Magnetic Resonance Imaging 63 (2019) 244–249

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.elsevier.com/locate/mri

Original contribution

Comparison of machine learning classifiers for differentiation of grade 1 from higher gradings in meningioma: A multicenter radiomics study

Gordian Hamerla[a,*], Hans-Jonas Meyer[b], Stefan Schob[a], Daniel T. Ginat[c], Ashley Altman[c], Tchoyoson Lim[d], Georg Alexander Gihr[e], Diana Horvath-Rizea[e], Karl-Titus Hoffmann[a], Alexey Surov[b]

[a] Department of Neuroradiology, University of Leipzig, Leipzig, Germany
[b] Department of Diagnostic and Interventional Radiology, University of Leip-zig, Leipzig, Germany
[c] University of Chicago, Pritzker School of Medicine, Chicago, IL, USA
[d] Department of Neuroradiology, National Neuroscience Institute, Singapore
[e] Clinic for Neuroradiology, Katharinenhospital Stuttgart, Stuttgart, Germany

ARTICLE INFO

Keywords:
Random forest
Support vector machine
Multilayer perceptron
XGBoost
Machine learning
Meningioma
Grading
Feature selection

ABSTRACT

Background and purpose: Advanced imaging analysis for the prediction of tumor biology and modelling of clinically relevant parameters using computed imaging features is part of the emerging field of radiomics research. Here we test the hypothesis that a machine learning approach can distinguish grade 1 from higher gradings in meningioma patients using radiomics features derived from a heterogenous multicenter dataset of multi-paramedic MRI.

Methods: A total of 138 patients from 5 international centers that underwent MRI prior to surgical resection of intracranial meningiomas were included. Segmentation was performed manually on co-registered multi-parametric MR images using apparent diffusion coefficient (ADC) maps, T1-weighted (T1), post-contrast T1-weighted (T1c), subtraction maps (Sub, T1c – T1), T2-weighted fluid-attenuated inversion recovery (FLAIR) and T2-weighted (T2) images. Feature selection was performed and using cross-validation to separate training from testing data, four machine learning classifiers were scored on combinations of MRI modalities: random forest (RF), extreme gradient boosting (XGBoost), support vector machine (SVM) and multilayer perceptron (MLP).

Results: The best AUC of 0.97 (1.0 and 0.97 for sensitivity and specificity) was observed for the combination of ADC, ADC of the peritumoral edema, T1, T1c, Sub and FLAIR-derived features using only 16 of the 10,914 possible features and XGBoost.

Conclusions: Machine learning using radiomics features derived from multi-parametric MRI is capable of high AUC scores with high sensitivity and specificity in classifying meningiomas between low and higher gradings despite heterogeneous protocols across different centers. Feature selection can be performed effectively even when extracting a large amount of data for radiomics fingerprinting.

1. Introduction

Meningiomas are the most frequent tumors of the central nervous system, representing about 37% of all tumors. They are generally considered to be of benign nature [1]. According to the WHO classification they can be subdivided into three grades reflecting their histological features as well as recurrence rates and clinical outcome [2]. The most common grade I meningioma (81%) has a recurrence rate of 20–39% in 10 years, grade II shows a recurrence rate of 50% at five years. Notably, the anaplastic grade III meningiomas are rare (2%) and show an outstanding morbidity and mortality with a median overall survival from 2.6–5.8 years [3–5]. The most common treatment strategies involve watch and wait, surgical resection and radiotherapy, and treatment planning is based on the diagnosis by imaging studies [6].

Recently, advanced imaging analysis using the extraction of computed features together with machine learning techniques to assess genomic, proteomic and clinical phenotypes of various solid tumor entities were evaluated, exceeding the qualitative imaging analysis by

Abbreviations: MLP, multilayer perceptron; RF, random forest; Sub, subtraction map (T1c – T1); SVM, support vector machine; T1c, contrast-enhanced T1-weighted imaging; WHO, World Health Organization

* Corresponding author.
E-mail address: gordian.hamerla@medizin.uni-leipzig.de (G. Hamerla).

https://doi.org/10.1016/j.mri.2019.08.011
Received 16 February 2019; Received in revised form 14 July 2019; Accepted 15 August 2019

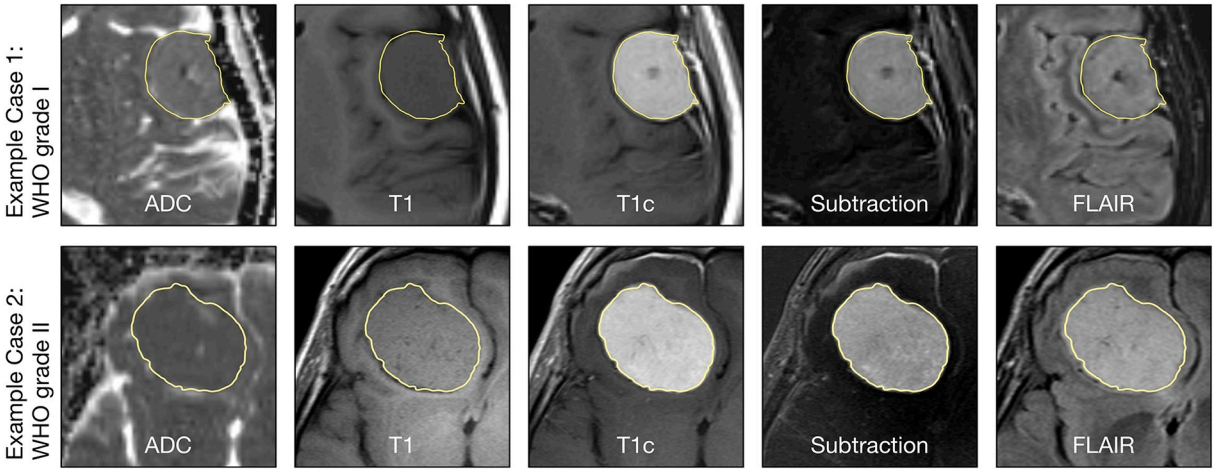

Fig. 1. Illustrative example of the tumor segmentation on co-registered MRI modalities. In meningioma, visual distinction of low-grade and high-grade tumors can be challenging, as there are no universal imaging markers. Both cases of meningioma, originating from the temporal dura and the frontal falx respectively, presented with intense contrast enhancement, relatively homogenous signal intensities and no relevant perifocal cerebral edema. Histopathological examination of the surgical specimen revealed different tumor gradings with consequence to further treatment.

the radiologist [7–10].

Numerous studies observed associations between radiomics features and tumor biology or were able to model clinically relevant parameters. For instance, Aerts et al. were able to predict survival in two different cohorts of lung and head and neck cancer patients, outperforming the TNM system as the current gold standard [11]. Grossmann et al. could derive treatment-relevant tumor biology from volumetric features using multi-parametric MRI in glioblastoma [12].

Regarding meningiomas, mean apparent diffusion coefficient (ADC) values derived from diffusion weighted imaging [13,14] could be shown to inversely correlate with cellularity and Ki-67 index [15–17]. Apart from the mean value other, first-order histogram-derived features were proposed as surrogates for histopathological grading [18]. However, there are also recent studies on other tumor entities indicating that even conventional, morphological sequences, such as T1- and T2-weighted images are correlated with cellularity and Ki-67 index, when investigated with histogram based analysis [19–21]. This suggests that the combination of conventional MRI techniques with diffusion-derived ADC or other functional parameters may potentially be useful for tumor grading.

Recent advances in artificial neural networks opened the possibility to analyze medical imaging data. Deep neural networks can be used for both image classification as well as generation of quantitative features [22,23].

The hypothesis of this study is that a machine learning approach using radiomics features derived from multi-parametric MRI can predict tumor grading in meningioma patients despite heterogeneous imaging protocols. Furthermore, the relation of the score to the use of different combinations of MRI modalities will be explored.

2. Materials and methods

2.1. Patients and imaging

161 cases of meningioma who underwent preoperative imaging followed by surgical resection of an intracranial meningioma were retrospectively collected from 5 international centers (Center 1: Martin-Luther-University Halle-Wittenberg, Halle/Saale, Germany; Center 2: Katharinenhospital Stuttgart, Stuttgart, Germany; Center 3: University of Leipzig, Leipzig, Germany; Center 4: University of Chicago, Pritzker School of Medicine, Chicago, Illinois, USA; Center 5: National Neuroscience Institute, Singapore). In the initial screening, 23 cases were excluded due to incomplete data, insufficient imaging quality (e.g. movement artifacts), divergent imaging parameters unsuitable for analysis (e.g. non-axial orientation, fat saturation), small tumor size (< 2 cm in diameter) or imaging suggesting recurrence on tumor site (evidence of craniotomy). A total of 147 patients could be included in this study (mean age: 61.9 years; SD: 12.0; male/female: 39/108). Information on the histopathological grading of the tumor according to routine oncologic care was acquired and WHO grading was dichotomized to grade I vs. grades II and III (G1/G2-3: 102/45). While all scans were performed on 1.5 T MRI scanners, the imaging protocols were subject to the standards of the individual study centers, resulting in a largely inhomogeneous dataset (see Supplementary Table S1).

2.2. Preprocessing, segmentation and feature extraction

To minimize the influence of the inhomogeneous dataset due to different scanners and protocols, the images underwent preprocessing: Correction for magnetic field inhomogeneities [24], intensity normalization by centering at mean with standard deviation [25] and resampling to $1 \times 1 \times 1$ mm voxel spacing. For each case, the available subset of the following imaging series were co-registered per subject: ADC map, T1-weighted (T1), contrast-enhanced T1-weighted (T1c), T2-weighted, fluid-attenuated inversion recovery (FLAIR) and T2-weighted (T2) images. Subtraction maps (Sub, T1c – T1) were computed automatically if applicable imaging series of equal parameters were available. Registration was executed manually with visual inspection (rigid transformation, 6 degrees of freedom) and focus on good alignment on the tumor borders using MeVisLab 3.0.2 software (MeVis Medical Solutions AG, Bremen, Germany). Following this first step, manual segmentation of the whole tumor volume was performed (GH, resident Neuroradiology, 3 years of experience) using a digital pen tablet as input device (Wacom Europe GmbH, Düsseldorf, Germany). During segmentation, care was taken to include only voxels that resemble tumor tissue, avoiding intraosseus spread, cysts, surrounded vessels or other structures as well as partial volume effects (see Fig. 1). It has been shown earlier that this strategy results in good inter-observer reliability in meningioma segmentation [26]. For semi-automatic segmentation of the peritumoral white matter edema, a threshold-based region growing algorithm was used visually checked on FLAIR or T2 images, where available.

The extraction of radiomics features was carried out using the previously reported PyRadiomics library implemented in Python, preserving the settings for possible future reproducibility [25,27]. A full set of tumor image features was calculated for each MRI sequence,

including shape, first-order (histogram-derived) and texture features using the original as well as derived images (square, square root, logarithm, exponential, Laplacian of Gaussian, wavelet reconstructions, gradient and logical binary pattern), resulting in 1819 features per MRI sequence. For the peritumoral edema region only ADC map-derived features were extracted (EdemaADC). Altogether, up to 12,733 features in total per case were possible.

2.3. Supervised machine learning, feature selection and scoring

For the analysis of the radiomics dataset an in-house developed Python-based software was used together with open-source statistics modules [28].

A random forest classifier (RF) consists of a set of tree-shaped classifiers that contain of a number of independent identically distributed random vectors [29,30]. Within the decision tree each node is a split function that uses the best in a subset of the available features randomly chosen at that node to split the dataset. This way a set of the largest possible trees is grown and thus trained for the classification task at hand. Subsequent classification of a testing case by the trained tree set is determined by the majority vote of the ensemble.

XGBoost (eXtreme Gradient Boosting) is another tree-based classification algorithm where an ensemble of decision trees is build [31]. During training, instead of splitting the dataset at random new trees are added to the ensample that correct the errors of the existing tree models. A computationally efficient stochastic gradient decent algorithm is used to minimize the error when adding new trees.

A support vector machine (SVM) classifier iteratively forms a hyperplane or a set of hyperplanes in high-dimensional feature space that separates the clinical classes [30,32]. During the process the distance of the separating planes to the data points of the different classes is maximized in order to minimize the classification error.

Multilayer perceptron (MLP) refers to a feed-forward deep artificial neural network. Due to the non-linear activation function of the individual units, the MLP is able to model complex non-linear relationships. For this classification task an MLP with two densely-connected hidden layers was constructed. The dimension of the layers was selected according to prior explorative testing and set to a modest size of 128 and 64 units in order to gain a model that is able to generalize well under the given circumstances and to avoid overfitting. Adam was selected as Optimizer, error function was binary crossentropy on the internal validation dataset, a random sample of 20% out of the training data.

Training machine learning algorithms on imbalanced datasets might lead to biased classifiers. If the validation dataset suffers from the same imbalance of classes, an overly optimistic estimation of the classification performance is likely. To address this potential issue, Synthetic Minority Oversampling Technique (SMOTE) was applied when training all four classifiers [33].

Training and testing of the patient sample was performed by splitting the dataset into n = 10 equally-sized subsets where one is reserved for testing and the others are combined for training. The process is repeated n times and the average score is taken. This process ensures no overlap between training and testing datasets in each iteration and the classifier is always scored on unseen test data. This is commonly referred to as n-fold cross validation (CV) and can be used to achieve reliable results in small datasets. For scoring the classifier, receiver operating characteristic (ROC) analysis was performed and the mean area under the curve (AUC) of 100 repetitions with 10-fold CV used for comparison.

For feature selection, all available features were ranked according to Mann-Whitney *U* test first. The optimal number of features was determined by the maximum mean AUC of 4-fold CV random forest classification after stepwise increment of the number of used features in increasing order of test p value. To optimize computation time, the process to determine the optimal number of features stopped when a declining score over 10 iterations was reached. Due to the large number of features obtained from the multimodal imaging dataset, feature selection was conducted prior to the classifier cross validation. While this may introduce potential for information leak from the testing to the training dataset, this approach was chosen in order to avoid considerable overfitting and hence a loss in the ability to generalize during cross validation due to the variation in feature selection. This way, a selection of the minimum required features could be ensured while preserving diversity over the numerous types of radiomics features.

Since not all MRI sequences were available for each case, various subsets were selected and analyzed separately in order to keep the maximum possible number of cases in the analysis. For each of the selected subsets a list of the most important features was generated.

This whole analysis was conducted for all combinations of the different classifiers and the chosen imaging modality subsets.

3. Results

The best performance in grading classification was observed for the XGBoost classifier for the combination of features derived from ADC, EdemaADC, FLAIR, Sub, T1, T1c (n = 43, AUC = 0.97, 95% CI: 0.85–1.00), yielding 0.90 and 0.97 as sensitivity and specificity respectively. For each of the four classifiers, this is the combination with the highest score. The combination of ADC, Sub, T1 and T1c performs second best for random forest, XGBoost and SVM and on par with the best combination for MLP. The lowest scores were achieved by the SVM using EdemaADC (n = 75, AUC = 0.29, 95% CI: 0.26–0.33) and the MLP using Sub alone (n = 99, AUC = 0.62, 95% CI: 0.45–0.76) and ADC and EdemaADC combined (n = 74, AUC = 0.64, 95% CI: 0.53–0.74). The highest sensitivities of 1.0 were produced by the combination of ADC, EdemaADC, FLAIR, Sub, T1 and T1c (XGBoost), and T2 (XGBoost, MLP, SVM). The results are summarized in Tables 2 and 3 and Fig. 2.

Of the radiomics features derived from EdemaADC only one wavelet-based feature was selected (EdemaADC_wavelet-HHH_glszm_ZoneEntropy). Since this limits the use of the machine learning classifiers, additional analysis was performed: in a separately conducted ROC analysis the single EdemaADC feature scored an AUC of 0.70 with a sensitivity and specificity of 0.62 and 0.78 respectively.

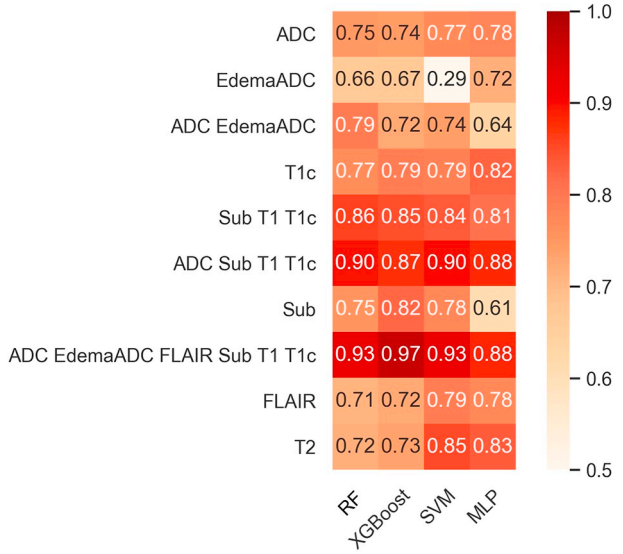

Fig. 2. Heatmap of AUC scores of the four machine learning classifiers by combination of MRI modalities.

Table 1
Demographic and histopathological data of the study cohort by center.

	Center 1	Center 2	Center 3	Center 4	Center 5
n	63	47	12	16	9
Age (years ± SD)	61.1 ± 11.9	61.0 ± 10.4	62.8 ± 9.7	62.0 ± 16.7	65.1 ± 10.6
Gender (male/female)	17/46	10/37	2/10	7/9	3/6
Histopathological grading					
WHO Grade I	56	35	11	0	0
WHO Grade II	7	12	1	16	9

4. Discussion

The best results for grading classification were achieved using a combination of radiomics features derived from ADC, T1, T1c, Sub and FLAIR imaging of the tumor together with ADC of the peritumoral white matter edema (see Table 1). The analysis of this combination comes at the cost of case numbers since the MRI protocols most commonly used for preoperative evaluation of tumor location and extent mainly focus on contrast-enhanced T1. Other techniques might be more subject to individual decisions and depend on the clinically relevant questions at hand. Consequently, only 43 of the overall 138 Cases could be included for the analysis of the best-performing combination of MRI sequences. The best AUC of 0.97 (0.91–1.0 CI) with sensitivity and specificity of 100% and 97% exceeds previously reported grading scores: Using first-order ADC-derived features, Gihr et al. observed 0.91 AUC in a series of 37 patients [18]. Surov et al. reported an AUC of 0.8 using mean tumor ADC alone in a sample of 389 cases [34]. By means of diffusion kurtosis imaging (DKI), a diffusion model different to ADC, and positron emission tomography (PET) AUCs of 0.84 and 0.81 were reported [35,36]. Using deep learning-based imaging features of post-contrast T1 and a random forest classifier, Zhu et al. gained an AUC of 0.81 for meningioma grading [23].

Coroller et al. proposed a model for prediction of meningioma grading based on a combination of semantic and radiomics features derived from contrast-enhanced T1w [37]. Notably, the AUC of 0.77 reported for using radiomics features only matches the score in this study (random forest classifier, T1c) despite the fact that different feature selection algorithms were applied. The additional use of semantic imaging features yields insights into the relevant characteristics of high-grade tumors that are recognizable and interpretable by human readers without requiring additional software. In their univariate analysis Coroller et al. could show that visually assessed intratumoral heterogeneity yields the biggest odds radio for identification of high-grade tumors. Radiomics analysis can be used to objectify and quantify tissue heterogeneity in MRI imaging studies. The combined use of radiomics and visual features showed an AUC of 0.86.

In this study we aimed to relate the classification score to different combinations of MR imaging modalities. In contrast to visual features, diverse MRI sequences may reflect different, complementary information on tissue pathophysiology. For instance the ADC could be related to Ki-67 expression in different tumor entities [38] and post-contrast T1 imaging may contain information on tumor heterogeneity on a cellular level due to the distribution of the contrast agent (both gadobutrol and gadopentetat-dimeglumine were used) in the extracellular space [39]. The MRI signal alterations as well as the change in diffusivity patterns of peritumoral white matter might reflect different growth and infiltration patterns of the tumor, thus, yielding additional information [40–42].

To avoid false positive results in radiomics studies, a sample size of at least 10–15 patients per feature has been suggested [43]. This way, using a large number of radiomics features can quickly result in a practically unfeasible amount of cases for study planning. In our sample this requirement is met by some combinations of MRI sequences (see Table 2) where the AUC of 0.90 is the best and observed in combination of features derived from ADC, T1, T1c and Sub using only 5 features in 86 patients.

From the subset of EdemaADC-derived features only one wavelet-based feature was selected (EdemaADC_wavelet-HHH_glszm_ZoneEntropy). While this is not a standard scenario for machine learning classification, the MLP performs on par with the separately conducted ROC analysis. SVM is the only classifier to perform below random baseline in this case.

There are several limitations to this study: First, its retrospective design with the consecutive selection bias. Second, the imbalance of the two classes (WHO Grade I:103, WHO Grade II/III: 44) presents a potential drawback regarding classification performance. While this was addressed by application of SMOTE oversampling, residual classification bias towards the majority class (WHO grade I) is possible. Third, due to the limited sample size, there is the possibility of overfitting and likely a considerable generalization error when tested on out of sample cases. Still, high AUC scores could be achieved despite the fact that two of the five centers contributed only high-grade tumors. This suggests sufficient ability of the classifiers to generalize within this multicenter dataset and identify high-grade tumors without strict limitation by diverse imaging protocols. Fourth, there is still a wide range in the analyses and preprocessing steps applied by various researchers, including model selection and optimization as well as feature selection so results might not be interchangeable. There is a need for standardization in radiomics research in order to obtain results that are more comparable while allowing for dataset-specific model optimization. Furthermore, in order to include a diverse range of MRI modalities, studies with a slice thickness of up to 7 mm had to be resampled to isotropic voxel sizes. This is likely to have limited the classification performance and might introduce a bias. Hence, different preprocessing strategies might lead to

Table 2
Group composition and best classification results per subgroup.

MRI modalities	n	WHO Grades I/II–III	Radiomics features	Selected features	Best AUC	Best classifier
ADC	136	103/33	1819	4	0.78	MLP
EdemaADC	75	46/29	1819	1	0.72	MLP
ADC EdemaADC	74	45/29	3638	26	0.79	RF
T1c	147	103/44	1819	4	0.82	MLP
Sub T1 T1c	98	66/32	5457	27	0.86	RF
ADC Sub T1 T1c	86	66/20	7276	5	0.90	SVM & RF
Sub	99	66/33	1819	8	0.82	XGBoost
ADC EdemaADC FLAIR Sub T1 T1c	43	25/18	10,914	16	0.97	XGBoost
FLAIR	91	59/32	1819	17	0.79	SVM
T2	41	30/11	1819	6	0.85	SVM

Table 3
Classification results per subgroup. AUC: area under the receiver operating characteristic curve. CI: confidence interval for 95% confidence level. Sens.: sensitivity. Spec.: specificity.

MRI modalities	Random forest				XGBoost				Support vector machine				Multilayer perceptron			
	AUC	CI	Sens.	Spec.	AUC	CI	Sens.	Spec.	AUC	CI	Sens.	Spec.	AUC	CI	Sens.	Spec.
ADC	0.75	[0.67–0.83]	0.90	0.72	0.74	[0.65–0.83]	0.85	0.75	0.77	[0.74–0.80]	0.91	0.73	0.78	[0.70–0.86]	0.85	0.79
EdemaADC	0.66	[0.54–0.79]	0.83	0.72	0.67	[0.51–0.83]	0.83	0.67	0.29	[0.26–0.33]	0.33	0.80	0.69	[0.58–0.79]	0.76	0.81
ADC EdemaADC	0.79	[0.69–0.89]	0.90	0.77	0.72	[0.58–0.87]	0.70	0.89	0.74	[0.70–0.78]	0.83	0.78	0.64	[0.53–0.74]	0.87	0.65
T1c	0.77	[0.69–0.85]	0.82	0.73	0.79	[0.73–0.86]	0.75	0.89	0.79	[0.76–0.81]	0.80	0.82	0.82	[0.77–0.88]	0.86	0.77
Sub T1 T1c	0.86	[0.77–0.95]	0.92	0.83	0.85	[0.76–0.93]	0.83	0.91	0.84	[0.81–0.87]	0.87	0.85	0.81	[0.71–0.91]	0.87	0.81
ADC Sub T1 T1c	0.90	[0.79–1.00]	0.90	0.94	0.87	[0.78–0.97]	0.90	0.91	0.90	[0.87–0.93]	0.95	0.90	0.88	[0.80–0.96]	0.95	0.87
Sub	0.75	[0.64–0.87]	0.81	0.78	0.82	[0.71–0.93]	0.89	0.81	0.78	[0.74–0.81]	0.80	0.84	0.61	[0.45–0.76]	0.72	0.77
ADC EdemaADC FLAIR Sub T1 T1c	0.93	[0.85–1.00]	0.90	0.97	0.97	[0.91–1.00]	1.00	0.97	0.93	[0.90–0.95]	0.95	0.94	0.88	[0.80–0.97]	0.85	0.97
FLAIR	0.71	[0.58–0.84]	0.94	0.70	0.72	[0.63–0.82]	0.88	0.67	0.79	[0.76–0.82]	0.91	0.76	0.78	[0.69–0.86]	0.83	0.82
T2	0.72	[0.53–0.90]	0.90	0.80	0.73	[0.56–0.90]	1.00	0.73	0.85	[0.81–0.89]	1.00	0.85	0.83	[0.70–0.97]	1.00	0.83

improved results and lower generalization error. Fifth, this study only included patients without history of previous treatment. The setting of recurrent tumor is of particular clinical interest and further studies are needed to assess the possible use of machine learning and radiomics in post-treatment imaging studies.

5. Conclusion

Supervised machine learning using radiomics features derived from multi-parametric MRI is capable of high AUC scores with high sensitivity and specificity in classifying meningiomas between low and higher grades despite heterogeneous protocols across different centers. Feature selection and reduction is an essential step that can be performed effectively, even when extracting a large number of features for radiomics fingerprinting.

Author contributions

Conception and experimental design: A.S., G.H. Data Collection: A.S., H.-J.M., S.S., D.T.G., A.A., T.L., G.A.G., D.H.R. Image processing (registration and segmentation): GH. Data analysis and interpretation: G.H., A.S. Drafting of the main manuscript: G.H. Critical revision of the manuscript: H.J.M., S.S., D.T.G., T.L., K.T.H., A.S., All authors reviewed the manuscript.

Declaration of competing interest

The Authors disclosed no relevant relationships.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mri.2019.08.011.

References

[1] Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C, et al. CBTRUS Statistical Report: primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. Neuro Oncol 2017;19:v1–88. https://doi.org/10.1093/neuonc/nox158.

[2] 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. https://doi.org/10.1007/s00401-016-1545-1.

[3] Cao X, Hao S, Wu Z, Wang L, Jia G, Zhang L, et al. Survival rates, prognostic factors and treatment of anaplastic meningiomas. J Clin Neurosci 2015;22:828–33. https://doi.org/10.1016/j.jocn.2014.11.022.

[4] Moliterno J, Cope WP, Vartanian ED, Reiner AS, Kellen R, Ogilvie SQ, et al. Survival in patients treated for anaplastic meningioma. J Neurosurg 2015;123:23–30. https://doi.org/10.3171/2014.10.JNS14502.

[5] Rosenberg LA, Prayson RA, Lee J, Reddy C, Chao ST, Barnett GH, et al. Long-term experience with World Health Organization grade III (malignant) meningiomas at a single institution. Int J Radiat Oncol Biol Phys 2009;74:427–32. https://doi.org/10.1016/j.ijrobp.2008.08.018.

[6] Apra C, Peyre M, Kalamarides M. Current treatment options for meningioma. Expert Rev Neurother 2018;0:1–24. https://doi.org/10.1080/14737175.2018.1429920.

[7] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278:563–77. https://doi.org/10.1148/radiol.2015151169.

[8] Lambin P, Rios Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441–6. https://doi.org/10.1016/j.ejca.2011.11.036.

[9] Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton JD, Snyder A, et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. Clin Radiol 2016:1–8. https://doi.org/10.1016/j.crad.2016.09.013.

[10] Yip SSF, Aerts HJWL. Applications and limitations of radiomics. Phys Med Biol 2018:150–66. https://doi.org/10.1088/0031-9155/61/13/R150.

[11] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006https://doi.org/10.1038/ncomms5006.

[12] Grossmann P, Gutman DA, Dunn WD, Holder CA, Aerts HJWL. Imaging-genomics reveals driving pathways of MRI derived volumetric tumor phenotype features in Glioblastoma. BMC Cancer 2016:1–10. https://doi.org/10.1186/s12885-016-2659-5.

[13] Le Bihan D, Iima M. Diffusion magnetic resonance imaging: what water tells us about biological tissues. PLoS Biol 2015;13:e1002203–13. https://doi.org/10.1371/journal.pbio.1002203.

[14] Le Bihan D. Apparent diffusion coefficient and beyond: what diffusion MR imaging can tell us about tissue structure. Radiology 2013;268:318–22. https://doi.org/10.1148/radiol.13130420.

[15] Surov A, Caysa H, Wienke A, Spielmann RP, Fiedler E. Correlation between different ADC fractions, cell count, Ki-67, total nucleic areas and average nucleic areas in meningothelial meningiomas. Anticancer Res 2015;35:6841–6. https://doi.org/10.1016/S0140-6736(04)16153-9.

[16] Ginat DT, Mangla R, Yeaney G, Wang HZ. Correlation of diffusion and perfusion MRI with Ki-67 in high-grade meningiomas. AJR Am J Roentgenol 2010;195:1391–5. https://doi.org/10.2214/AJR.10.4531.

[17] Surov A, Meyer HJ, Wienke A. Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: a meta-analysis. Oncotarget 2017;8:59492–9. https://doi.org/10.18632/oncotarget.17752.

[18] Gihr GA, Horvath-Rizea D, Garnov N, Kohlhof-Meinecke P, Ganslandt O, Henkes H, et al. Diffusion profiling via a histogram approach distinguishes low-grade from high-grade meningiomas, can reflect the respective proliferative potential and progesterone receptor status. 2018. p. 1–9. https://doi.org/10.1007/s11307-018-1166-2.

[19] Meyer HJ, Schob S, Höhn AK, Surov A. MRI texture analysis reflects histopathology parameters in thyroid cancer - a first preliminary study. Transl Oncol 2017;10:911–6. https://doi.org/10.1016/j.tranon.2017.09.003.

[20] Chang Y-W, Kwon KH, Choi DL, Lee DW, Lee MH, Lee HK, et al. Magnetic resonance imaging of breast cancer and correlation with prognostic factors. Acta Radiol 2009;50:990–8. https://doi.org/10.3109/02841850903225180.

[21] Meyer HJ, Schob S, Münch B, Frydrychowicz C, Garnov N, Quäschling U, et al. Histogram analysis of T1-weighted, T2-weighted, and postcontrast T1-weighted images in primary CNS lymphoma: correlations with histopathological findings-a preliminary study. Mol Imaging Biol 2017;113:97https://doi.org/10.1007/s11307-017-1115-5.

[22] Shaver MM, Kohanteb PA, Chiou C, Bardis MD, Chantaduly C, Bota D, et al. Optimizing neuro-oncology imaging: a review of deep learning approaches for glioma imaging. Cancers (Basel) 2019;11:829. https://doi.org/10.3390/cancers11060829.

[23] Zhu Y, Man C, Gong L, Dong D, Yu X, Wang S, et al. A deep learning radiomics model for preoperative grading in meningioma. Eur J Radiol 2019;116:128–34. https://doi.org/10.1016/j.ejrad.2019.04.022.

[24] Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK:

improved N3 bias correction. IEEE Trans Med Imaging 2013;29:1310–20. https://doi.org/10.1109/TMI.2010.2046908.
[25] Hosny A, van Griethuysen JJM, Parmar C, Aerts HJWL, Fedorov A, Beets-Tan RGH, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104–7. https://doi.org/10.1158/0008-5472.can-17-0339.
[26] Surov A, Hamerla G, Meyer HJ, Winter K, Schob S, Fiedler E. Whole lesion histogram analysis of meningiomas derived from ADC values. Correlation with several cellularity parameters, proliferation index KI 67, nucleic content, and membrane permeability. Magn Reson Imaging 2018;51:158–62. https://doi.org/10.1016/j.mri.2018.05.009.
[27] Wu S, Zheng J, Li Y, Wu Z, Shi S, Huang M, et al. Development and validation of an MRI-based radiomics signature for the preoperative prediction of lymph node metastasis in bladder cancer. EBioMedicine 2018;34:76–84. https://doi.org/10.1016/j.ebiom.2018.07.029.
[28] Jones E, Oliphant E, Peterson P. SciPy: open source scientific tools for python. http://www.scipy.org/; 2001.
[29] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[30] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[31] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min New York, NY, USA: ACM; 2016. p. 785–94. https://doi.org/10.1145/2939672.2939785.
[32] Chang C, Lin C, Tieleman T. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2008;307:1–39. https://doi.org/10.1145/1961189.1961199.
[33] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321–57. https://doi.org/10.1613/jair.953.
[34] Surov A, Ginat DT, Sanverdi E, Lim CCT, Hakyemez B, Yogi A, et al. Use of diffusion weighted imaging in differentiating between maligant and benign meningiomas. a multicenter analysis. World Neurosurg 2016;88:598–602. https://doi.org/10.1016/j.wneu.2015.10.049.
[35] Okuchi S, Okada T, Yamamoto A, Kanagaki M, Fushimi Y, Okada T, et al. Grading meningioma. Medicine (Baltimore) 2015;94:e549-6https://doi.org/10.1097/MD.0000000000000549.
[36] Lin L, Bhawana R, Xue Y, Duan Q, Jiang R, Chen H, et al. Comparative analysis of diffusional kurtosis imaging, diffusion tensor imaging, and diffusion-weighted imaging in grading and assessing cellular proliferation of meningiomas. Am J Neuroradiol 2018;39:1032–8. https://doi.org/10.3174/ajnr.A5662.
[37] Coroller TP, Bi WL, Huynh E, Abedalthagafi M, Aizer AA, Greenwald NF, et al. Radiographic prediction of meningioma grade by semantic and radiomic features. PLoS One 2017;12:e0187908–15. https://doi.org/10.1371/journal.pone.0187908.
[38] Surov A, Meyer HJ, Wienke A. Correlation between minimum apparent diffusion coefficient (ADCmin) and tumor cellularity: a meta-analysis. Anticancer Res 2017;37:3807–10. https://doi.org/10.21873/anticanres.11758.
[39] Bellin MF, Van Der Molen AJ. Extracellular gadolinium-based contrast media: an overview. Eur J Radiol 2008;66:160–7. https://doi.org/10.1016/j.ejrad.2008.01.023.
[40] Zakaria R, Das K, Radon M, Bhojak M, Rudland PR, Sluming V, et al. Diffusion-weighted MRI characteristics of the cerebral metastasis to brain boundary predicts patient outcomes. BMC Med Imaging 2014;14:613–49. https://doi.org/10.1186/1471-2342-14-26.
[41] Berghoff AS, Rajky O, Winkler F, Bartsch R, Furtner J, Hainfellner JA, et al. Invasion patterns in brain metastases of solid cancers. Neuro Oncol 2013;15:1664–72. https://doi.org/10.1093/neuonc/not112.
[42] Lee EJ, terBrugge K, Mikulis D, Choi DS, Bae JM, Lee SK, et al. Diagnostic value of peritumoral minimum apparent diffusion coefficient for differentiation of glioblastoma Multiforme from solitary metastatic lesions. Am J Roentgenol 2011;196:71–6. https://doi.org/10.2214/AJR.10.4752.
[43] Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. PLoS One 2015;10:e0124165-18https://doi.org/10.1371/journal.pone.0124165.