

Predicting the 1p/19q Codeletion Status of Presumed Low-Grade Glioma with an Externally Validated Machine Learning Algorithm

van der Voort, Sebastian R.; Incekara, Fatih; Wijnenga, Maarten M.J.; Kapas, Georgios; Gardeniers, Mayke; Schouten, Joost W.; French, Pim J.; Niessen, Wiro J.; Klein, Stefan; More Authors

DOI

[10.1158/1078-0432.CCR-19-1127](https://doi.org/10.1158/1078-0432.CCR-19-1127)

Publication date

2019

Document Version

Final published version

Published in

Clinical cancer research : an official journal of the American Association for Cancer Research

Citation (APA)

van der Voort, S. R., Incekara, F., Wijnenga, M. M. J., Kapas, G., Gardeniers, M., Schouten, J. W., French, P. J., Niessen, W. J., Klein, S., & More Authors (2019). Predicting the 1p/19q Codeletion Status of Presumed Low-Grade Glioma with an Externally Validated Machine Learning Algorithm. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 25(24), 7455-7462. <https://doi.org/10.1158/1078-0432.CCR-19-1127>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Predicting the 1p/19q Codeletion Status of Presumed Low-Grade Glioma with an Externally Validated Machine Learning Algorithm



Sebastian R. van der Voort¹, Fatih Incekara^{2,3}, Maarten M.J. Wijnenga⁴, Georgios Kapas², Mayke Gardeniers², Joost W. Schouten³, Martijn P.A. Starmans¹, Rishie Nandoe Tewarie⁵, Geert J. Lycklama⁶, Pim J. French⁴, Hendrikus J. Dubbink⁷, Martin J. van den Bent⁴, Arnaud J.P.E. Vincent³, Wiro J. Niessen^{1,8}, Stefan Klein¹, and Marion Smits²

Abstract

Purpose: Patients with 1p/19q codeleted low-grade glioma (LGG) have longer overall survival and better treatment response than patients with 1p/19q intact tumors. Therefore, it is relevant to know the 1p/19q status. To investigate whether the 1p/19q status can be assessed prior to tumor resection, we developed a machine learning algorithm to predict the 1p/19q status of presumed LGG based on preoperative MRI.

Experimental Design: Preoperative brain MR images from 284 patients who had undergone biopsy or resection of presumed LGG were used to train a support vector machine algorithm. The algorithm was trained on the basis of features extracted from post-contrast T1-weighted and T2-weighted MR images and on patients' age and sex. The performance of the algorithm compared with tissue

diagnosis was assessed on an external validation dataset of MR images from 129 patients with LGG from The Cancer Imaging Archive (TCIA). Four clinical experts also predicted the 1p/19q status of the TCIA MR images.

Results: The algorithm achieved an AUC of 0.72 in the external validation dataset. The algorithm had a higher predictive performance than the average of the neurosurgeons (AUC 0.52) but lower than that of the neuroradiologists (AUC of 0.81). There was a wide variability between clinical experts (AUC 0.45–0.83).

Conclusions: Our results suggest that our algorithm can noninvasively predict the 1p/19q status of presumed LGG with a performance that on average outperformed the oncological neurosurgeons. Evaluation on an independent dataset indicates that our algorithm is robust and generalizable.

Introduction

Low-grade glioma (LGG) is a primary brain tumor that originates from glial cells. The World Health Organization (WHO) 2016 criteria recognize three subtypes based on molecular and

histologic features: (1) diffuse isocitrate dehydrogenase (IDH) wild-type astrocytoma (IDH wild-type, 1p/19q intact), (2) diffuse IDH mutant astrocytoma (IDH mutated, 1p/19q intact), and (3) oligodendroglioma (IDH mutated, 1p/19q codeleted; refs. 1, 2).

Studies have shown that the distinction between these three categories is clinically relevant in terms of prognosis and management: in patients treated with optimal surgical resection, followed by radiotherapy with or without chemotherapy, median survival is longest of those with oligodendroglioma (3, 4). In addition, studies have suggested that residual tumor has a more negative impact on survival in 1p/19q intact, IDH-mutated astrocytomas than on 1p/19q codeleted, IDH-mutated oligodendrogliomas (5, 6). Therefore, the ability to predict the molecular subtypes of LGG at an early stage could provide better guidance of risk-benefit assessment and clinical decision making.

The recent shift from histopathology-based glioma classification to the molecular subtype-based WHO 2016 classification gave rise to neuro-oncologic radiogenomics research in which features seen on preoperative MR images are used to predict the genetic mutation status of glioma (7–9). Features such as frontal tumor localization, indistinct tumor borders, heterogeneous signal intensity (SI) on T2-weighted images, and both cortical and subcortical tumor infiltration all suggest the presence of 1p/19q codeletion (7).

One way of linking MRI features to 1p/19q codeletion is through machine learning. Although several studies have applied this method to datasets of patients with high-grade glioma, few studies have developed radiogenomics methodology in

¹Biomedical Imaging Group Rotterdam, Departments of Radiology and Nuclear Medicine Medical Informatics, Erasmus MC-University Medical Centre Rotterdam, Rotterdam, the Netherlands. ²Department of Radiology and Nuclear Medicine, Erasmus MC-University Medical Centre Rotterdam, Rotterdam, the Netherlands. ³Department of Neurosurgery, Brain Tumor Center, Erasmus MC-University Medical Centre Rotterdam, Rotterdam, the Netherlands. ⁴Department of Neurology, Brain Tumor Center, Erasmus MC Cancer Institute, Rotterdam, the Netherlands. ⁵Department of Neurosurgery, Haaglanden MC, the Hague, the Netherlands. ⁶Department of Radiology, Haaglanden MC, the Hague, the Netherlands. ⁷Department of Pathology, Brain Tumor Center at Erasmus MC Cancer Institute, Rotterdam, the Netherlands. ⁸Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, Delft, the Netherlands.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

S.R. van der Voort and F. Incekara contributed equally as first co-authors.

S. Klein and M. Smits contributed equally as senior authors to this article.

Corresponding Author: Marion Smits, Erasmus MC, PO Box 2040, 3000 CA Rotterdam, the Netherlands. Phone: +31 10 703 5372; Fax: +31 10 703 4033; E-mail: marion.smits@erasmusmc.nl

Clin Cancer Res 2019;25:7455–62

doi: 10.1158/1078-0432.CCR-19-1127

©2019 American Association for Cancer Research.

Translational Relevance

This study is, to the best of our knowledge, one of the first to train an algorithm that predicts the 1p/19q codeletion status in an unselected, real-world patient population with presumed low-grade glioma and to validate this algorithm on an independent, external dataset. It shows that the predictive performance of the algorithm outperforms the predictive performance of surgical neuro-oncology experts but not that of the radiologists. It indicates that together with age and sex, the location of the tumor and heterogeneity as seen on T2-weighted MRI are important for the prediction. The study discusses the potential impact of machine learning algorithm on clinical decision making. Our algorithm provides the treating physician and the patient with a nonenhancing, presumed low-grade glioma with a prediction of the tumor's 1p/19q codeletion status prior to surgery, allowing better informed decision making on treatment.

LGG (10–15). Of the ones that have, most have not used an independent test set and, therefore, it is difficult to estimate their actual performance in the real-world clinical setting (10, 11, 13, 14). Lu and colleagues (12) did use an independent test set, but this set contained a very limited number of LGG cases ($N = 12$). Zhou and colleagues (15) used a test set consisting of IDH-mutated LGG and high-grade glioma to evaluate the 1p/19q codeletion prediction performance. This is not an ideal test set as 1p/19q codeletion status is not clinically relevant for high-grade glioma, and there is a selection bias of IDH-mutated tumors only.

The aim of this retrospective study was to develop a radiogenomics approach to predict the 1p/19q codeletion status of presumed LGG based on preoperative MRI features, with a machine learning algorithm that was validated on a large external dataset.

Materials and Methods

EMC/HMC dataset

Study participants. All patients aged 18 years or older newly diagnosed with presumed LGG and who underwent tumor resection or biopsy between October 2002 and March 2017 at the Erasmus MC, University Medical Centre Rotterdam (EMC) or the Haaglanden Medical Centre (HMC) were retrospectively included in the EMC/HMC dataset. Patients were eligible if histopathologic diagnosis with molecular subclassification of the 1p/19q codeletion status and preoperative post-contrast T1-weighted and T2-weighted MR images were available. The study was approved by the Medical Ethical Committee of Erasmus MC, who waived the need for written informed consent from the patients due to the retrospective nature of this study and the (emotional) burden that would result from contacting the patients or their relatives to obtain consent. The study was performed in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Histopathologic diagnosis and molecular subclassification. Tumor samples were obtained from patients who underwent surgical

resection or biopsy. Histopathologic examination was performed by neuropathologists and further molecular subclassification of the 1p/19q codeletion and/or IDH mutation status was performed as part of the diagnostic routine by molecular biologists using fluorescence *in situ* hybridization (FISH), loss of heterozygosity analysis, and targeted next-generation sequencing panel using an Ion Torrent Personal Genome Machine (Life Technologies) or Ion S5XL or a Multiplex Ligation Probe Assay (MRC-Holland; refs. 4, 16–18). All tumors were subclassified on the basis of the WHO 2016 criteria.

Imaging acquisition and postprocessing. MR images were used that were acquired in the routine diagnostic process. T1-weighted and T2-weighted MRI sequences were used for the algorithm. In many, but not all, patients, T2-weighted fluid-attenuated inversion recovery (T2w-FLAIR) imaging was also available. As images were acquired at a number of sites, the imaging data were heterogeneous with a wide range of acquisition settings in voxel spacing, matrix size, echo time, repetition time, number of slices, slice thickness, and field strengths on scanners from three different manufacturers (General Electric, Philips, and Siemens). An overview of the scanning settings is given in the Supplementary Materials, Appendix 1.

All images were visually inspected by M. Smits and excluded if MRI artifacts were present. Presumed LGG was defined as nonenhancing tumor, as seen on the presurgical post-contrast T1-weighted MR image. Therefore, all post-contrast T1-weighted images were reviewed and excluded if clear or solid enhancement was present. When available, T1-weighted precontrast images were inspected for hemorrhage to prevent false-positive assessment of enhancement. Although tumors with evident contrast enhancement were excluded, minimal enhancement was tolerated. Minimal enhancement was defined as punctiform (<1 mm in diameter) or poorly defined faint enhancement, similar to Pallud and colleagues (19).

Tumor segmentation was performed by two independent observers (F. Incekara and G. Kapas) using ITK-Snap (20). Segmentation was done on T2w-FLAIR when available ($N = 119$), otherwise on the T2-weighted images ($N = 165$). Because in our institution LGG segmentations are preferably performed on T2w-FLAIR images, we did not enforce the assessors to segment on T2-weighted images in order to stick to the real-world clinical practice. The segmentations were then transformed to the T2-weighted images (in the case of T2w-FLAIR segmentation) and the T1-weighted images, using the image registration software SimpleElastix (21). For all patients, brain masks were automatically constructed using FSL's BET tool with a fractional intensity threshold of 0.5 (22). These brain masks were subsequently used to normalize the intensity of the MR images. Details can be found in Supplementary Materials, Appendix 2.

The Cancer Imaging Archive dataset

Patients from The Cancer Imaging Archive (TCIA) "LGG-1p19qDeletion" dataset were screened for eligibility on the basis of previously described inclusion and exclusion criteria and used as the external validation dataset (10, 23, 24).

This data collection is a publicly available dataset that consists of histopathologically proven LGG with coregistered T1- and T2-weighted preoperative MRI images as well as biopsy-proven 1p/19q codeletion status. Molecular analysis of the 1p/19q codeletion status was performed with FISH for all tumors; IDH

mutation status was not determined. All MRI images were visually inspected by M. Smits as previously described. An overview of the MRI settings is listed in the Supplementary Materials, Appendix 1. All tumors were semiautomatically segmented by M. Smits on the T2-weighted images using ITK-Snap. Because the T1-weighted and T2-weighted images were already coregistered in this study, the segmentation could be directly used for the T1-weighted images without the need for registration. Brain masks were made using FSL's BET tool, with the same settings as for the EMC/HMC dataset.

Classification algorithm

To predict the 1p/19q status of the tumors based on MRI features, the PREDICT toolbox was used. This toolbox was used to extract a total of 78 image features (such as image intensity, tumor texture, tumor shape, and tumor location) from the T1-weighted and T2-weighted MR image. These features, as well as the age and sex of the patient, were then used to train a support vector machine (SVM), resulting in a total of 80 features. All parameter optimization and classifier training was performed on the EMC/HMC training set dataset using 100 iterations of stratified random-split cross-validation, with 80% of the dataset used for training and 20% used for validation. Once the algorithm was optimized, no more changes were made to the algorithm and it was then evaluated on the TCIA dataset. To evaluate the algorithm, the accuracy, sensitivity (1p/19q codeletion prediction), specificity (1p/19q intact prediction), area under the ROC curve (AUC), weighted F1 score, and precision were determined by comparing the predicted labels with the reference labels obtained from tissue diagnosis. Full details of the algorithm can be found in the Supplementary Materials, Appendix 2 with more information about the evaluation metrics in Supplementary Materials, Appendix 3. An overview of the classification algorithm is provided in Supplementary Materials, Fig. S1.

To minimize the variance due to randomness in the algorithm training, an ensemble of five SVMs, which averages the predictions of the five independently trained models, was also constructed; the details can be found in Supplementary Materials, Appendix 2. One hundred different ensembles were constructed and were evaluated on the TCIA dataset using the evaluation metrics described previously. Mean and standard deviation of the metrics over the 100 ensembles were computed.

To evaluate the contribution of the different features to the final prediction, a sensitivity analysis using polynomial chaos expansions was performed, resulting in Sobol indices for each feature (25). The total Sobol index was used to determine the relative feature importance of the individual features. The total Sobol index is relative measure of the sensitivity of the algorithm to the input features. The OpenPC toolbox was used to create the polynomial chaos expansions and to calculate the Sobol indices (26, 27).

We also determined which patients from the TCIA dataset were considered as representative examples for the 1p/19q codeleted and 1p/19q intact class by the algorithm. This was achieved by counting the number of times the algorithm correctly predicted the class for a specific patient in the 100 ensembles that were constructed.

We also evaluated the performance of the algorithm when the EMC/HMC and TCIA dataset were mixed instead of used as a separate train and validation set to evaluate the effect of adding additional training data.

Prediction of 1p/19q status by clinical experts

To compare the results of the algorithm with expert performance, the 1p/19q status of the TCIA tumors was also predicted by two neuroradiologists and two neurosurgeons at the Erasmus MC Brain Tumor Centre. They were presented with the T1-weighted and T2-weighted images side by side for each patient as well as the sex and age to ensure that the algorithm and the raters had access to the same information. For each tumor, the raters were then asked to choose whether they thought it was 1p/19q codeleted or intact to provide a confidence score ranging from 1 to 5 (1 indicating very unsure and 5 indicating very sure). This confidence score was then turned into a prediction "score" by dividing it by 5 and multiplying it by 1 if the predicted label was 1p/19q codeleted or by -1 if the predicted label was 1p/19q intact. In this way, an AUC could be determined for the manual classification. The accuracy, sensitivity, and specificity were determined in the same way as for the algorithm.

Statistical analysis

Statistical analyses to test differences between the two datasets were performed with SPSS 21.0 statistical software (IBM Corp.). We tested whether the two datasets differed significantly from each other using the Mann-Whitney U test for continuous, nonnormally distributed variables (age and volume) and the χ^2 test for all other categorical variables (sex, genetic analysis, presence of mild enhancement, codeletion status). Predictive performances [mean, 95% confidence interval (CI)] between the EMC/HMC training set and TCIA validation set were tested with the Welch t test. Accuracy between the clinical experts and the algorithm was tested with the McNemar test. A P value of <0.05 was considered statistically significant. The 95% CIs were calculated such that if the entire experiment of training on EMC/HMC and prediction on TCIA would be repeated in 95% of the repetitions, the result would lie within that interval.

Data sharing

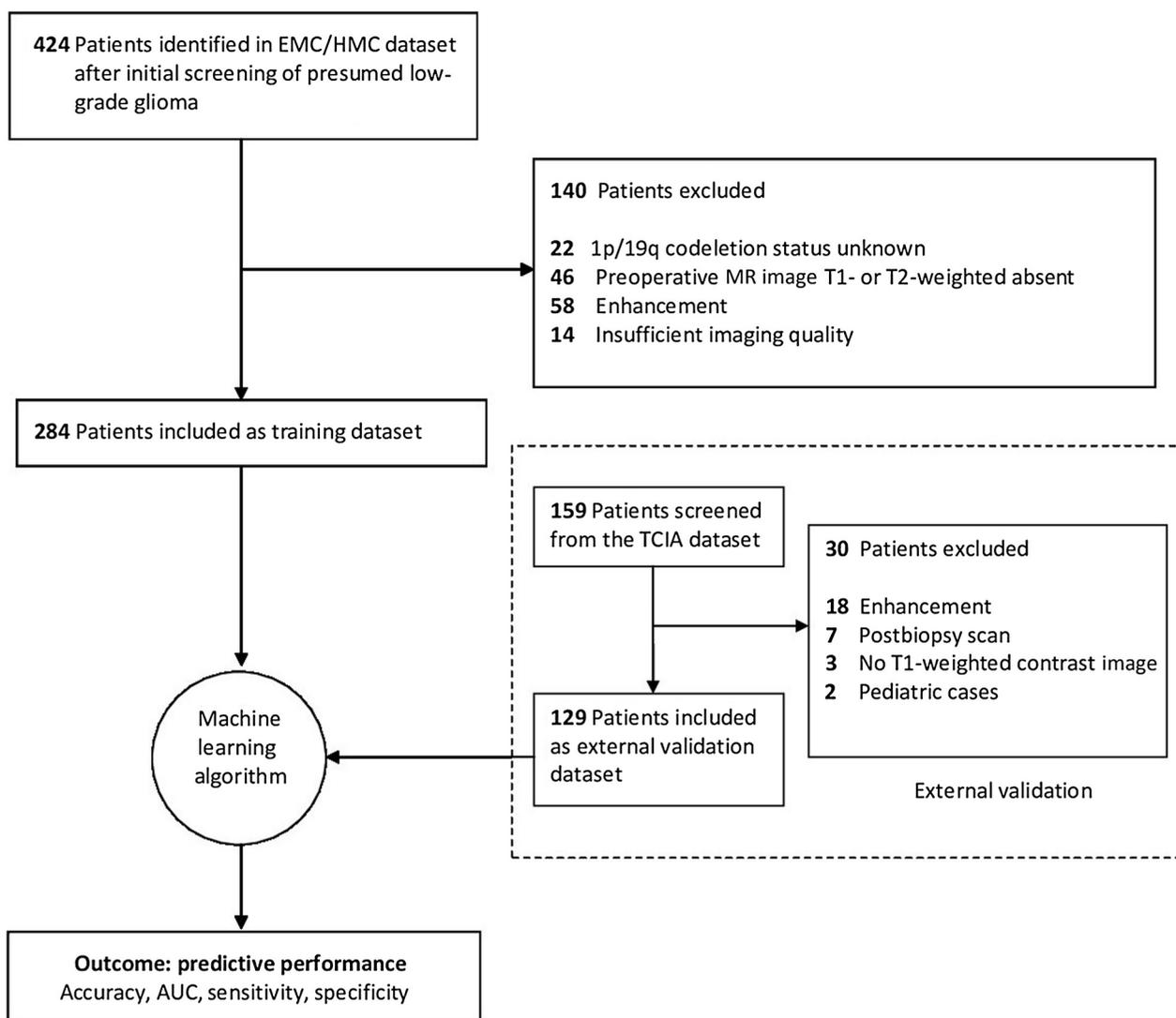
The data used in this study are available on Mendeley Data (<http://dx.doi.org/10.17632/rssf5nxxby.1>). The code for the construction and evaluation of the prediction algorithm is available on GitHub (<https://github.com/Svdvoort/PREDICT>). The code used to construct the polynomial chaos expansions and calculate the Sobol indices is available on GitHub as well (<https://github.com/Svdvoort/OpenPC>).

Results

In the EMC/HMC dataset, 424 LGGs were identified and screened for eligibility. Cases were excluded because of unknown 1p/19q codeletion status ($N = 22$), absence of T1- and/or T2-weighted MRI images ($N = 46$), enhancement ($N = 58$), and unacceptable image quality ($N = 14$), which resulted in 284 patients included for final analysis (flowchart, Fig. 1).

From the TCIA database, all 159 patients were screened for eligibility. Patients were excluded because of enhancement ($N = 18$), signs of prior biopsy/surgical procedure ($N = 7$), no post-contrast T1-weighted imaging available ($N = 3$), and patients being younger than 18 years ($N = 2$), resulting in 129 patients included in the external validation dataset (flowchart, Fig. 1). An overview of the excluded patients from the TCIA database as well as the reason for exclusion is available as Supplementary Materials, Appendix 4.

van der Voort et al.

**Figure 1.**

Flow diagram of the inclusion procedure for both the EMC/HMC training dataset and TCIA validation dataset.

There was no significant difference between the EMC/HMC and TCIA datasets for median age [43.0 years, interquartile range (IQR): 17.0, vs. 39 years, IQR: 19.5, respectively; $P = 0.11$] and sex distribution (56.7% vs. 52.7% male, respectively; $P = 0.45$). Median tumor volume in the EMC/HMC dataset was significantly larger than in the TCIA dataset (median: 47.80 cm³, IQR: 58.65 vs. median 35.70 cm³, IQR: 49.10), $P = 0.04$). There were fewer 1p/19q codeleted tumors in the EMC/HMC compared with the TCIA dataset (35.20% vs. 65.40%, $P < 0.0001$). Patient and tumor characteristics of both datasets are further presented in Table 1.

The predictive performance of the algorithm on the EMC/HMC training dataset, obtained from the cross validation, and the TCIA validation dataset is given in terms of accuracy, AUC, F1 score, precision, sensitivity, and specificity in Table 2. The accuracy, AUC, and sensitivity did not differ significantly between training and validation datasets ($P = 0.886$, $P = 0.746$, and $P = 0.146$, respectively), whereas the specificity was significantly lower in the validation dataset ($P = 0.038$).

The predictive performances of the clinical experts compared with the algorithm can be found in Table 3, and their ROC curves in Fig. 2. The algorithm had a higher AUC when compared with the average performance of the neurosurgeons but a lower AUC when compared with the neuroradiologists. There was high variability in predictive performance between the clinical experts (AUC of 0.449–0.830).

The results of mixing the EMC/HMC dataset and the TCIA dataset are shown in Supplementary Materials, Appendix 5. Mixing the datasets leads to a slightly improved performance but still within the CI of the EMC/HMC dataset cross-validation results.

According to the algorithm, the most important features for accurate 1p/19q codeletion status prediction were the cranial/caudal location of the tumor, the skewness of the T2-weighted SI histogram, and one of the texture features, together with age and sex (Supplementary Materials, Fig. S2). The algorithm identified a typical 1p/19q codeleted glioma as a frontal heterogeneous

Table 1. Patient and tumor characteristics

	EMC/HMC—Training set (n = 284)	TCIA—Validation set (n = 129)	P
Clinical			
Age median [IQR] in years	n (%) 43 [17]	n (%) 39 [19.5]	0.11
Sex			0.45
Male	161 (56.7)	68 (52.7)	
Female	123 (43.3)	61 (47.3)	
Imaging			
Volume median [IQR] in cm ³	47.8 [58.7]	35.7 [49.1]	0.04
Mild enhancement			0.005
Yes	27 (9.5)	25 (19.4)	
No	257 (90.5)	104 (80.6)	
Histopathology (WHO, 2016)			<0.0001
Oligodendroglioma	100	85	
Astrocytoma	181	44	
Glioblastoma	3	0	
Genetic			
1p/19q codeletion			<0.0001
Yes	100 (35.2)	85 (65.9)	
No	184 (64.8)	44 (34.1)	
IDH mutation			n/a
Yes	214 (75.4)	0 (0.0)	
No	35 (12.3)	0 (0.0)	
Unknown	35 (12.3)	129 (100.0)	
Method of analysis			<0.0001
NGS	214 (75.4)	0 (0)	
FISH	45 (15.8)	129 (100)	
MLPA	25 (8.8)	0 (0)	

Abbreviations: FISH, fluorescence *in situ* hybridization; MLPA, multiplex ligation probe assay; NGS, next-generation sequencing.

Table 2. Predictive performances of the algorithm on the EMC/HMC training and TCIA validation datasets

	EMC/HMC—Training set Mean (95% CI)	TCIA—Validation set Mean (95% CI)	P
Accuracy	0.698 (0.636–0.760)	0.693 (0.657–0.729)	0.872
AUC	0.755 (0.694–0.817)	0.723 (0.708–0.737)	0.313
F1 score	0.701 (0.640–0.761)	0.697 (0.661–0.733)	0.896
Precision	0.570 (0.491–0.649)	0.787 (0.754–0.820)	<0.001
Sensitivity	0.657 (0.562–0.752)	0.732 (0.689–0.775)	0.123
Specificity	0.721 (0.628–0.813)	0.617 (0.544–0.691)	0.027

NOTE: The performances on the EMC/HMC training dataset were obtained by cross-validation; the performances on the TCIA validation dataset were obtained by training on the EMC/HMC dataset and then testing on the TCIA dataset.

tumor as seen on T1-weighted and T2-weighted scans, whereas a typical 1p/19q intact glioma was identified as a parietal homogeneous tumor, as shown in Fig. 3.

Discussion

In this study, we developed an algorithm that predicted the 1p/19q codeletion status of presumed LGG noninvasively based on preoperative MR images with an AUC of approximately 0.75.

We tested the algorithm on an external, independent validation dataset. To the best of our knowledge, this is the first time that this has been done in presumed LGG and thus sets a benchmark for the expected performance in the real-world clinical setting. The algorithm had a higher AUC than the averaged AUC of the neurosurgeons but lower than the averaged AUC of the neuroradiologists.

To the best of our knowledge, this is the first study performing a radiogenomics-based machine learning study in LGG from the perspective of real-world clinical practice: we included all patients with presumed, non-contrast-enhancing LGG, rather than a selection of patients with histopathologically defined LGG. This is important, because in a clinical setting the genetic mutation is unknown at first symptomatic presentation. Because it is known only after surgery and molecular analysis, we aimed to mirror this real-world situation as best as possible by not selecting patients on the basis of histologic tumor features but on the imaging features that are available at the time of presentation. Note that subsequently all lesions were surgically resected to obtain the ground truth data based on confirmed histologic and molecular analysis. We trained the algorithm on a heterogeneous training dataset and used a separate, completely independent, publicly available dataset with data from an entirely different institute to validate the

Table 3. Predictive performance of four clinical experts compared with the algorithm on the TCIA validation dataset

	Neurosurgeon 1	Neurosurgeon 2	Average of surgeons	Neuroradiologist 1	Neuroradiologist 2	Average of radiologists	Algorithm
Accuracy, with P value ^a	0.520–0.073	0.457–0.002	0.489	0.690–0.720	0.574–0.266	0.632	0.693
AUC	0.580	0.449	0.515	0.830	0.792	0.811	0.723
Sensitivity	0.370	0.459	0.415	0.610	0.459	0.535	0.732
Specificity	0.820	0.455	0.638	0.840	0.795	0.818	0.617

^aStatistical comparison (McNemar) of accuracy between clinical experts and algorithm.

van der Voort et al.

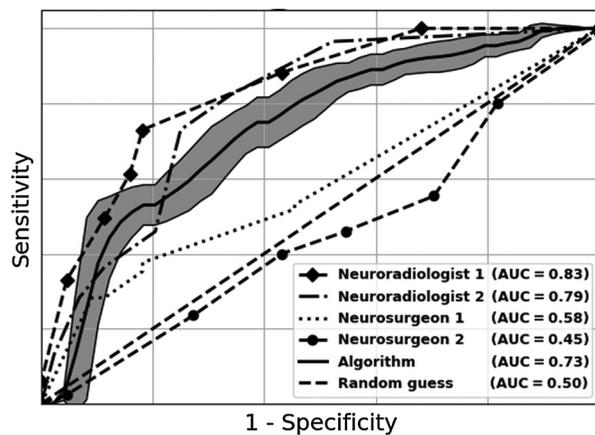


Figure 2.

ROC curves of clinical expert and algorithm performance. For the performance of the algorithm, the 95% confidence interval is plotted as well (shaded gray), representing the uncertainty due to randomness in model training.

algorithm. As such, this study is the first to demonstrate that the performance of a radiogenomics algorithm in predicting the 1p/19q codeletion status of presumed LGG based on MR images was robust and matched expert clinical performance. Furthermore, we were also able to show which image features were important in the classification, increasing the clinical understand-

ing of the machine learning algorithm and potentially aiding better acceptance, as well as furthering fundamental research into understanding of glioma pathophysiology.

Although other studies did already investigate the noninvasive prediction of the molecular subtype of LGG, these often focused on IDH mutations only and did not consider the 1p/19q codeletion status (11, 28, 29). In comparison with studies that did look at the 1p/19q codeletion, we used a larger cohort and an external validation dataset (10, 13, 14, 30, 31), which makes our results more robust and generalizable, respectively. Although one study by Lu and colleagues (12) did use an independent dataset, this study used only 5 patients to externally validate the 1p/19q codeletion predictive performance of the algorithm, which severely limits the reliability of its predictive performance. In addition, that specific study retrospectively selected patients with histopathologically defined LGG only, which represents the diagnosis-treatment workflow in clinical practice less accurately. The starting point of decision making on the optimal treatment strategy for LGG is the initial diagnosis on first MRI, when a non-contrast-enhancing space occupying lesion is seen, at which point knowledge on the histopathologic grade is not yet available.

The optimal timing and effect of surgical treatment of LGG are extensively being debated within literature and have recently been reevaluated in the light of molecular subclassification after the introduction of WHO 2016 criteria (5, 6, 32, 33). Currently, the molecular subtype based on 1p/19q codeletion and IDH mutation can be diagnosed only after obtaining tissue with biopsy or surgery. Indeed, as our results suggest, it is even for experienced

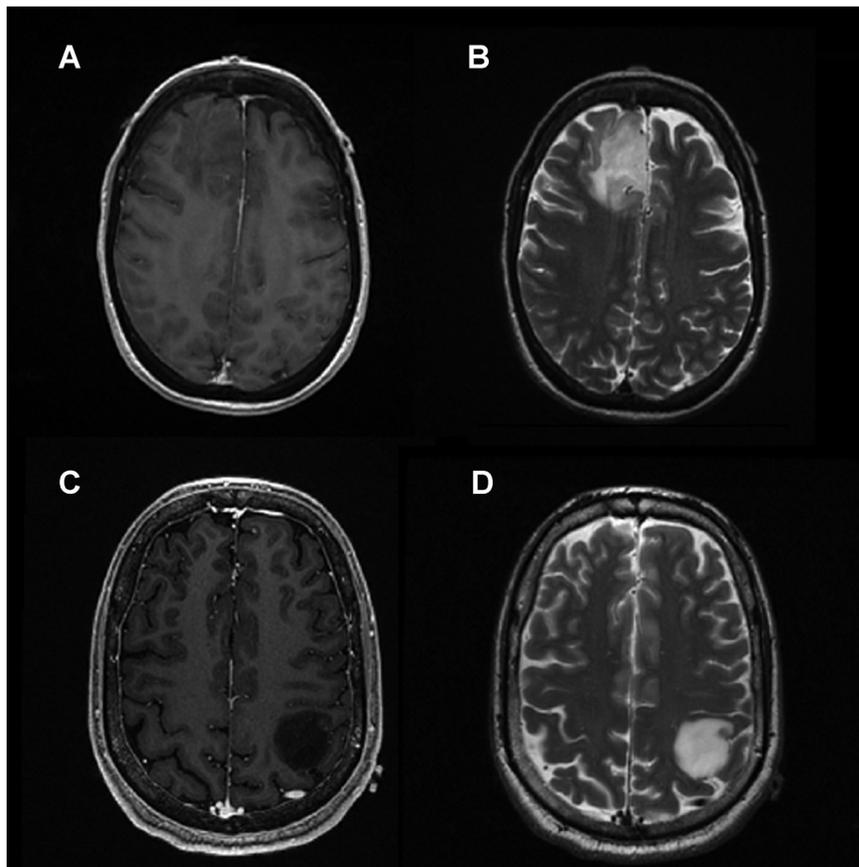


Figure 3.

A and B, Frontally located glioma. It is nonenhancing on post-contrast T1-weighted MR image (**A**). A heterogeneous signal intensity with indistinct border is visible on the T2-weighted MR image (**B**). Correctly predicted as a 1p/19q codeleted glioma (oligodendroglioma) by the algorithm. **C and D,** Parietally located glioma. It is nonenhancing on post-contrast T1-weighted MR image (**C**). A homogeneous signal intensity with sharply demarcated border is visible on the T2-weighted MR image (**D**). Correctly predicted as a 1p/19q intact glioma (astrocytoma) by the algorithm.

neuro-oncologic surgeons and radiologists a challenge to accurately predict the codeletion status of nonenhancing tumors based on preoperative MR images (AUC of 0.45–0.83).

There are two scenarios in which preoperative, noninvasive prediction of the 1p/19q codeletion status based on MRI would be clinically relevant. First, some patients are not eligible for surgical resection or diagnostic biopsy due to older age, poor neurologic condition, or tumor localization in eloquent brain areas or basal ganglia (33). However, knowledge of the molecular LGG subtype might add to a more appropriate (timing of) chemo- and/or radiotherapy regimes (immediate postoperative therapy vs. watchful waiting; ref. 34). Therefore, noninvasive, accurate prediction of the molecular subtype on imaging could help clinicians select the optimal treatment when tissue diagnosis is difficult to obtain. Second, it is suggested that postsurgical residual 1p/19q intact, IDH-mutated tumor has a more negative impact on survival than residual 1p/19q codeleted, IDH-mutated (oligodendroglioma) tumor (5, 6). With presurgical knowledge of the specific molecular subtype, the surgeon can make a better informed decision on whether or not to push the limits of resection at the time of surgery, avoiding on the one hand reresection in case of residual 1p/19q intact, IDH-mutated tumor and less-justified postsurgical deficits in 1p/19q codeleted tumor on the other hand. Clearly, the diagnostic accuracy of our algorithm is as yet too low to rely on for clinical practice. However, the results are promising because they generalize through multiple datasets, encouraging future research in this direction.

Our study had a few limitations. First, for this study, only the T1-weighted and T2-weighted images were used, whereas diffusion-weighted and perfusion imaging also contain relevant features for the 1p/19q status. These sequences were not included in the development of the present algorithm, as these were scarcely available in both datasets.

Second, the IDH mutation status was undetermined in all of the TCIA cases and in 35 cases of the EMC/HMC dataset. Because molecular subclassification according to the WHO 2016 guidelines is based on both the 1p/19q codeletion status and IDH mutation status, it is important to predict both. Therefore, for our future work, we are expanding our database with more patients in whom the tumor IDH status is known to eventually be able to predict all clinically relevant subtypes of presumed LGG.

There was an imbalance between the EMC/HMC dataset and the TCIA dataset in terms of the number of codeleted and intact cases. Despite this imbalance, our algorithm still shows similar performance between the cross-validation result of the EMC/HMC dataset and the performance on the TCIA test dataset.

In conclusion, our results suggest that our algorithm can noninvasively predict the 1p/19q codeletion status of presumed LGG with a performance that in general outperforms oncological neurosurgeons. We evaluated our algorithm on an independent, multicenter dataset, which demonstrated that our algorithm is robust and generalizable. The prediction of the 1p/19q codeletion status by our algorithm can eventually add value to clinical decision making by tailoring the treatment strategy for patients with presumed LGG even prior to surgery.

Disclosure of Potential Conflicts of Interest

W.J. Niessen is an employee/paid consultant for and holds ownership interest (including patents) in Quantib. M. Smits is an employee/paid consultant for Parexel Ltd. and reports receiving speakers bureau honoraria from GE Healthcare. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: S.R. van der Voort, F. Incekara, J.W. Schouten, M.J. van den Bent, S. Klein, M. Smits

Development of methodology: S.R. van der Voort, F. Incekara, M.P.A. Starmans, S. Klein, M. Smits

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.R. van der Voort, F. Incekara, M.M.J. Wijnenga, G. Kapas, J.W. Schouten, H.J. Dubbink, M.J. van den Bent, A.J.P.E. Vincent

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.R. van der Voort, F. Incekara, M. Gardeniers, M.P.A. Starmans, G.J. Lycklama, P.J. French, S. Klein, M. Smits

Writing, review, and/or revision of the manuscript: S.R. van der Voort, F. Incekara, M.M.J. Wijnenga, J.W. Schouten, M.P.A. Starmans, R. Nandoe Tewarie, G.J. Lycklama, P.J. French, H.J. Dubbink, M.J. van den Bent, A.J.P.E. Vincent, W.J. Niessen, S. Klein, M. Smits

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.R. van der Voort, F. Incekara, J.W. Schouten
Study supervision: S.R. van der Voort, H.J. Dubbink, W.J. Niessen, S. Klein, M. Smits

Acknowledgments

The authors thank the patients who participated in this study and Claudine Nogarede-Bloemendaal for assistance with data collection at the Haaglanden MC.

S.R. van der Voort and F. Incekara were funded by the Dutch Cancer Society (KWF project number EMCR 2015-7859). M.P.A. Starmans and S. Klein were funded by the Netherlands Organisation for Scientific Research (NWO project number 14929-14930).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 9, 2019; revised June 12, 2019; accepted September 6, 2019; published first September 23, 2019.

References

- Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* 2016;131:803–20.
- van den Bent MJ, Smits M, Kros JM, Chang SM. Diffuse infiltrating oligodendroglioma and astrocytoma. *J Clin Oncol* 2017;35:2394–401.
- Cairncross JC, Wang M, Jenkins RB, Shaw EG, Giannini C, Brachman DG, et al. Benefit from procarbazine, lomustine, and vincristine in oligodendroglial tumors is associated with mutation of IDH. *J Clin Oncol* 2014;32:783–90.
- Dubbink HJ, Atmodimedjo PN, Kros JM, French PJ, Sanson M, Idbaih A, et al. Molecular classification of anaplastic oligodendroglioma using next-generation sequencing: a report of the prospective randomized EORTC Brain Tumor Group 26951 phase III trial. *Neuro Oncol* 2016;18:388–400.
- Wijnenga MMJ, French PJ, Dubbink HJ, Dinjens WNM, Atmodimedjo PN, Kros JM, et al. The impact of surgery in molecularly defined low-grade glioma: an integrated clinical, radiological, and molecular analysis. *Neuro Oncol* 2018;20:103–12.
- Clark VE, Cahill DP. Extent of resection versus molecular classification: what matters when? *Neurosurg Clin N Am* 2019;30:95–101.

van der Voort et al.

7. Smits M, van den Bent MJ. Imaging correlates of adult glioma genotypes. *Radiology* 2017;284:316–31.
8. Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, et al. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. *Radiology* 2014;273:168–74.
9. Gutman DA, Cooper LA, Hwang SN, Holder CA, Gao J, Aurora TD, et al. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. *Radiology* 2013;267:560–9.
10. Akkus Z, Ali I, Sedlar J, Agrawal JP, Parney IF, Giannini C, et al. Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence. *J Digit Imaging* 2017;30:469–76.
11. 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.
12. Lu CF, Hsu FT, Hsieh KL, Kao YJ, Cheng SJ, Hsu JB, et al. Machine learning-based radiomics for molecular subtyping of gliomas. *Clin Cancer Res* 2018;24:4429–36.
13. Chang P, Grinband J, Weinberg BD, Bardis M, Khy M, Cadena G, et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *Am J Neuroradiol* 2018;39:1201–7.
14. Han Y, Xie Z, Zang Y, Zhang S, Gu D, Zhou M, 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;140:297–306.
15. Zhou H, Chang K, Bai HX, Xiao B, Su C, Bi WL, et al. Machine learning reveals multimodal MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low-and high-grade gliomas. *J Neurooncol* 2019;142:299–307.
16. Dubbink HJ, Atmodimedjo PN, van Marion R, Krol NMG, Riegman PHJ, Kros JM, et al. Diagnostic detection of allelic losses and imbalances by next-generation sequencing: 1p/19q co-deletion analysis of gliomas. *J Mol Diagn* 2016;18:775–86.
17. Riemenschneider MJ, Jeuken JW, Wesseling P, Reifenberger G. Molecular diagnostics of gliomas: state of the art. *Acta Neuropathol* 2010;120:567–84.
18. Bieńkowski M, Wöhrer A, Moser P, Kitzwögerer M, Ricken G, Ströbel T, et al. Molecular diagnostic testing of diffuse gliomas in the real-life setting: a practical approach. *Clin Neuropathol* 2018;37:166–77.
19. Pallud J, Capelle L, Taillandier L, Fontaine D, Mandonnet E, Guillemin R, et al. Prognostic significance of imaging contrast enhancement for WHO grade II gliomas. *Neuro Oncol* 2009;11:176–82.
20. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006;31:1116–28.
21. Marstal K, Berendsen F, Staring M, Klein S. SimpleElastix: a user-friendly, multi-lingual library for medical image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Workshops*; 2016:134–42.
22. Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp* 2002;17:143–55.
23. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 2013;26:1045–57.
24. Erickson B, Akkus Z, Sedlar J, Kofiat P. Data from LGG-1p19qDeletion. The Cancer Imaging Archive 2017.
25. Sudret B. Global sensitivity analysis using polynomial chaos expansions. *Reliability Eng Syst Saf* 2008;93:964–79.
26. Perko Z, van der Voort SR, van de Water S, Hartman CM, Hoogeman M, Lathouwers D. Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion. *Phys Med Biol* 2016;61:4646–64.
27. van der Voort S, van de Water S, Perko Z, Heijmen B, Lathouwers D, Hoogeman M. Robustness recipes for minimax robust optimization in intensity modulated proton therapy for oropharyngeal cancer patients. *Int J Radiat Oncol Biol Phys* 2016;95:163–70.
28. Ren Y, Zhang X, Rui W, Pang H, Qiu T, Wang J, 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;49:808–17.
29. Yu J, Shi Z, Lian Y, Li Z, Liu T, Gao Y, et al. Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. *Eur Radiol* 2017;27:3509–22.
30. Park YW, Han K, Ahn SS, Bae S, Choi YS, Chang JH, et al. Prediction of IDH1-mutation and 1p/19q-codeletion status using preoperative MR imaging phenotypes in lower grade gliomas. *AJNR Am J Neuroradiol* 2018;39:37–42.
31. Shofty B, Artzi M, Ben Bashat D, Liberman G, Haim O, Kashanian A, et al. MRI radiomics analysis of molecular alterations in low-grade gliomas. *Int J Comput Assist Radiol Surg* 2018;13:563–71.
32. Jakola AS, Skjulsvik AJ, Myrnes KS, Sjøvik K, Unsgård G, Torp SH, et al. Surgical resection versus watchful waiting in low-grade gliomas. *Ann Oncol* 2017;28:1942–8.
33. Jiang B, Chaichana K, Veeravagu A, Chang SD, Black KL, Patil CG. Biopsy versus resection for the management of low-grade gliomas. *Cochrane Database Syst Rev* 2017;4:CD009319.
34. Ricard D, Kaloshi G, Amiel-Benouaich A, Lejeune J, Marie Y, Mandonnet E, et al. Dynamic history of low-grade gliomas before and after temozolomide treatment. *Ann Neurol* 2007;61:484–90.

Clinical Cancer Research

Predicting the 1p/19q Codeletion Status of Presumed Low-Grade Glioma with an Externally Validated Machine Learning Algorithm

Sebastian R. van der Voort, Fatih Incekara, Maarten M.J. Wijnenga, et al.

Clin Cancer Res 2019;25:7455-7462. Published OnlineFirst September 23, 2019.

Updated version Access the most recent version of this article at:
[doi:10.1158/1078-0432.CCR-19-1127](https://doi.org/10.1158/1078-0432.CCR-19-1127)

Supplementary Material Access the most recent supplemental material at:
<http://clincancerres.aacrjournals.org/content/suppl/2019/09/21/1078-0432.CCR-19-1127.DC1>

Cited articles This article cites 32 articles, 4 of which you can access for free at:
<http://clincancerres.aacrjournals.org/content/25/24/7455.full#ref-list-1>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://clincancerres.aacrjournals.org/content/25/24/7455>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.