems. *Comput Methods Prog Biomed*. 2024;250:108200.
66. Guan H, Liu M. Domain Adaptation for Medical Image Analysis: A Survey. *IEEE Trans Biomed Eng*. 2022;69(3):1173–1185.
67. Sun YZ, Yan LF, Han Y, et al. Differentiation of Pseudoprogression from True Progression in Glioblastoma Patients after Standard Treatment: A Machine Learning Strategy Combined with Radiomics Features from T₁-weighted Contrast-enhanced Imaging. *BMC Med Imaging*. 2021;21(1):17.
68. Zhou J, Park S, Dong S, Tang X, Wei X. Artificial intelligence-driven transformative applications in disease diagnosis technology. *Med Rev*. 2025;5(5):353–377.
69. Gomaa A, Huang Y, Stephan P, et al. A self-supervised multimodal deep learning approach to differentiate post-radiotherapy progression from pseudoprogression in glioblastoma. *Sci Rep*. 2025;15(1):17133.
70. Ariful Islam MD, Mridha MF, Safran M, Alfarhood S, Mohsin Kabir Md. Revolutionizing brain tumor detection using explainable AI in MRI images. *NMR Biomed*. 2025;38(3):e70001.
71. Nazir MI, Akter A, Hussien Wadud MA, Uddin MA. Utilizing customized CNN for brain tumor prediction with explainable AI. *Heliyon*. 2024;10(20):e38997.
72. Akbari H, Rathore S, Bakas S, et al. Histopathology-validated machine learning radiographic biomarker for noninvasive discrimination between true progression and pseudo-progression in glioblastoma. *Cancer*. 2020;126(11):2625–2636.
73. Abdel Razek AAK, Alksas A, Shehata M, et al. Clinical applications of artificial intelligence and radiomics in neuro-oncology imaging. *Insights Imaging*. 2021;12(1):152.
74. Jorg T, Halfmann MC, Stoeck F, et al. A novel reporting workflow for automated integration of artificial intelligence results into structured radiology reports. *Insights Imaging*. 2024;15(1):80.
75. Mañas-García A, González-Valverde I, Camacho-Ramos E, et al. Radiological structured report integrated with quantitative imaging biomarkers and qualitative scoring systems. *J Digit Imaging*. 2022;35(3):396–407.
76. Cohen RY, Sodickson AD. An Orchestration Platform that Puts Radiologists in the Driver's Seat of AI Innovation: a Methodological Approach. *J Digit Imaging*. 2022;36(2):700–714.
77. Wenderott K, Krups J, Luetkens JA, Weigl M. Radiologists' perspectives on the workflow integration of an artificial intelligence-based computer-aided detection system: a qualitative study. *Appl Ergon*. 2024;117:104243.
78. Huang Z, Yang E, Shen J, et al. A pathologist-AI collaboration framework for enhancing diagnostic accuracies and efficiencies. *Nat Biomed Eng*. 2024;9(4):455–470.
79. Calisto FM, Santiago C, Nunes N, Nascimento JC. BreastScreening-AI: Evaluating medical intelligent agents for human-AI interactions. *Artif Intell Med*. 2022;127:102285.
80. Budd S, Robinson EC, Kainz B. A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Med Image Anal*. 2021;71:102062.
81. Brady AP, Allen B, Chong J, et al. Developing, Purchasing, Implementing and Monitoring AI Tools in Radiology: Practical Considerations. A Multi-Society Statement from the ACR, CAR, ESR, RANZCR and RSNA. *Radiol Artif Intell*. 2024;6(1):e230513.
82. Sheller MJ, Edwards B, Reina GA, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep*. 2020;10(1):12598.
83. Leming MJ, Bron EE, Bruffaerts R, et al. Challenges of implementing computer-aided diagnostic models for neuroimages in a clinical setting. *NPJ Digit Med*. 2023;6(1):129.
84. Sahiner B, Chen W, Samala RK, Petrick N. Data drift in medical machine learning: implications and potential remedies. *Br J Radiol*. 2023;96(1150):20220878.
85. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *N Engl J Med*. 2018;378(11):981–983.
86. Feng J, Phillips RV, Malenica I, et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. *NPJ Digit Med*. 2022;5(1):66.
87. Ebrahimian S, Kalra MK, Agarwal S, et al. FDA-regulated AI Algorithms: Trends, Strengths, and Gaps of Validation Studies. *Acad Radiol*. 2022;29(4):559–566.
88. Yearley AG, Goedmakers CMW, Panahi A, et al. FDA-approved machine learning algorithms in neuroradiology: A systematic review of the current evidence for approval. *Artif Intell Med*. 2023;143:102607.
89. Pesapane F, Hauglid MK, Fumagalli M, et al. The translation of in-house imaging AI research into a medical device ensuring ethical and regulatory integrity. *Eur J Radiol*. 2025;182:111852.
90. Gilbert S. The EU passes the AI Act and its implications for digital medicine are unclear. *NPJ Digit Med*. 2024;7(1):135.
91. Kalodanis K, Feretzakis G, Rizomiliotis P, et al. Evaluating the Impact of the EU AI Act on Medical Device Regulation. *Stud Health Technol Inform*. 2025;323:40–44.
92. Klement W, El Emam K. Consolidated Reporting Guidelines for Prognostic and Diagnostic Machine Learning Modeling Studies: Development and Validation. *J Med Internet Res*. 2023;25:e48763.
93. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.

94. Kwak SG, Kim J. Comprehensive reporting guidelines and checklist for studies developing and utilizing artificial intelligence models. *Korean J Anesthesiol*. 2025;78(3):199–214.
95. Li M, Xu P, Hu J, Tang Z, Yang G. From challenges and pitfalls to recommendations and opportunities: Implementing federated learning in healthcare. *Med Image Anal*. 2025;101:103497.
96. Li X, Gu Y, Dvornek N, Staib LH, Ventola P, Duncan JS. Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results. *Med Image Anal*. 2020;65:101765.
97. Stripelis D, Gupta U, Saleem H, et al. A federated learning architecture for secure and private neuroimaging analysis. *Patterns*. 2024;5(8):101031.
98. Mehmood R, Lazar M, Liang X, Corchado JM, See S. Editorial: Protecting privacy in neuroimaging analysis: balancing data sharing and privacy preservation. *Front Neuroinform*. 2025;18:1543121.
99. Ueda D, Walston SL, Fujita S, et al. Climate change and artificial intelligence in healthcare: Review and recommendations towards a sustainable future. *Diagn Interv Imaging*. 2024;105(11):453–459.
100. Herington J, McCradden MD, Creel K, et al. Ethical Considerations for Artificial Intelligence in Medical Imaging: Data Collection, Development, and Evaluation. *J Nucl Med*. 2023;64(12):1848–1854.
101. Nousi SC, Uren V, Jariwala S. Evaluating accountability, transparency, and bias in AI-assisted healthcare decision-making: a qualitative study of healthcare professionals' perspectives in the UK. *BMC Med Ethics*. 2025;26(1):89.
102. Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. *NPJ Digit Med*. 2020;3(1):41.
103. Moons KGM, Damen JAA, Kaul T, et al. PROBAST + AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*. 2025;388:e082505.
104. Mitchell M, Wu S, Zaldivar A, et al. Model cards for model reporting. *Proc Conf Fairness Account Transpar*. 2019:220–229.
105. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26(6):565–574.
106. Geethanath S, Vaughan Jr JT. Accessible magnetic resonance imaging: a review. *J Magn Reson Imaging*. 2019;49(7):e65–e77.
107. Dinsdale NK, Bluemke E, Sundaresan V, Jenkinson M, Smith SM, Namburete AIL. Challenges for machine learning in clinical translation of big data imaging studies. *Neuron*. 2022;110(23):3866–3881.
108. Badrulhisham F, Pogatzki-Zahn E, Segelcke D, Spisak T, Vollert J. Machine learning and artificial intelligence in neuroscience: A primer for researchers. *Brain Behav Immun*. 2024;115:470–479.
109. Bhagwat N, Barry A, Dickie EW, et al. Understanding the impact of preprocessing pipelines on neuroimaging cortical surface analyses. *GigaScience*. 2021;10(1):giaa155.
110. Boyd A, Ye Z, Prabhu SP, et al. Stepwise transfer learning for expert-level pediatric brain tumor MRI segmentation in a limited data scenario. *Radiol Artif Intell*. 2024;6(4):e230254.
111. Bangert P, Moon H, Woo JO, Didari S, Hao H. Active learning performance in labeling radiology images is 90% effective. *Front Radio*. 2021;1:748968.
112. Karimi D, Dou H, Warfield SK, Gholipour A. Deep learning with noisy labels: exploring techniques and remedies in medical image analysis. *Med Image Anal*. 2020;65:101759.
113. Kim D, Chung J, Choi J, et al. Accurate auto-labeling of chest X-ray images based on quantitative similarity to an explainable AI model. *Nat Commun*. 2022;13(1):1867.
114. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206–215.
115. Seoni S, Jahmunah V, Salvi M, Barua PD, Molinari F, Acharya UR. Application of uncertainty quantification to artificial intelligence in healthcare: a review of last decade (2013–2023). *Comput Biol Med*. 2023;165:107441.
116. Guo KH, Chaudhari NN, Jafar T, et al. Anatomic interpretability in neuroimage deep learning: saliency approaches for typical aging and traumatic brain injury. *Neuroinform*. 2024;22(4):591–606.
117. Joseph J. Algorithmic bias in public health AI: a silent threat to equity in low-resource settings. *Front Public Health*. 2025;13:1643180.
118. Schwarz CG, Kremers WK, Therneau TM, et al. Identification of anonymous MRI research participants with face-recognition software. *N Engl J Med*. 2019;381(17):1684–1686.
119. Markiewicz CJ, Gorgolewski KJ, Feingold F, et al. The OpenNeuro resource for sharing of neuroscience data. *eLife*. 2021;10:e71774.
120. Van Essen DC, Ugurbil K, Auerbach E, et al. The human connectome project: a data acquisition perspective. *NeuroImage*. 2012;62(4):2222–2231.
121. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. *Radiol Artif Intell*. 2022;4(3):e210064.
122. Joshi S, Urteaga I, van Amsterdam WAC, et al. AI as an intervention: improving clinical outcomes relies on a causal approach to AI development and validation. *J Am Med Inform Assoc*. 2025;32(3):589–594.
123. Ketola JHJ, Inkinen SI, Mäkelä T, et al. Testing process for artificial intelligence applications in radiology practice. *Phys Med*. 2024;128:104842.
124. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell*. 2020;2(6):305–311.
125. Gupta V, Erdal B, Ramirez C, et al. Current state of community-driven radiological AI deployment in medical imaging. *JMIR AI*. 2024;3:e55833.