Received: 23 June 2022 | Revised: 16 March 2023 | Accepted: 21 March 2023

DOI: 10.1111/vru.13242

ORIGINAL INVESTIGATION

WILEY

Machine learning predicts histologic type and grade of canine gliomas based on MRI texture analysis

Pablo Barge[1] | **Anna Oevermann**[2] | **Arianna Maiolini**[3] | **Alexane Durand**[1]

[1]Division of Clinical Radiology, Department of Clinical Veterinary Science, Vetsuisse Faculty, University of Bern, Bern, Switzerland

[2]Division of Neurological Sciences, Department of Clinical Research and Veterinary Public Health, Vetsuisse Faculty, University of Bern, Bern, Switzerland

[3]Division of Clinical Neurology, Department of Clinical Veterinary Science, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Correspondence
Pablo Barge, Division of Clinical Radiology, Department of Clinical Veterinary Science, Vetsuisse Faculty, University of Bern, Länggassstrasse 128, 3012 Bern, Switzerland.
Email: pablo.barge@unibe.ch

Abstract

Conventional MRI features of canine gliomas subtypes and grades significantly overlap. Texture analysis (TA) quantifies image texture based on spatial arrangement of pixel intensities. Machine learning (ML) models based on MRI-TA demonstrate high accuracy in predicting brain tumor types and grades in human medicine. The aim of this retrospective, diagnostic accuracy study was to investigate the accuracy of ML-based MRI-TA in predicting canine gliomas histologic types and grades. Dogs with histopathological diagnosis of intracranial glioma and available brain MRI were included. Tumors were manually segmented across their entire volume in enhancing part, non-enhancing part, and peri-tumoral vasogenic edema in T2-weighted (T2w), T1-weighted (T1w), FLAIR, and T1w postcontrast sequences. Texture features were extracted and fed into three ML classifiers. Classifiers' performance was assessed using a leave-one-out cross-validation approach. Multiclass and binary models were built to predict histologic types (oligodendroglioma vs. astrocytoma vs. oligoastrocytoma) and grades (high vs. low), respectively. Thirty-eight dogs with a total of 40 masses were included. Machine learning classifiers had an average accuracy of 77% for discriminating tumor types and of 75.6% for predicting high-grade gliomas. The support vector machine classifier had an accuracy of up to 94% for predicting tumor types and up to 87% for predicting high-grade gliomas. The most discriminative texture features of tumor types and grades appeared related to the peri-tumoral edema in T1w images and to the non-enhancing part of the tumor in T2w images, respectively. In conclusion, ML-based MRI-TA has the potential to discriminate intracranial canine gliomas types and grades.

KEYWORDS
artificial intelligence, dog, glial cell neoplasm, radiomics, tumor heterogeneity

1 | INTRODUCTION

Conventional MRI features of canine gliomas types and grades largely overlap.[1–4] These intra-axial tumors have a predominant localization within the frontal, temporal, and parietal lobes.[5] They most commonly show ovoid to irregular shape, well to poorly defined margins, T2-weighted (T2w) hyperintense, T1-weighted (T1w) iso- to hypointense signal, variable degrees of contrast enhancement, intratumoral hemorrhage, peri-tumoral edema, and mass effect.[1,6] Oligodendrogliomas have been reported to more likely contact the

brain surface and distort the ventricular system, to have smoother margins and T1w hypointense signal compared to astrocytomas.[2–4,7] Neoplastic spread into neighboring brain structures and contrast enhancement have been reported more commonly in high-grade than low-grade gliomas.[2–4,7] However, despite the broad description of canine gliomas MRI features, the evaluation of previously defined conventional MRI predictors[3,4,8] yielded a sensitivity and specificity of 58.8% and 68.8%, respectively, for the diagnosis of astrocytoma, and 67.2% and 57.1% for the diagnosis of a high-grade gliomas in dogs.[2] The low sensitivities and specificities may be explained by the restricted ability of the human eye to objectively detect subtle heterogeneous features of specific tissues or lesions.[9]

Texture analysis (TA), on the contrary, mathematically extracts quantitative information from medical images, providing objective measurements of tumor heterogeneity.[10,11] It is based on the assumption that biomedical images contain information reflecting specific diseases or lesions,[10,12,13] describing the spatial arrangement of pixel intensities and providing a representation of tumor heterogeneity within a specific region of interest (ROI).[9] Texture features can be divided into shape-based, first-order, and second-order texture features.[12] Shape-based features describe the geometrical characteristics of ROIs. First-order texture features are histogram intensity-based features that describe the distribution of each pixel or voxel intensity without considering their spatial relationships and second-order texture features study pixels or voxels interrelationships.[12] These texture features can serve as an input to machine learning (ML) models,[14] and offer an alternative to the prediction of different types of pathologies. Machine learning is a subfield of artificial intelligence that provides computers the ability to learn from experience by pattern recognition,[15] aiming to find hard-to-discern patterns in large amounts of data.[12]

In small animal medicine, TA, with or without ML, has proven to be valuable in differentiating radiographic lung patterns[16], histologic grade of meningiomas on MRI[17], healthy from abnormal lung[18] or malignant from benign hepatic lesions[19] on computed tomography, and inflammatory brain lesions from gliomas on MRI.[20] However, a TA-based random forest (RF) classifier was unable to correctly discriminate canine gliomas types and grades.[20] In this study,[20] a single segment across a single slice including the entire tumor and excluding the perilesional edema was used for texture features extraction. In human medicine, multiple studies have demonstrated the benefit of using multiple segments across the entire tumor volume, as well as using different ML models to classify neoplasia or predict their outcome, especially in the field of neuroradiology.[21–24] Different ML models were able to classify human glioma grades based on TA with an average area under the curve (AUC) of 0.9.[23]

We hypothesized that using multiple segments across the entire tumor volume, which has not been previously performed in canine patients, would allow a more accurate ML classification of canine glioma types and grades based on MRI-TA compared to conventional MRI. Therefore, the aim of our study was to test different ML classifiers based on MRI-TA for this purpose.

2 | MATERIALS AND METHODS

The study design is summarized in Figure 1.

2.1 | Selection and description of subjects

For this retrospective diagnostic accuracy study, the medical records of a single academic institution (Vetsuisse Faculty, University of Bern) were searched for dogs with histopathological diagnosis of intra-axial glioma, for which a brain MRI was performed. Approval for use of medical data was provided by the owner's informed consent prior hospital

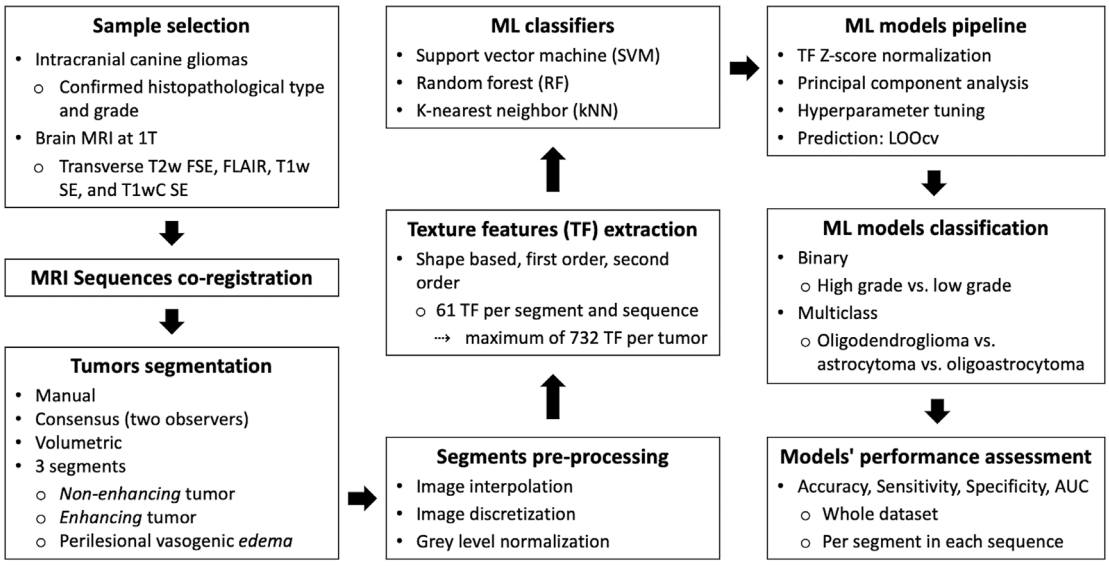

FIGURE 1 Flowchart of the study design.

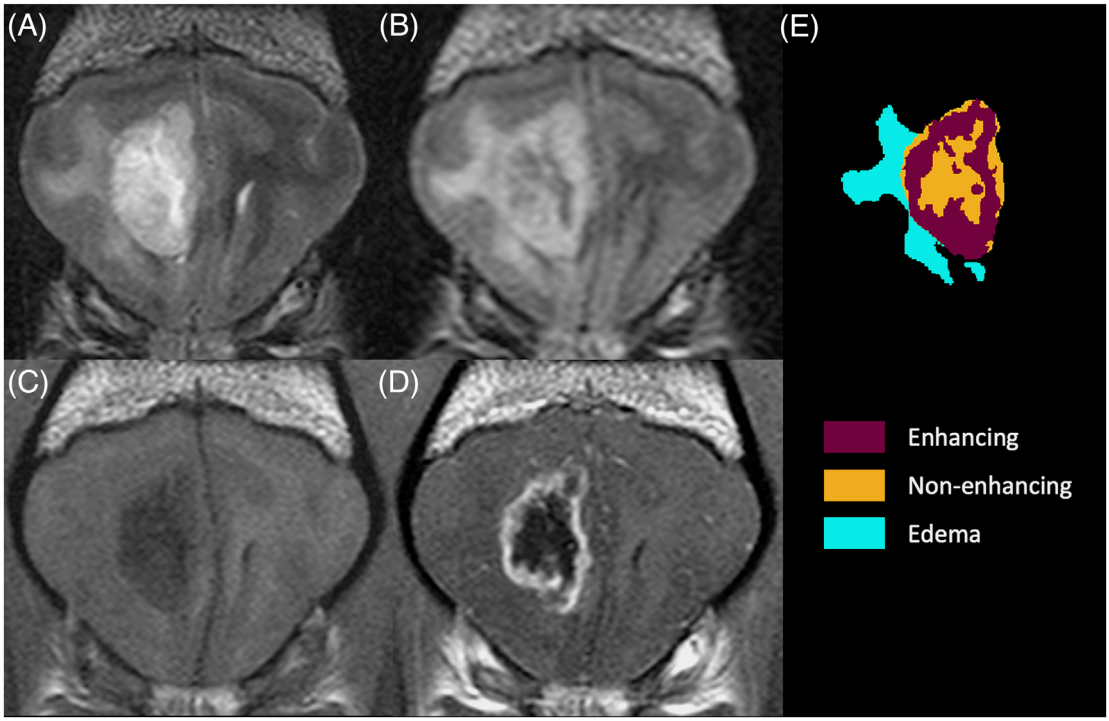

FIGURE 2 T2-weighted (A) (TE, 100 ms; TR, 5921.69 ms; slice thickness, 3.5 mm), FLAIR (B) (TE, 140 ms; TR, 11000 ms; TI, 2600 ms; slice thickness, 3.5 mm), and T1-weighted pre- (C) and post-contrast (D) (TE, 15 ms; TR, 411.58 ms; slice thickness, 3.5 mm) transverse images at the level of the caudate nuclei, of a dog with a high-grade oligoastrocytoma. In each MRI slice, the tumor was segmented in enhancing, non-enhancing, and vasogenic edema segments (E) to extract the texture features. [Color figure can be viewed at wileyonlinelibrary.com]

admission. Patients were included if the histopathological diagnosis was obtained by a board-certified pathologist based on postmortem examination or stereotactic brain biopsy. The pathology report had to include the glioma type (oligodendroglioma, astrocytoma, oligoastrocytoma) and grade (II, III, IV) based on the World Health Organization classification.[5,25] MRI examinations had to be performed using the same scanner (Panorama High Field Open 1.0 Tesla; Philips Medical Systems), and transverse T2w fast spin echo, fluid-attenuated inversion recovery (FLAIR), pre- and postcontrast (obtained immediately after intravenous administration of 0.2 ml/kg of gadoterate meglumine [ClariscanTM, 0.5 mmol/mL, GE Healthcare Pharmaceutical, Chicago, IL, USA]) T1w spin echo sequences had to be available for inclusion. The type of coil, time of echo (TE), repetition time (TR), time of inversion (TI), slice thickness, and presence or absence of artifacts in each sequence were recorded for each dog. If a dog demonstrated more than one mass with confirmed histopathological diagnosis of glioma, each mass was included independently.

2.2 | Data recording and analysis

2.2.1 | Magnetic resonance images co-registration and segmentation

Images were exported as Digital Imaging and Communications in Medicine files and converted into Neuroimaging Informatics Technology Initiative format. T2w, FLAIR, and T1w precontrast images were co-registered to the T1w post-contrast sequence using an available software (3DSlicer, v4.11.2, slicer.org).[26]

Three-dimensional segmentations of the enhancing and non-enhancing parts of the tumor, and of the perilesional vasogenic edema were manually performed by consensus among a second-year diagnostic imaging resident (P.B.) and a board-certified veterinary radiologist (A.D.) using 3DSlicer Segment Editor module. The "enhancing" segment was defined as areas within the tumor that displayed T1w postcontrast enhancement compared to T1w precontrast sequence. The "non-enhancing" segment was defined as tumor areas that did not exhibit T1w post-contrast enhancement. The vasogenic "edema" segment was defined as T2w and FLAIR hyperintense areas within the peri-tumoral white matter (normal or deformed due to mass effect) that was more extensive than the lesion itself. Meninges, large vessels, intratumoral cystic regions with a fully suppressing FLAIR signal, and areas of partial volume averaging artifact were not included in any segment. The volumetric segmentations were then exported as label-maps and used to extract texture features (Figure 2).

2.2.2 | Pre-processing and texture features extraction

Pre-processing techniques and texture features extraction were performed with an available freeware software (LIFEx;v6;

www.lifexsoft.org).[27] Before extraction, each segment in each sequence was pre-processed, with the goal of homogenizing images with respect to pixel spacing, gray-level intensities, and bins of gray-level histograms. Image interpolation was performed setting the common in-plane resolution to 1×1 mm while maintaining the original slice thickness. For image discretization, a fixed bin number of 32 was applied,[28–30] and the MRI gray levels within the segmentations were normalized to the mean ± 3 standard deviations.[9] Sixty-one texture features were then extracted per segment in each sequence, leading to a maximum of 732 texture values per tumor.

2.2.3 | Machine learning models

Supervised ML algorithms require labeled data with the known ground truth (i.e., histopathology). Support Vector Machine (SVM), RF, and k-nearest neighbors (kNN) are supervised ML classifiers used for both classification and regression problems. An SVM maps each data point into a dimensional feature space and aims to find a hyperplane that will separate the data points into different classes (e.g., high-grade vs. low-grade).[31] RF classifiers consist of a combination of decision trees that classify data into different classes,[32] and kNN models cluster data points into different groups and assigns a class to unseen data based on similarities with prior examples from the training set.[33] For each of these classifiers, a high-level programming language script (Python, v3.6.9) was written by a second-year diagnostic imaging resident with prior training in artificial intelligence in medical imaging (P.B.), using different Python libraries (numpy, pandas, and scikit-learn[34]) and run on Google Colab.

Before feeding the models, normalization of the texture values and dimensionality reduction were performed. Because texture values do not have unit and their ranges vary among different types, a *z*-score normalization was applied to improve their numerical stability, such as their distribution would have a mean of 0 and a standard deviation of 1.[28] Principal Component Analysis (PCA) was performed to reduce the dimensionality of the dataset (e.g., number of texture variables) while maximizing the variance and minimizing the information loss.[35,36] It transforms the data into a new coordinate system while retaining the most useful information,[36] with the aim to make the classifier's task substantially easier. Given the expected low sample size, a leave-one-out cross-validation (LOOcv) technique was chosen to assess the classifiers performance. A grid search cross-validation was implemented using LOOcv to normalize the data, find the required number of principal components and best hyperparameters for the three classifiers.

Different datasets were prepared to feed the models. One included all extracted texture features, from the three different segments and four different sequences (*whole dataset*). Smaller datasets included texture features of independent segments from independent sequences (e.g., non-enhancing segment in T2w sequence). This led to a total of 13 datasets.

TABLE 1 Distribution of 40 glioma types and grades among the 38 dogs included.

	Low-grade	High-grade	Total
Oligodendroglioma	6	19	**25**
Astrocytoma	4	3	**7**
Oligoastrocytoma	3	3	**6**
Oligosarcoma	0	2	**2**
Total	**13**	**27**	**40**

2.3 | Statistics and assessment of ML classifiers performance

Once the best combination of principal components and hyperparameters was optimized based on the supplied dataset using LOOcv, predictions were obtained for each classifier. Multiclass classification models aimed to predict gliomas histologic type (oligodendroglioma vs. astrocytoma vs. oligoastrocytoma). The predictions were compared to the histopathological diagnosis and overall accuracy, sensitivity, and specificity for each class were calculated from confusion matrices.[37] Binary classification models aimed to predict gliomas histologic grade (high-grade [grade III or IV] vs. low-grade [grade II]).[5,25] Based on the predictions, from confusion matrices, accuracy, sensitivity, specificity, and AUC were calculated.[37]

3 | RESULTS

Thirty-eight dogs met the inclusion criteria. The population mean age was 8.6 years (median: 9 years; range: 5 months to 13 years). Thirteen French Bulldogs, eight Boxers, three Labrador Retrievers, two German Shepherds, two Jack Russell Terriers, two mixed breed dogs, and one each of Border Terrier, Bolonka Zwetna, Continental Bulldog, Boston Terrier, Irish Setter, Bullmastiff, Cane Corso, and Pomeranian, respectively, were included. Twenty patients were females (15 spayed) and 18 were males (8 castrated). Two dogs had two independent masses, leading to the inclusion of 40 tumors. Their types and grades based on histopathology (32 postmortem and 8 in vivo stereotactic brain biopsy) are summarized in Table 1. The single glioblastoma of the study population was included as a high-grade astrocytoma.[25]

In six of 38 MRI studies, the slice thickness of the transverse FLAIR sequence was thicker (4.0 mm) than the remaining sequences (3.5 mm). Motion artifact was identified in four different cases affecting a single sequence per MRI study. A 3-channel transmitter and receiver knee coil was used in 32/38 cases; a 4-channel transmitter and receiver head coil was used in 6/38 cases. In FLAIR, TR (11000 ms), TE (140 ms), and TI (2600 ms) were equal among all MRI studies. In T1w sequences, TR ranged from 400 to 411.63 ms, and TE was constant (15 ms). In T2w sequences, TR ranged from 4372.5 to 5979.06 ms, and TE was constant (100 ms).

TABLE 2 Machine learning multiclass classification (oligodendrogliomas vs. astrocytomas vs. oligoastrocytomas) test performance measures on the *whole dataset*.

	Overall accuracy(%)				Sensitivity(%)				Specificity(%)			
	SVM	RF	kNN	*Average*	SVM	RF	kNN	*Average*	SVM	RF	kNN	*Average*
Oligodendrogliomas	79	76	76	*77*	96	92	92	*93*	53	46	61	*53*
Astro-cytomas					29	43	28	*33*	96	96	100	*97*
Oligoastro-cytomas					67	50	66	*61*	96	97	87	*93*

Abbreviations: SVM, Support vector machine; RF, random forest; kNN, k-nearest neighbor.

TABLE 3 Machine learning binary classification (high-grade vs. low-grade gliomas) test performance measures on the *whole dataset*.

Classifier	Accuracy(%)	Sensitivity(%)	Specificity(%)	AUC
SVM	80	74	92	0.83
RF	67	81	38	0.60
kNN	80	85	69	0.83
Average	*76*	*80*	*66*	*0.75*

Abbreviations: AUC, area under the curve; SVM, Support vector machine; RF, random forest; kNN, k-nearest neighbor.

All tumors (40/40) demonstrated non-enhancing segments. Perilesional edema was present in 21 of 40 tumors. In four (3 high-grade oligodendrogliomas, 1 oligosarcoma), the edema segment was too small to extract texture features. Of the remaining 17, 13 were high-grade (10 oligodendrogliomas, 1 astrocytoma, 1 oligoastrocytoma, 1 oligosarcoma), and four were low-grade (2 oligodendrogliomas, 1 astrocytoma, 1 oligoastrocytoma). An enhancing segment was present in 26 of 40 tumors. In five (3 low-grade astrocytomas, 2 high-grade oligodendrogliomas), the enhancing segment was too small to extract texture features. Of the remaining 21, 20 were high-grade (16 oligodendrogliomas, 2 oligosarcomas, 1 astrocytoma, 1 oligoastrocytoma), and one was a low-grade oligodendroglioma. Given the low number of cases, oligosarcomas were excluded from the multiclass but included in the binary classification models.

On the whole dataset, ML multiclass classifiers had an average accuracy of 77% for predicting tumor types. Their average sensitivity and specificity were respectively 93% and 53% for predicting oligodendrogliomas, 33% and 97% for predicting astrocytomas, and 61% and 93% for predicting oligoastrocytomas (Table 2).

On the whole dataset, ML binary classifiers had an average accuracy of 76%, sensitivity of 80%, specificity of 66%, and AUC of 0.75 for predicting high-grade gliomas. The SVM and kNN classifiers performed better than RF, having similar accuracy (80%) and AUC (0.83), with sensitivities of 74% and 85%, and specificities of 92% and 69%, respectively, for predicting high-grade gliomas (Table 3).

The enhancing segment was not fed into the binary or multiclass models as individual segment, as only one low-grade glioma, one astrocytoma, and one oligoastrocytoma showed enhancement with available texture features. Therefore eight of 12 small datasets were used for further analysis.

The SVM multiclass classifier had the best performance for predicting tumor types on *edema* segments in T1w precontrast sequence (overall accuracy 94%), with respective sensitivities and specificities of 100% and 75% for predicting oligodendrogliomas, 50% and 100% for predicting astrocytomas, and 100% and 100% for predicting oligoastrocytomas. The three classifiers had overall poor performances on the remaining combination of sequences and segments to predict tumor types, with overall accuracies ranging between 62% and 82%, and variable sensitivities and specificities (Table 4).

The SVM (accuracy 87%, sensitivity 89%, specificity 85%, AUC 0.87) and kNN (accuracy 80%, sensitivity 81%, specificity 76%, AUC 0.79) binary classifiers achieved the best performances for predicting high-grade gliomas on non-enhancing segments in T2w sequence, followed by the SVM (AUC 0.77) using non-enhancing segments in T1w postcontrast sequence. The three classifiers had overall poor performances on the remaining combination of sequences and segments to predict high-grade gliomas, with accuracies ranging between 43% and 82%, sensitivities between 58% and 100%, and low specificities (Table 5).

4 | DISCUSSION

The results of our study support the use of multiple segments across the entire tumor volume to allow a more accurate ML classification of canine glioma types and grades based on MRI-TA compared to conventional MRI. Our TA-based SVM and kNN models demonstrated a higher accuracy (80%), sensitivity (74%–85%), and specificity (69%–92%) on the whole dataset to discriminate between high-grade and low-grade intracranial canine gliomas than conventional MRI.[2] This suggests that intrinsic texture features discriminate intracranial canine glioma grades and supports the use of TA in combination with ML for

TABLE 4 Machine learning multiclass classification (oligodendrogliomas vs. astrocytomas vs. oligoastrocytomas) test performance measures on each combination of segments and sequences.

		Overall Accuracy(%)			Sensitivity(%)			Specificity(%)		
Sequence	Tumor type	SVM	RF	kNN	SVM	RF	kNN	SVM	RF	kNN
Non-enhancing Segment										
T2w	Oligodendroglioma	82	71	74	100	92	96	53	38	38
	Astrocytoma				57	43	14	96	93	100
	Oligoastrocytoma				33	17	50	100	97	94
FLAIR	Oligodendroglioma	76	61	68	92	92	96	84	46	15
	Astrocytoma				43	0	14	87	87	100
	Oligoastrocytoma				33	0	17	96	87	97
T1w	Oligodendroglioma	74	68	68	88	96	96	53	23	15
	Astrocytoma				43	28	28	87	93	97
	Oligoastrocytoma				33	0	0	96	100	100
T1wC	Oligodendroglioma	71	61	71	88	92	88	53	15	38
	Astrocytoma				43	0	28	87	87	97
	Oligoastrocytoma				33	0	50	96	100	94
Edema Segment										
T2w	Oligodendroglioma	75	62	75	100	83	100	0	0	0
	Astrocytoma				0	0	0	N/A	100	100
	Oligoastrocytoma				0	0	0	N/A	86	100
FLAIR	Oligodendroglioma	81	69	75	92	91	100	50	0	0
	Astrocytoma				50	0	0	100	100	100
	Oligoastrocytoma				50	0	0	92	93	100
T1w	Oligodendroglioma	94	69	75	100	91	100	75	0	0
	Astrocytoma				50	0	0	100	93	100
	Oligoastrocytoma				100	0	0	100	100	100
T1wC	Oligodendroglioma	75	75	75	100	100	100	0	0	0
	Astrocytoma				0	0	0	N/A	100	100
	Oligoastrocytoma				0	0	0	N/A	100	100

Abbreviations: SVM, Support vector machine; RF, random forest; kNN, k-nearest neighbor; N/A, not applicable.

TABLE 5 Machine learning binary classification (high-grade vs. low-grade gliomas) test performance measures on each combination of segments and sequences.

	Accuracy(%)			Sensitivity(%)			Specificity(%)			AUC		
	SVM	RF	kNN	SVM	RF	kNN	SVM	RF	kNN	SVM	RF	kNN
Non-enhancing segment												
T2w	87	75	80	89	85	81	85	53	76	0.87	0.70	0.79
FLAIR	75	70	70	100	85	85	23	38	39	0.62	0.62	0.62
T1w	72	70	70	88	81	96	38	46	15	0.64	0.64	0.56
T1wC	80	60	70	85	74	96	69	31	15	0.77	0.52	0.56
Edema segment												
T2w	76	59	76	100	77	100	0	0	0	0.50	0.38	0.50
FLAIR	75	65	76	100	84	100	0	0	0	0.50	0.42	0.50
T1w	82	43	75	100	58	100	25	0	0	0.62	0.29	0.50
T1wC	82	65	76	100	77	100	25	25	0	0.62	0.51	0.50

Abbreviations: AUC, area under the curve; SVM, Support vector machine; RF, random forest; kNN, k-nearest neighbor; N/A, not applicable.

their grading. In a recent study,[20] RF models failed to accurately discriminate canine gliomas into types and grades. In that study,[20] a single 2D ROI was drawn to feed the model, including both enhancing and non-enhancing tumor parts while excluding perilesional edema. Our results may also indicate that SVM perform overall better compared to RF and kNN for TA-based canine glioma classification.

Texture metrics are difficult to interpret, but their use in combination with ML offers an alternative for quantitative MRI assessment[38] and lesion classification. In dogs, an association between glioma grade and prognosis has been suggested but not yet proven.[39–41] Texture analysis-based ML may have the potential to improve treatment planning and prognosis assessment, particularly in cases in which antemortem diagnosis cannot be obtained, and may contribute to a better understanding of the biological behavior of canine gliomas. However, it is unknown whether glioma types, regardless of the grade, play a role in prognosis in dogs. When fed with the whole dataset, the three ML classifiers demonstrated a high overall accuracy for differentiating tumor types (76%–79%), but a lower sensitivity (28%–43%) and higher specificity (96%–100%) than conventional MRI features for predicting astrocytomas,[2] which likely indicates that texture features extracted from the whole dataset overlap between tumor types.

High-performance measures were achieved for predicting high-grade gliomas with the SVM and kNN classifiers using *non*-enhancing segments in T2w sequence. Therefore, discriminative texture features, having the potential to correlate with histopathological findings not captured on conventional MRI, are likely present in the non-enhancing part of the tumor. The histopathological diagnosis of oligodendrogliomas grade III and of astrocytomas grade IV relies on the presence of microvascular proliferation and necrosis.[5] Presumably, the non-enhancing segments did not contain microvascular proliferation as this would be expected in the *enhancing* segments. Only four low-grade gliomas were enhancing and only one had available texture values. This is in accordance with previous studies which report a higher prevalence of enhancement in high-grade gliomas.[2,3,7] However, the low number of low-grade gliomas with an enhancing segment and only one with available texture features did not allow these segments to be fed in to the ML classifiers.

Perilesional edema secondary to compression, considered purely vasogenic, is typical in low-grade gliomas, while vasogenic edema due to blood–brain barrier disruption by malignant cell infiltration, referred as the peritumoral zone, is typical in high-grade gliomas.[42] Whereas ML classifiers were unable to discriminate tumor grades using edema segments, the SVM classifier performance was higher using *edema* segments in T1w sequence to differentiate tumor types compared to the whole dataset. This may indicate underlying texture differences within the white matter tracks adjacent to the tumors due to differences in growing pattern or presence of a peritumoral zone. It is likely that the peritumoral zone was included within the edema segment, as MRI underestimates the size and extent of brain tumors.[3]

Using different pre-processing techniques, including one or multiple segments, and using a single 2D ROI or volumetric segmentations, can produce large variability in texture feature metrics. Similarly, the signal-to-noise ratio is one of the most influential factors in TA. A signal-to-noise ratio greater than four is necessary to adequately represent the textural behavior of the human brain.[9] Multiple TA studies in human medicine apply different pre-processing techniques and texture feature extraction methods.[43–46] Voxel size, TR, TE, field strength inhomogeneity, and noise, largely influence the result of texture metrics.[29,38,47] Important limitations in our study, due to its retrospective nature, are the use of different slice thicknesses and coils according to patients size, and the TR variability particularly in T2w sequences. These differences may have influenced the TA, but pre-processing methods were implemented before texture feature extraction according to other publications,[28–30] aiming to homogenize images across all segments, sequences, and subjects. These pre-processing techniques may facilitate multicenter studies, allowing to increase the sample size and improve the generalizability of the results.

As part of the TA process, feature selection and dimensionality reduction methods are performed to reduce the data complexity, keeping only texture features that help differentiate classes. Dimensionality reduction methods, such as PCA, project the relevant texture features into new variables (principal components), which are also challenging to interpret.[23,48] Further studies assessing the different texture metrics across glioma types and grades and their correlation to histopathological findings may yield further information.

Different methods for training and testing samples in machine learning exist. In our study, LOOcv was the chosen method given the small sample size, for which the model is applied once for the selected sample as a single-item test set, using all other samples as the training set, and this process is repeated as many times as number of samples in the dataset.[18] Sample size is critical in ML. It has been shown that *k*-fold cross-validation techniques can produce biased results, are more likely to overfit ML models, and are therefore less likely to generalize well when applied to different datasets.[49] However, the aim of our study was to demonstrate the feasibility of ML and TA synergy for canine glioma classification in a single academic institution, rather than creating a ML tool that could be applied elsewhere.

This study had other limitations. Motion and partial volume averaging artifacts were identified in four sequences, and few slices were more rostrally or caudally positioned than others, likely due to respiratory motion in the z-axis during MRI acquisition. These cannot be fixed with pre-processing techniques or rigid coregistration in 2D sequences and may influence TA results. The dataset was imbalanced (higher proportion of oligodendrogliomas and high-grade gliomas), which is known to be problematic for ML but reflects the distribution of intracranial glioma types and grades in the canine population.[2–4] Whereas different ML techniques could have been performed to counteract this limitation, such as up-sampling the minority classes,[50] this would have led to a rather artificial dataset and would have increased the risk of model overfitting. The histopathological criteria for grading gliomas depend on their tumor type.[25] Not differentiating the type to assess the grade can therefore have introduced bias in the TA and ML performance results. Studies with larger sample sizes aiming to grade gliomas based on their types are encouraged to further assess texture feature differences. Another important limitation of our study is that we did not aim to differentiate gliomas from benign mass lesions, such as granulomas, as it has been proved that brain granulomas share common MRI

features with canine gliomas[51]. Although another study[20] demonstrated an accuracy of 85% for differentiation between meningoencephalitis and gliomas using texture analysis, further research for differentiation between solid benign intra-axial masses and gliomas is encouraged. Lastly, the segmentations were manually performed by two observers by mutual agreement. It is known that tumor margin delineation between non-enhancing tumor and perilesional edema cannot always be made accurately, especially in T2w sequences, and that tumor margins often exceed the margins established by MRI.[3] This may have led to a large number of outliers in the edema region due to inclusion of the peritumoral zone. A possible solution would include margin shrinking to avoid wrong labelling, at the cost of potentially excluding relevant texture metrics. Manual segmentation is time consuming with a large inter and intra-observer variability.[51] In human medicine, the use of automated techniques provide an unbiased brain tumor segmentation with comparable estimates to human raters.[51] Recently, deep learning-based automatic segmentation of head and neck organs was successfully achieved in dogs for radiotherapy planning.[52] Similarly, automated tumor segmentation could be implemented in the future in dogs for TA and brain lesion volumetry.

5 | CONCLUSIONS

Machine learning models based on MRI-TA, particularly SVM, have the ability to discriminate intracranial canine glioma types and grades with respective accuracies of up to 94% and 87%. The most discriminative texture features for differentiating tumor types and grades appear related to peri-tumoral edema in T1w images and to the non-enhancing tumor part in T2w images, respectively. Further multicenter studies with larger sample sizes are needed to corroborate these results.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Yannick Suter, Artorg Center, University of Bern, for his valuable support in the design of the study.

LIST OF AUTHOR CONTRIBUTIONS

Category 1

(a) Conception and Design: Barge, Durand
(b) Acquisition of Data: Barge, Oevermann, Maiolini, Durand
(c) Analysis and Interpretation of Data: Barge, Durand

Category 2

(a) Drafting the Article: Barge, Durand
(b) Revising Article for Intellectual Content: Barge, Oevermann, Maiolini, Durand

Category 3

(a) Final Approval of the Completed Article: Barge, Oevermann, Maiolini, Durand

Category 4

(a) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: Barge, Oevermann, Maiolini, Durand

CONFLICT OF INTEREST DISCLOSURE

The authors have declared no conflict of interest.

PREVIOUS PRESENTATION

Preliminary results presented as an oral presentation at the 2021 Online EVDI Annual Scientific Conference.

Reporting checklist disclosure: The CLAIM checklist was used in the preparation of this manuscript.

DATA ACCESSIBILITY STATEMENT

Available from the corresponding author upon reasonable request.

ORCID

Pablo Barge https://orcid.org/0000-0002-7070-341X

REFERENCES

1. Wisner ER, Dickinson PJ, Higgins RJ. Magnetic resonance imaging features of canine intracranial neoplasia. *Vet Radiol Ultrasound*. 2011;52:S52-S61.
2. José-López R, Gutierrez-Quintana R, Fuente C, et al. Clinical features, diagnosis, and survival analysis of dogs with glioma. *J Vet Intern Med*. 2021;35:1902-1917.
3. Young BD, Levine JM, Porter BF, et al. Magnetic resonance imaging features of intracranial astrocytomas and oligodendrogliomas in dogs. *Vet Radiol Ultrasound*. 2011;52:132-141.
4. Bentley RT, Ober CP, Anderson KL, et al. Canine intracranial gliomas: Relationship between magnetic resonance imaging criteria and tumor type and grade. *Vet J*. 2013;198:463-471.
5. Koehler JW, Miller AD, Miller CR, et al. A Revised diagnostic classification of canine glioma: towards validation of the canine glioma patient as a naturally occurring preclinical model for human glioma. *J Neuropathol Exp Neurol*. 2018;77:1039-1054.
6. Kraft SL, Gavin PR, Dehaan C, Moore M, Wendling LR, Leathers CW. Retrospective review of 50 canine intracranial tumors evaluated by magnetic resonance imaging. *J Vet Intern Med*. 1997;11:218-225.
7. Bentley RT. Magnetic resonance imaging diagnosis of brain tumors in dogs. *Vet J*. 2015;205:204-216.
8. Stadler KL, Ruth JD, Pancotto TE, Werre SR, Rossmeisl JH. Computed tomography and magnetic resonance imaging are equivalent in mensuration and similarly inaccurate in grade and type predictability of canine intracranial gliomas. *Front Vet Sci*. 2017;4:157.
9. Larroza A, Bodí V, Moratal D. Texture Analysis in Magnetic Resonance Imaging: Review and Considerations for Future Applications. In: Constantinides C (ed): *Assessment of Cellular and Organ Function and Dysfunction using Direct and Derived MRI Methodologies*. Internet:London: IntechOpen, 2016.
10. Van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging-"how-to" guide and critical reflection. *Insights Imaging*. 2020;11:91.
11. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clin Radiol*. 2004;59:1061-1069.
12. Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp*. 2018;2:36.
13. Sanduleanu S, Woodruff HC, De Jong EEC, et al. Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. *Radiother Oncol*. 2018;127:349-360.

14. Giger ML. Machine Learning in Medical Imaging. *J Am Coll Radiol*. 2018;15:512-520.
15. Choy G, Khalilzadeh O, Michalski M, et al. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology*. 2018;288:318-328.
16. Yoon Y, Hwang T, Choi H, Lee H. Classification of radiographic lung pattern based on texture analysis and machine learning. *J Vet Sci*. 2019;20:e44.
17. Banzato T, Bernardini M, Cherubini GB, Zotti A. Texture analysis of magnetic resonance images to predict histologic grade of meningiomas in dogs. *Am J Vet Res*. 2017;78:1156-1162.
18. Marschner CB, Kokla M, Amigo JM, Rozanski EA, Wiinberg B, Mcevoy FJ. Texture analysis of pulmonary parenchymateous changes related to pulmonary thromboembolism in dogs - a novel approach using quantitative methods. *BMC Vet Res*. 2017;13:219.
19. Shaker R, Wilke C, Ober C, Lawrence J. Machine learning model development for quantitative analysis of CT heterogeneity in canine hepatic masses may predict histologic malignancy. *Vet Radiol Ultrasound*. 2021;62:711-719.
20. Wanamaker MW, Vernau KM, Taylor SL, Cissell DD, Abdelhafez YG, Zwingenberger AL. Classification of neoplastic and inflammatory brain disease using MRI texture analysis in 119 dogs. *Vet Radiol Ultrasound*. 2021;62:445-454.
21. Alis D, Bagcilar O, Senli YD, et al. Machine learning-based quantitative texture analysis of conventional MRI combined with ADC maps for assessment of IDH1 mutation in high-grade gliomas. *Jpn J Radiol*. 2020;38:135-143.
22. Chang Y, Lafata K, Sun W, Wang C, Chang Z, Kirkpatrick JP, Yin F-F An investigation of machine learning methods in delta-radiomics feature analysis. *PLoS One*. 2019;14:e0226348.
23. Cho H-Ho, Lee S-H, Kim J, Park H. Classification of the glioma grading using radiomics analysis. *PeerJ*. 2018;6:e5982.
24. Zacharaki EI, Wang S, Chawla S, et al. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn Reson Med*. 2009;62:1609-1618.
25. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol*. 2016;131:803-820.
26. Klein S, Staring M, Murphy K, Viergever MA, Pluim J. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29:196-205.
27. Nioche C, Orlhac F, Boughdad S, et al. LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res*. 2018;78:4786-4789.
28. Carré A, Klausner G, Edjlali M, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Sci Rep*. 2020;10:12340.
29. Bologna M, Corino V, Mainardi L. Technical Note: Virtual phantom analyses for preprocessing evaluation and detection of a robust feature set for MRI-radiomics of the brain. *Med Phys*. 2019;46:5116-5123.
30. Kjær L, Ring P, Thomsen C, Henriksen O. Texture analysis in quantitative MR imaging. Tissue characterisation of normal brain and intracranial tumours at 1.5 T. *Acta Radiol*. 1995;36:127-135.
31. Uddin S, Khan A, Hossain MdE, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak*. 2019;19:281.
32. Sarica A, Cerasa A, Quattrone A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Front Aging Neurosci*. 2017;9:329.
33. Abu Alfeilat HA, Hassanat ABA, Lasassmeh O, et al. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data*. 2019;7:221-248.
34. Pedregosa F, Varoquax G, Gramfort A, Michel V, Thirion B. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-2830.
35. Lever J, Krzywinski M, Altman N. Principal component analysis. *Nature Methods*. 2017;14:641-642.
36. Hung C-C, Song E, Lan Y. Image Texture Analysis. Foundations, Models and Algorithms. China:Spinger Nature Switzerland AG, 2019.
37. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Information Processing and Management*. 2009;45:427-437.
38. Hagiwara A, Fujita S, Ohno Y, Aoki S. Variability and standardization of quantitative imaging: monoparametric to multiparametric quantification, radiomics, and artificial intelligence. *Invest Radiol*. 2020;55:601-616.
39. Hu H, Barker A, Harcourt-Brown T, Jeffery N. Systematic review of brain tumor treatment in dogs. *J Vet Intern Med*. 2015;29:1456-1463.
40. Dickinson PJ. Advances in diagnostic and treatment modalities for intracranial tumors. *J Vet Intern Med*. 2014;28:1165-1185.
41. Lecouteur RA. Current concepts in the diagnosis and treatment of brain tumours in dogs and cats. *J Small Anim Pract*. 1999;40:411-416.
42. Lemée J-M, Clavreul A, Menei P. Intratumoral heterogeneity in glioblastoma: Don't forget the peritumoral brain zone. *Neuro Oncol*. 2015;17:1322-1332.
43. Ditmer A, Zhang B, Shujaat T, Pavlina A, Luibrand N, Gaskill-Shipley M, Vagal A Diagnostic accuracy of MRI texture analysis for grading gliomas. *J Neurooncol*. 2018;140:583-589.
44. Skogen K, Schulz A, Dormagen JB, Ganeshan B, Helseth E, Server A. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. *Eur J Radiol*. 2016;85:824-829.
45. Suárez-García JG, Hernández-López JM, Moreno-Barbosa E, De Celis-Alonso B. A simple model for glioma grading based on texture analysis applied to conventional brain MRI. *PLoS One*. 2020;15:e0228972.
46. Tian Q, Yan L-F, Zhang Xi, et al. Radiomics strategy for glioma grading using texture features from multiparametric MRI. *J Magn Reson Imaging*. 2018;48:1518-1528.
47. Ford J, Dogan N, Young L, Yang F. Quantitative radiomics: Impact of pulse sequence parameter selection on MRI-based textural features of the brain. *Contrast Media Mol Imaging*. 2018;2018:1729071.
48. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374:20150202.
49. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS One*. 2019;14:e0224365.
50. Varghese BA, Cen SY, Hwang DH, Duddalwar VA. Texture analysis of imaging: What radiologists need to know. *AJR Am J Roentgenol*. 2019;212:520-528.
51. Meier R, Knecht U, Loosli T, Bauer S, Slotboom J, Wiest R, Reyes M Clinical evaluation of a fully-automatic segmentation method for longitudinal brain tumor volumetry. *Sci Rep*. 2016;6:23376.
52. Park J, Choi B, Ko J, et al. Deep-learning-based automatic segmentation of head and neck organs for radiation therapy in dogs. *Front Vet Sci*. 2021;8:721612.

How to cite this article: Barge P, Oevermann A, Maiolini A, Durand A. Machine learning predicts histologic type and grade of canine gliomas based on MRI texture analysis. *Vet Radiol Ultrasound*. 2023;1-9. https://doi.org/10.1111/vru.13242