Author/s:
Bhandari, AP;Liong, R;Koppen, J;Murthy, SV;Lasocki, A

Title:
Noninvasive determination of IDH and 1p19q status of lower-grade gliomas using MRI radiomics: A systematic review

Date:
2021-01-01

Citation:
Bhandari, A. P., Liong, R., Koppen, J., Murthy, S. V. & Lasocki, A. (2021). Noninvasive determination of IDH and 1p19q status of lower-grade gliomas using MRI radiomics: A systematic review. American Journal of Neuroradiology, 42 (1), pp.94-101. https://doi.org/10.3174/ajnr.A6875.

Persistent Link:
https://hdl.handle.net/11343/280210

Noninvasive Determination of *IDH* and 1p19q Status of Lower-grade Gliomas Using MRI Radiomics: A Systematic Review

A.P. Bhandari, R. Liong, J. Koppen, S.V. Murthy, and A. Lasocki

ABSTRACT

BACKGROUND: Determination of *isocitrate dehydrogenase* (*IDH*) status and, if *IDH*-mutant, assessing 1p19q codeletion are an important component of diagnosis of World Health Organization grades II/III or lower-grade gliomas. This has led to research into noninvasively correlating imaging features ("radiomics") with genetic status.

PURPOSE: Our aim was to perform a diagnostic test accuracy systematic review for classifying *IDH* and 1p19q status using MR imaging radiomics, to provide future directions for integration into clinical radiology.

DATA SOURCES: Ovid (MEDLINE), Scopus, and the Web of Science were searched in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Diagnostic Test Accuracy guidelines.

STUDY SELECTION: Fourteen journal articles were selected that included 1655 lower-grade gliomas classified by their *IDH* and/or 1p19q status from MR imaging radiomic features.

DATA ANALYSIS: For each article, the classification of *IDH* and/or 1p19q status using MR imaging radiomics was evaluated using the area under curve or descriptive statistics. Quality assessment was performed with the Quality Assessment of Diagnostic Accuracy Studies 2 tool and the radiomics quality score.

DATA SYNTHESIS: The best classifier of *IDH* status was with conventional radiomics in combination with convolutional neural network–derived features (area under the curve = 0.95, 94.4% sensitivity, 86.7% specificity). Optimal classification of 1p19q status occurred with texture-based radiomics (area under the curve = 0.96, 90% sensitivity, 89% specificity).

LIMITATIONS: A meta-analysis showed high heterogeneity due to the uniqueness of radiomic pipelines.

CONCLUSIONS: Radiogenomics is a potential alternative to standard invasive biopsy techniques for determination of *IDH* and 1p19q status in lower-grade gliomas but requires translational research for clinical uptake.

ABBREVIATIONS: AI = artificial intelligence; AUC = area under the curve; CNN = convolutional neural network; *IDH* = *isocitrate dehydrogenase*; *IDH*-mut = *IDH*-mutant; LGG = lower-grade gliomas; ML = machine learning; PRISMA-DTA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Diagnostic Test Accuracy; QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies 2; RQS = radiomics quality score; SVM = support vector machine; VASARI = Visually Accessible Rembrandt Images; WHO = World Health Organization

Lower-grade gliomas (LGG), World Health Organization (WHO) grades II/III, are diffusely infiltrative tumors of the CNS. With time, these tumors typically progress to glioblastoma (WHO grade IV), which has a median survival of only 12–18 months despite treatment.[1] A growing understanding of the prognostic and therapeutic importance of molecular markers has led to their incorporation into the 2016 WHO classification, and they now constitute a key component of the diagnosis of LGG.[2] The 2 key markers of LGG are *isocitrate dehydrogenase* (*IDH*), with tumors classified as either *IDH*-mutant (*IDH*-mut) or *IDH*-

Received May 2, 2020; accepted after revision August 17.

From the Department of Anatomy (A.P.B.) and College of Medicine and Dentistry (S.V.M.), James Cook University, Townsville, Queensland, Australia; Townsville University Hospital (A.P.B., J.K.), Douglas, Queensland, Australia; Department of Medical Imaging Research Office (R.L.), Royal Brisbane and Women's Hospital, Herston, Queensland, Australia; Department of Cancer Imaging (A.L.), Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia; Griffith University School of Medicine (R.L.), Gold Coast, Queensland, Australia; and Sir Peter MacCallum Department of Oncology (A.L.), The University of Melbourne, Melbourne, Victoria, Australia.

Please address correspondence to Abhishta Bhandari, MBBS MMI, Department of Anatomy, 1 James Cook Dr, James Cook University, Douglas, Townsville, 4814, Queensland, Australia; e-mail: Abhishta.bhandari@my.jcu.edu.au

Indicates open access to non-subscribers at www.ajnr.org
Indicates article with supplemental on-line tables.
Indicates article with supplemental on-line photos.
http://dx.doi.org/10.3174/ajnr.A6875

wild-type, and 1p19q, with 1p19q-codeletion representing a combined loss of both the short arm of chromosome 1 and the long arm of chromosome 19.

Determining *IDH* and 1p19q status is invasive, requiring a tissue specimen via stereotactic biopsy or definitive resection, with the associated operative risks[3] and possibility of sampling error. While the possibility of sampling error is perhaps of greatest relevance to the determination of tumor grade,[4] it is also relevant to the determination of tumor genetic status.[5,6] For example, *IDH* sequencing may be falsely negative if there are few glioma cells within the sample,[5] and intratumoral genetic heterogeneity can occur.[6] These considerations have led to research into characterizing *IDH* and 1p19q status by imaging, known as "radiogenomics" or "imaging genomics." The most specific visual MR imaging feature is the "T2-FLAIR mismatch sign," which has been shown to predict an *IDH*-mut 1p19q-codeletion gliomas with 100% specificity and high interobserver correlation ($\kappa = 0.38–0.88$).[7-9] Other useful features include the presence of calcification (suggestive of a 1p19q-codeletion glioma)[8,9] and homogeneous signal (likely 1p19q-intact).[10] While some features such as >50% T2-FLAIR mismatch and the presence of calcification have high interobserver correlation, other features are limited by greater variability in interpretation. Furthermore, a substantial proportion (29%–37%) of gliomas do not exhibit these features, limiting sensitivity.[8]

Artificial intelligence (AI) is emerging as a solution to the limitations of conventional visual assessment. AI techniques may identify features hidden to the naked eye by extracting data from images and relating them to outcomes. Given the inherent signal and volume heterogeneity of gliomas, a perceived signature or pattern may be modelled to genetic, clinical, and biochemical outcomes.[11] Features can be learned from the image or predefined. The field of radiomics involves the extraction of predefined features such as shape, intensity, and texture from a segmented (tumor) volume of interest.[12] This is opposed to deep learning–derived features, which are identified without human predefinition. Radiomic features can be correlated with genetic status through a subset of AI known as machine learning (ML). The ML algorithm is trained to a clinical outcome via a training dataset and validated using a testing/validation dataset. Extracted radiomic features undergo selection and can then be related to molecular markers such as *IDH* and 1p19q, providing a more objective method of radiogenomic correlation.

Radiomic analysis has several advantages compared with human observers, including the ability to rapidly assess multiple imaging features, less interobserver variability,[13] and potentially higher sensitivity and specificity. The aim of this article was to perform a systemic review of the use of MR imaging radiomics for the classification of *IDH* and 1p19q status in LGG.

MATERIALS AND METHODS

Search methodology and study synthesis were performed in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Diagnostic Test Accuracy (PRISMA-DTA) checklist.[14] The search was performed on the Web of Science, Ovid (MEDLINE), and Scopus on April 18, 2020. Online Table 1 summarizes the search strategy. Search terms were developed from the PICO framework and Medical Subject Headings, which included terms relating to radiomics or radiogenomics, gliomas, and *IDH*/1p19q status. The PRISMA flowchart is available in Online Fig 1.

Study Selection

Studies were included if they were original research articles relating radiomic features to *IDH* and/or 1p19q status in LGG (WHO grades II/III) with pathologic confirmation. Studies were excluded under the following circumstances: 1) They investigated the effects of radiogenomic pipelines on factors that affect imaging quality rather than assessing diagnostic potential or 2) they included imaging modalities other than MR imaging because recent literature has not shown superior outcomes.[15] There was no restriction on study date.

The references were imported from the Web of Science, Ovid (MEDLINE), and Scopus into EndNote (Version X9; https://www.endnote.com/product-details/). Duplicates were removed using the "Find Duplicates" function in EndNote and manual review of the reference list. Two independent authors (A.P.B. and J.K.) screened the titles and abstracts for eligibility. The full texts were then screened. When questions arose regarding inclusion of articles, these were resolved through discussion between both authors responsible for data extraction (experience: A.P.B., medical doctor with a master's degree in medical imaging analysis, and J.K., medical doctor with 4 years' clinical experience). Ties were to be reviewed together with the senior author, but none were encountered.

Data Collection and Analysis

The primary outcome was the classification of *IDH* and/or 1p19q status by MR imaging radiomics. This was based on the receiver operating or precision recall curve and associated sensitivity (%), specificity (%), and area under the curve (AUC) if available. The AUC is presented as a value between 0.5 and 1, with 1 representing perfect classification (and 100% sensitivity and specificity). For studies that did not include ML in the pipeline, descriptive statistics (for example, mean and SD with *t* testing) were also included. Only significant findings for descriptive statistics were reported or the highest AUC for ML classifiers, given that some studies related numerous radiomic features to genetic status (*IDH* and/or 1p19q) or reported a considerable number of ML classifiers. If training and validation set data were reported, only the validation set was used. Secondary outcome measures were related to pipeline features and included the number of lesions, imaging sequences and segmentation method, features and their selection method, ML classifier, genetic status, and WHO tumor grade. A meta-analysis using random effects[16] was performed on AUC values with 95% confidence intervals when available in MedCalc (MedCalc Software). A Higgins I^2 index of heterogeneity was reported, in which 0% represents no heterogeneity and 100% represents maximum heterogeneity.

Quality Assessment

Quality assessment was performed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool and the radiomics quality score (RQS).[12] The QUADAS-2 scoring system

Table 1: Derived aims and key findings of studies comparing *IDH*-mut and *IDH* wild-type LGG

First Author and Year	Derived Aim	Key Findings
Fukuma 2019[22]	To integrate CNN deep learning features with conventional radiomic features	Conventional radiomic features: accuracy (mean ± 95% CI) = 71.7% ± 8.3%; AUC (± 95% CI) = 0.718 ± 0.139 CNN features: accuracy = 69.6% ± 5.6%; AUC = 0.619 ± 0.132 CNN and conventional radiomic features: accuracy = 73.1% ± 9.4%; AUC = 0.699 ± 0.145
Gihr 2020[23]	To determine if intensity features relate to *IDH* status	Entropy, a second-order histogram parameter of the ADC volume was significant: *IDH*-mut versus *IDH* wild-type, mean ± SD = 5.5 ± 0.63 vs 4.75 ± 0.69; $P = .0144$
Jakola 2018[24]	To determine if texture features can predict *IDH* status on FLAIR	Homogeneity and volume could classify *IDH* status with an AUC = 0.940 (85% sensitivity, 100% specificity) using the generalized linear model
Kim 2020[25]	To determine if DWI- and DSC perfusion-based image integration with standard imaging (T1WI postcontrast and FLAIR) can improve classification	Integration increased the AUC (95% CI) = 0.747 (0.663–0.832); (53.6% sensitivity and 86.7% specificity) from 0.705 (0.613–0.796) (43.9% sensitivity and 88.8% specificity) compared with conventional MR imaging radiomics
Li 2017[27]	To determine if integration of deep learning features into the radiogenomic pipeline improves classification	Conventional radiomics produced an AUC = 0.85 (sensitivity of 82.9%, specificity of 73.5%) CNN deep learning–derived features plus conventional radiomic features with feature selection produced an AUC = 0.95 (sensitivity of 94.4%; specificity of 86.7%)
Lu 2018[28]	To determine the best ML classifier	Linear SVM classified *IDH* status with an AUC = 0.936 (sensitivity of 85.7%, specificity of 93.0%)
Park 2020[29]	To determine if DTI improves classification when added to conventional radiomics	Addition of DTI radiomic features to conventional imaging radiomics increased the AUC (95% CI) = 0.900 (0.855–0.945) from 0.835 (0.773–0.896)
Ren 2019[30]	To compare radiomic, VASARI, and radiomic plus VASARI features derived from FLAIR, ADC, eADC, and CBF	Radiomics: AUC (95% CI) = 0.931 (0.842–1); sensitivity of 100%, specificity of 85.71% VASARI: AUC = 0.843 (sensitivity of 91.67%; specificity of 61.90%) Radiomics plus VASARI: AUC = 0.888 (0.786–0.989); sensitivity of 94.44% and specificity of 71.43%
Yu 2017[32]	To classify using the improved genetic algorithm for feature selection and leave-one-out cross-validation method in WHO grade II LGG	Using the proposed method and the SVM ML classifier, an AUC = 0.71 (sensitivity = 56% and specificity = 74%) was achieved
Zhou 2017[34]	To determine if VASARI annotations were superior to standard radiomic classification analysis	*IDH* classification through texture features found an AUC (± 95% CI) = 0.79 ± 0.02; sensitivity 90%, specificity of 89% *IDH* classification through VASARI features, AUC = 0.73 ± 0.02; sensitivity of 69%, specificity of 69%
Zhang 2018[33]	To classify by conventional radiomics	AUC = 0.830 (sensitivity = 82%, specificity = 92%) using SVM

Note:—eADC indicates exponential ADC.

was developed to assess bias and the applicability of diagnostic-accuracy studies.[17] The RQS is specific to radiomics and is based on the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis initiative, which examines domains of application for predictive models.[18] Application of RQS and QUADAS-2 was performed by discussion between A.P.B. and J.K. A κ statistic[19] was considered for the RQS, similar to that used in previous studies;[20] however, for quantitatively-defined criteria, it was determined that resolution by discussion would be superior.[21]

RESULTS

The initial search obtained 610 articles; 431 articles were from Ovid (MEDLINE); 111, from Scopus; and 68, from the Web of Science. After duplicates were removed, a total of 532 articles

Table 2: Derived aims and key findings of studies examining 1p19q status of *IDH*-mut LGG

First Author and Year	Derived Aim	Key Findings
Han 2020[35]	To determine if clinical and standard imaging factors improve classification	The AUC (95% CI) = 0.753 (0.654–0.852) for clinical plus radiomic features versus AUC = 0.760 (0.663-0.857) for just radiomic features; radiomic features were superior to clinical features alone, AUC = 0.627 (0.551–0.703)
Kocak 2020[26]	To determine the best ML classifier	The neural network produced the highest AUC (95% CI) = 0.869 (0.751–0.981); sensitivity of 87.5%, specificity of 75.8%
Lu 2018[28]	To determine the best ML classifier	Classification occurred with an AUC = 0.92 (sensitivity of 88.5%, specificity of 86.2%) using quadratic SVM
Shofty 2018[31]	To determine the best ML classifier	Classification occurred with an AUC = 0.87 (sensitivity of 92%, specificity of 83%) using ensemble bagged trees classifier
Zhou 2017[34]	To determine if VASARI annotations were superior to standard radiomic analysis for classification	Texture features classified with an AUC (± 95% CI) = 0.96 ± 0.01; sensitivity of 90% ± 2%, specificity of 89% ± 2% VASARI features classified with an AUC = 0.78 ± 0.02; sensitivity of 72% ± 3%, specificity of 67% ±3%
Fukuma 2019[22]	To determine if integration of CNN deep learning with radiomic features improved classification	Conventional radiomic features (± 95% CI): accuracy = 59.0 ± 9.0%; AUC = 0.656 ± 0.113 CNN features: accuracy = 84.0 ± 9.3%; AUC = 0.868 ± 0.099 CNN and conventional radiomic features: accuracy = 79.8 ± 11.0%; AUC = 0.861 ± 0.116

remained. The articles were screened by title and abstract, and 18 remained. Full texts were reviewed, and 14 articles[22-35] fit the review question and inclusion criteria. The publication dates of the 14 included studies[22-35] ranged from 2017 to 2020. A total of 1655 LGG were analyzed. Online Table 2 summarizes the pipeline features for each study.

All segmentations incorporated manual components except for 2 studies, both of which used convolutional neural network (CNN)-based segmentation.[28,32] Standard imaging sequences included pre- and postcontrast T1WI, T2WI, and FLAIR. ADC,[23,25,30] cerebral blood flow/volume,[28,30] DTI,[29] and exponential ADC[30] were used as adjuncts in some studies. Radiomic features were extracted most commonly by programs developed in-house on the Matlab software platform (MathWorks).[23,25,27] AlexNet (https://www.mygreatlearning.com/blog/alexnet-the-first-cnn-to-win-image-net/) was used in 1 study for deep learning–derived features in the highest discriminating pipeline.[27] The most common method of feature selection was support vector machine (SVM)–recursive feature elimination,[25,30,33,34] followed by a Student *t* test.[27-29] All categories of radiomic features were used. Two studies did not use ML.[23,24] Most studies assessed WHO grade II and III LGG,[22-30,32-34] apart from one that assessed only WHO grade II LGG.[31] Table 1 demonstrates the derived aims and key findings of studies that examined the *IDH* status of LGG, while Table 2 summarizes studies examining the 1p19q status of *IDH*-mut LGG. Figures 1 and 2 provide the associated forest plots for studies assessing *IDH* and 1p19q, respectively. Further details are provided in the online material. A meta-analysis on *IDH* status was performed on 5 studies[22,26,29,30,34] that had sufficient data with a pooled value of 0.827 (95% CI, 0.760–0.894; I^2 = 88.55%). For 1p19q status, a meta-analysis was performed on 4 studies[22,26,34,35] that had sufficient data with a pooled value of 0.872 (95% CI, 0.789–0.954; I^2 = 86.19%).

The QUADAS-2 score showed low bias and high applicability (see Online Fig 2 for individual studies). The radiomic-specific RQS average score was low, with a mean of 10 (range, 2–14). On average, the RQS was 29% (range, 6%–39%) of the highest possible score. There were no studies that reported on cost-effectiveness, imaging used on phantom models, a prospectively validated radiomic signature in an appropriate clinical trial or performed clinical utility statistics (beyond just discussion of uses).[21] Further details are provided in Online Table 3.

DISCUSSION

The systemic literature review found that the highest classifier for *IDH* status was conventional radiomics with CNN deep learning–derived features, which achieved an AUC = 0.95 (sensitivity of 94.4%, specificity of 86.7%).[27] For classification of 1p19q status, conventional texture-based radiomics was optimal, with an AUC = 0.96 (sensitivity of 90%, specificity of 89%).[34]

Segmentation had manual components in both studies[28,32] and was generally performed by trained personnel and approved by neuroradiologists or neurosurgeons. Manual segmentation is time-consuming, resource-intensive and introduces interobserver variability. Automation of segmentation is being actively progressed by the Brain Tumour Segmentation Challenge, and ongoing improvements have the potential to address the limitations of manual segmentation and thus improve the accuracy and efficiency of radiomic methods.[36-38] For the whole tumor, the 2018 winning team achieved a Sørensen–Dice coefficient of 0.88, in which a value of 1 represents perfect consistency between manual (ground truth) and automated segmentation.[39]

For *IDH* status, the literature indicates that a standard sequence image acquisition, use of texture-based features (most common being gray-level co-occurrence matrix,[23-25,27-29,34,40]

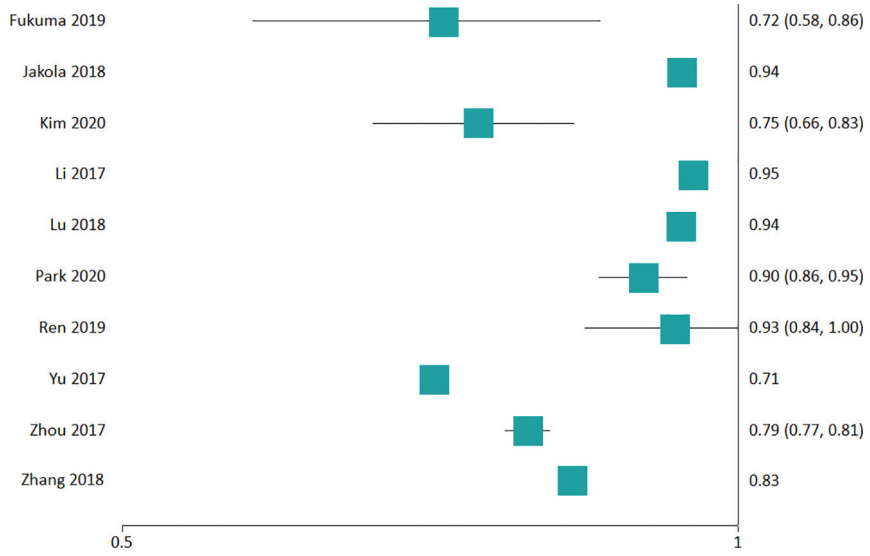

FIG 1. *IDH* status forest plot of included studies with an AUC.

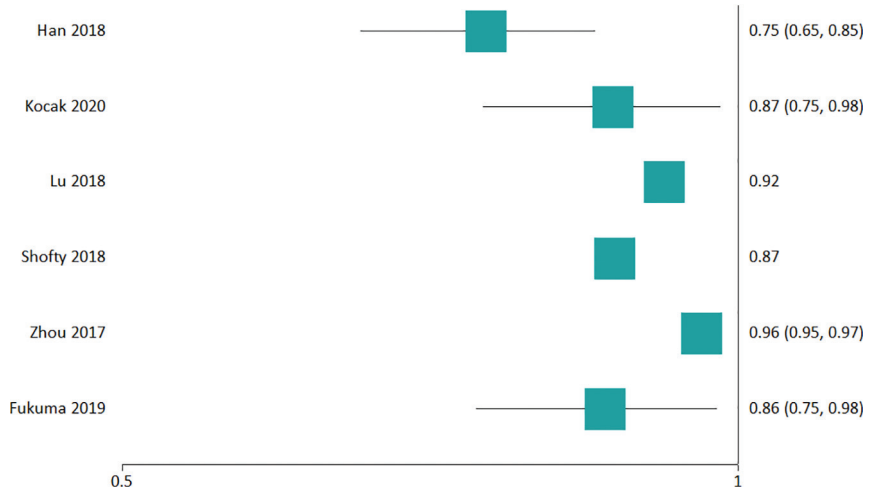

FIG 2. 1p19q status forest plot of included studies with an AUC.

followed by the gray-level run-length matrix[25,27-30,34,40]) with deep learning–derived features, and an SVM machine learning model may result in an optimal radiomic pipeline. One study classified solely using texture-based radiomic features and achieved an AUC = 0.79.[34] Integration of deep learning with radiomic features did not increase the AUC in 1 study[22] but produced the highest AUC = 0.95 in another study.[27] Features derived from qualitative visual inspection (Visually Accessible Rembrandt Images; VASARI) did not increase the AUC compared with just radiomic features.[34] Four studies examined multiparametric imaging.[23,25,29,30] The entropy (randomness of voxel intensities) feature derived from ADC images was significantly different between *IDH*-mut and *IDH* wild-type LGG,[23] suggesting that heterogeneity of ADC values may be helpful in predicting *IDH* status. Nevertheless, while integration of diffusion/perfusion imaging showed improved classification in 3 studies,[25,29,30] ultimately it was not superior to using standard sequences with a different radiogenomic pipeline.[28]

For 1p19q status, the literature indicates that standard image sequences, use of texture-based features (the most common being grey-level run length matrix[26,28,34,35] followed by gray-level co-occurrence matrix[26,28,34]), and a linear SVM machine learning model may result in an optimal radiomic pipeline. The highest AUC = 0.96 was achieved solely using texture-based radiomic features.[34] Clinical and imaging-feature (such as age, sex, and the presence of bleeding or enhancement) integration did not improve the classification performance,[38] nor did solely examining visually-created features.[34] Deep learning feature integration with radiomic features increased classification performance; however, solely examining deep features was superior.[22] The best-performing ML model classifier was achieved by a linear SVM.[28]

For studies included in the meta-analysis, there was high heterogeneity, given the variation in the unique elements of each radiomic pipeline. Heterogeneity is inevitable with any meta-analysis; however, acceptable levels may be a Higgins I^2 of 0%–40%.[41] The meta-analysis found 88.55% and 86.19% heterogeneity for *IDH* and 1p19q status, respectively. Although the QUADAS-2 showed low bias and high applicability, the radiomic-specific RQS assessment showed an overall inadequate clinical applicability of studies, identifying issues, including a lack of cost-effectiveness analysis, clinical utility statistics, or prospective validation. This is consistent with other neuro-oncologic radiomic studies in the literature.[21] The RQS has some limitations, however. For example, greater emphasis is placed on the image-acquisition parameters[12] than on the image-normalization process (making the voxel, section thickness, and matrix size similar among MR imaging scans), despite the latter being important for optimal translation into multi-institutional contexts. Of note, a perceived advantage of the AI algorithms is greater objectivity and thus a more consistent diagnosis, but this has yet to be convincingly proved in the literature.[42]

Classification of LGG for *IDH* status followed by further classification of 1p19q status (when *IDH*-mut) will have multiplicative

effects. There was a sensitivity of 94.4% and specificity of 86.7%[27] for *IDH* status, with a sensitivity of 90% and specificity of 89%[34] for 1p19q. Thus, by using multiplication, we can find the maximum literature prediction of 1p19q status in an *IDH*-mut LGG to have a sensitivity of 85.0% (94.4% × 90%) and a specificity of 77.2% (86.7% × 89%). The conventional radiogenomic pipelines assume that the features assessed are independent, though they are not. For example, to take an example from the visual-feature literature, ill-defined tumor margins have been correlated with *IDH* wild-type LGG,[43] but if the tumor is *IDH*-mut, it is more likely 1p19q-codeletion.[10] There is also uncertainty regarding the interaction between radiomic and conventional visual MR imaging features. For example, if the T2-FLAIR mismatch sign is present, the literature would suggest that this can predict an *IDH*-mut 1p19q-intact glioma with greater confidence than radiomics.[8,44] Yet, when these conventional features with the greatest predictive value are absent, one could expect that radiomics would predict the genotype better than other conventional MR imaging features. Thus, optimal classification may be achieved using a combination of conventional and radiomic features.

Acceptance of AI into clinical practice remains an issue. Much of the literature on integration is opinion-based,[45-48] and research related to understanding challenges is in its early stages.[49-51] Acceptance by patients also remains an issue; a recent study by Palmisciano et al[52] found that only 66.3% of patients found it acceptable for AI to be used during imaging interpration.[53] Issues raised by patients include distrust, lack of knowledge, a lack of personal interaction, questions about the efficacy of the AI algorithm, and the importance of being properly informed of its uses.[54] Similar relevant issues were identified by a computer science literature review[55] on human-AI interaction, such as task allocation, lack of knowledge and/or trust, incorrect use due to confusion, and integration issues due to a potentially radically different work practice.

Future directions for integration into the clinical sphere may come in the form of examining the nonmedical sphere,[55] given successful implementation in other fields such as failure detection in truck engines and welding robots.[56] One specific issue is that some AI programs used were developed in-house and may not be readily available to other institutions; important next steps include comparisons between programs and subsequent validation on larger external cohorts. There is also a lack of clinical trials assessing the integration of radiomic analysis into clinical practice,[44] which was confirmed on our RQS assessment. Guidelines have recently been developed to address these issues, which may provide a framework for integration. For example, Microsoft has recently released a set of 18 general principles for integration into systems, such as explaining to the user (clinician) what the AI algorithm can do, how well it can be done and making it clear how it is performed.[57] A thinking paradigm that may solve this is treating radiomic analysis as a new intervention or drug and applying ideas from existing protocols such as Phase I–IV clinical trials.[58] The Food and Drug Administration has also recently released guidelines for AI integration into health care systems.[59] Given that radiomic analysis is rapidly progressing and combining AI with standard radiologist assessment may show superior outcomes, there needs to be greater effort to translate findings into an interpretable format for clinical radiology.

CONCLUSIONS

The greatest classifier of *IDH* status in LGG was achieved with conventional radiomics in combination with convolutional neural network–derived features, providing a sensitivity of 94.4% and specificity of 86.7% (AUC = 0.95). Optimal classification of 1p19q status occurred using texture-based radiomics, with a sensitivity of 90% and a specificity of 89% (AUC = 0.96). The literature is limited by the use of manual segmentation, suboptimal study design, and the lack of translational work to integrate radiogenomic analysis into clinical practice.

Disclosures: Arian Lasocki—*RELATED*: *Grant*: Peter MacCallum Cancer Foundation, *Comments*: Arian Lasocki was supported by a Peter MacCallum Cancer Foundation Discovery Partner Fellowship, providing clinical backfill to allow dedicated research time.* *Money paid to the institution.

REFERENCES

1. Stark AM, van de Bergh J, Hedderich J, et al. **Glioblastoma: clinical characteristics, prognostic factors and survival in 492 patients.** *Clin Neurol Neurosurg* 2012;114:840–45 CrossRef Medline
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;131:803–20 CrossRef Medline
3. Akay A, Rüksen M, Islekel S. **Magnetic resonance imaging-guided stereotactic biopsy: a review of 83 cases with outcomes.** *Asian J Neurosurg* 2019;14:90–95 CrossRef Medline
4. Lasocki A, Tsui A, Tacey MA, et al. **MRI grading versus histology: predicting survival of World Health Organization grade II-IV astrocytomas.** *AJNR Am J Neuroradiol* 2015;36:77–83 CrossRef Medline
5. Horbinski C. **What do we know about IDH 1/2 mutations so far, and how do we use it?** *Acta Neuropathol* 2013;125:621–36 CrossRef Medline
6. Preusser M, Wöhrer A, Stary S, et al. **Value and limitations of immunohistochemistry and gene sequencing for detection of the IDH 1-R132H mutation in diffuse glioma biopsy specimens.** *J Neuropathol Exp Neurol* 2011;70:715–23 CrossRef Medline
7. Broen MPG, Smits M, Wijnenga MM, et al. **The T2-FLAIR mismatch sign as an imaging marker for non-enhancing IDH-mutant, 1p/19q-intact lower-grade glioma: a validation study.** *Neuro Oncol* 2018;20:1393–99 CrossRef Medline
8. Lasocki A, Gaillard F, Gorelik A, et al. **MRI features can predict 1p/19q status in intracranial gliomas.** *AJNR Am J Neuroradiol* 2018;39:687–92 CrossRef Medline
9. Patel SH, Poisson LM, Brat DJ, et al. **T2-FLAIR mismatch, an imaging biomarker for IDH and 1p/19q status in lower-grade gliomas: a TCGA/TCIA Project.** *Clin Cancer Res* 2017;23:6078–85 CrossRef Medline
10. Johnson DR, Diehn FE, Giannini C, et al. **Genetically defined oligodendroglioma is characterized by indistinct tumor borders at MRI.** *AJNR Am J Neuroradiol* 2017;38:678–84 CrossRef Medline
11. Gillies RJ, Kinahan PE, Hricak H. **Radiomics: images are more than pictures: they are data.** *Radiology* 2016;278:563–77 CrossRef Medline
12. Lambin P, Leijenaar RT, Deist TM, et al. **Radiomics: the bridge between medical imaging and personalized medicine.** *Nat Rev Clin Oncol* 2017;14:749–62 CrossRef Medline
13. van den Bent MJ. **Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective.** *Acta Neuropathol* 2010;120:297–304 CrossRef Medline

14. McInnes MD, Moher D, Thombs BD, et al; PRISMA-DTA Group. **Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: the PRISMA-DTA Statement.** *JAMA* 2018;319:388–96 CrossRef Medline
15. Matsui Y, Maruyama T, Nitta M, et al. **Prediction of lower-grade glioma molecular subtypes using deep learning.** *J Neurooncol* 2020;146:321–27 CrossRef Medline
16. Zhou XH, McClish DK, Obuchowski NA, et al. *Statistical Methods in Diagnostic Medicine*. John Wiley & Sons; 2011
17. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. **QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.** *Ann Intern Med* 2011;155:529–36 CrossRef Medline
18. Collins GS, Reitsma JB, Altman DG, et al. **Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD).** *Ann Intern Med* 2015;162:735–36 CrossRef Medline
19. McHugh ML. **Interrater reliability: the kappa statistic.** *Biochem Med (Zagreb)* 2012;22:276–82 CrossRef Medline
20. Ursprung S, Beer L, Bruining A, et al. **Radiomics of computed tomography and magnetic resonance imaging in renal cell carcinoma-a systematic review and meta-analysis.** *Eur Radiol* 2020;30:3558–66 CrossRef Medline
21. Park JE, Kim HS, Kim D, et al. **A systematic review reporting quality of radiomics research in neuro-oncology: toward clinical utility and quality improvement using high-dimensional imaging features.** *BMC Cancer* 2020;20:29 CrossRef Medline
22. 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;9:20311 CrossRef Medline
23. Gihr GA, Horvath-Rizea D, Hekeler E, et al. **Histogram analysis of diffusion weighted imaging in low-grade gliomas: in vivo characterization of tumor architecture and corresponding neuropathology.** *Front Oncol* 2020;10:206 CrossRef Medline
24. Jakola AS, Zhang YH, Skjulsvik AJ, et al. **Quantitative texture analysis in the prediction of IDH status in low-grade gliomas.** *Clin Neurol Neurosurg* 2018;164:114–20 CrossRef Medline
25. Kim M, Jung SY, Park JE, et al. **Diffusion- and perfusion-weighted MRI radiomics model may predict isocitrate dehydrogenase (IDH) mutation and tumor aggressiveness in diffuse lower grade glioma.** *Eur Radiol* 2020;30:2142–51 CrossRef Medline
26. 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;30:877–86 CrossRef Medline
27. Li Z, Wang Y, Yu J, et al. **Deep Learning-Based Radiomics (DLR) and its usage in noninvasive IDH 1 prediction for low grade glioma.** *Sci Rep* 2017;7:5467 CrossRef Medline
28. Lu CF, Hsu FT, Hsieh KLC, et al. **Machine learning–based radiomics for molecular subtyping of gliomas.** *Clin Cancer Res* 2018;24:4429–36 CrossRef Medline
29. Park CJ, Choi YS, Park YW, et al. **Diffusion tensor imaging radiomics in lower-grade glioma: improving subtyping of isocitrate dehydrogenase mutation status.** *Neuroradiology* 2020;62:319–26 CrossRef Medline
30. Ren Y, Zhang X, Rui W, et al. **Noninvasive prediction of IDH 1 mutation and ATRX expression loss in low-grade gliomas using multiparametric MR radiomic features.** *J Magn Reson Imaging* 2019;49:808–17 CrossRef Medline
31. 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;13:563–71 CrossRef Medline
32. Yu J, Shi Z, Lian Y, et al. **Noninvasive IDH 1 mutation estimation based on a quantitative radiomics approach for grade II glioma.** *Eur Radiol* 2017;27:3509–22 CrossRef Medline
33. 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;48:916–26 CrossRef Medline
34. Zhou H, Vallieres M, Bai H, et al. **MRI features predict survival and molecular markers in diffuse lower-grade gliomas.** *Neuro Oncol* 2017;19:862–70 CrossRef Medline
35. Han Y, Wang W, Yang Y, et al. **Amide proton transfer imaging in predicting isocitrate dehydrogenase 1 mutation status of grade II/III gliomas based on support vector machine.** *Front Neurosci* 2020;14:144 CrossRef Medline
36. Bakas S, Akbari H, Sotiras A, et al. **Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features.** *Sci Data* 2017;4:170117 CrossRef Medline
37. Bakas S, Reyes M, Jakab A, et al. **Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge.** 2018 https://arxiv.org/abs/1811.02629. Accessed April 18, 2020
38. Menze BH, Jakab A, Bauer S, et al. **The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS).** *IEEE Trans Med Imaging* 2015;34:1993–2024 CrossRef Medline
39. Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. *International MICCAI Brainlesion Workshop:* Springer; 2018:311–20
40. Zhang Z, Xiao J, Wu S, et al. **Deep convolutional radiomic features on diffusion tensor images for classification of glioma grades.** *J Digit Imaging*.2020;33:826–37 CrossRef Medline
41. Higgins JP, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons; 2019
42. van der Voort SR, Incekara F, Wijnenga MM, et al. **Predicting the 1p/19q codeletion status of presumed low-grade glioma with an externally validated machine learning algorithm.** *Clin Cancer Res* 2019;25:7455–62 CrossRef Medline
43. 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* 2017;135:601–09 CrossRef Medline
44. Zhou Q, Cao YH, Chen ZH. **Lack of evidence and criteria to evaluate artificial intelligence and radiomics tools to be implemented in clinical settings.** *Eur J Nucl Med Mol Imaging* 2019;46:2812–13 CrossRef Medline
45. Lew C. **The promise of clinical AI: an adaptive future.** *Applied Radiol* 2018;47:3–6 https://www.appliedradiology.com/articles/the-promise-of-clinical-ai-an-adaptive-future#:~:text=The%20Promise%20of%20Clinical%20AI%3A%20An%20Adaptive%20Future.,predictive%20analytics%20for%20population%20health%20management%2C%20and%20beyond. Accessed April 18, 2020
46. Sana M. **Machine learning and artificial intelligence in radiology.** *J Am Coll Radiol* 2018;15:1139–42 CrossRef Medline
47. Shader RI. **Some thoughts about artificial intelligence.** *Clin Ther* 2019;41:1401–03 CrossRef Medline
48. Topol EJ. **High-performance medicine: the convergence of human and artificial intelligence.** *Nat Med* 2019;25:44–56 CrossRef Medline
49. European Society of Radiology. **Impact of artificial intelligence on radiology: a EuroAIM survey among members of the European Society of Radiology.** *Insights imaging* 2019;10:105 CrossRef Medline
50. van Hoek J, Huber A, Leichtle A, et al. **A survey on the future of radiology among radiologists, medical students and surgeons: students and surgeons tend to be more skeptical about artificial intelligence and radiologists may fear that other disciplines take over.** *Eur J Radiol* 2019;121:108742 CrossRef Medline
51. Waymel Q, Badr S, Demondion X, et al. **Impact of the rise of artificial intelligence in radiology: what do radiologists think?** *Diagn Interv Imaging* 2019;100:327–36 CrossRef Medline
52. Palmisciano P, Jamjoom AA, Taylor D, et al. **Attitudes of patients and their relatives toward artificial intelligence in neurosurgery.** *World Neurosurg* 2020;138:e627–33 CrossRef Medline

53. Likert R. **A technique for the measurement of attitudes.** *Archives of Psychology* 1932;22:55
54. Ongena YP, Haan M, Yakar D, et al. **Patients' views on the implementation of artificial intelligence in radiology: development and validation of a standardized questionnaire.** *Eur Radiol* 2020;30:1033–40 CrossRef Medline
55. Janssen CP, Donker SF, Brumby DP, et al. **History and future of human-automation interaction.** *International Journal of Human-Computer Studies* 2019;131:99–107 CrossRef
56. IBM Services. **Beyond the hype: a guide to understanding and successfully implementing artificial intelligence within your business.** https://www.ibm.com/downloads/cas/8ZDXNKQ4. 2018. Accessed April 18, 2020
57. Amershi S, Weld D, Vorvoreanu M, et al. **Guidelines for human-AI interaction.** In: *Proceedings of the Cultural Homestay International Conference on Human Factors in Computing System*, Glasgow, UK; May 4–9, 2019
58. Friedman LM. Furberg CD, DeMets Dl, et al. *Fundamentals of Clinical Trials*. Springer-Verlag; 2010
59. He J, Baxter SL, Xu J, et al. **The practical implementation of artificial intelligence technologies in medicine.** *Nat Med* 2019;25:30–36 CrossRef Medline