Diagn Interv Radiol 2021; 27:716–724

HEAD AND NECK IMAGING

REVIEW

Diagnostic performance of radiomics using machine learning algorithms to predict MGMT promoter methylation status in glioma patients: a meta-analysis

Huan Huang
Fei-fei Wang
Shigang Luo
Guangxiang Chen
Guangcai Tang

PURPOSE
We aimed to assess the diagnostic performance of radiomics using machine learning algorithms to predict the methylation status of the O^6-methylguanine-DNA methyltransferase (MGMT) promoter in glioma patients.

METHODS
A comprehensive literature search of PubMed, EMBASE, and Web of Science until 27 July 2021 was performed to identify eligible studies. Stata SE 15.0 and Meta-Disc 1.4 were used for data analysis.

RESULTS
A total of 15 studies with 1663 patients were included: 5 studies with training and validation cohorts and 10 with only training cohorts. The pooled sensitivity and specificity of machine learning for predicting MGMT promoter methylation in gliomas were 85% (95% CI 79%–90%) and 84% (95% CI 78%–88%) in the training cohort (n=15) and 84% (95% CI 70%–92%) and 78% (95% CI 63%–88%) in the validation cohort (n=5). The AUC was 0.91 (95% CI 0.88–0.93) in the training cohort and 0.88 (95% CI 0.85–0.91) in the validation cohort. The meta-regression demonstrated that magnetic resonance imaging sequences were related to heterogeneity. The sensitivity analysis showed that heterogeneity was reduced by excluding one study with the lowest diagnostic performance.

CONCLUSION
This meta-analysis demonstrated that machine learning is a promising, reliable and repeatable candidate method for predicting MGMT promoter methylation status in glioma and showed a higher performance than non-machine learning methods.

Glioma is the most common type of primary malignant central nervous system (CNS) tumor and accounts for approximately 75% of primary malignant CNS tumors (1). Despite developments in surgery, chemotherapy, and radiotherapy, patients with glioma still suffer an unpleasant prognosis (2). In recent years, increasing attention has been given to molecular markers in patients with glioma. O^6-methylguanine-DNA methyltransferase (MGMT) is a key gene that encodes a DNA repair enzyme. The methylated MGMT promoter is usually related to better overall survival in temozolomide (TMZ)-treated gliomas (3–5). In addition, the MGMT gene is a potential attractive therapeutic target in the molecularly targeted therapy field (6, 7). Moreover, it has been reported that MGMT promoter methylation status is significantly associated with glioma pseudo-progression in recent studies (8, 9).

At present, the approaches for determining MGMT promoter methylation status in glioma are based on surgical sampling (10), which is an invasive procedure and may induce severe complications. The results always take a relatively long period, which may delay important therapeutic decisions and be influenced by intra-tumoral heterogeneity. Thus, identifying a noninvasive, preoperative, and robust means to detect MGMT promoter status is of great significance. Radiomics, which is an advanced imaging analysis technique, utilizes algorithms to automatically extract a large number of data features to convert imaging data into a high-dimensional and mineable feature space (11). Machine learning algorithms have been used to create credible statistical models for classification in radiomics (12), and

From the Department of Radiology (H.H., F.F.W., S.L., G.C. ✉ *cgx23ly2002@163.com*, G.T. ✉ *168345315@qq.com*) Affiliated Hospital of Southwest Medical University, Sichuan, China.

Received 23 February 2021; revision requested 21 May 2021; last revision received 9 August 2021; accepted 28 August 2021.

Published online 15 October 2021.

DOI 10.5152/dir.2021.21153

You may cite this article as: Huang H, Wang FF, Luo S, Chen G, Tang G. Diagnostic performance of radiomics using machine learning algorithms to predict MGMT promoter methylation status in glioma patients: a meta-analysis. Diagn Interv Radiol 2021; 27:716–724.

they have already had a powerful influence on radiology practice and could further change the area of radiology (13–15). For the CNS, magnetic resonance imaging (MRI) is the most common and noninvasive preoperative diagnostic imaging method. Increasing evidence in neuro-oncology has indicated that radiomics features based on MRI can predict the molecular subtype of glioma (16–18).

In recent years, a few studies have demonstrated that machine learning performs well in predicting MGMT promoter methylation status in glioma (16, 18–22). However, to our knowledge, no study has performed a systematic assessment of the diagnostic accuracy of machine learning for predicting MGMT promoter methylation status. Thus, our meta-analysis aimed to systematically evaluate the diagnostic efficacy of machine learning for predicting MGMT promoter methylation status in patients with glioma.

Methods

The present meta-analysis was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (23).

Literature search

A comprehensive literature search of PubMed, EMBASE, and Web of Science from 1 January 2000 to 27 July 2021 was performed to identify related eligible studies. The specific search strategy combining the following keywords was used: ((Radiomics) OR (machine learning) OR (deep learning) OR (neural network) OR (radiomics nomogram) OR (Algorithms) OR (Artificial Intelligence) OR (computer-assisted diagnosis) OR (texture analysis)) AND ((glioma) OR (glioblastoma) OR (astrocytoma) OR (astroglioma) OR (oligodendroglia)) AND ((O^6-methylguanine-DNA-methyltransferase) OR (MGMT) OR (methylguanine-DNA methyltransferase)). In addition, the reference lists of the included original articles and relevant papers were manually reviewed for studies that were not found during the database searches. The literature search had no language or date limitations.

Main points

- This meta-analysis demonstrated that machine learning is a promising, reliable, and repeatable candidate method for predicting the methylation status of the MGMT promoter.
- Machine learning utilizing conventional MRI sequences showed a higher diagnostic performance than advanced MRI sequences in predicting MGMT promoter methylation status.
- Our results revealed that machine learning has a relatively higher performance than non-machine learning methods.

Inclusion criteria

Studies were considered on the basis of the following criteria: 1) patients with grade II, III or IV gliomas confirmed by histopathological analysis; 2) methylation-specific polymerase chain reaction used as the reference standard to identify MGMT promoter methylation status; 3) machine learning used to predict the MGMT promoter status; and 4) true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) values available from the original studies to generate 2 × 2 tables for determining the diagnostic efficacy of machine learning in prediction of MGMT promoter status.

Exclusion criteria

Studies were excluded if they 1) were conference abstracts, reviews, letters, comments or case reports/case series involving <10 patients; 2) were not focusing on the diagnostic efficacy of machine learning for predicting the MGMT promoter even though machine learning was used to classify the MGMT promoter status; and 3) included overlapping patient cohorts.

Data extraction and quality assessment

We extracted the following data from the included studies: sample size, number of patients in the training and validation cohorts, number of patients with MGMT promoter methylation, World Health Organization (WHO) grade of glioma, mean age, male/female ratio, authors, year of publication, patient recruitment period, study design, MRI field strength (T) and scanning sequence, model methods, and method of region-of-interest (ROI) delineation. These data were independently extracted by two investigators, and any disagreements were resolved by discussion.

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) (24) was applied to assess the quality of the included studies. Domains including patient selection, index test, reference standards, flow and timing were evaluated:

Statistical analysis

All statistical analyses were performed by using Stata SE 15.0 and Meta-Disc 1.4. For each study, 2×2 tables were reconstructed to calculate the pooled sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR). A bivariate random effect model was used to calculate the pooled sensitivity, specificity, PLR, NLR, DOR and their 95% confidence intervals (95% CIs). An integrated hierarchical summary receiver operating characteristic (HSROC) plot and the curve area under the HSROC curve (AUC) were used to evaluate the diagnostic performance of machine learning for predicting MGMT promoter methylation status.

Heterogeneity was evaluated according to the following means: (1) Cochran's Q test, with $p > 0.1$ suggesting no heterogeneity; (2) Higgins inconsistency index (I^2) test with a value >50% suggesting substantial heterogeneity; (3) visual assessment of the difference between the 95% confidence region in the HSROC curve; and (4) a Spearman correlation coefficient >0.6 suggesting the presence of a threshold effect. Publication bias was evaluated by using Deeks' funnel plot asymmetry test, and a $p < 0.05$ showed potential publication bias.

Furthermore, we performed a subgroup analysis of the included studies with training cohorts to explain the reasons for heterogeneity. The covariates that were covered were as follows: 1) MRI field strength; 2) glioma grade; 3) number of patients in the training cohort; 4) MRI sequences; 5) publication year; and 6) blinding of MRI readers to the reference standard.

Results

A total of 456 studies were initially confirmed using the above-described search strategies. Ultimately, according to the inclusion criteria, 15 original articles (16, 19–22, 25–34) with 1663 patients were involved in this meta-analysis. Among these included studies, 10 studies included only training cohorts (16, 25–28, 30–34), and the other 5 studies included training and validation cohorts (19–22, 29). The number of patients were 1432 and 231 in the training and validation cohorts, respectively.

All the included studies were retrospective cohort studies. Seven of the 15 studies used 3T MRI (16, 19–22, 26, 30, 31), 2 studies used 1.5T or 3T MRI (28, 32), and 4 studies did

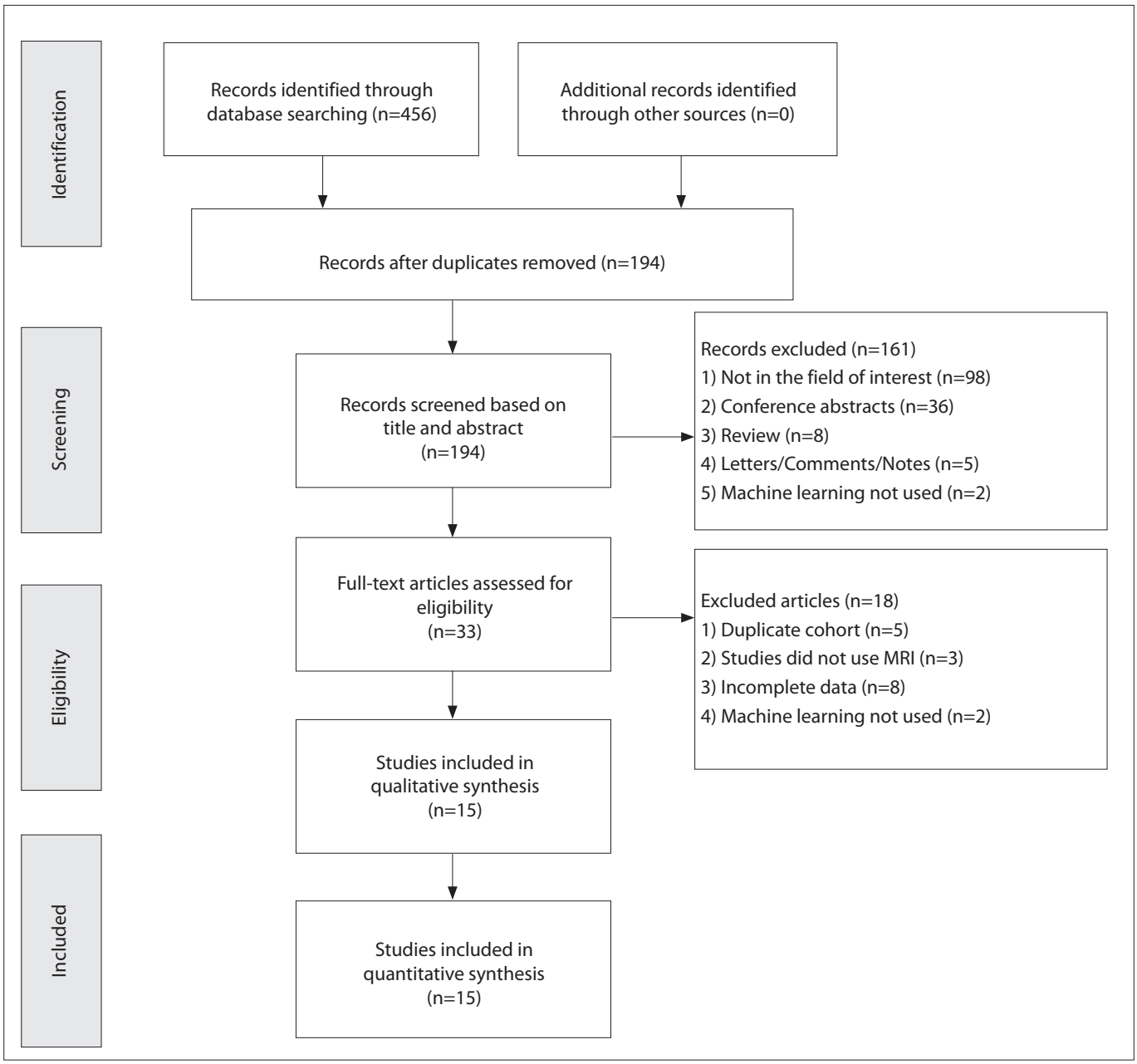


Figure 1. Flow diagram of the literature selection process.

not report the MRI strength (26, 29, 33, 34). Regarding the MRI sequences, 8 studies utilized conventional MRI sequences, namely T1-weighted imaging, T2-weighted imaging (T2WI), fluid-attenuated inversion recovery, and contrast-enhanced (19–22, 28, 29, 33, 34), and the remaining 7 studies utilized advanced sequences, namely diffusion-weighted imaging, dynamic susceptibility contrast, diffusion tensor imaging, susceptibility-weighted imaging (SWI), resting-state functional MRI, inflow-based vascular space occupancy (16, 20, 25, 27, 30–32) in addition to conventional MRI sequences. Seven studies included patients with grade IV gliomas (19, 21, 22, 26–29), 1 study included patients with grades III and IV gliomas (25), 6 studies included patients with gliomas of grades II, III and IV (20, 30–34), and 1 study included patients with gliomas of grades II and III (16). Seven of 9 studies were published before 2019 (19, 20, 22, 25–28), and the other 8 studies were published after 2019 (16, 21, 29–34). Only 1 of 15 studies used ROI measurement and tumor segmentation to delineate ROIs (16), and the remaining 14 studies used tumor segmentation (19–22, 25–34). Various machine-learning models were used in all included studies, including multilabel nonlinear matrix completion (MNMC), deep learning, decision tree (DT), naive Bayes (NB), multilayer perceptron (MLP), support vector machine (SVM), lasso and elastic net regularized generalized linear model (GLMNET), random forest (RF), k-nearest neighbors (KNN), stochastic gradient boosting machines (sGBM), and eXtreme gradient boosting (XGBoost). The process of the literature selection is shown in Fig. 1. The patient and study characteristics are shown in Table 1.

Table 1. Studies and patient characteristics

Study no.	Author (year of publication)	No. of total patients	No. of patients in training cohorts	No. of patients in validation cohorts	WHO grade	MGMT (+)	MGMT (-)	Mean age (years)	Male: Female	Duration of patients	Study design	Blinding to reference standard	Magnet field strength (T)	Scanning sequence	Model methods	Method of ROI delineation
1	Chen et al. (2018)	47	47	--	III, IV	26	20	48.3	26:20	2010–2015	Retrospective	NR	3.0T	T1WI, DTI, RS-functional MRI	MNMC	Tumor segmentation and ROI measurement
2	Crisi et al. (2020)	59	59	--	II, III	39	20	59.4	39:20	2013–2017	Retrospective	Yes	3.0T	T1-GRE, T2-GRE, FLAIR, DWI, DSC	NB, DT, MLP	Tumor segmentation
3	Jiang et al. (2019)	122	87	35	IV	88	34	45.1	62:60	2010–2018	Retrospective	Yes	3.0T	3D-CE-T1, T2WI	SVM, RF, AdaBoost	Tumor segmentation
4	Kanas et al. (2017)	86	86	--	IV	43	43	58.0	59:27	NR	Retrospective	Yes	NR	T1WI, FLAIR, CE-T1	DT, KNN, Gaussian NB	Tumor segmentation
5	Kickinge-reder et al. (2016)	152	152	--	IV	68	77	63.0	84:68	2009–2016	Retrospective	NR	3.0T	T2-FLAIR, DWI, SWI, DSC	PLR, RF, sGBM	Tumor segmentation
6	Korfiatis et al. (2016)	155	155	--	IV	66	89	NR	NR	2007–2015	Retrospective	Yes	1.5T/ 3.0T	T2-FSE, T1WI	SVM, RF	Tumor segmentation
7	Li et al. (2018)	193	133	60	IV	86	107	53.3	118:68	2011–2016	Retrospective	NR	3.0T	T1WI, T2WI, FLAIR	RF	Tumor segmentation
8	Wei et al. (2018)	105	74	31	II, III, IV	73	32	31.0	60:45	2011–2017	Retrospective	Yes	3.0T	CE-T1, FLAIR, DWI	Logistic regression	Tumor segmentation
9	Xi et al. (2017)	118	98	20	IV	56	62	NR	70:48	2012–2016	Retrospective	Yes	3.0T	T1WI, T2WI, CE-T1	SVM	Tumor segmentation

ROI, region of interest; NR, not reported; MNMC, multilabel nonlinear matrix completion; NB, naive Bayes; DT, decision tree; MLP, multilayer perceptron; SVM, support vector machine; RF, random forest; GLMNET, lasso and elastic net regularized generalized linear model; random forest; KNN, k-nearest neighbors; PLR, positive likelihood ratio; sGBM, stochastic gradient boosting machines; XGBoost, eXtreme gradient boosting; iVASO, inflow-based vascular-space-occupancy.

The results of the quality assessment of the included studies using the QUADAS-2 assessment checklist are described in Fig. 2.

Our sample sizes contained 15 studies assessing the diagnostic performance of machine learning for predicting MGMT promoter status in patients with glioma. The sensitivity of the individual included studies ranged from 57% to 96% and 70% to 94% in the training cohort and validation cohort, respectively. The specificity of the individual included studies ranged from 65% to 96% and 54% to 88% in the training cohort and validation cohort, respectively. The pooled sensitivity and specificity of machine learning for predicting MGMT promoter methylation in the training cohorts (n=15) were 85% (95% CI 79%–90%) and 84% (95% CI 78%–88%), respectively (Fig. 3). The pooled PLR, NLR, and DOR were 5.3 (95% CI 3.7–7.6), 0.18 (95% CI 0.12–0.26), and 30 (95% CI 14–62), respectively. The AUC was 0.91 (95% CI 0.88–0.93, standard error (SE)=0.024, $p < 0.001$) (Fig. 4). In the validation cohort (n=4), the pooled sensitivity, specificity, PLR, NLR, and DOR were 84% (95% CI 70%–92%), 78% (95% CI 63%–88%), 3.8 (95% CI 2.2–6.6), 0.21 (95% CI 0.11–0.38), and 18 (95% CI 8–43), respectively. The AUC was 0.88 (95%CI 0.85–0.91, SE=0.033, $p < 0.001$).

Cochran's Q test demonstrated that heterogeneity was absent (Q=0.20, $p = 0.45$) in all included studies, but substantial heterogeneity in the sensitivity (I^2=81.22%, $p < 0.001$) and specificity (I^2=79.66%, $p < 0.001$) was measured by the Higgins I^2 statistic. The Spearman correlation coefficient was -0.618 ($p = 0.014$), indicating the absence of a threshold effect. The likelihood of publication bias was low ($p = 0.89$; Fig. 5), as demonstrated by Deeks' funnel plot.

A meta-regression was conducted to explore the source of the heterogeneity. Among the covariates, MRI sequences were associated with heterogeneity. Magnetic field strength, glioma grade, and blinding of MRI readers to the reference standard were not demonstrated to be significant causes influencing the heterogeneity.

The sample sizes of the validation cohort were inadequate to make credible conclusions, and we only performed a subgroup analysis in the training cohort (Table 2). Magnetic field strength, grade, sample size, MRI sequences, blinding to the reference standard, and publication year influenced the diagnostic performance of machine learning for predicting MGMT promoter

Table 1. Studies and patient characteristics (Cont'd)

Study no.	Author (year of publication)	No. of total patients	No. of patients in training cohorts	No. of patients in validation cohorts	WHO grade	MGMT (+)	MGMT (-)	Mean age (years)	Male: Female	Duration of patients	Study design	Blinding to reference standard	Magnet field strength (T)	Scanning sequence	Model methods	Method of ROI delineation
10	Yogananda et al. (2021)	247	247		II, III, IV	163	84	49.4	NR	NR	Retrospective	Yes	NR	T2WI	Deep learning	Tumor segmentation
11	Huang et al. (2021)	53	53		II, III, IV	21	32	NR	33:20	2013–2018	Retrospective	Yes	3.0T	CE-T1, FLAIR, DWI	Logistic regression	Tumor segmentation
12	Le et al. (2020)	53	53		II, III, IV	26	27	60.0	27:26	NR	Retrospective	NR	NR	T1WI, T2WI, FLAIR, CE-T1	KNN, RF, SVM, XGBoost	Tumor segmentation
13	Chen et al. (2021)	66	51	15	IV	32	34	NR	NR	NR	Retrospective	NR	NR	FLAIR, CE-T1	Deep learning	Tumor segmentation
14	Kihira et al. (2021)	91	91		II, III, IV	42	49	57.0	54:37	2016–2018	Retrospective	Yes	1.5T/ 3.0T	T2WI, CE-T1, DWI	Logistic regression	Tumor segmentation
15	He et al. (2021)	46	46		II, III, IV	20	26	43.9	33:13	2017–2019	Retrospective	Yes	3.0T	FLAIR, CE-T1, 3D-iVASO	Logistic regression, RF	Tumor segmentation

ROI, region of interest; NR, not reported; MNMC, multilabel nonlinear matrix completion; NB, naive Bayes; DT, decision tree; MLP, multilayer perceptron; SVM, support vector machine; RF, random forest; GLMNET, lasso and elastic net regularized generalized linear model; random forest; KNN, k-nearest neighbors; PLR, positive likelihood ratio; sGBM, stochastic gradient boosting machines; XGBoost, eXtreme gradient boosting; iVASO, inflow-based vascular-space-occupancy.

methylation status. Studies using conventional MRI sequences showed better diagnostic efficacy in predicting MGMT promoter methylation with a pooled sensitivity of 90% and specificity of 86%; however, studies using conventional and advanced MRI sequences had a lower pooled sensitivity of 73% and specificity of 74%. Studies involving low-grade glioma had better sensitivity (88% vs. 79%) and specificity (85% vs. 79%) and a higher diagnostic odds ratio (38 vs. 20). In our analysis, the studies in which the sample size of the training group was more than 100 patients had a similar sensitivity (84% vs. 85%) and specificity (81% vs. 82%) but a slightly lower DOR (24 vs. 28). In addition, the diagnostic efficacy of machine learning also correlated with publication year, and studies published after 2019 had better values of sensitivity (89% vs. 80%), specificity (87% vs. 77%) and DOR (53 vs. 15) than studies published before 2019.

In the sensitivity analysis (Supplementary material 1), the study with the lowest diagnostic performance (sensitivity of 57% and specificity of 61%) among the included studies showed significant influence on heterogeneity, when it was removed the heterogeneity of sensitivity decreased from $I^2 = 81.2\%$ to $I^2 = 64.2\%$, and specificity decreased from $I^2 = 79.6\%$ to $I^2 = 60.2\%$.

Discussion

In this meta-analysis, we found that machine learning showed high diagnostic efficacy (AUC=0.91) in noninvasively predicting MGMT promoter methylation status. The summary sensitivity and specificity were 85% and 84%, respectively. The meta-regression demonstrated that MRI sequences were associated with heterogeneity. The sensitivity analysis showed that the study with the lowest diagnostic performance among the included studies showed significant influence on heterogeneity. Moreover, subgroup analysis revealed that the diagnostic efficacy of machine learning in the prediction of MGMT promoter methylation status was affected by magnetic field strength, glioma grade, blinding of MRI readers to the reference standard, number of training groups, MRI sequences, and publication year.

Based on the results of this meta-analysis, machine learning studies showed a relatively higher performance in predicting MGMT promoter methylation status than that of non-machine learning studies using

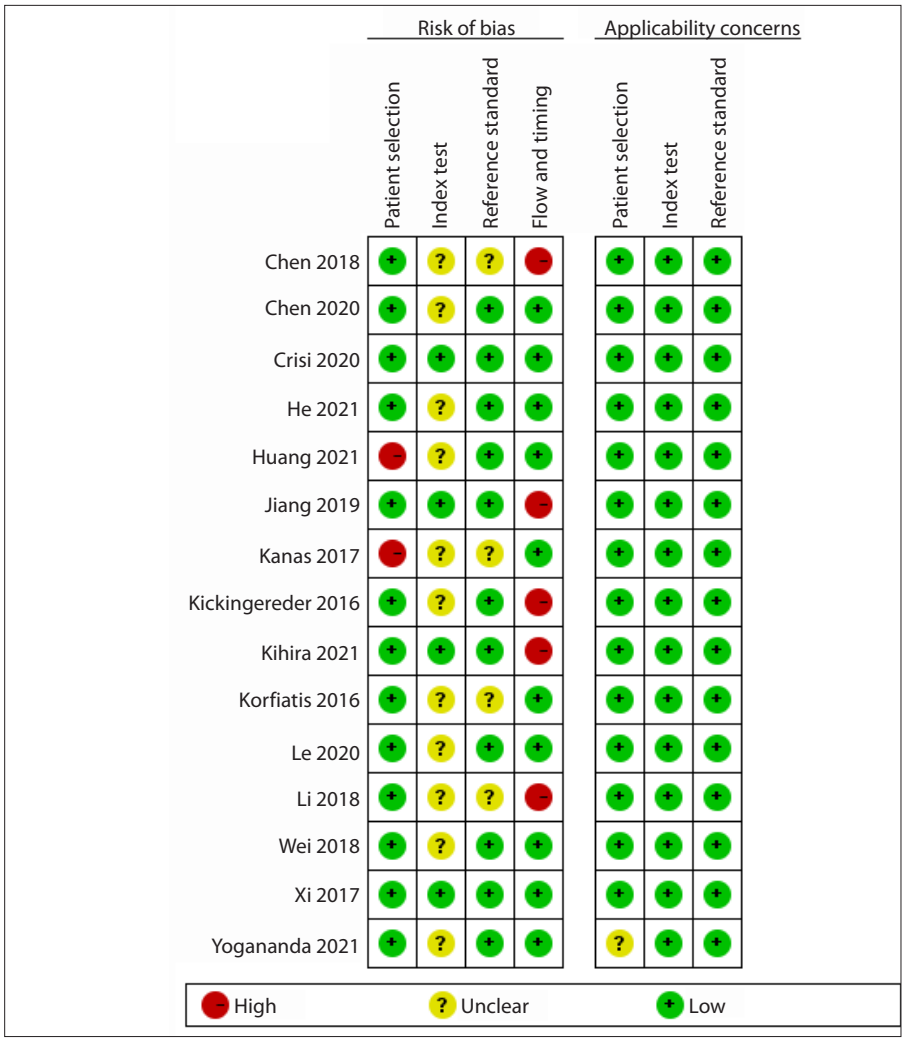


Figure 2. Risk of bias and applicability concerns summary: review of authors' judgements about each domain for each included study.

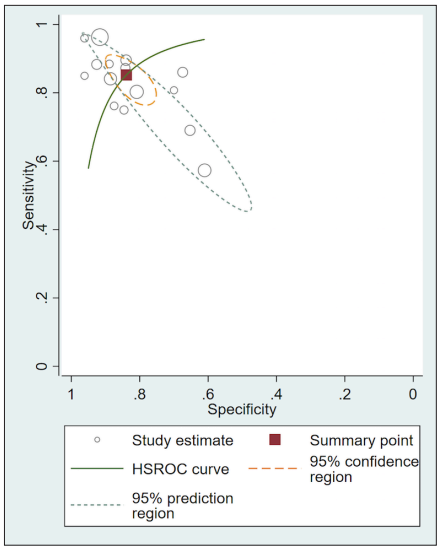


Figure 4. HSROC curve of the diagnostic efficacy of machine learning in the prediction of MGMT promoter methylation status in glioma patients.

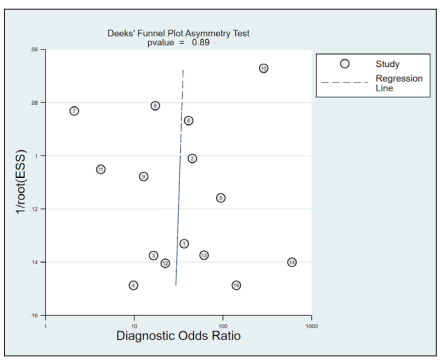


Figure 5. Deeks' funnel plots indicated no publication bias ($p = 0.48$).

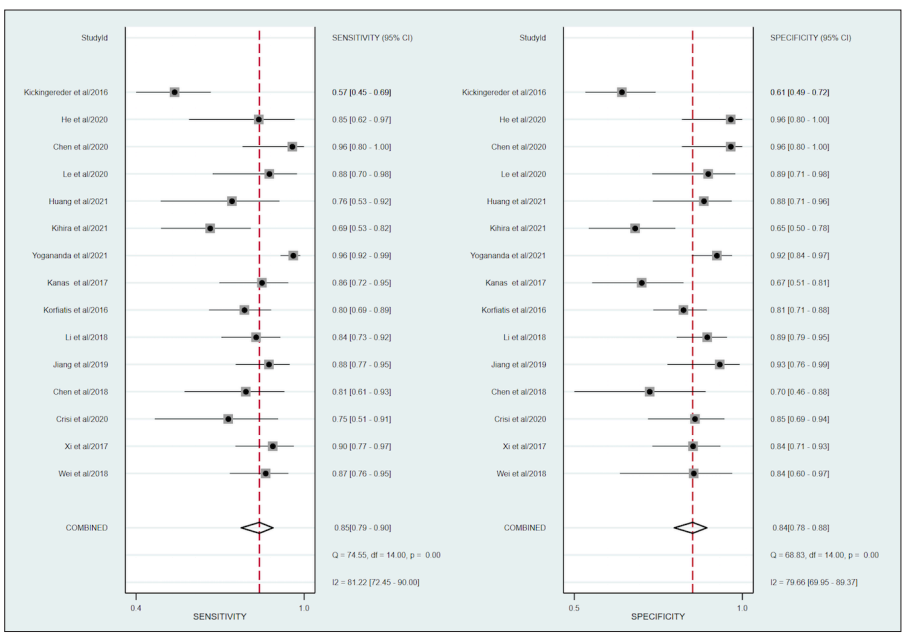


Figure 3. Coupled forest plots of the pooled sensitivity and specificity for the diagnostic efficacy of machine learning in the prediction of MGMT promoter methylation status.

perfusion sequence studies (the sensitivity and specificity ranged from 56.3% to 84.2% and 75.0% to 85.7%, respectively; Supplementary material 2) (35–38). Although some non-machine learning studies using ADC values showed good diagnostic efficacy in predicting MGMT promoter methylation status (sensitivity 81%–84%, specificity 82%–91%) (37, 39), the results of non-machine learning studies using ADC values remained controversial. Han et al. (37) reported that the ADC value from diffusion-weighted MRI of MGMT methylated glioblastomas was higher than that of unmethylated glioblastomas. However, Pope et al. (40) demonstrated that the ADC value of MGMT-methylated gliomas was lower than that of gliomas without MGMT methylation. In addition, a few studies (36, 41, 42) revealed that the ADC value was not significantly different between MGMT methylated and unmethylated gliomas. Furthermore, as machine learning

Table 2. Results of subgroup analysis

Analysis	No. of studies	Number of patients	Pooled sensitivity Value (95% CI)	Pooled specificity Value (95% CI)	Pooled PLR Value (95% CI)	Pooled NLR Value (95% CI)	Pooled DOR Value (95% CI)
Overall training group	15	1432	0.85 (0.79–0.90)	0.84 (0.78–0.88)	5.3 (3.7–7.6)	0.18 (0.12–0.26)	30 (14–62)
Magnet field strength (T)							
3 T	9	840	0.80 (0.76–0.84)	0.811 (0.77–0.85)	5.04 (2.73–9.3)	0.22 (0.13–0.38)	24.08 (8.45–68.60)
1.5 or 3.0T	2	246	0.76 (0.67–0.84)	0.75 (0.67–0.82)	2.89 (1.37–6.00)	0.34 (0.18–0.66)	8.62 (2.16–34.4)
NR	4	346	0.94 (0.90–0.96)	0.86 (0.80–0.90)	7.70 (2.41–24.63)	0.089 (0.04–0.23)	91.60 (14.7–570.8)
Glioma grade							
HGG	7	762	0.79 (0.75–0.84)	0.79 (0.74–0.83)	4.17 (2.38–7.31)	0.23 (0.12–0.43)	20.34 (6.73–61.48)
LGG/LGG and HGG	8	670	0.88 (0.85–0.91)	0.85 (0.80–0.89)	5.99 (2.97–12.01)	0.17 (0.09–0.31)	38.35 (11.51–127.78)
Blinding to reference standard							
Yes	10	996	0.87 (0.84–0.89)	0.83 (0.79–0.86)	5.09 (3.24–7.99)	0.18 (0.12–0.28)	31.00 (13.38–71.82)
Not explicit	5	436	0.77 (0.71–0.83)	0.78 (0.72–0.84)	4.76 (1.77–12.78)	0.21 (0.08–0.56)	24.83 (4.20–146.84)
Sample size of training group							
≥100	4	687	0.84 (0.80–0.88)	0.81 (0.76–0.85)	4.66 (1.68–12.93)	0.19 (0.06–0.63)	24.76 (3.33–184.15)
<100	11	745	0.85 (0.80–0.88)	0.82 (0.78–0.86)	4.88 (3.13–7.62)	0.20 (0.14–0.29)	27.74 (13.18–58.39)
Sequence							
CS MRI	8	910	0.90 (0.87–0.92)	0.86 (0.82–0.89)	6.39 (3.93–10.40)	0.13(0.08–0.21)	28.42 (15.13–53.38)
CS +advanced MRI	7	522	0.73 (0.68–0.79)	0.74 (0.69–0.80)	3.38 (1.96–5.81)	0.31(0.19–0.51)	54.06 (23.29–125.51)
Publication year							
Before 2019	7	590	0.80 (0.75–0.84)	0.77 (0.72–0.81)	3.56 (2.18–5.81)	0.24 (0.13–0.42)	15.62 (5.82–41.91)
After 2019	8	842	0.89 (0.85–0.92)	0.87 (0.83–0.90)	7.57 (3.48–16.47)	0.16 (0.08–0.31)	53.53 (14.39–199.09)

CI, confidence interval; PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio; NR, not reported; HGG, higher grade gliomas; LGG, lower grade gliomas; CS, conventional sequences; MRI, magnetic resonance imaging.

algorithms become more advanced and we learn more about the molecular subgroups of glioma, the ability of machine learning to predict the methylation status of the MGMT promoter has significantly improved. Of the included studies, those published in the last two years had better diagnostic efficacy than those published before that, with a summary sensitivity of 89%, specificity of 87% and AUC of 0.94 (95% CI 0.92–0.96; SE=0.027, $p < 0.01$). The results demonstrated that machine learning approaches showed a steadier and better diagnostic accuracy than these non-machine learning approaches. In addition, the highly predictive value of machine learning methods was also demonstrated in identifying other important biomarkers of gliomas (43–45). The diagnostic efficacy of machine learning would be further improved with increase in training data and further development of machine learning algorithms. Based on the results of the present meta-analysis, we cautiously recommend the inclusion of machine learning in daily radiology practice to improve the identification of MGMT promoter methylation status. However, further prospective validation studies are critical.

This meta-analysis revealed that machine learning studies utilizing conventional MRI sequences showed a higher diagnostic performance than advanced MRI sequences in predicting MGMT promoter methylation status. One possible reason for this discrepancy is that advanced MRI sequences may increase the number of redundant features that may reduce its diagnostic sensitivity. However, none of our included studies directly compared the diagnostic efficacy between advanced and conventional sequences, so this should be taken into consideration in future studies. Studies with sample sizes of less than 100 patients in the training group had slightly higher DOR than assessments performed with a larger sample (≥100). This could be attributed to the fact that a small sample size might lead to overfitting, which is equal to confounding or selection bias in machine learning (46–48). As a common problem in machine learning, overfitting leads to the machine learning algorithm becoming highly experienced in handling specific situations of the training set but lacking the capacity to successfully settle slightly differing cases except when it is trained utilizing a highly heterogeneous population (48, 49). Thus, to avoid overfitting, larger sample sizes, cross validation and multiple imaging methods should be involved in the proper training cohort, and a separate test cohort should also be considered (50, 51).

Significant heterogeneity was noted in the sensitivity and specificity in this meta-analysis. Although some reasons for the heterogeneity were explained by the meta-regression and sensitivity analysis, further underlying factors remain unexplained. Huang et al. (52) demonstrated that different machine learning algorithms would influence the diagnostic performance; however, because various machine learning algorithms were used in

our selected studies, we did not conduct a subgroup analysis of algorithms. Therefore, more studies are needed to explore the potential influencing factors of machine learning in identifying MGMT promoter methylation status in gliomas.

It is worth noting that this meta-analysis has a few limitations. First, all of the included studies were retrospective; thus, prospectively designed studies are needed to validate these findings. Second, obvious heterogeneity was observed in the sensitivity and specificity. To explain the influencing factors of the heterogeneity, we performed subgroup analysis, meta-regression, and sensitivity analysis. Third, ten of the included studies did not have a validation set, which might cause false high-performance results. Fourth, although five of our included studies simultaneously predicted the status of IDH mutation and MGMT methylation (21, 25, 31–33), all included studies graded gliomas without using the criteria of the WHO classification of 2016, which were based on genetic biomarkers such as IDH mutation and 1p/19q codeletion. In addition, only one study had a clinical integrated model (21), which is also a limitation because the inclusion of clinical characteristics might be conducive to improving diagnostic performance.

In conclusion, our meta-analysis demonstrated that machine learning is a promising and credible method to predict MGMT promoter methylation status in glioma and showed a higher performance than non-machine learning methods. However, more prospective and large sample size studies are required to verify the diagnostic efficacy and explore the most suitable MRI sequences and machine learning algorithms.

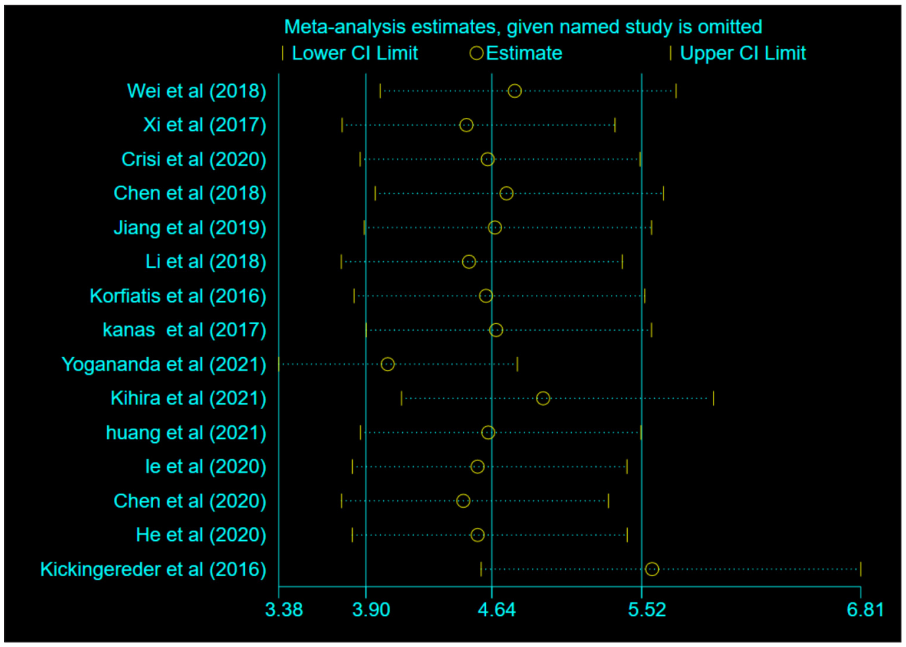
Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Lapointe S, Perry A and Butowski NA. Primary brain tumours in adults. Lancet 2018; 392:432–446. [Crossref]
2. Rivera AL, Pelloski CE, Gilbert MR, et al. MGMT promoter methylation is predictive of response to radiotherapy and prognostic in the absence of adjuvant alkylating chemotherapy for glioblastoma. Neuro Oncol 2010; 12:116–121. [Crossref]
3. Bell EH, Zhang P, Fisher BJ, et al. Association of MGMT promoter methylation status with survival outcomes in patients with high-risk glioma treated with radiotherapy and temozolomide: an analysis from the NRG oncology/RTOG 0424 trial. JAMA Oncol 2018; 4:1405–1409. [Crossref]
4. Everhard S, Kaloshi G, Criniere E, et al. MGMT methylation: a marker of response to temozolomide in low-grade gliomas. Ann Neurol 2006; 60:740–743. [Crossref]
5. Binabaj MM, Bahrami A, ShahidSales S, et al. The prognostic value of MGMT promoter methylation in glioblastoma: A meta-analysis of clinical trials. J Cell Physiol 2018; 233:378–386. [Crossref]
6. Bobustuc GC, Baker CH, Limaye A, et al. Levetiracetam enhances p53-mediated MGMT inhibition and sensitizes glioblastoma cells to temozolomide. Neuro Oncol 2010; 12:917–927. [Crossref]
7. Johnson DR, Guerin JB, Ruff MW, et al. Glioma response assessment: Classic pitfalls, novel confounders, and emerging imaging tools. Brit J Radiol 2019; 92. [Crossref]
8. 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; 26:2192–2197. [Crossref]
9. Li H, Li J, Cheng G, et al. IDH mutation and MGMT promoter methylation are associated with the pseudoprogression and improved prognosis of glioblastoma multiforme patients who have undergone concurrent and adjuvant temozolomide-based chemoradiotherapy. Clin Neurol Neurosurg 2016; 151:31–36. [Crossref]
10. Quillien V, Lavenu A, Sanson M, et al. Outcome-based determination of optimal pyrosequencing assay for MGMT methylation detection in glioblastoma patients. J Neurooncol 2014; 116: 487–496. [Crossref]
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016; 278:563–577. [Crossref]
12. Kocak B, Durmaz ES, Ates E, et al. Radiomics with artificial intelligence: a practical guide for beginners. Diagn Interv Radiol 2019; 25:485–495. [Crossref]
13. Erickson BJ, Korfiatis P, Akkus Z, et al. Machine learning for medical imaging. Radiographics 2017; 37:505–515. [Crossref]
14. Wang S, Summers RM. Machine learning and radiology. Med Image Anal 2012; 16:933–951. [Crossref]
15. Thrall JH, Li X, Li Q, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. J Am Coll Radiol 2018; 15:504–508. [Crossref]
16. Crisi G, Filice S. Predicting MGMT promoter methylation of glioblastoma from dynamic susceptibility contrast perfusion: a radiomic approach. J Neuroimaging 2020; 30:458–462. [Crossref]
17. Li ZC, 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; 7:5999–6009. [Crossref]
18. 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. [Crossref]
19. Li ZC, 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; 28:3640–3650. [Crossref]
20. Wei J, Yang G, Hao X, et al. A multi-sequence and habitat-based MRI radiomics signature for preoperative prediction of MGMT promoter methylation in astrocytomas with prognostic implication. Eur Radiol 2019; 29:877–888. [Crossref]
21. 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. [Crossref]
22. Xi YB, Guo F, Xu ZL, et al. Radiomics signature: A potential biomarker for the prediction of MGMT promoter methylation in glioblastoma. J Magn Reson Imaging 2018; 47:1380–1387. [Crossref]
23. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009; 6: e1000100. [Crossref]
24. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011; 155:529–536. [Crossref]
25. 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; 37:1775–1787. [Crossref]
26. Kanas VG, Zacharaki EI, Thomas GA, et al. Learning MRI-based classification models for MGMT methylation status prediction in glioblastoma. Comput Methods Programs Biomed 2017; 140:249–257. [Crossref]
27. Kickingereder P, Bonekamp D, Nowosielski M, et al. Radiogenomics of glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. Radiology 2016; 281:907–918. [Crossref]
28. Korfiatis P, Kline TL, Coufalova L, et al. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. Med Phys 2016; 43:2835–2844. [Crossref]
29. Chen X, Zeng M, Tong Y, et al. Automatic prediction of MGMT status in glioblastoma via deep learning-based MR image analysis. Biomed Res Int 2020; 2020:9258649. [Crossref]
30. He W, Li X, Hua J, et al. Noninvasive assessment of O(6)-methylguanine-DNA methyltransferase promoter methylation status in World Health Organization Grade II-IV glioma using histogram analysis of inflow-based vascular-space-occupancy combined with structural magnetic resonance imaging. J Magn Reson Imaging 2021; 54:227–236. [Crossref]
31. huang wy, wen lh, wu g, et al. comparison of radiomics analyses based on different magnetic resonance imaging sequences in grading and molecular genomic typing of glioma. J Comput Assist Tomogr 2021; 45:110–120. [Crossref]
32. Kihira S, Tsankova NM, Bauer A, et al. Multiparametric MRI texture analysis in prediction of glioma biomarker status: added value of MR diffusion. Neurooncol Adv 2021; 3:vdab051. [Crossref]

33. Le NQK, Do DT, Chiu FY, et al. XGBoost improves classification of MGMT promoter methylation status in IDH1 wildtype glioblastoma. J Pers Med 2020; 10:128. [Crossref]
34. Yogananda CGB, Shah BR, Nalawade SS, et al. MRI-based deep-learning method for determining glioma MGMT promoter methylation status. AJNR Am J Neuroradiol 2021; 42:845–852. [Crossref]
35. Zhang HW, Lyu GW, He WJ, et al. DSC and DCE histogram analyses of glioma biomarkers, including IDH, MGMT, and TERT, on differentiation and survival. Acad Radiol 2020; 27:e263–e271. [Crossref]
36. Ahn SS, Shin NY, Chang JH, et al. Prediction of methylguanine methyltransferase promoter methylation in glioblastoma using dynamic contrast-enhanced magnetic resonance and diffusion tensor imaging. J Neurosurg 2014; 121:367–373. [Crossref]
37. Han Y, Yan LF, Wang XB, et al. Structural and advanced imaging in predicting MGMT promoter methylation of primary glioblastoma: a region of interest based analysis. BMC Cancer 2018; 18:215. [Crossref]
38. Ryoo I, Choi SH, Kim JH, et al. Cerebral blood volume calculated by dynamic susceptibility contrast-enhanced perfusion MR imaging: preliminary correlation study with glioblastoma genetic profiles. PLoS One 2013; 8:e71704. [Crossref]
39. Romano A, Calabria LF, Tavanti F, et al. Apparent diffusion coefficient obtained by magnetic resonance imaging as a prognostic marker in glioblastomas: correlation with MGMT promoter methylation status. Eur Radiol 2013; 23:513–520. [Crossref]
40. Pope WB, Lai A, Mehta R, et al. Apparent diffusion coefficient histogram analysis stratifies progression-free survival in newly diagnosed bevacizumab-treated glioblastoma. AJNR Am J Neuroradiol 2011; 32:882–889. [Crossref]
41. Gupta A, Prager A, Young RJ, et al. Diffusion-weighted MR imaging and MGMT methylation status in glioblastoma: a reappraisal of the role of preoperative quantitative ADC measurements. AJNR Am J Neuroradiol 2013; 34:E10–11. [Crossref]
42. Kanazawa T, Minami Y, Takahashi H, et al. Magnetic resonance imaging texture analyses in lower-grade gliomas with a commercially available software: correlation of apparent diffusion coefficient and T2 skewness with 1p/19q codeletion. Neurosurg Rev 2020; 43:1211–1219. [Crossref]
43. Wu S, Meng J, Yu Q, et al. Radiomics-based machine learning methods for isocitrate dehydrogenase genotype prediction of diffuse gliomas. J Cancer Res Clin Oncol 2019; 145: 543–550. [Crossref]
44. 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; 49:808–817. [Crossref]
45. Shofty B, Artzi M, Ben Bashat D, et al. MRI radiomics analysis of molecular alterations in low-grade gliomas. Int J Comput Assist Radiol Surg 2018; 13:563–571. [Crossref]
46. Subramanian J, Simon R. Overfitting in prediction models - is it a problem only in high dimensions? Contemp Clin Trials 2013; 36:636–641. [Crossref]
47. Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci 2004; 44:1–12. [Crossref]
48. Lemm S, Blankertz B, Dickhaus T, et al. Introduction to machine learning for brain imaging. Neuroimage 2011; 56:387–399. [Crossref]
49. Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. Biomed Eng Online 2014; 13:94. [Crossref]
50. Erickson BJ, Korfiatis P, Kline TL, et al. Deep learning in radiology: does one size fit all? J Am Coll Radiol 2018; 15:521–526. [Crossref]
51. Kohli M, Prevedello LM, Filice RW, et al. Implementing machine learning in radiology practice and research. AJR Am J Roentgenol 2017; 208:754–760. [Crossref]
52. Huang J, Tian W, Zhang L, et al. Preoperative prediction power of imaging methods for microvascular invasion in hepatocellular carcinoma: a systemic review and meta-analysis. Front Oncol 2020; 10:887. [Crossref]



Supplementary Material 1. Result of sensitivity analysis.

Supplementary Material 2

With a systematic literature search, we included four non-machine learning studies using perfusion sequence studies, and the sensitivity and specificity of the individual studies ranged from 56.3% to 84.2% and 75.0% to 85.7%, respectively. Six non-machine learning studies using ADC values were found, but among them, only two studies demonstrated that the ADC value was significantly different between MGMT-methylated and unmethylated gliomas (sensitivity ranged from 81% to 84% and specificity 82% to 91%). Thus, we only performed a systematic comparison in machine learning studies and non-machine learning studies using perfusion sequences in predicting MGMT status. The results are shown in the following table.

Method	Pooled sensitivity	Pooled specificity	Pooled AUC
Machine learning studies	0.85 (0. 79-0.90)	0.84 (0.78-0.88)	0.91 (0.88-0.93)
Non-machine learning studies using perfusion sequences	0.74 (0.64-0.82)	0.81 (0.71-0.87)	0.84 (0.81-0.87)