Article

Fully Automated MR Based Virtual Biopsy of Cerebral Gliomas

Johannes Haubold [1,*], René Hosch [1], Vicky Parmar [1], Martin Glas [2], Nika Guberina [3], Onofrio Antonio Catalano [4], Daniela Pierscianek [5], Karsten Wrede [5], Cornelius Deuschl [1], Michael Forsting [1], Felix Nensa [1], Nils Flaschel [1,†] and Lale Umutlu [1,†]

1   Department of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Essen, Hufelandstr. 55, D-45147 Essen, Germany; rene.hosch@uk-essen.de (R.H.); vicky.parmar@uk-essen.de (V.P.); Cornelius.Deuschl@uk-essen.de (C.D.); Michael.Forsting@uk-essen.de (M.F.); Felix.Nensa@uk-essen.de (F.N.); Nils.Flaschel@uk-essen.de (N.F.); lale.umutlu@uk-essen.de (L.U.)
2   Department of Neurology, Division of Clinical Neurooncology, University Hospital Essen, D-45147 Essen, Germany; Martin.Glas@uk-essen.de
3   Department of Radiotherapy, University Hospital Essen, D-45147 Essen, Germany; nika.guberina@uk-essen.de
4   Department of Radiology, Division of Abdominal Imaging, A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard University Medical School, Boston, MA 02114, USA; ocatalano@mgh.harvard.edu
5   Department of Neurosurgery, University Hospital Essen, D-45147 Essen, Germany; Daniela.Pierscianek@uk-essen.de (D.P.); Karsten.Wrede@uk-essen.de (K.W.)
*   Correspondence: johannes.haubold@uk-essen.de; Tel.: +49-201-723-84528; Fax: +49-201-723-1548
†   Contributed equally.

Simple Summary: Over the past few years, radiomics-based tissue characterization has demonstrated its potential for non-invasive prediction of the genetic profile and grading in cerebral gliomas using multiparametric MRI. The aim of our study was to investigate the feasibility and diagnostic accuracy of a fully automated radiomics analysis based on a simplified MR protocol derived from various scanner systems to prospectively ease the transition of radiomics-based non-invasive tissue sampling into clinical practice. Using an MRI with non-contrast and post-contrast T1-weighted sequences and FLAIR, our workflow automatically predicts the IDH1/2 mutation, the ATRX expression loss, the 1p19q co-deletion and the MGMT methylation status. It also effectively differentiates low-grade from high-grade gliomas. In summary, the present study demonstrated that a fully automated prediction of grading and the genetic profile of cerebral gliomas could be performed with our proposed method using a simplified MRI protocol that is robust to variations in scanner systems, imaging parameters and field strength.

Abstract: Objective: The aim of this study was to investigate the diagnostic accuracy of a radiomics analysis based on a fully automated segmentation and a simplified and robust MR imaging protocol to provide a comprehensive analysis of the genetic profile and grading of cerebral gliomas for everyday clinical use. Methods: MRI examinations of 217 therapy-naïve patients with cerebral gliomas, each comprising a non-contrast T1-weighted, FLAIR and contrast-enhanced T1-weighted sequence, were included in the study. In addition, clinical and laboratory parameters were incorporated into the analysis. The BraTS 2019 pretrained DeepMedic network was used for automated segmentation. The segmentations generated by DeepMedic were evaluated with 200 manual segmentations with a DICE score of 0.8082 ± 0.1321. Subsequently, the radiomics signatures were utilized to predict the genetic profile of ATRX, IDH1/2, MGMT and 1p19q co-deletion, as well as differentiating low-grade glioma from high-grade glioma. Results: The network provided an AUC (validation/test) for the differentiation between low-grade gliomas vs. high-grade gliomas of $0.981 \pm 0.015/0.885 \pm 0.02$. The best results were achieved for the prediction of the ATRX expression loss with AUCs of $0.979 \pm 0.028/0.923 \pm 0.045$, followed by $0.929 \pm 0.042/0.861 \pm 0.023$ for the prediction of IDH1/2. The prediction of 1p19q and MGMT achieved moderate results, with AUCs of $0.999 \pm 0.005/0.711 \pm 0.128$ for 1p19q and $0.854 \pm 0.046/0.742 \pm 0.050$ for MGMT. Conclusion: This fully automated approach utilizing simplified MR protocols to predict the genetic profile and grading of cerebral gliomas provides an easy and efficient method for non-invasive tumor decoding.

1. Introduction

Recent developments in radiomics based image analysis have enabled new possibilities to non-invasively determine the histopathology and/or genetic profile of a number of different tumor entities [1]. This is especially important for cerebral gliomas, as the latest WHO classification of 2016 puts the focus on the genetic profile to classify tumors precisely and plan personalized therapy regimens accordingly [2]. While this change in tumor classification aims to improve patient management and treatment, it implies the requirement for invasive tissue sampling to the detriment of potentially severe biopsy and/or surgical complications, resulting in a mortality rate of 2.8% [3]. As a result, there is a high demand for non-invasive tissue analysis to initiate adequate therapy regimens without potential biopsy-associated side effects. In particular, this is true for inoperable tumors that would be solely biopsied. Over the past few years, a number of studies demonstrated the great potential of MRI-based image analyses for the tumor decoding of gliomas [4]. Here, various approaches to image analysis for predicting the genetic profile of gliomas have emerged in recent years [5–7]. On the one hand, radiomics features can be extracted from segmentations and be used to train machine learning algorithms [8–11]. On the other hand, the features can be extracted from a segmentation using convolutional neural networks. Here, Guta et al. were able to show that features extracted using CNN lead to better predictions for the grading of cerebral gliomas than features extracted using Radiomics [5]. Third, various networks have been established for direct image analysis of cerebral gliomas utilizing a variety of deep learning networks [6,12]. All approaches come with their own advantages and disadvantages. While the traditional approach with radiomics has traceable features, allowing them to be replicated in studies [13], the segmentation required and the often poorer prediction compared to CNN approaches is a major point of criticism [5]. In contrast, feature extraction by means of CNN often delivers better results [5], but these are less comprehensible and cannot be compared in studies to the same extent; simultaneously, the often manual or semi-automatic segmentation is a major point of criticism here as well. Direct image analysis using deep learning does not require manual segmentation and is therefore probably the easiest to use in day-to-day business. Still in this case there is only limited understanding of how the prediction is made. In addition to these major differences between methods of image analysis, there are common limitations that hinder the clinical use of these systems.

Most studies put the focus of the analysis on a limited number of mutations instead of providing a comprehensive analysis of the genetic profile [9,10,14]. Furthermore, whilst some recent studies show excellent accuracies for the prediction of mutations [4,14,15], the study protocols entail high-profile, dedicated multiparametric MR imaging as imaging platforms. These protocols can be challenging to perform and, moreover, difficult to generalize for common clinical practice. Lastly, there has been a shift from manual segmentation [16] to semi-automatic [14] to fully automated algorithms [17]. Nevertheless, there is still a great number of recent publications that apply manual/semi-automatic segmentation, again impairing the generalization and clinical applicability of radiomics.

Hence, to address the above-mentioned limitations, the aim of this study was to establish a radiomics analysis based on fully automated segmentations and a simplified and robust MR imaging protocol to provide a comprehensive analysis of the genetic profile and grading of cerebral gliomas for everyday clinical use.

2. Methods

2.1. Ethics Statement

This study was performed in adherence to all of the guidelines defined by the approving institutional review board of the investigating hospital. Written informed consent was waived by the Institutional Review Board due to the retrospective nature of the study. Complete anonymization of all of the data was performed prior to their inclusion in the study.

2.2. Study Design and Cohort

217 MRI studies of 217 patients with histopathologically confirmed therapy naïve cerebral gliomas were included in the study. The collective comprised 28 low-grade gliomas (WHO 2) and 187 high-grade gliomas (WHO III + WHO IV). Two cerebral gliomas could not be conclusively classified and were excluded from the analysis. The distribution of genetic parameters on the training and test collective, as well as the age and sex distribution, are shown in Table 1. Exclusion criteria were previous brain surgery, pretreatment of the brain tumor or lack of histopathologic confirmation. The following sequences were included in the radiomics analysis: (1) FLAIR, (2) non-contrast T1-weighted sequence and (3) a contrast-enhanced T1-weighted sequence. All patients with incomplete MRI examinations were excluded from the study. The distribution of genetic parameters in the collective is shown in Table 1. In addition to the MRI data, clinical and laboratory parameters were included in the analysis (please refer to the clinical feature section). Prior to the inclusion of MRI data and clinical features, all data were anonymized. Subsequent to this, the preprocessing of the MRIs was performed, including co-registration and skull stripping. Thereafter, the data were used to segment the gliomas in an automated process using DeepMedic [18]. Example results of the automated segmentation are shown in Figure 1. Radiomics features were subsequently extracted from the segmented data and used in conjunction with the clinical features to predict the genetic profile and grading of cerebral gliomas using machine learning. The workflow is shown in Figure 2.
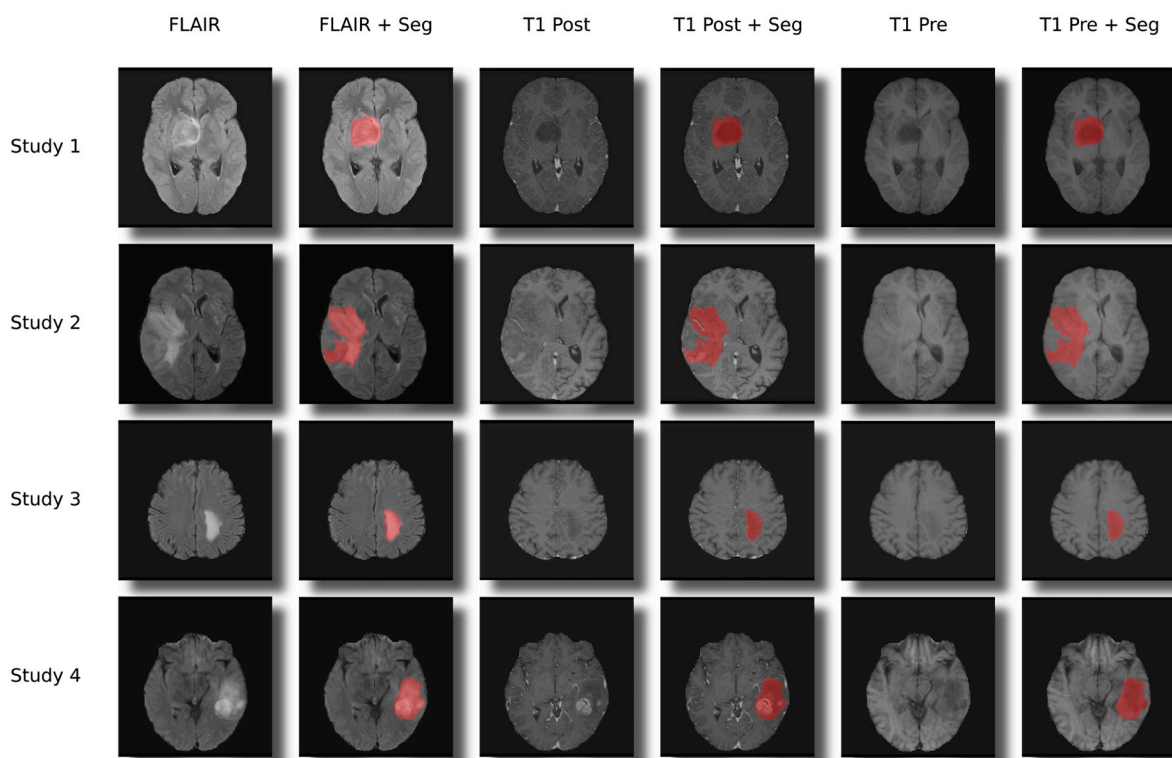

Figure 1. Examples for the segmentations and co-registration with other sequences.

Table 1. Initial patient collective: Distribution of the training, testing and total cohort of the respective genetic parameters, age and gender. A positive label is defined as having an IDH1/2 mutation, an ATRX expression loss, a 1p19q co-deletion and a positive MGMT methylation status. For the WHO label, LGG is defined as positive and HGG as negative. The gender distribution is expressed as the percentage of females in the collective.

| Label	Total (pos/neg)	Train (pos/neg), %Female, Mean Age + SD	Test (pos/neg) %Female, Mean Age + SD
MGMT	164 (81/83)	131 (65/66), 45.0%, 57.0 ± 14.3 years	33 (16/17), 30.3%, 60.6 ± 13.1 years
IDH1/2	145 (34/111)	116 (27/89), 37.9%, 55.5 ± 16.9 years	29 (7/22), 41.4%, 50.2 ± 17.8 years
1p19q	30 (5/25)	24 (4/20), 45.8%, 45.3 ± 14 years	6 (1/5), 33.3%, 56.7 ± 6.9 years
WHO	215 (28/187)	172 (22/150), 41.9%, 56.0 ± 16.2 years	43 (6/37), 41.9%, 52.6 ± 15.4 years
ATRX	67 (13/54)	53 (10/43), 32.1%, 56.8 ± 13.3 years	14 (3/11), 64.3%, 56.9 ± 9.5 years
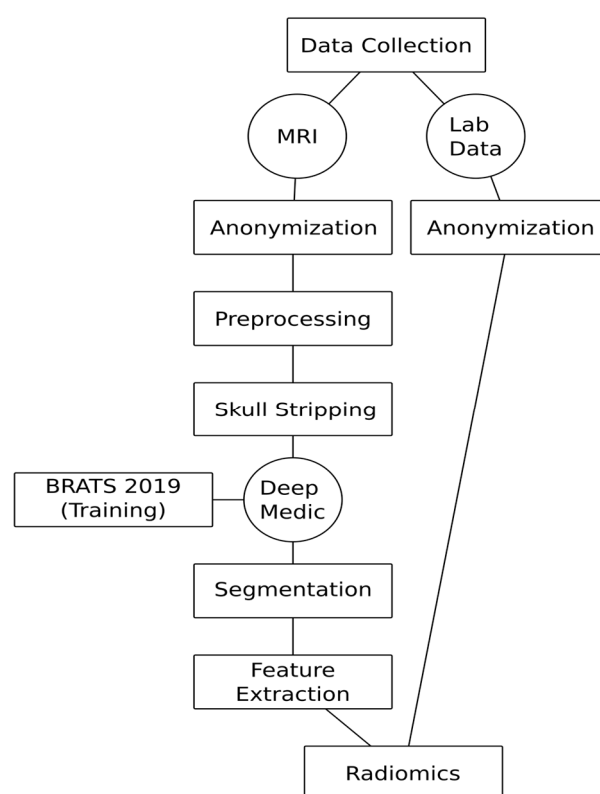
Figure 2. Automated virtual biopsy workflow.

2.3. Magnetic Resonance Imaging

The MRI examinations were performed at a single center on assorted 1.5 T and 3 T MR-machines (MAGNETOM Symphony, MAGNETOM Sonata, MAGNETOM Avanto, MAGNETOM Aera, MAGNETOM Skyra) from a single vendor (Siemens Healthineers, Munich, Germany). The distribution among the different MR scanners is shown in Table 2, and the distribution of the train and test set among the different field strengths is shown in Table 3.

Table 2. Distribution of examinations among MR scanners (all scanners are from Siemens Healthineers).

Scanner	1.5 T Magnetom Aera	3 T Magnetom Skyra	1.5 T Magnetom Espree	3 T Bio-graph_mMR	1.5 T Magnetom Avanto	1.5 T Magnetom Sonata	1.5 T Magnetom Symphony
n	68	57	35	19	19	12	10

Table 3. Distribution of the train and test among magnetic field strength.

Labels	Magnetic Field Strength	Genetic Profile				
		MGMT	IDH1/2	ATRX	1p19Q	WHO_HIGH
0 (train)	1.5	45	53	21	11	15
	3.0	21	36	22	9	7
1 (train)	1.5	38	14	6	2	76
	3.0	27	13	4	2	74
0 (test)	1.5	14	15	7	3	3
	3.0	3	7	4	2	3
1 (test)	1.5	10	4	2	0	26
	3.0	6	3	1	1	11

2.4. Preprocessing

Prior to executing the preprocessing, a brain extraction was performed to ensure an anonymized data set and to clean the image from unneeded structures. Due to its state-of-the-art performance, the publicly available HD-BET artificial neural network-based algorithm was used [19]. To spatially align the different series of a study, the SimpleITK extension SimpleElastix was used to co-register the series of a study as part of the preprocessing pipeline. A rigid registration, which is a form of a linear transformation, was carried out for all images using the post-contrast series as a fixed reference. The fully automated segmentation of the tumor regions was carried out using DeepMedic trained on the BraTS 2019 dataset, a Convolutional Neural Network (CNN) based algorithm [18].

2.5. Train Test Split

The full dataset was split into a training set and a test set with a ratio of 80:20, resulting in the number of studies in each set as stated in Table 1.

2.6. Clinical Features

The clinical features included gender, hemoglobin, c-reactive protein (CRP), sodium, calcium, total protein, platelets, aspartate aminotransferase (ASAT), lactate dehydrogenase (LDH) and bilirubin (total). The laboratory values were obtained within a time frame of 4 weeks prior to the MR examination. Listed laboratory data were available for all patients.

2.7. Radiomics Feature Extraction

To extract radiomics related features from the brain tumor images, the PyRadiomics package was used [20]. The extracted features comprise first-order statistics features, shape-based features, Gray Level Cooccurence Matrix (GLCM) features, Gray Level Run Length Matrix (GLRLM) features, Gray Level Size Zone Matrix (GLSZM) features, Neighbouring Gray Tone Difference Matrix (NGTDM) features and Gray Level Dependence Matrix (GLDM) features. Relevant features were additionally extracted from filter transformed images using Wavelet transformation, Laplacian of Gaussian (LoG) transformation, Local Binary Pattern 3D (LBP3D) transformation and Gradient transformation. In total, 1562 features were extracted for each series, adding up to a total of 4686 features per study.

2.8. Feature Selection

The number of extensive features leads to the necessity of selecting a subset of features with the highest predictive power to reduce the noise added by redundant features. For that purpose, the Boruta algorithm was utilized—a wrapper method that was combined with the Gradient Boosting algorithm XGBoost [21,22]. Subsets of features were selected with Boruta tuned to different thresholds, resulting in a relaxed selection process. Two different approaches of the Boruta Algorithm were used, one utilizing the permutation importance and the other based on the SHAP importance of the features in the underlying tree method. Table 4 shows the distribution of the three important features—volume, flatness and surface area—among the respective genetic profiles.

Table 4. Lesion-specific features for positive and negative samples and the Bonferroni corrected Mann-Whitney U statistics grouped by label.

	MGMT	**WHO**	**IDH 1/2**	**1p19q**	**ATRX**
Volume	neg: $9.9 \times 10^4 \pm$ 5.5×10^4 pos: $1.1 \times 10^5 \pm$ 6.3×10^4 $p = 0.340$	neg: $1 \times 10^5 \pm$ 5.8×10^4 pos: $5.8 \times 10^4 \pm$ 6.8×10^4 $p < 0.0001$	neg: $9.9 \times 10^4 \pm$ 5.8×10^4 pos: $9.1 \times 10^4 \pm$ 8.2×10^4 $p = 0.070$	neg: $9.1 \times 10^4 \pm$ 7×10^4 pos: $1 \times 10^5 \pm$ 1.2×10^5 $p = 0.434$	neg: $1.1 \times 10^5 \pm$ 6.5×10^4 pos: $9.2 \times 10^4 \pm$ 6.4×10^4 $p = 0.194$
Flatness	neg: 0.59 ± 0.12 pos: 0.58 ± 0.12 $p = 0.276$	neg: 0.58 ± 0.12 pos: 0.65 ± 0.1 $p = 0.002$	neg: 0.59 ± 0.13 pos: 0.63 ± 0.11 $p = 0.013$	neg: 0.57 ± 0.13 pos: 0.65 ± 0.091 $p = 0.110$	neg: 0.59 ± 0.13 pos: 0.62 ± 0.1 $p = 0.177$
Surface Area	neg: $2.1 \times 10^4 \pm$ 9.9×10^3 pos: $2.2 \times 10^4 \pm$ 1.2×10^4 $p = 0.358$	neg: $2.2 \times 10^4 \pm$ 1.1×10^4 pos: $1.3 \times 10^4 \pm$ 1.4×10^4 $p < 0.0001$	neg: $2.1 \times 10^4 \pm$ 1.1×10^4 pos: $1.8 \times 10^4 \pm$ 1.6×10^4 $p = 0.012$	neg: $2.1 \times 10^4 \pm$ 1.5×10^4 pos: $2.2 \times 10^4 \pm$ 2.4×10^4 $p = 0.348$	neg $= 2.2 \times 10^4 \pm$ 1.2×10^4 pos: $1.9 \times 10^4 \pm$ 1.2×10^4 $p = 0.177$

2.9. Parameter Optimization

The XGBoost parameters were tuned using the Tree-structured Parzen Estimator (TPE) sampler of the Optuna framework [23]. Each optimization run comprised 200 initial iterations with parameters picked randomly from a given parameter space before the final 2000 TPE steps were conducted. Each iteration of the optimization contained a bootstrapping-based cross-validation maximizing the logloss score of the validation data. The XGBoost model used with the Python API was initialized using the gbtree booster and the binary logistic objective within the parameter space given in Table 5.

Table 5. Parameter space used for hyperparameter optimization with the Tree-structured Parzen Estimator (TPE).

Hyperparameter	
n_estimators	[100, 1500] stepsize: 100
max_depth	[1, 6] stepsize: 1
learning_rate	[0.05, 0.03] stepsize: loguniform
gamma	[0, 20] stepsize: uniform
min_child_weight	[1, 20] stepsize: 1
subsample	[0.5, 1.0] stepsize: 0.05
colsample_bytree	[0.1, 1.0] stepsize: 0.05
reg_landa	$[1 \times 10^{-8}, 1.0]$ stepsize: loguniform
reg_landa	$[1 \times 10^{-8}, 1.0]$ stepsize: loguniform

2.10. Model Evaluation

For each label, the best result after the parameter optimization in terms of the AUC score of the validation set was picked. All model performance relevant metrics pre-

sented in this paper were calculated in a 1000fold bootstrapping cross-validation to obtain uncertainty intervals of the model performance. As the XGBoost algorithm returns an uncalibrated probability of a data point being of a given class, the sensitivity and specificity—depending on discrete outcomes—were calibrated using the geometric mean.

2.11. Validation of Segmentations

The performance of the DeepMedic segmentation Model was evaluated on 200 manually segmented patients. The cerebral gliomas were segmented manually by two radiology residents and inspected—and corrected—by the consensus of two radiology consultants. This resulted in a mean DICE score of 0.8082 ± 0.1321.

3. Results

Overall, all of the genetic parameters could be predicted with a good performance. The ROC curves of the predictions are shown in Figures 3 and 4, while the AUC values, accuracy, precision, sensitivity and specificity of the train, validation and test set for each genetic parameter and grading (LGG vs. HGG) are shown in Tables 6 and 7. All predictions were compared with the prediction of a dummy classifier in order to avoid possible errors due to an unbalanced collective.
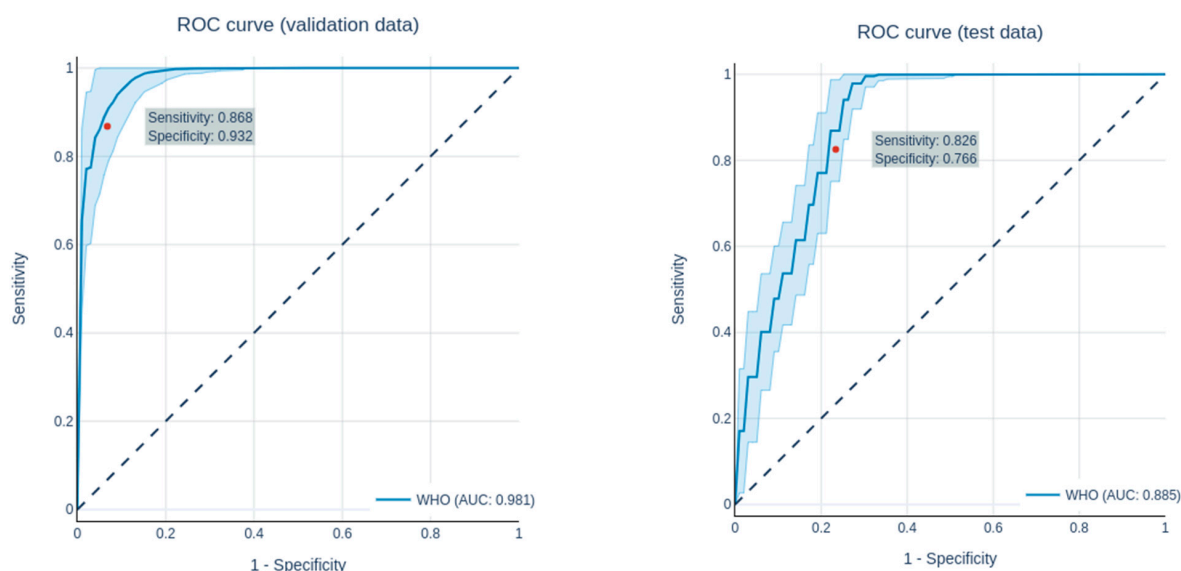

Figure 3. ROC curves of the prediction of the grading (LGG vs. HGG) in the validation data set (**left**) and the test data set (**right**).

Table 6. Performance in predicting the grading.

Dataset	AUC	Accuracy	Precision	Sensitivity	Specificity
Grading-Validation	0.981 ± 0.015	0.925 ± 0.042	0.685 ± 0.168	0.868 ± 0.035	0.932 ± 0.048
Grading-Test	0.885 ± 0.021	0.774 ± 0.032	0.369 ± 0.042	0.826 ± 0.035	0.766 ± 0.040

Table 7. Performance in predicting the genetic parameters.

| Dataset	AUC	Accuracy	Precision	Sensitivity	Specificity
ATRX-Validation	0.981 ± 0.028	0.914 ± 0.048	0.844 ± 0.193	0.721 ± 0.089	0.955 ± 0.060
ATRX-Test	0.923 ± 0.045	0.810 ± 0.066	0.593 ± 0.174	0.656 ± 0.059	0.853 ± 0.086
1p19q-Validation	0.999 ± 0.005	0.889 ± 0.032	0.521 ± 0.497	0.271 ± 0.260	0.999 ± 0.012
1p19q-Test	0.711 ± 0.128	0.611 ± 0.094	0.000 ± 0.000	0.000 ± 0.000	0.733 ± 0.113
IDH1/2-Validation	0.929 ± 0.042	0.888 ± 0.045	0.751 ± 0.131	0.798 ± 0.087	0.913 ± 0.059
IDH1/2-Test	0.861 ± 0.023	0.769 ± 0.053	0.544 ± 0.113	0.689 ± 0.095	0.795 ± 0.091
MGMT-Validation	0.854 ± 0.046	0.786 ± 0.044	0.807 ± 0.076	0.756 ± 0.094	0.815 ± 0.091
MGMT-Test	0.742 ± 0.050	0.699 ± 0.046	0.705 ± 0.077	0.684 ± 0.125	0.714 ± 0.117
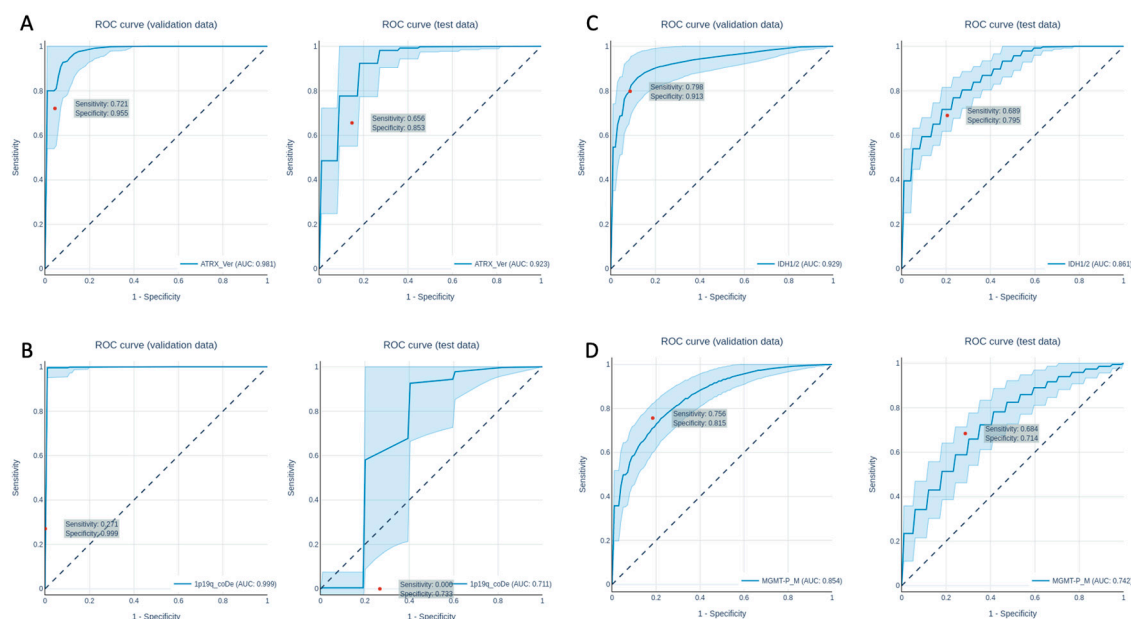
Figure 4. ROC curves of the prediction of the ATRX expression loss (**A**), the 1p19q co-deletion (**B**), the IDH1/IDH2 mutation (**C**) and the MGMT-status (**D**) in the validation data set (left) and the test data set (right).

The results of the dummy classifier are interpreted as the best results that can be achieved without training a model by guessing the prediction based on the class imbalance. A classifier was chosen that predicts the data by randomly classifying each data point as positive or negative while taking a possible class imbalance into account (stratified). Overall, all predictions of the dummy classifier were very close to an AUC of 0.5 (0.497, 0.498, 0.513, 0.500, 0.503), and, therefore, no capability of separating the classes from each other was observed.

3.1. Radiomics Analysis to Predict Grading

Grading in this study was limited to the clinically highly relevant differentiation between LGG (WHO 2) and HGG (WHO 3 + 4). Following the feature selection, 56 features were selected for the predictions. All in all, with an AUC of $1.0 \pm 0.000/0.981 \pm 0.015/0.885 \pm 0.02$ (training/validation/test), the network showed a good performance in differentiating between LGG and HGG. The performance in predicting grading in the validation and test datasets are shown in Table 6, while the ROC curves for the predictions of the validation and test datasets are shown in Figure 3.

3.2. Radiomics Analysis to Predict the Genetic Profile

The analysis of the genetic parameters showed overall good to very good results. In detail, the network performed well in predicting the ATRX expression with an AUC of $1.0 \pm 0.000/0.979 \pm 0.028/0.923 \pm 0.045$ (training/validation/test). A total of 18 features were used for prediction, which were selected by feature selection. The performance to predict the genetic parameters of the validation and test datasets are shown in Table 7. The ROC curves for the predictions of the validation and test datasets are shown in Figure 4.

For the prediction of the 1p19q co-deletion, the network showed a heterogeneous performance, with an AUC of $1.0 \pm 0.000/0.999 \pm 0.005/0.711 \pm 0.128$ (training/validation/test). The ROC curves for validation and test dataset predictions are plotted in Figure 4B. The predictions were based on 11 features, which were selected by feature selection.

The networks performance in predicting IDH1/2 mutation was quite good, with an AUC of $1.0 \pm 0.000/0.929 \pm 0.042/0.861 \pm 0.023$ (training/validation/test). The ROC curves for the validation and test dataset predictions are shown in Figure 4C. For predic-

tions, a number of 36 features were used. For the prediction of the MGMT methylation status, the network performed well, with an AUC of $1.0 \pm 0.000, 0.854 \pm 0.046, 0.742 \pm 0.050$ (training/validation/test) using 25 features, which were previously selected using feature selection. The ROC curves for the validation and test dataset predictions are shown in Figure 4D.

4. Discussion

In recent years, radiomics-based tissue characterization has demonstrated its potential for non-invasive prediction of the genetic profile and grading in cerebral gliomas using multiparametric MRI [9,17,24]. The benefits of this innovative tissue characterization form include its non-invasive nature and whole-tumor assessment when compared to focal stereotactic biopsy sampling, with an associated mortality rate of 2.8% [3]. Nevertheless, a number of challenges have remained: the majority of published studies are limited to the prediction of a genetic parameter and/or grading, lacking the convenience and comparability to an extensive invasive workup [9,25–27]. Additionally, previous studies are predominantly based on manual or semi-automatic segmentation methods, which significantly limits reproducibility and inter-study comparability [28]. Furthermore, these studies generally require a complex MRI protocol and standardized sequences and are mainly based on specific field strengths or specific scanners, impeding the transition from local unicenter radiomics-based tumor assessment to a more universal approach.

Hence, the main aim of our study was to address the majority of these limitations to prospectively ease the transition of radiomics-based non-invasive tissue sampling into clinical practice. Thus, our study comprised a collective of 217 patients acquired on seven different MRI scanners, entailing both 1.5 as well as 3 Tesla scanner systems. The radiomics analysis was based on a basic imaging protocol consisting of three universally applied sequences: a FLAIR, a non-contrast and a contrast-enhanced T1-weighted sequence with variable imaging parameters. Previous publications indicate a direct correlation of the performance power of radiomics to the homogeneity and complexity of MR protocols [4]. Nevertheless, neither homogeneous imaging parameters nor complex protocols reflect the global clinical approach to MR imaging. Thus, to understand whether radiomics-based tumor assessment can be applied in a broader scope and more universal approach, we took a potential loss in predictive power into account.

To account for reproducibility and to remove potential human bias or influence of any kind, we applied DeepMedic for the automated segmentation of cerebral gliomas, which had been pretrained with the BraTS 2019 dataset [18,29–31]. On this basis, feature selections of radiomics parameters were performed and used for training machine learning algorithms. During the feature selection, gender and laboratory parameters dropped out because of their poor predictive power, so only imaging parameters were used to train the final model. While gender is certainly a risk factor for cerebral gliomas [32], this feature does not appear to be relevant for differentiating genetic parameters and grading in our study collective. No data on association with grading and genetic parameters are available for the laboratory parameters used. Nevertheless, there are various studies on the improvement of predictive power in a variety of other tumor entities by laboratory parameters [33], so we wanted to test the predictive power for our study setting.

One of the most compelling challenges for radiomics-based tumor assessment is the comparability to the extensive invasive work up, as the majority of previously published studies focused on single parameter analyses [9,14]. Therefore, we included an extensive work up including the differentiation of high-grade versus low-grade gliomas and ATRX, IDH1/2, 1p19q, MGMT in our analysis. The clinical management of a patient is crucially dependent on the genetic profile and the differentiation of LGG vs. HGG [34,35]. For this purpose, in this study, the network was able to differentiate LGG from HGG with an AUC of $1.0 \pm 0.000/0.981 \pm 0.015/0.885 \pm 0.02$ (train/validation/test), yielding very good results, which are on a similar level or even better compared to other studies with more complex protocols and more homogeneous collectives [8,36]. Overall, good results were

obtained in predicting the genetic parameters, although these varied depending on the type of genetic parameter, as described previously in the literature [4,9,27,37].

For the predictions of ATRX, IDH1/2, 1p19q and MGMT, this resulted in an AUC (train/validation/test) of $1.0 \pm 0.000/0.979 \pm 0.028/0.923 \pm 0.045$ (ATRX), $1.0 \pm 0.000/0.929 \pm 0.042/0.861 \pm 0.023$ (IDH1/2), $1.0 \pm 0.000/0.999 \pm 0.005/0.711 \pm 0.128$ (1p19q) and $1.0 \pm 0.000, 0.854 \pm 0.046, 0.742 \pm 0.050$ (MGMT). Our results for the grading and prediction of ATRX and IDH1/2 are comparable and equally good in comparison to a previously published study on a highly-complex 18F-FET-PET/MRI study involving MR Fingerprinting as a new relaxometry technique [4]. This could be due to the fact that, in our current study, the collective was significantly larger. Thus, the predictive power of the ATRX expression loss and IDH1/2 mutation is similar to or higher than other recent publications, although fewer sequences were used in our study [9,27,38].

However, following the excellent accuracies for prediction of 1p19q co-deletion and IDH1/2 mutation, our results reveal a substantial drop as well as noticeably strong variation in the prediction of the 1p19q co-deletion in our test and validation collective. This is probably due to the fact that the 1p19q co-deletion is the least present mutation in the collective. Hence, this issue should be further investigated in a larger collective, entailing more patients with the 1p19q co-deletion.

Overall, the small number of patients with a 1p19q co-deletion in our collective reduces the predictive power. To address this, we performed a 1000fold bootstrapping cross-validation inside the hyperparameter optimization on all predictions. Overall, however, this remains a limitation, particularly regarding the 1p19q co-deletion, which can only be fully addressed in larger follow-up studies.

Furthermore, a substantial drop from the test to the validation collective is often observed in radiomics studies [27]. In our study, this might have been caused by the very heterogeneous collective consisting of scans from seven different MRI scanners. While some might see this as a limitation, it is also an advantage, as it improves transferability. Nevertheless, this can also be a sign of overfitting. To avoid overfitting, we used early stopping of the training process based on the validation score (early_stopping = 10), restricted the tree depth to a max of 6 and used a cross-validation inside the hyperparameter optimization.

Alternatively, according to the literature, additional T2 mapping may significantly increase the predictive power for the 1p19q co-deletion [4]. Lastly, in line with current literature, our network only enabled a moderate prediction of the MGMT mutation with AUCs of 0.742, hence, deeming the prediction of this parameter as non-clinical [39].

Overall, despite our promising results, our study is not free of limitations. One limitation is that, while our study was performed on different MRI scanners with 1.5 and 3 T, there is a lack of diversity from different manufacturers, which may alter the results. In addition, recent studies have shown that different methods of data normalization can improve the robustness of radiomics predictions in a heterogeneous collective, a technique that may further increase the efficiency of our workflow in the future [40]. Further limitations are the retrospective setting and the partly small sample size, e.g., for the 1p19q co-deletion, which was addressed by our extensive validation using the train/validation/test collective. These issues could be easily addressed in a follow-up multi-center setup using the proposed fully automated workflow.

Finally, a biopsy cannot be waived if a requirement to use the network is that a glioma is present. Although this is the largest group of primary cerebral neoplasms, other rarer entities must be differentiated from this. Therefore, further networks should be used to differentiate important differential diagnoses such as cerebral lymphoma. Although a biopsy in the setting of initial diagnosis of cerebral neoplasm cannot be avoided without the help of additional networks, the prediction of the genetic profile could help to early detect mutations from low-grade gliomas to higher-grade gliomas in follow-up examinations.

5. Conclusions

In summary, the present study demonstrated that a fully automated prediction of grading and the genetic profile of cerebral gliomas could be performed with our proposed method using a simplified MRI protocol that is robust to variations in scanner systems, imaging parameters and field strength.

Author Contributions: Conceptualization, J.H., K.W., M.F. and F.N.; data curation, J.H., D.P. and N.F.; formal analysis, R.H., V.P. and N.F.; investigation, J.H. and N.F.; methodology, J.H., R.H., F.N. and N.F.; project administration, J.H., M.F. and L.U.; resources, M.F. and L.U.; software, N.F.; supervision, F.N. and L.U.; validation, N.F. and L.U.; writing—original draft, J.H., N.F. and L.U.; writing—review & editing, J.H., R.H., M.G., V.P., N.G., O.A.C., D.P., K.W., C.D., M.F., F.N., N.F. and L.U. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the University Hospital Essen (Approval Code: 17-7395-BO; Approval date: 14 February 2017).

Informed Consent Statement: Written informed consent was waived by the Institutional Review Board due to the retrospective nature of the study. All data were anonymized before inclusion in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available for data protection reasons.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lohmann, P.; Galldiks, N.; Kocher, M.; Heinzel, A.; Filss, C.P.; Stegmayr, C.; Mottaghy, F.M.; Fink, G.R.; Jon Shah, N.; Langen, K.-J. Radiomics in neuro-oncology: Basics, workflow, and applications. *Methods* **2020**, *188*, 112–121. [CrossRef]
2. Louis, D.N.; Perry, A.; Reifenberger, G.; von Deimling, A.; Figarella-Branger, D.; Cavenee, W.K.; Ohgaki, H.; Wiestler, O.D.; Kleihues, P.; Ellison, D.W. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathol.* **2016**, *131*, 803–820. [CrossRef] [PubMed]
3. Malone, H.; Yang, J.; Hershman, D.L.; Wright, J.D.; Bruce, J.N.; Neugut, A.I. Complications Following Stereotactic Needle Biopsy of Intracranial Tumors. *World Neurosurg* **2015**, *84*, 1084–1089. [CrossRef] [PubMed]
4. Haubold, J.; Demircioglu, A.; Gratz, M.; Glas, M.; Wrede, K.; Sure, U.; Antoch, G.; Keyvani, K.; Nittka, M.; Kannengiesser, S.; et al. Non-invasive tumor decoding and phenotyping of cerebral gliomas utilizing multiparametric 18F-FET PET-MRI and MR Fingerprinting. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 1435–1445. [CrossRef]
5. Gutta, S.; Acharya, J.; Shiroishi, M.S.; Hwang, D.; Nayak, K.S. Improved Glioma Grading Using Deep Convolutional Neural Networks. *Am. J. Neuroradiol.* **2020**, *42*, 233–239. [CrossRef]
6. Cluceru, J.; Interian, Y.; Phillips, J.J.; Molinaro, A.M.; Luks, T.L.; Alcaide-Leon, P.; Olson, M.P.; Nair, D.; LaFontaine, M.; Shai, A.; et al. Improving the noninvasive classification of glioma genetic subtype with deep learning and diffusion-weighted imaging. *Neuro. Oncol.* **2021**, *23*, noab238. [CrossRef]
7. Kawaguchi, R.K.; Takahashi, M.; Miyake, M.; Kinoshita, M.; Takahashi, S.; Ichimura, K.; Hamamoto, R.; Narita, Y.; Sese, J. Assessing Versatile Machine Learning Models for Glioma Radiogenomic Studies across Hospitals. *Cancers* **2021**, *13*, 3611. [CrossRef] [PubMed]
8. Cho, H.; Lee, S.; Kim, J.; Park, H. Classification of the glioma grading using radiomics analysis. *PeerJ* **2018**, *6*, e5982. [CrossRef] [PubMed]
9. Li, Y.; Liu, X.; Qian, Z.; Sun, Z.; Xu, K.; Wang, K.; Fan, X.; Zhang, Z.; Li, S.; Wang, Y.; et al. Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature. *Eur. Radiol.* **2018**, *28*, 2960–2968. [CrossRef] [PubMed]
10. Arita, H.; Kinoshita, M.; Kawaguchi, A.; Takahashi, M.; Narita, Y.; Terakawa, Y.; Tsuyuguchi, N.; Okita, Y.; Nonaka, M.; Moriuchi, S.; et al. Lesion location implemented magnetic resonance imaging radiomics for predicting IDH and TERT promoter mutations in grade II/III gliomas. *Sci. Rep.* **2018**, *8*, 11773. [CrossRef] [PubMed]
11. Shofty, B.; Artzi, M.; Ben Bashat, D.; Liberman, G.; Haim, O.; Kashanian, A.; Bokstein, F.; Blumenthal, D.T.; Ram, Z.; Shahar, T. MRI radiomics analysis of molecular alterations in low-grade gliomas. *Int. J. Comput Assist. Radiol. Surg.* **2018**, *13*, 563–571. [CrossRef] [PubMed]

12. Yogananda, C.G.B.; Shah, B.R.; Nalawade, S.S.; Murugesan, G.K.; Yu, F.F.; Pinho, M.C.; Wagner, B.C.; Mickey, B.; Patel, T.R.; Fei, B.; et al. MRI-Based Deep-Learning Method for Determining Glioma MGMT Promoter Methylation Status. *AJNR Am. J. Neuroradiol.* **2021**, *42*, 845–852. [CrossRef] [PubMed]

13. Pati, S.; Verma, R.; Akbari, H.; Bilello, M.; Hill, V.B.; Sako, C.; Correa, R.; Beig, N.; Venet, L.; Thakur, S.; et al. Reproducibility analysis of multi-institutional paired expert annotations and radiomic features of the Ivy Glioblastoma Atlas Project (Ivy GAP) dataset. *Med. Phys.* **2020**, *47*, 6039–6052. [CrossRef]

14. Akbari, H.; Bakas, S.; Pisapia, J.M.; Nasrallah, M.P.; Rozycki, M.; Martinez-Lage, M.; Morrissette, J.J.D.; Dahmane, N.; O'Rourke, D.M.; Davatzikos, C. In vivo evaluation of EGFRvIII mutation in primary glioblastoma patients via complex multiparametric MRI signature. *Neuro-Oncology* **2018**, *20*, 1068–1079. [CrossRef] [PubMed]

15. Kebir, S.; Weber, M.; Lazaridis, L.; Deuschl, C.; Schmidt, T.; Mönninghoff, C.; Keyvani, K.; Umutlu, L.; Pierscianek, D.; Forsting, M.; et al. Hybrid 11C-MET PET/MRI Combined With "Machine Learning" in Glioma Diagnosis According to the Revised Glioma WHO Classification 2016. *Clin. Nucl. Med.* **2019**, *44*, 214–220. [CrossRef] [PubMed]

16. Zinn, P.O.; Singh, S.K.; Kotrotsou, A.; Hassan, I.; Thomas, G.; Luedi, M.M.; Elakkad, A.; Elshafeey, N.; Idris, T.; Mosley, J.; et al. A co-clinical radiogenomic validation study—Conserved magnetic resonance radiomic appearance of Periostin expressing Glioblastoma in patients and xenograft models. *Clin. Cancer Res.* **2018**, *24*, 6288–6299. [CrossRef]

17. Zhuge, Y.; Ning, H.; Mathen, P.; Cheng, J.Y.; Krauze, A.V.; Camphausen, K.; Miller, R.W. Automated glioma grading on conventional MRI images using deep convolutional neural networks. *Med. Phys.* **2020**, *47*, 3044–3053. [CrossRef] [PubMed]

18. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.J.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [CrossRef] [PubMed]

19. Isensee, F.; Schell, M.; Pflueger, I.; Brugnara, G.; Bonekamp, D.; Neuberger, U.; Wick, A.; Schlemmer, H.-P.; Heiland, S.; Wick, W.; et al. Automated brain extraction of multisequence MRI using artificial neural networks. *Hum. Brain Mapp.* **2019**, *40*, 4952–4964. [CrossRef]

20. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [CrossRef] [PubMed]

21. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13. [CrossRef]

22. Chen, T.; Guestrin, C. *XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.

23. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 2623–2631.

24. Zhao, S.-S.; Feng, X.-L.; Hu, Y.-C.; Han, Y.; Tian, Q.; Sun, Y.-Z.; Zhang, J.; Ge, X.-W.; Cheng, S.-C.; Li, X.-L.; et al. Better efficacy in differentiating WHO grade II from III oligodendrogliomas with machine-learning than radiologist's reading from conventional T1 contrast-enhanced and fluid attenuated inversion recovery images. *BMC Neurol.* **2020**, *20*, 48. [CrossRef] [PubMed]

25. Peng, H.; Huo, J.; Li, B.; Cui, Y.; Zhang, H.; Zhang, L.; Ma, L. Predicting Isocitrate Dehydrogenase (IDH) Mutation Status in Gliomas Using Multiparameter MRI Radiomics Features. *J. Magn. Reson. Imaging* **2020**, *53*, 1399–1407. [CrossRef]

26. Kong, Z.; Jiang, C.; Zhang, Y.; Liu, S.; Liu, D.; Liu, Z.; Chen, W.; Liu, P.; Yang, T.; Lyu, Y.; et al. Thin-Slice Magnetic Resonance Imaging-Based Radiomics Signature Predicts Chromosomal 1p/19q Co-deletion Status in Grade II and III Gliomas. *Front. Neurol.* **2020**, *11*, 551771. [CrossRef]

27. Li, Z.-C.; Bai, H.; Sun, Q.; Li, Q.; Liu, L.; Zou, Y.; Chen, Y.; Liang, C.; Zheng, H. 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] [PubMed]

28. Qiu, Q.; Duan, J.; Duan, Z.; Meng, X.; Ma, C.; Zhu, J.; Lu, J.; Liu, T.; Yin, Y. Reproducibility and non-redundancy of radiomic features extracted from arterial phase CT scans in hepatocellular carcinoma patients: Impact of tumor segmentation variability. *Quant. Imaging Med. Surg.* **2019**, *9*, 45364–45464. [CrossRef] [PubMed]

29. Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R.T.; Berger, C.; Ha, S.M.; Rozycki, M.; et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. *arXiv* **2019**, arXiv:1811.02629.

30. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [CrossRef]

31. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.S.; Freymann, J.B.; Farahani, K.; Davatzikos, C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **2017**, *4*, 170117. [CrossRef]

32. Lopes-Ramos, C.M.; Quackenbush, J.; DeMeo, D.L. Genome-Wide Sex and Gender Differences in Cancer. *Front. Oncol.* **2020**, *10*, 2486. [CrossRef] [PubMed]

33. Liu, L.; Yi, X.; Lu, C.; Qi, L.; Zhang, Y.; Li, M.; Xiao, Q.; Wang, C.; Zhang, L.; Pang, Y.; et al. Applications of radiomics in genitourinary tumors. *Am. J. Cancer Res.* **2020**, *10*, 2293–2308.

34. Forst, D.A.; Nahed, B.V.; Loeffler, J.S.; Batchelor, T.T. Low-Grade Gliomas. *Oncologist* **2014**, *19*, 403–413. [CrossRef] [PubMed]

35. Sim, H.-W.; Morgan, E.R.; Mason, W.P. Contemporary management of high-grade gliomas. *CNS Oncol.* **2018**, *7*, 51–65. [CrossRef] [PubMed]

36. Zhang, X.; Yan, L.-F.; Hu, Y.-C.; Li, G.; Yang, Y.; Han, Y.; Sun, Y.-Z.; Liu, Z.-C.; Tian, Q.; Han, Z.-Y.; et al. Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features. *Oncotarget* **2017**, *8*, 47816–47830. [CrossRef]

37. Zhang, S.; Sun, H.; Su, X.; Yang, X.; Wang, W.; Wan, X.; Tan, Q.; Chen, N.; Yue, Q.; Gong, Q. Automated machine learning to predict the co-occurrence of isocitrate dehydrogenase mutations and O6 -methylguanine-DNA methyltransferase promoter methylation in patients with gliomas. *J. Magn. Reson. Imaging* **2021**, *54*, 197–205. [CrossRef]

38. Ren, Y.; Zhang, X.; Rui, W.; Pang, H.; Qiu, T.; Wang, J.; Xie, Q.; Jin, T.; Zhang, H.; Chen, H.; 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] [PubMed]

39. Xi, Y.-B.; Guo, F.; Xu, Z.-L.; Li, C.; Wei, W.; Tian, P.; Liu, T.-T.; Liu, L.; Chen, G.; Ye, J.; 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] [PubMed]

40. Carré, A.; Klausner, G.; Edjlali, M.; Lerousseau, M.; Briend-Diop, J.; Sun, R.; Ammari, S.; Reuzé, S.; Alvarez Andres, E.; Estienne, T.; et al. Standardization of brain MR images across machines and protocols: Bridging the gap for MRI-based radiomics. *Sci. Rep.* **2020**, *10*, 12340. [CrossRef]