Machine Learning for the Prediction of Molecular Markers in Glioma on Magnetic Resonance Imaging: A Systematic Review and Meta-Analysis

Anne Jian, MD [*] [‡]
Kevin Jang, MBBS [*] [§]
Maurizio Manuguerra, PhD [¶]
Sidong Liu, PhD [*] [‖]
John Magnussen, MBBS, PhD, FRANZCR [*] [#]
Antonio Di Ieva, MD, PhD, FRACS [*] [**]

[*]Computational NeuroSurgery (CNS) Lab, Department of Clinical Medicine, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, Australia; [‡]Melbourne Medical School, University of Melbourne, Melbourne, Australia; [§]Discipline of Surgery, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia; [¶]Department of Mathematics and Statistics, Faculty of Science and Engineering, Macquarie University, Sydney, Australia; [‖]Centre for Health Informatics, Macquarie University, Sydney, Australia; [#]Macquarie Medical Imaging, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, Australia; [**]Macquarie Neurosurgery, Macquarie University, Sydney, Australia

Correspondence:
Antonio Di Ieva, MD, PhD, FRACS,
Macquarie Neurosurgery,
Macquarie University,
Suite 201, Level 2, 2 Technology Place,
Sydney, NSW 2109, Australia.
Email: antonio.diieva@mq.edu.au

Received, October 6, 2020.
Accepted, January 24, 2021.

BACKGROUND: Molecular characterization of glioma has implications for prognosis, treatment planning, and prediction of treatment response. Current histopathology is limited by intratumoral heterogeneity and variability in detection methods. Advances in computational techniques have led to interest in mining quantitative imaging features to noninvasively detect genetic mutations.

OBJECTIVE: To evaluate the diagnostic accuracy of machine learning (ML) models in molecular subtyping gliomas on preoperative magnetic resonance imaging (MRI).

METHODS: A systematic search was performed following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines to identify studies up to April 1, 2020. Methodological quality of studies was assessed using the Quality Assessment for Diagnostic Accuracy Studies (QUADAS)-2. Diagnostic performance estimates were obtained using a bivariate model and heterogeneity was explored using metaregression.

RESULTS: Forty-four original articles were included. The pooled sensitivity and specificity for predicting isocitrate dehydrogenase (IDH) mutation in training datasets were 0.88 (95% CI 0.83-0.91) and 0.86 (95% CI 0.79-0.91), respectively, and 0.83 to 0.85 in validation sets. Use of data augmentation and MRI sequence type were weakly associated with heterogeneity. Both O^6-methylguanine-DNA methyltransferase (MGMT) gene promoter methylation and 1p/19q codeletion could be predicted with a pooled sensitivity and specificity between 0.76 and 0.83 in training datasets.

CONCLUSION: ML application to preoperative MRI demonstrated promising results for predicting IDH mutation, MGMT methylation, and 1p/19q codeletion in glioma. Optimized ML models could lead to a noninvasive, objective tool that captures molecular information important for clinical decision making. Future studies should use multicenter data, external validation and investigate clinical feasibility of ML models.

KEY WORDS: Artificial intelligence, Genetic markers, Glioma, Machine learning, MRI, Radiomics

Neurosurgery 0:1–14, 2021 DOI:10.1093/neuros/nyab103 www.neurosurgery-online.com

Gliomas account for 75% of malignant primary brain tumors.[1] In addition to clinical factors and tumor grading, molecular characterization has demonstrated wide-ranging implications for patient prognosis, treatment planning, and prediction of treatment response, leading to its integration in the 2016 World Health Organization (WHO)

ABBREVIATIONS: AI, artificial intelligence; **ATRX,** alpha thalassemia/mental retardation syndrome X-linked; **DKI,** diffusion kurtosis imaging; **DL,** deep learning; **EGFR,** epidermal growth factor receptor; **FLAIR,** fluid-attenuated inversion recovery; **FN,** false negatives; **FP,** false positive; **HGG,** high-grade glioma; **IDH,** isocitrate dehydrogenase; **ML,** machine learning; **PRISMA-DTA,** Preferred Reporting Items for Systematic Reviews and Meta-Analysis of Diagnostic Test Accuracy Studies; **PTEN,** phosphatase and tensin homolog; **QUADAS,** Quality Assessment for Diagnostic Accuracy Studies; **SCC,** Spearman correlation coefficient; **SROC,** summary receiver operating curve; **SVM,** support vector machine; **TERT,** telomerase reverse transcriptase; **TN,** true negative; **TP,** true positive; **VASARI,** Visually Accessible Rembrandt Imaging; **WHO,** World Health Organization

Supplemental digital content is available for this article at www.neurosurgery-online.com.

Classification system.[2] An accurate means of determining hallmark mutations in glioma is thus desirable to aid patients' and surgeons' decision making in the era of personalized medicine. While histopathology is the current diagnostic gold standard, intratumoral spatial heterogeneity and changes in the tumor environment over time cannot be captured by limited samples at a single timepoint.[3] Moreover, different detection methods[4] and the unavailability of sequencing facilities in some centers cause variability and delays in glioma classification.

Magnetic resonance imaging (MRI) contains abundant information reflecting tumor physiology and microenvironment at a voxel (volume pixel) level, which can be quantified using computational tools such as texture analysis, a process known as *radiomics*.[5-7] Importantly, we can harness the powerful ability of machine learning (ML) to recognize patterns among thousands of imaging features to make predictions. ML has led progress in artificial intelligence (AI) by allowing machines to automatically learn how to act in different conditions and improve on their own performance without being explicitly programed.[8] Deep learning (DL) models, a subset of ML, further mimic the human visual cortex neural networks to learn abstract representation of data. Such models could augment decision making in neurosurgery, with applications in diagnosis, tumor grading, and prediction of surgical outcomes.[8-10] Their use in *radiogenomics*,[11] the process of detecting genetic alterations corresponding with imaging features, is of particular interest here.

A previous systematic review[12] summarized studies exploring radiogenomic markers; however, the use of ML models was not examined. Another review assessed the accuracy of ML in predicting isocitrate dehydrogenase (IDH) mutations,[13] but additional molecular markers have diagnostic and prognostic significance, including 1p/19q codeletion, O^6-methylguanine-DNA methyltransferase (MGMT) gene promoter methylation, p53 mutation, epidermal growth factor receptor (EGFR) amplification, phosphatase and tensin homolog (PTEN) loss, alpha thalassemia/mental retardation syndrome X-linked (ATRX), and telomerase reverse transcriptase (TERT) mutations. We sought to systematically review the literature and, firstly, evaluate the diagnostic accuracy of ML models to predict these molecular markers in glioma on MRI and, secondly, determine methodological and clinical factors that affect their performance.

METHODS

This study was conducted in concordance with our protocol (available upon request) and the Preferred Reporting Items for Systematic Reviews and Meta-Analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines.[14]

Eligibility Criteria

Studies were included if the following criteria were met: (1) aimed to predict at least 1 of 8 aforementioned molecular markers using any MRI sequence; (2) utilized an ML algorithm for classification; (3) included patients with histopathologically confirmed glioma of any WHO grade; (4) contained sufficient information to reconstruct 2 × 2 tables (we contacted authors if this was insufficient); and (5) English language. Studies were excluded if they were (1) commentaries, editorials, letters, review articles, case reports, conference abstracts, or (2) animal studies.

Study Selection

We performed a systematic search of Medline Ovid, EMBASE, Scopus, and Web of Science to identify studies published up to April 1, 2020 using key terms for ML, brain tumors, genetic mutations, and MRI (see **Methods**, **Supplemental Digital Content 1** for search strategy). The titles and abstracts of deduplicated articles were independently screened (A.J. and K.J.) and disagreements were resolved by discussion. Full text review identified studies satisfying the eligibility criteria. A secondary search involved handsearching the reference lists of included studies.

Data Extraction

We collected data about the study design, patient characteristics, MRI sequence, segmentation, feature selection, ML classifier, histopathology, and validation using piloted form. Sensitivity, specificity, true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) were extracted from training and validation datasets for each molecular marker. We defined performance from cross-validation as a training dataset since it involved data "seen" by the machine, while "unseen" data from a held-out test set or external cohort were treated as validation. If multiple models were compared within a study, the top-performing model was selected.

Risk of Bias Assessment

Risk of Bias (RoB) was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool[15] and tailored through adding 2 questions: whether the study avoided a severe imbalance of genotype classes, and reduced variability in segmentation or studied feature robustness, an item adopted from the radiomics quality score (RQS).[16] Since a data-driven approach is generally used in ML to derive an optimal cutoff, a study was deemed unclear with regard to prespecified threshold, unless its reliability was improved through other strategies.[17] Histopathology is interpreted before ML results, thus all studies answered "yes" to the question pertaining to review bias. A conservative approach was used for overall judgement—high RoB if there was a "no" to at least one question, and uncertain if any item was "unclear." Authors were contacted if RoB was unclear, with 18 of 42 providing further information.

Applicability was evaluated considering the patient population, index test, reference standard, and RQS items, such as the risk of overfitting, an issue of ML model performance degradation when applied to a cohort different to its training samples.

Meta-Analysis

We constructed a bivariate random effects model to obtain pooled sensitivity, specificity, and 95% CI where there were ≥5 studies, and derived a summary receiver operating curve (SROC) using the hierarchical model. To assess heterogeneity, we examined (1) overlap in 95% CIs on forest plots; (2) deviation of 95% prediction region from confidence region on SROC curve; (3)

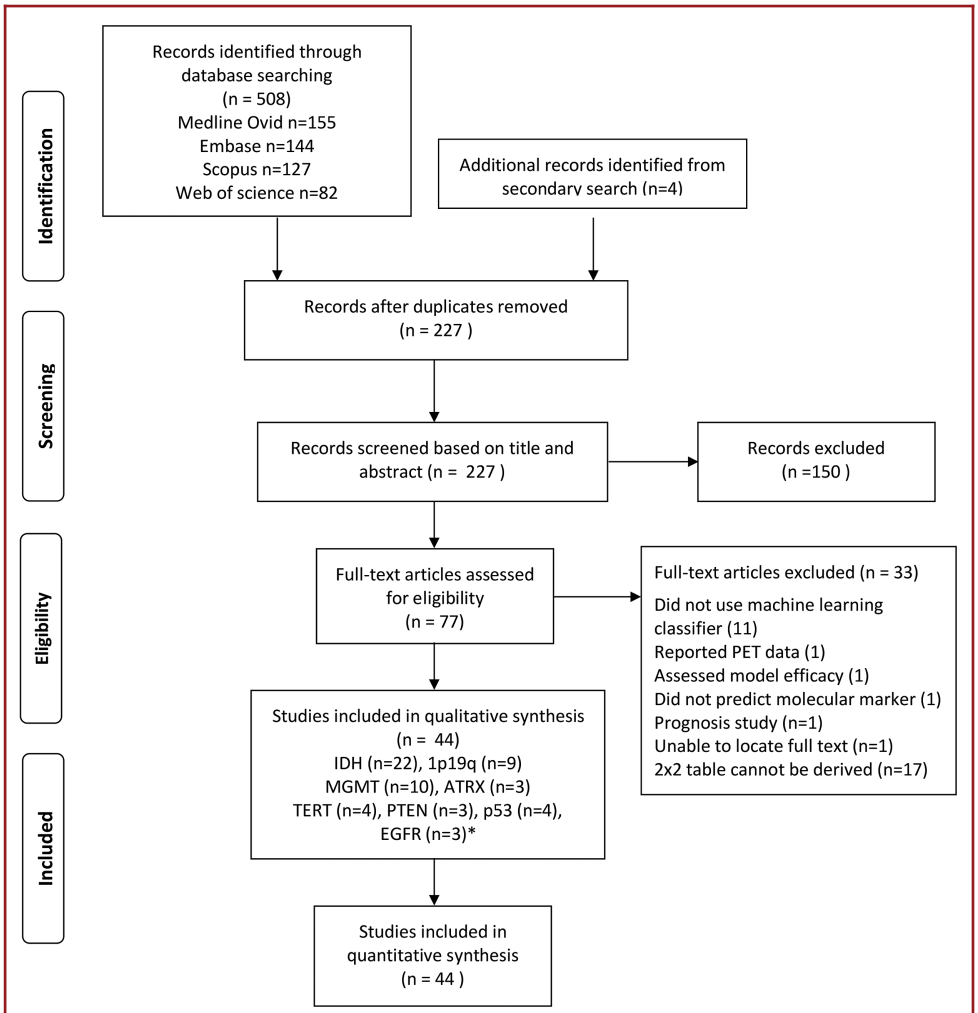

FIGURE 1. *Flow diagram of the study selection process. * Studies investigating more than one genetic subtype were counted separately for each gene.*

between-study variance τ^2 in logit sensitivity and false positive rate (FPR = 1 − specificity).[18] In addition, a Spearman correlation coefficient (SCC) between sensitivity and FPR of >0.6 indicates threshold effect,[19] as well as a V-shaped forest plot ordered by sensitivity.

We investigated heterogeneity by subgroup analysis and metaregression using prespecified covariates where there were at least 10 studies[20]: study setting (single/multicenter), glioma grade, conventional (T1-weighted pre- and post-gadolinium contrast, T2-weighted, and fluid-attenuated inversion recovery [FLAIR]) or advanced MRI (eg, diffusion weighted imaging, MR spectroscopy), clinical information, ML algorithm, and data augmentation (ie, any technique used to increase the amount of input data for the training of the model aimed to reduce overfitting). Publication bias was assessed by Deeks funnel plot asymmetry test.[21] Statistical analyses were conducted using "mada" (v0.5.10) and "metafor" (v2.4-0) packages in R (R Stats v4.0.0).

RESULTS

We identified 512 articles from primary and secondary search. After preliminary screening and full text review excluding studies for reasons outlined (Figure 1), 44 studies met all eligibility criteria.[22-65]

Study Characteristics

Key study characteristics are summarized in Table 1. The majority employed a retrospective design except 2 prospective studies.[24,55] Sample size ranged from 13 to 463 patients. Twenty studies drew patients from multiple institutions, commonly using the Cancer Genome Atlas Lower Grade Glioma and Glioblastoma databases.

TABLE 1. Characteristics of Included Studies

Study	Country	Total no. pts	Genes	% mutation	Glioma grade	MRI sequences	Clinical information	Machine learning classifier(s)	Reference standard	Externally validated
Akbari et al, 2018[22]	USA	129	EGFR	29.5	IV	T1, T1CE, T2, FLAIR, DTI, DSC	No	Linear SVM	Next-generation sequencing	No
Akkus et al, 2017[23]	USA	159	1p/19q	35.8	II, III	T1CE, T2	No	CNN, SVM	FISH	No
Alis et al, 2020[34]	Turkey	142	IDH	33.8	II, III	T1CE, FLAIR, DWI	No	Random forest	NR	No
Bisdas et al, 2018[55]	UK	37	IDH	70.3	II, III	FLAIR, DKI	No	RBF-SVM	IHC, Sanger sequencing	No
Chen et al, 2018[57]	China	47	IDH, MGMT	27.7 (IDH), 55.3 (MGMT)[a]	III, IV	DTI, RS-fMRI	Yes	Dictionary learning, random forest	NR	No
Choi et al, 2019[56]	Korea	463	IDH	27.0	II, III, IV	T1, T1CE, T2, FLAIR, DSC	No	RNN	IHC, Sanger sequencing	No
De Looze et al, 2018[61]	Ireland	381	IDH	19.9	II, III, IV	T1, T1CE, T2, FLAIR, DWI	No	Random forest	IHC, pyrosequencing	No
Eichinger et al, 2017[58]	Germany	79	IDH	75.9	II, III	DTI	No	CNN	IHC ± sequencing	No
Fellah et al, 2013[62]	France	50	1p/19q	38.0	II, III	DWI, MRS, DSC	No	Random forest	FISH	No
Fukuma et al, 2019[59]	Japan	164	IDH, TERT	65.9 (IDH), 50 (TERT)	II, III	T1, T1CE, T2, FLAIR	Yes	SVM	Sanger/ pyrosequencing	No
Ge et al, 2020[63]	Sweden	167	IDH	32.9	II, III, IV	T1, T1CE, T2, FLAIR	No	CNN	Genome sequencing	No
Hajianfar, 2019[60]	Iran	82	MGMT	54.9	IV	T1CE, FLAIR	No	Random forest, SVM, k-NN, naïve-Bayes, decision tree, AdaBoost, logistic regression, stochastic gradient descent	DNA methylation probes, BeadChip	No
Han et al, 2018[64]	China	277	1p/19q	39.4	II, III	T2	Yes	Random forest	FISH	Yes
Haubold et al, 2019[65]	Germany	42	IDH, 1p/19q, MGMT, ATRX	53.3 (IDH), 16.7 (1p/19q), 20 (MGMT), 20 (ATRX)	I-IV	T1, T1CE, T2, FLAIR, DWI, SWI, MR fingerprinting FET-PET	No	Linear SVM, random forest	NR	No
Hu et al, 2017[24]	USA	13	EGFR, PTEN, p53	43.8 (EGFR), 75 (PTEN), 35.4 (p53)	IV	T1, T2, DSC, DTI	No	Decision tree	Array comparative genomic hybridization, exome sequencing	No
Jiang et al, 2019[25]	China	122	MGMT	72.1	II, III	T1CE, T2	Yes	SVM, random forest, AdaBoost	Pyrosequencing, methylation BeadChip	Yes
Jiang et al, 2020[26]	China	116	TERT	48.3	II, III	T1CE, T2	No	SVM, random forest, AdaBoost	Genome sequencing	Yes

TABLE 1. Continued

Study	Country	Total no. pts	Genes	% mutation	Glioma grade	MRI sequences	Clinical information	Machine learning classifier(s)	Reference standard	Externally validated
Kanas et al, 2017[27]	Greece	86	MGMT	50.0	IV	T1, T1CE, T2, FLAIR	No	Random forest, k-NN, naïve-Bayes, decision tree	DNA methylation probes, BeadChip	No
Kickingereder et al, 2016[28]	Germany	152	MGMT, EGFR, PTEN	45(MGMT), 51(EGFR), 87 (PTEN)[a]	IV	T1, T1CE, T2, FLAIR, DWI, DSC, SWI	No	Stochastic gradient boosting, random forest, penalized logistic regression	Methylation BeadChip, pyrosequencing	No
Kocak et al, 2020[29]	Turkey	107	1p/19q	67.3	II, III	T1CE, T2	No	Adaptive boosting, k-NN, naïve-Bayes, neural network, random forest, stochastic gradient descent, SVM	TCGA	No
Korfiatis et al, 2016[30]	USA	155	MGMT	42.6	IV	T1CE, T2	No	RBF-SVM, random forest	NR	No
Korfiatis et al, 2017[31]	USA	155	MGMT	42.6	IV	T2	No	ResNet	NR	No
Li Y et al, 2018[33]	China	186	ATRX	34.9	II, III	T2	No	SVM	IHC, whole exome sequencing	Yes
Li Y et al, 2018[35]	China	272	p53	44.5	II, III	T2	No	SVM	IHC	No
Li Y et al, 2019[36]	China	109	PTEN	29.4	IV	T1CE, T2	No	SVM	Whole exome sequencing, Sanger sequencing	Yes
Li Z et al, 2017[37]	China	151	IDH	74.2	II	T1CE, FLAIR	No	SVM	Genome sequencing, Sanger sequencing	Yes
Li Z-C et al, 2018[45]	China	225	IDH	8.9	IV	T1, T1CE, T2, FLAIR	Yes	Random forest	Genome sequencing, pyrosequencing	Yes
Li Z-C et al, 2018[38]	China	193	MGMT	44.6	IV	T1, T1CE, T2, FLAIR	Yes	Random forest	Pyrosequencing	Yes
Liang et al, 2018[32]	China	167	IDH	31.7	II, III, IV	T1, T1CE, T2, FLAIR	No	CNN	Genome sequencing	No
Liu et al, 2012[39]	China	31	p53	51.6	I-IV	T1 & T2 FLAIR	No	RBF-SVM	IHC	No
Lo et al, 2020[40]	Taiwan	39	IDH	17.9	IV	T1CE	No	Logistic regression, k-NN, SVM	Genome sequencing	No
Lu et al, 2018[41]	Taiwan	284	IDH, 1p/19q	44.4 (IDH), 35.6(1p19q)	II, III, IV	T1CE, T2, FLAIR, DWI	Yes	SVM, ensemble classifiers	NR	Yes

TABLE 1. Continued

Study	Country	Total no. pts	Genes	% mutation	Glioma grade	MRI sequences	Clinical information	Machine learning classifier(s)	Reference standard	Externally validated
Ozturk-Isik et al, 2019[42]	Turkey	112	IDH, TERT	65.2 (IDH), 44.2 (TERT)	II, III, IV	T1, T1CE, T2, DSC, MRS	No	SVM, decision trees, logistic regression, k-NN	Minisequencing, Sanger sequencing	No
Ren et al, 2019[43]	China	57	IDH, ATRX	63.2 (IDH), 52.8 (ATRX)	II	T1CE, FLAIR, DWI, 3D-ASL	Yes	SVM	IHC	No
Shofty et al, 2018[44]	Israel	47	1p/19q	55.3	II	T1CE, T2, FLAIR, DWI	No	SVM, k-NN, ensemble classifiers	Microsatellite analysis	No
Tan et al, 2019[46]	China	105	IDH	48.6	II, III, IV	T1CE, FLAIR, DWI	Yes	Linear SVM, logistic regression	Sanger sequencing	No
van der Voort et al, 2019[47]	Netherlands	413	1p/19q	44.8	II, III, IV	T1CE, T2	Yes	SVM	FISH, multiplex ligation probe assay, next-generation sequencing	Yes
Wu et al, 2018[48]	China	105	IDH	68.6	NR	T1CE, T2	No	Dictionary learning	NR	No
Xi et al, 2018[49]	China	118	MGMT	47.5	IV	T1, T1CE, T2W	No	SVM	Pyrosequencing	No
Yamashita et al, 2019[50]	Japan	112	TERT	61.6	IV	T1, T1CE, DWI	Yes	SVM	PCR	No
Yogananda et al, 2020[51]	USA	214	IDH	43.9	II, III, IV	T1CE, T2, FLAIR, DWI	No	CNN	Sanger sequencing, exome sequencing	No
Yu et al, 2017[52]	China	140	IDH	72.1	II	FLAIR	No	SVM, AdaBoost	Sanger sequencing	No
Zhang et al, 2018[53]	China	103	IDH, p53	68.0 (IDH), 50.2 (p53)	II, III	T1, T1CE, T2, FLAIR	No	SVM	Genome sequencing	No
Zhou et al, 2017[54]	China	84	IDH, 1p/19q	75 (IDH), 25.4 (1p19q)	II, III	T1, T1CE, T2, FLAIR	Yes	Random forest	Genome sequencing	No

[a] Dataset includes 1 or more patients with genotype not classified.

ASL, arterial spin labeling; ATRX, alpha thalassemia/mental retardation syndrome X-linked; CNN, convolutional neural network; DKI, diffusion kurtosis imaging; DSC, dynamic susceptibility contrast; DTI, diffusion tensor imaging; DWI, diffusion-weighted imaging; EGFR, epidermal growth factor receptor; FISH, fluorescence in situ hybridization; FLAIR, fluid-attenuated inversion recovery; IDH, isocitrate dehydrogenase; IHC, immunohistochemistry; k-NN, k-nearest neighbor; MGMT, O^6-methylguanine-DNA methyltransferase; NR, not reported; PCR, polymerase chain reaction; PTEN, phosphatase and tensin homolog; RBF, radial basis function kernel; RNN, residual neural network; SVM, support vector machine; SWI, susceptibility weighted imaging; TERT, telomerase reverse transcriptase; T1, T1-weighted; T1CE, T1 contrast-enhanced; T2, T2-weighted.

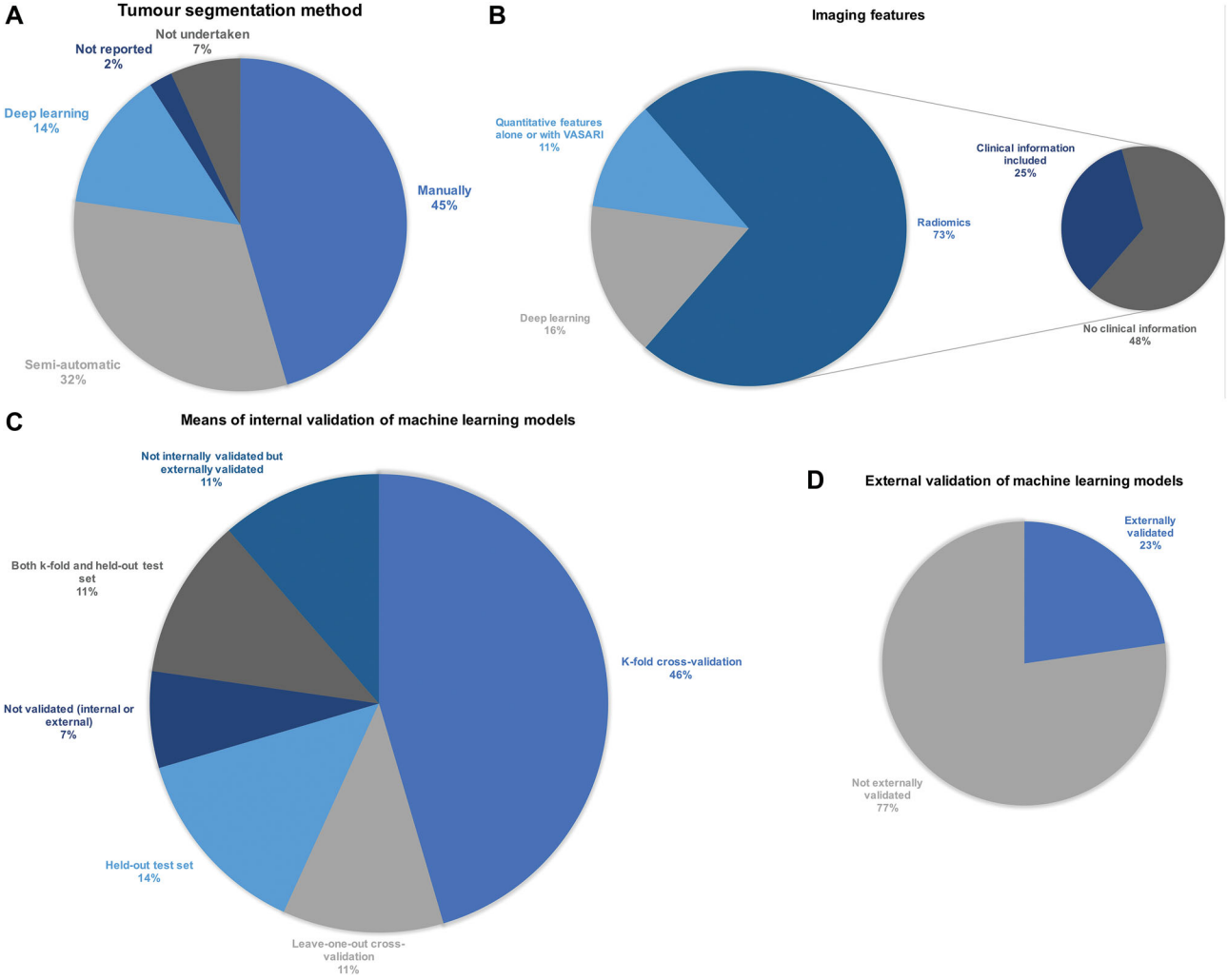

FIGURE 2. *Summary of tumor segmentation methods* **A**, *types of imaging features* **B**, *means of internal validation* **C**, *and external validation* **D** *used by studies (n = 44) investigating machine learning models for predicting genetic subtypes of glioma. VASARI, Visually Accessible Rembrandt Imaging.*

In total, 16 of 44 studies used advanced MR sequences. Image segmentation was undertaken manually or semiautomatically, while 6 studies adopted DL (Figure 2A).[37,38,45,48,52,56] In total, 32 studies extracted radiomics features such as tumor shape, intensity and texture, 7 used DL, and 5 used quantitative parameters alone, such as MR spectroscopy metabolite concentration, or combined with Visually Accessible Rembrandt Imaging (VASARI) features (Figure 2B). Several ML classifiers were investigated, the most common being random forest and support vector machines (SVM). Means of internal validation are summarized in Figure 2C, while only 10 studies externally validating their models (Figure 2D). Imbalanced genotype classes were addressed in several studies by selecting a balanced number of image slices,[23] setting class weights,[63] or using oversampling techniques to increase the minority class.[28,42,45,61] Data augmentation techniques used commonly included flipping, shifting, rotation,[32,51,59] and synthetic generation of images during model training.[63]

Risk of Bias

Quality assessment results are summarized in Figure 3. In the patient selection domain, most studies did not report exclusion criteria and/or sequence of patient enrolment, while 2 studies had inappropriate patient exclusions.[46,50] Two studies scored high risk in index test due to concerns with selection of regions of interest,[40,42] and it was unclear in many instances if investigators were blinded to the genotype. Reference standard was overall low risk. Finally, 14 of the 44 studies had a high RoB with respect to patient flow as not all patients were genotyped or underwent the

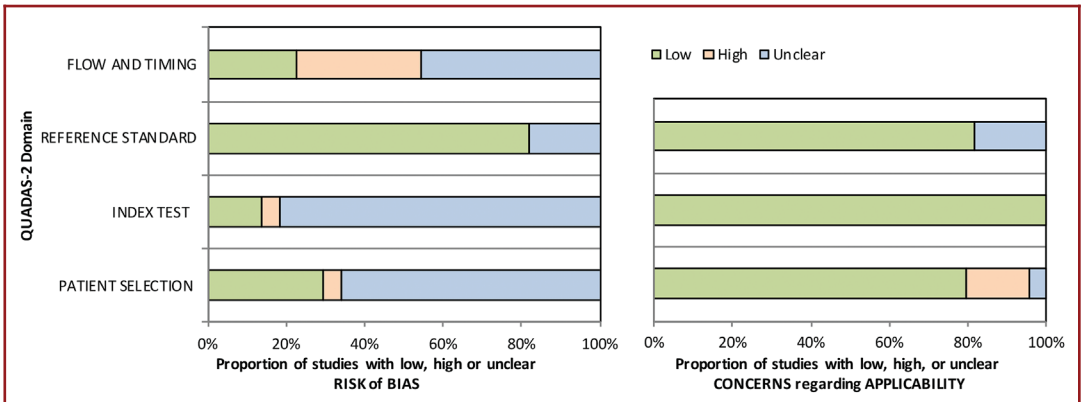

FIGURE 3. *Assessment of risk of bias and applicability based on QUADAS-2.*

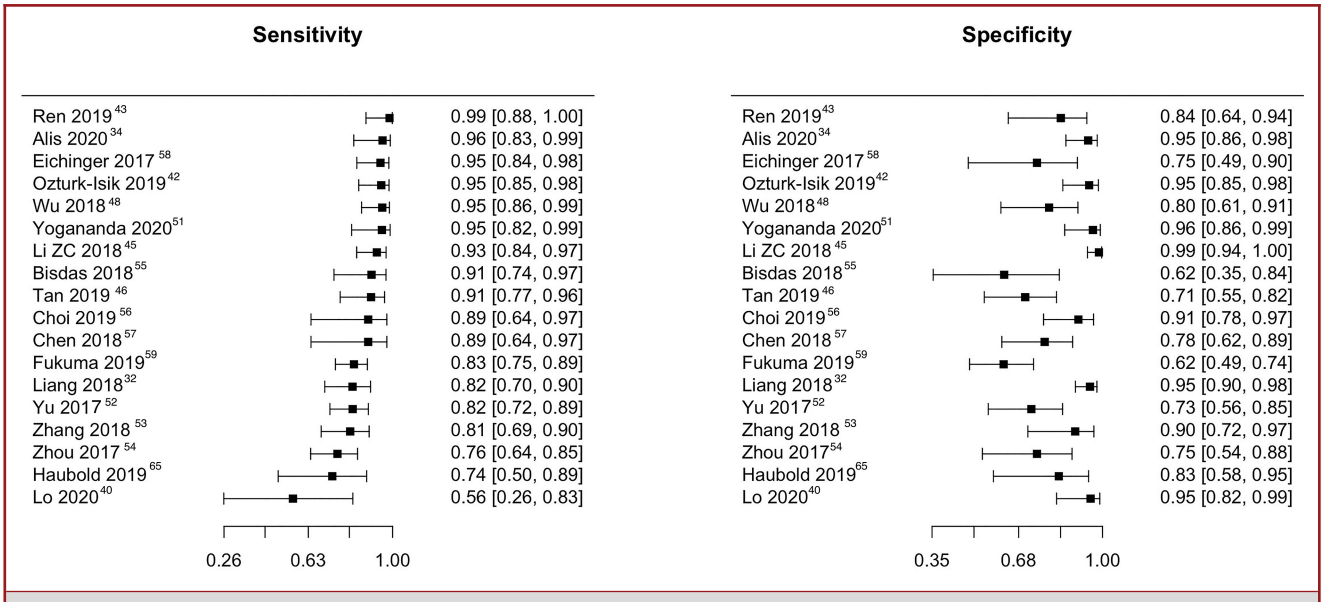

FIGURE 4. *Coupled forest plots of included studies using machine learning to predict isocitrate dehydrogenase (IDH) status in training dataset. Numbers are pooled estimates with 95% CI in brackets and indicated by horizontal lines.*

same histopathology detection. The unreported interval between MRI and surgery/biopsy left most with an unclear RoB.

There were no concerns with applicability except 7 studies that selected patients based on information unavailable preoperatively,[37,43,44,46,50,52,62] 3 of which included only grade II glioma,[37,43,52] thus less applicable to the heterogeneous lower grade (II and III) glioma.

IDH Mutation

The training diagnostic performance of IDH prediction is displayed in Figure 4, which revealed no threshold effect between sensitivity and specificity, confirmed by SCC of 0.03 (95%CI–0.45 to 0.49). Substantial between-study heterogeneity was indicated by poorly overlapping CIs and the deviation of prediction from confidence region (Figure 5A). This was more prominent in specificity, confirmed by the between-study variance, τ^2 in logit FPR of 0.76 and logit sensitivity 0.27. The pooled sensitivity and specificity were 0.88 (95% CI 0.83-0.91) and 0.86 (95% CI 0.79-0.91), respectively, with an area under curve (AUC) of 0.92.

In validation datasets, the pooled sensitivity, specificity, and AUC were 0.85 (95% CI 0.77-0.90), 0.83 (95% CI 0.72-0.90) and 0.90. A threshold effect was suggested by the studies lying close to the SROC curve (Figure 5B) and a SCC of 0.74 (95%CI 0.28-0.92). Heterogeneity was likewise greater in specificity (τ^2 in logit FPR 0.69, sensitivity 0.27). Examination of outliers revealed imbalanced genotype classes.[34,40,52] For validation performance, the study with the lowest sensitivity contained 33% of IDH

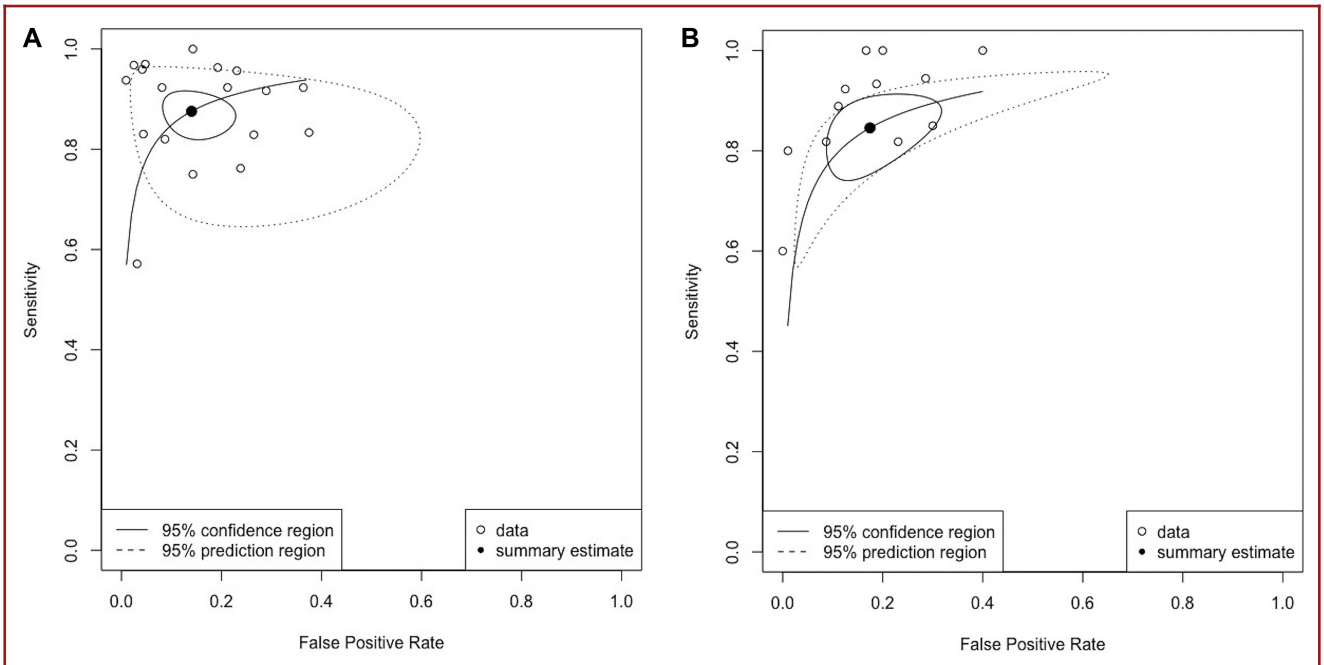

FIGURE 5. *Summary Receiver Operating Characteristic curve of the diagnostic performance of machine learning models for predicting isocitrate dehydrogenase (IDH) status. Each study is represented by an open circle with the shaded circle representing the pooled sensitivity 0.876 (95% CI 0.832-0.909) and specificity 0.860 (95% CI 0.792-0.908) in training dataset* **A** *and 0.846 (95% CI 0.765-0.902) and 0.826 (95% CI 0.715-0.899), respectively, in validation dataset* **B**. *In total, 95% confidence ellipse is indicated in solid line and prediction ellipse in dashed line.*

mutants.[34] Other studies with lower proportion of mutants addressed imbalanced learning.[45,61,63]

1p/19q Codeletion

In total, 6 of the 9 studies investigating 1p/19q codeletion reported training results, yielding a pooled sensitivity and specificity of 0.83 (95% CI 0.72-0.90) and 0.76 (95% CI 0.71-0.81), respectively, and AUC of 0.83 (see **Figure**, **Supplemental Digital Contents 2** and **3**). Validation performance across 5 studies gave a sensitivity of 0.70 (95%CI 0.45-0.86), specificity 0.72 (95% CI 0.63-0.80), and AUC 0.75. In both instances, heterogeneity was observed primarily in sensitivity.

MGMT Methylation and Other Markers

In total, 10 studies evaluated ML prediction of MGMT methylation, demonstrating a pooled sensitivity of 0.81 (95% CI 0.72-0.87), specificity 0.80 (95% CI 0.73-0.86), and AUC 0.87 in training dataset (see **Figure**, **Supplemental Digital Contents 4** and **5**). There was poor overlap in 95% CIs across studies. A negative correlation of sensitivity with FPR was detected, as with 1p/19q codeletion training dataset, suggesting possibly different implicit thresholds across studies that contributed to the variability.[66] Only 3 studies reported validation performance, with sensitivity and specificity ranging 0.70-0.88. Ranges for other molecular markers were reported in Table 2. Of these, the reported sensitivity and specificity for ATRX mutation were $\geq$0.75, followed by TERT promoter with sensitivity of 0.75-0.86. The prediction performance of EGFR, p53 mutation, and PTEN loss were largely variable (sensitivity and specificity 0.55-1).

Metaregression and Subgroup Analysis

For IDH prediction in the training dataset, data augmentation was weakly associated with heterogeneity in specificity, and MRI modality with sensitivity ($P = .05$). Studies using augmentation achieved a higher specificity (0.95, 95%CI 0.81-0.99) than those without (0.82, 95%CI 0.75-0.87), while those investigating advanced MRI (0.92, 95%CI 0.86-0.95) showed higher sensitivity compared to conventional sequences (0.85, 95%CI 0.81-0.88) (Table 3). For the validation set, high grade glioma (HGG) contributed to heterogeneity in specificity ($P = .002$), reducing τ^2 in logit FPR from 0.69 to 0.12, although only one study[45] included solely HGG patients.

With respect to MGMT methylation (training), MRI sequence contributed significantly to heterogeneity in sensitivity ($P < .001$) and specificity ($P < .05$), both of which were respectively higher in the conventional MRI group (0.86, 95%CI 0.81-0.89 and 0.84, 95%CI 0.76-0.90)[25,27,30,31,49] than advanced MRI (0.65, 95%CI 0.40-0.84 and 0.66, 95%CI 0.56-0.74).[28,57,65]

TABLE 2. Sensitivity and Specificity for Machine Learning Prediction of Major Molecular Markers in Glioma

Gene	Dataset	No. of studies	No. of patients	Sensitivity	95% CI	Specificity	95% CI
IDH	Training	18	1496	0.88	0.83-0.91	0.86	0.79-0.91
	Validation	12	500	0.85	0.77-0.90	0.83	0.72-0.90
1p/19q	Training	6	719	0.83	0.72-0.90	0.76	0.71-0.81
	Validation	5	367	0.70	0.45-0.86	0.72	0.63-0.80
MGMT	Training	10	1024	0.81	0.72-0.87	0.80	0.73-0.86
	Validation[a]	3	115	0.70-0.88	–	0.75-0.86	–
ATRX[a]		3	129	0.84-0.95	–	0.75-0.90	–
TERT[a]		4	382	0.75-0.86	–	0.55-0.93	–
EGFR[a]		3	259	0.69-1	–	0.55-0.86	–
P53[a]		4	293	0.67-1	–	0.64-1	–
PTEN[a]		3	253	0.56-1	–	0.84-1	–

[a]Ranges for sensitivity and specificity across training and validation datasets were given where there were insufficient studies (n < 5) to undertake a meta-analysis to obtain pooled values.

ATRX, alpha thalassemia/mental retardation syndrome X-linked; CI, confidence intervals; EGFR, epidermal growth factor receptor; IDH, isocitrate dehydrogenase; MGMT, O^6-methylguanine-DNA methyltransferase; PTEN, phosphatase and tensin homolog; TERT, telomerase reverse transcriptase.

TABLE 3. Investigation of Heterogeneity Through Metaregression for Prediction of Isocitrate Dehydrogenase Mutation in Training Set

Covariates	Subgroup	No. of studies	Sensitivity (95% CI)	*P* value	Specificity (95% CI)	*P* value
Glioma grade	LGG	8	0.86 (0.80-0.91)	.54	0.78 (0.68-0.86)	.12
	HGG	3	0.86 (0.57-0.97)	.47	0.95 (0.72-0.99)	.51
	LGG & HGG	7	0.91 (0.83-0.95)		0.90 (0.81-0.95)	
MRI sequence	Conventional	9	0.85 (0.81-0.88)	.05	0.90 (0.78-0.96)	.49
	Advanced	9	0.92 (0.86-0.95)		0.84 (0.74-90)	
ML algorithm	Random forest	4	0.87 (0.73-0.94)	.49	0.92 (0.75-0.98)	.90
	Support vector machine	7	0.86 (0.82-0.90)	.69	0.78 (0.66-0.86)	.06
	Other	3	0.89 (0.54-0.98)	.57	0.87 (0.66-0.96)	.32
	Deep learning	4	0.92 (0.81-0.97)		0.92 (0.82-0.97)	
Data augmentation	With	5	0.90 (0.84-0.94)	.46	0.95 (0.81-0.99)	.05
	Without	13	0.87 (0.80-0.91)		0.82 (0.75-0.87)	
Setting	Single center	11	0.91 (0.86-0.94)	.07	0.82 (0.75-0.88)	.17
	Multiple centers	7	0.84 (0.79-0.88)		0.93 (0.79-0.98)	
Clinical information	Included	2	0.80 (0.71-0.87)	.12	0.68 (0.51-0.81)	.07
	Not included	16	0.89 (0.84-0.92)		0.88 (0.81-0.92)	

CI, confidence interval; HGG, higher grade glioma (defined as including glioblastoma patients); LGG, lower grade glioma (defined as inclusion of grade II and/or III); ML, machine learning.

Publication Bias

Deeks funnel plot asymmetry test revealed publication bias only in 1p/19q codeletion training dataset analysis ($P = .02$) (see **Figure**, **Supplemental Digital Content 6**).

DISCUSSION

This study found a robust performance of ML models in predicting IDH mutation in glioma with a sensitivity and specificity of 0.85-0.88 in training and 0.83-0.86 in validation set, comparable to previous findings.[13] Heterogeneity was partially explained by data augmentation and MRI sequence. MGMT methylation and 1p/19q codeletion could be predicted with moderate accuracy (sensitivity and specificity 0.76-0.83) in training datasets and should be further validated. Among the other genetic subtypes with limited data, detection of ATRX and TERT promoter mutation reported, more consistently, moderate-high performance compared to EGFR, p53, and PTEN.

Although many studies had unclear RoB, the only important limitation arises from lack of blinding to genotype where manual segmentation or radiologist scoring (eg, VASARI features) was involved, which could introduce subjectivity to the otherwise objective computational process. Patient selection based on

specific genotype class or glioma grade can also introduce imbalanced learning problems and reduce the applicability of findings. This affects the quality of evidence for 1p/19q codeletion,[44,62] TERT promoter,[50,59] and ATRX mutation.[43] Overall, the quality of evidence for use of ML models is strongest for IDH mutation and MGMT methylation.

Implications for Application of ML Models

Several imaging features assessed by radiologists could predict genetic subtypes with moderate accuracy.[67-70] However, the use of ML techniques offers additional value. Firstly, ML approaches are well-suited to recognizing patterns and integrating imaging features from multiple modalities, thereby enabling multimodal multiparametric assessment. Processes such as dimensionality reduction that eliminate irrelevant and redundant variables[27] allow automated selection of the optimal predictive features.

Secondly, applying high-throughput computational analyses to extract radiomics features captures heterogeneous tumor characteristics and overcomes interobserver variability of visual assessment. This is particularly useful for markers like TERT mutation where few imaging phenotypes could be found.[71] Jiang et al[26] demonstrated that a ML model using radiomic features predicted TERT promoter mutation with AUC 0.827. Lastly, DL is increasing favored as it allows learning directly from raw data without predefined features, and automated segmentation.[72] In our subgroup analysis, IDH prediction (training) showed a trend for superior performance (sensitivity and specificity 0.92) in DL group compared to other algorithms.

The clinical applications of optimized ML models are multifold. The prognostic significance of IDH mutation,[73] its implication for surgical resection in astrocytoma,[74] and the role of 1p/19q codeletion and MGMT methylation in predicting chemoradiotherapy response[75,76] highlight the benefits of molecular subtyping for treatment planning. The potential for distinguishing pseudoprogression from tumor progression in glioblastoma[77] is also advantageous. Moreover, a noninvasive model could provide a valuable tool for selecting patients and monitoring response to future targeted therapeutics.[78]

Challenges for Application of ML Models

Only one study explored using ML methods as an adjunct to radiologists in routine practice.[61] Several factors need to be considered for clinical application. We found that advanced MR sequences achieved greater sensitivity for IDH mutation prediction in training dataset, contrary to previous findings.[13] The discrepancy is likely due to the larger number of studies (18 vs 9) and wider range of MR modalities investigated here. However, conventional MRI was superior in predicting MGMT methylation, possibly explained by the greater requirements of advanced sequences for careful postprocessing, such as avoiding areas of necrosis in perfusion-weighted imaging, which affect feature extraction. Thus, different combinations of MRI modalities and feature types need to be tailored to individual molecular markers.

Biological interpretability of ML models was only explored by a few.[22,36] The feasibility of image-guide biopsy,[24] however, suggests it is possible to more accurately characterize radiogenomic correlates across tumor subregions. This is significant given that studies found varying contribution of radiomics features from tumor subregions, such as peritumoral area, in genotype prediction.[26,45,46] The intrinsic heterogeneity of ML studies in patient selection, image acquisition, processing, and ML algorithms further poses a challenge. However, the conglomerate of evidence suggests that advances such as in DL and adherence to high quality methodology, eg, evaluating stability of features across different conditions[16,79] can improve the precision and reliability of ML models.

Lastly, current findings are largely limited to retrospective analysis. Bisdas et al[55] prospectively examined patients with suspected low- or indeterminate grade gliomas who underwent multimodal MRI including diffusion kurtosis imaging (DKI) preoperatively, followed by texture analysis to classify IDH mutation using an SVM model. Hu et al[24] recruited patients with clinically suspected glioblastoma undergoing preoperative conventional MRI for stereotactic biopsies, and then built decision-tree models based on texture features for predicting core driver genes. However, both included a small number of patients. As current ML models are reliant on labeled data, ie, histopathology for model training, future studies should prospectively validate trained models using larger sample sizes in the preoperative setting for clinical translation.

Limitations of the Study

This review has some limitations. While direct within-study comparisons of ML algorithms are desirable, we selected the top-performing model due to the variable algorithms employed. Heterogeneity was partly explained by prespecified covariates in IDH mutation and resolved for MGMT methylation. Poor reporting in some studies also restricted our assessment of methodological quality, but correspondence with authors and tailoring the QUADAS-2 tool allowed more thorough assessment. Lastly, we did not search gray literature (ie, conference proceedings and theses) to identify unpublished studies, although we extended our search to 4 databases. Publication bias was detected for 1p/19q codeletion, which may be attributed to unexplained heterogeneity and selective reporting, but should be interpreted with caution due to the small number of studies.

CONCLUSION

Current evidence shows that IDH mutation can be predicted with moderate-high accuracy, while MGMT methylation and 1p/19q codeletion demonstrated promising results that need to be further validated. Interpretation of other markers is limited, although greater prediction performance was reported for ATRX and TERT promoter mutation. The quality of evidence is overall stronger for IDH mutation and MGMT methylation considering

the precision, applicability, lack of publication bias, and investigated heterogeneity, where data augmentation and choice of MRI sequences were contributing factors. These radiogenomic markers should be prioritized in future studies investigating feasibility in routine workflow to maximize clinical application.

Funding

This study did not receive any funding or financial support.

Disclosures

The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

REFERENCES

1. Dolecek TA, Propp JM, Stroup NE, Kruchko C. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005-2009. *Neuro Oncol*. 2012;14(Suppl 5):v1-v49.
2. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. 2016;131(6):803-820.
3. Patel AP, Tirosh I, Trombetta JJ, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 2014;344(6190):1396-1401.
4. Håvik AB, Brandal P, Honne H, et al. MGMT promoter methylation in gliomas-assessment by pyrosequencing and quantitative methylation-specific PCR. *J Transl Med*. 2012;10(1):1-13.
5. Jang K, Russo C, Di Ieva A. Radiomics in gliomas: clinical implications of computational modeling and fractal-based analysis. *Neuroradiology*. 2020;62(7):771-790.
6. Di Ieva A, Le Reste PJ, Carsin-Nicol B, Ferre JC, Cusimano MD. Diagnostic value of fractal analysis for the differentiation of brain tumors using 3-tesla magnetic resonance susceptibility-weighted imaging. *Neurosurgery*. 2016;79(6):839-846.
7. Zinn PO, Singh SK, Kotrotsou A, et al. Distinct radiomic phenotypes define glioblastoma TP53-PTEN-EGFR mutational landscape. *Clin Neurosurg*. 2017;64(CN_Suppl_1):203-210.
8. Senders JT, Arnaout O, Karhade AV, et al. Natural and artificial intelligence in neurosurgery: a systematic review. *Clin Neurosurg*. 2018;83(2):181-192.
9. Hale AT, Stonko DP, Wang L, Strother MK, Chambless LB. Machine learning analyses can differentiate meningioma grade by features on magnetic resonance imaging. *Neurosurg Focus*. 2018;45(5):E4.
10. Nguyen AV, Blears EE, Ross E, Lall RR, Ortega-Barnett J. Machine learning applications for the differentiation of primary central nervous system lymphoma from glioblastoma on imaging: a systematic review and meta-analysis. *Neurosurg Focus*. 2018;45(5):E5.
11. Kotrotsou A, Zinn PO, Colen RR. Radiomics in brain tumors: an emerging technique for characterization of tumor environment. *Magn Reson Imaging Clin N Am*. 2016;24(4):719-729.
12. Seow P, Wong JHD, Ahmad-Annuar A, Mahajan A, Abdullah NA, Ramli N. Quantitative magnetic resonance imaging and radiogenomic biomarkers for glioma characterisation: a systematic review. *Br J Radiol*. 2018;91(1092):20170930.
13. Zhao J, Huang Y, Song Y, et al. Diagnostic accuracy and potential covariates for machine learning to identify IDH mutations in glioma patients: evidence from a meta-analysis. *Eur Radiol*. 2020;30(8):4664-4674.
14. McInnes MDF, Moher D, Thombs BD, et al. Reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies the PRISMA-DTA statement. *JAMA*. 2018;319(4):388-396.
15. Whiting PF, Rutjes AW, Westwood ME, et al. The QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;154(4):253-260.
16. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762.
17. Leeflang MMG, Moons KGM, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. *Clin Chem*. 2008;54(4):729-737.
18. Zapf A. Appraising heterogeneity. In: Biondi-Zoccai G, ed. *Diagnostic Meta-Analysis*. 1st ed. Cham, Switzerland: Springer International Publishing; 2018:139-140.
19. Kim KW, Lee J, Choi SH, Huh J, Park SH. Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part I. General guidance and tips. *Korean J Radiol*. 2015;16(6):1175-1187.
20. Deeks JJ, Higgins JPT, Altman DG. Chapter 10: analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane; 2019. www.training.cochrane.org/handbook. Published 2019.
21. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882-893.
22. Akbari N, Bakas S, Pisapia JM, et al. In vivo evaluation of EGFRvIII mutation in primary glioblastoma patients via complex multiparametric MRI signature. *Neuro Oncol*. 2018;20(8):1068-1079.
23. Akkus Z, Ali I, Sedlář J, et al. Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence. *J Digit Imaging*. 2017;30(4):469-476.
24. Hu LS, Ning S, Eschbacher JM, et al. Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. *Neuro Oncol*. 2017;19(1):128-137.
25. Jiang C, Kong Z, Liu S, et al. Fusion radiomics features from conventional MRI predict MGMT promoter methylation status in lower grade gliomas. *Eur J Radiol*. 2019;121:108714. doi:10.1016/j.ejrad.2019.108714.
26. Jiang C, Kong Z, Zhang Y, et al. Conventional magnetic resonance imaging – based radiomic signature predicts telomerase reverse transcriptase promoter mutation status in grade II and III gliomas. *Neuroradiology*. 2020;62(7):803-813.
27. Kanas VG, Zacharaki EI, Thomas GA, Zinn PO, Megalooikonomou V, Colen RR. Learning MRI-based classification models for MGMT methylation status prediction in glioblastoma. *Comput Methods Programs Biomed*. 2017;140:249-257. doi:10.1016/j.cmpb.2016.12.018.
28. 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.
29. Kocak B, Durmaz ES, Ates E, et al. Radiogenomics of lower-grade gliomas: machine learning-based MRI texture analysis for predicting 1p/19q codeletion status. *Eur Radiol*. 2020;30(2):877-886.
30. Korfiatis P, Kline TL, Coufalova L, et al. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. *Med Phys*. 2016;43(6 pt 1):2835-2844.
31. Korfiatis P, Kline TL, Lachance DH, Parney IF, Buckner JC, Erickson BJ. Residual deep convolutional neural network predicts MGMT methylation status. *J Digit Imaging*. 2017;30(5):622-628.
32. Liang S, Zhang R, Liang D, et al. Multimodal 3D densenet for IDH genotype prediction in gliomas. *Genes (Basel)*. 2018;9(8):1-17.
33. Li Y, Liu X, Qian Z, et al. Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature. *Eur Radiol*. 2018;28(7):2960-2968.
34. Alis D, Bagcilar O, Senli YD, et al. Machine learning-based quantitative texture analysis of conventional MRI combined with ADC maps for assessment of IDH1 mutation in high-grade gliomas. *Jpn J Radiol*. 2020;38(2):135-143.
35. Li Y, Qian Z, Xu K, et al. MRI features predict p53 status in lower-grade gliomas via a machine-learning approach. *NeuroImage: Clinical*. 2018;17:306-311. doi:10.1016/j.nicl.2017.10.030.
36. Li Y, Liang Y, Sun Z, et al. Radiogenomic analysis of PTEN mutation in glioblastoma using preoperative multi-parametric magnetic resonance imaging. *Neuroradiology*. 2019;61(11):1229-1237.
37. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep*. 2017;7(5467):1-11.
38. Li Z-C, Bai H, Sun Q, et al. Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: a multicentre study. *Eur Radiol*. 2018;28(9):3640-3650.
39. Liu C, Zhang H, Pan Y, Huang F, Xia S. Towards MIB-1 and p53 detection in glioma magnetic resonance image: a novel computational image analysis method. *Phys Med Biol*. 2012;57(24):8393-8404.

40. Lo CM, Weng RC, Cheng SJ, Wang HJ, Hsieh KLC, Ding J. Computer-aided diagnosis of isocitrate dehydrogenase genotypes in glioblastomas from radiomic patterns. *Med (United States)*. 2020;99(8):1-8.
41. Lu C-F, Hsu F-T, Hsieh KL-C, et al. Machine learning-based radiomics for molecular subtyping of gliomas. *Clin Cancer Res*. 2018;24(18):4429-4436.
42. Ozturk-Isik E, Cengiz S, Ozcan A, et al. Identification of IDH and TERTp mutation status using 1H-MRS in 112 hemispheric diffuse gliomas. *J Magn Reson Imaging*. 2019;51(6):1799-1809.
43. Ren Y, Zhang X, Rui W, et al. Noninvasive prediction of IDH1 mutation and ATRX expression loss in low-grade gliomas using multiparametric MR radiomic features. *J Magn Reson Imaging*. 2019;49(3):808-817.
44. Shofty B, Artzi M, Ben Bashat D, et al. MRI radiomics analysis of molecular alterations in low-grade gliomas. *Int J CARS*. 2018;13(4):563-571.
45. Li Z-C, Bai H, Sun Q, et al. Multiregional radiomics profiling from multiparametric MRI: identifying an imaging predictor of IDH1 mutation status in glioblastoma. *Cancer Med*. 2018;7(12):5999-6009.
46. Tan Y, Zhang S-T, Wei J-W, et al. A radiomics nomogram may improve the prediction of IDH genotype for astrocytoma before surgery. *Eur Radiol*. 2019;29(7):3325-3337.
47. van der Voort SR, Incekara F, Wijnenga MMJ, et al. Predicting the 1p/19q codeletion status of presumed low-grade glioma with an externally validated machine learning algorithm. *Clin Cancer Res*. 2019;25(24):7455-7462.
48. Wu G, Chen Y, Wang Y, et al. Sparse representation-based radiomics for the diagnosis of brain tumors. *IEEE Trans Med Imaging*. 2018;37(4):893-905.
49. Xi Y-B, Guo F, Xu Z-L, et al. Radiomics signature: a potential biomarker for the prediction of MGMT promoter methylation in glioblastoma. *J Magn Reson Imaging*. 2018;47(5):1380-1387.
50. Yamashita K, Hatae R, Hiwatashi A, et al. Predicting TERT promoter mutation using MR images in patients with wild-type IDH1 glioblastoma. *Diagn Interv Imaging*. 2019;100(7-8):411-419.
51. Yogananda CGB, Shah BR, Vejdani-Jahromi M, et al. A novel fully automated MRI-based deep-learning method for classification of IDH mutation status in brain gliomas. *Neuro Oncol*. 2020;22(3):402-411.
52. Yu J, Shi Z, Lian Y, et al. Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. *Eur Radiol*. 2017;27(8):3509-3522.
53. Zhang X, Tian Q, Wang L, et al. Radiomics strategy for molecular subtype stratification of lower-grade glioma: detecting IDH and TP53 mutations based on multimodal MRI. *J Magn Reson Imaging*. 2018;48(4):916-926.
54. Zhou H, Vallières M, Bai HX, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol*. 2017;19(6):862-870.
55. Bisdas S, Shen H, Thust S, et al. Texture analysis- and support vector machine-assisted diffusional kurtosis imaging may allow in vivo gliomas grading and IDH-mutation status prediction: a preliminary study. *Sci Rep*. 2018;8(1):1-9.
56. Choi KS, Choi SH, Jeong B. Prediction of IDH genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network. *Neuro Oncol*. 2019;21(9):1197-1209.
57. Chen L, Zhang H, Lu J, et al. Multi-label nonlinear matrix completion with transductive multi-task feature selection for joint MGMT and IDH1 status prediction of patient with high-grade gliomas. *IEEE Trans Med Imaging*. 2018;37(8):1775-1787.
58. Eichinger P, Alberts E, Delbridge C, et al. Diffusion tensor image features predict IDH genotype in newly diagnosed WHO grade II/III gliomas. *Sci Rep*. 2017;7(1):13396.
59. Fukuma R, Yanagisawa T, Kinoshita M, et al. Prediction of IDH and TERT promoter mutations in low-grade glioma from magnetic resonance images using a convolutional neural network. *Sci Rep*. 2019;9(1):1-5.
60. Hajianfar G, Shiri I, Maleki H, et al. Noninvasive O^6 Methylguanine-DNA methyltransferase status prediction in glioblastoma multiforme cancer using magnetic resonance imaging radiomics features: univariate and multivariate radiogenomics analysis. *World Neurosurg*. 2019;132:e140-e161. doi:10.1016/j.wneu.2019.08.232.
61. De Looze C, Beausang A, Cryan J, et al. Machine learning: a useful radiological adjunct in determination of a newly diagnosed glioma's grade and IDH status. *J Neurooncol*. 2018;139(2):491-499.
62. Fellah S, Caudal D, De Paula AM, et al. Multimodal MR imaging (diffusion, perfusion, and spectroscopy): is it possible to distinguish oligodendroglial tumor grade and 1p/19q codeletion in the pretherapeutic diagnosis?. *AJNR Am J Neuroradiol*. 2013;34(7):1326-1333.
63. Ge C, Gu IYH, Jakola AS, Yang J. Enlarged training dataset by pairwise GANs for molecular-based brain tumor classification. *IEEE Access*. 2020;8:22560-22570. doi:10.1109/ACCESS.2020.2969805.
64. Han Y, Xie Z, Zang Y, et al. Non-invasive genotype prediction of chromosome 1p/19q co-deletion by development and validation of an MRI-based radiomics signature in lower-grade gliomas. *J Neurooncol*. 2018;140(2):297-306.
65. Haubold J, Demircioglu A, Gratz M, et al. Non-invasive tumor decoding and phenotyping of cerebral gliomas utilizing multiparametric 18F-FET PET-MRI and MR fingerprinting. *Eur J Nucl Med Mol Imaging*. 2020;47(6):1435-1445.
66. Rücker G, Schumacher M. Letter to the editor. *Biostatistics*. 2009;10(4):806-807.
67. Delfanti RL, Piccioni DE, Handwerker J, et al. Imaging correlates for the 2016 update on WHO classification of grade II/III gliomas: implications for IDH, 1p/19q and ATRX status. *J Neurooncol*. 2018;135(3):601-609.
68. Qi S, Yu L, Li H, et al. Isocitrate dehydrogenase mutation is associated with tumor location and magnetic resonance imaging characteristics in astrocytic neoplasms. *Oncol Lett*. 2014;7(6):1895-1902.
69. Kickingereder P, Sahm F, Radbruch A, et al. IDH mutation status is associated with a distinct hypoxia/angiogenesis transcriptome signature which is non-invasively predictable with rCBV imaging in human glioma. *Sci Rep*. 2015;5(Oct):1-9.
70. Kanazawa T, Minami Y, Jinzaki M, Toda M, Yoshida K, Sasaki H. Predictive markers for MGMT promoter methylation in glioblastomas. *Neurosurg Rev*. 2019;42(4):867-876.
71. Ersoy TF, Keil VC, Hadizadeh DR, et al. New prognostic factor telomerase reverse transcriptase promotor mutation presents without MR imaging biomarkers in primary glioblastoma. *Neuroradiology*. 2017;59(12):1223-1231.
72. Shaver MM, Kohanteb PA, Chiou C, et al. Optimizing neuro-oncology imaging: a review of deep learning approaches for glioma imaging. *Cancers (Basel)*. 2019;11(6):1-15.
73. Hartmann C, Hentschel B, Wick W, et al. Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: implications for classification of gliomas. *Acta Neuropathol*. 2010;120(6):707-718.
74. Beiko J, Suki D, Hess KR, et al. IDH1 mutant malignant astrocytomas are more amenable to surgical resection and have a survival benefit associated with maximal surgical resection. *Neuro Oncol*. 2014;16(1):81-91.
75. Chamberlain MC, Born D. Prognostic significance of relative 1p/19q codeletion in oligodendroglial tumors. *J Neurooncol*. 2015;125(2):249-251.
76. Weller M, Tabatabai G, Kästner B, et al. MGMT MGMT promoter methylation is a strong prognostic biomarker for benefit from dose-intensified temozolomide rechallenge in progressive glioblastoma: the DIRECTOR trial. *Clin Cancer Res*. 2015;21(9):2057-2064.
77. Brandes AA, Franceschi E, Tosoni A, et al. MGMT promoter methylation status can predict the incidence and outcome of pseudoprogression after concomitant radiochemotherapy in newly diagnosed glioblastoma patients. *J Clin Oncol*. 2008;26(13):2192-2197.
78. Jain KK. A critical overview of targeted therapies for glioblastoma. *Front Oncol*. 2018;8(Oct):1-19.
79. Cattell R, Chen S, Huang C. Robustness of radiomic features in magnetic resonance imaging: review and a phantom study. *Vis Comput Ind Biomed Art*. 2019;2(1):19.

Acknowledgments

We would like to acknowledge Marie Simons for her assistance in the search strategy and Philip Döbler for assisting with the use of R packages.

Supplemental digital content *is available for this article at www.neurosurgery-online.com.*

Supplemental Digital Content 1. Methods. Search Strategy.
Supplemental Digital Content 2. Figure. Coupled forest plots of included studies using machine learning to predict 1p/19q codeletion in training dataset.

Supplemental Digital Content 3. Figure. Sensitivity and false positive rate (1 – specificity) of studies that evaluated prediction of 1p19 codeletion in training dataset. Each study is represented by an open circle with the shaded circle representing the pooled sensitivity 0.828 (95% CI 0.722-0.899) and specificity 0.763 (95% CI 0.706-0.813) surrounded by a 95% confidence ellipse (solid line) and prediction ellipse (dashed line).

Supplemental Digital Content 4. Figure. Coupled forest plots of included studies using machine learning to predict MGMT methylation status in training dataset.

Supplemental Digital Content 5. Figure. Sensitivity and false positive rate (1 – specificity) of studies that evaluated MGMT methylation status prediction in training dataset. Each study is represented by an open circle with the shaded circle representing the pooled sensitivity 0.808 (95% CI 0.720-0.873) and specificity 0.800 (95% CI 0.726-0.858) surrounded by a 95% confidence ellipse (solid line) and prediction ellipse (dashed line).

Supplemental Digital Content 6. Figure. Funnel plot of publication bias for 1p/19q codeletion prediction in training dataset.