Magn Reson Med Sci 2019; 18; 44–52
doi:10.2463/mrms.mp.2017-0178
Published Online: May 16, 2018

MAJOR PAPER

Machine Learning-based Texture Analysis of Contrast-enhanced MR Imaging to Differentiate between Glioblastoma and Primary Central Nervous System Lymphoma

Akira Kunimatsu[1,2*], Natsuko Kunimatsu[3], Koichiro Yasaka[1,2], Hiroyuki Akai[1,2], Kouhei Kamiya[2], Takeyuki Watadani[4], Harushi Mori[5], and Osamu Abe[5]

Purpose: Although advanced MRI techniques are increasingly available, imaging differentiation between glioblastoma and primary central nervous system lymphoma (PCNSL) is sometimes confusing. We aimed to evaluate the performance of image classification by support vector machine, a method of traditional machine learning, using texture features computed from contrast-enhanced T_1-weighted images.

Methods: This retrospective study on preoperative brain tumor MRI included 76 consecutives, initially treated patients with glioblastoma ($n = 55$) or PCNSL ($n = 21$) from one institution, consisting of independent training group ($n = 60$: 44 glioblastomas and 16 PCNSLs) and test group ($n = 16$: 11 glioblastomas and 5 PCNSLs) sequentially separated by time periods. A total set of 67 texture features was computed on routine contrast-enhanced T_1-weighted images of the training group, and the top four most discriminating features were selected as input variables to train support vector machine classifiers. These features were then evaluated on the test group with subsequent image classification.

Results: The area under the receiver operating characteristic curves on the training data was calculated at 0.99 (95% confidence interval [CI]: 0.96–1.00) for the classifier with a Gaussian kernel and 0.87 (95% CI: 0.77–0.95) for the classifier with a linear kernel. On the test data, both of the classifiers showed prediction accuracy of 75% (12/16) of the test images.

Conclusions: Although further improvement is needed, our preliminary results suggest that machine learning-based image classification may provide complementary diagnostic information on routine brain MRI.

Keywords: *classification, glioblastoma, magnetic resonance imaging, primary central nervous system lymphoma, support vector machine*

Introduction

Glioblastoma is the most common malignant brain tumor in adults, and accounts for approximately 15% of all primary brain tumors.[1,2] Primary central nervous system lymphoma (PCNSL) is an uncommon form of extra-nodal non-Hodgkin's lymphoma that is confined to the central nervous system, and it constitutes 2–3.1% of all primary brain neoplasms.[2–4] Glioblastoma and PCNSL share some common clinical features: they preferentially affect middle-aged or older patients with a slight male predominance, and the progression of symptoms can be very rapid. However, treatment strategies are substantially different between glioblastoma and PCNSL. For glioblastoma patients, maximal tumor resection followed by chemo-radiotherapy with temozolomide is the current standard for treatment,[5] whereas for all patients with findings suggestive of PCNSL, a stereotactic biopsy followed by methotrexate-based chemotherapy is indicated.[4] Therefore, preoperative differentiation of glioblastoma and PCNSL is of high clinical relevance.[6]

Preoperative MRI for a brain tumor usually includes contrast-enhanced T_1-weighted images. Glioblastoma typically exhibits heterogeneous enhancement with central

[1]Department of Radiology, IMSUT Hospital, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
[2]Department of Radiology, The University of Tokyo Hospital, Tokyo, Japan
[3]Department of Radiology, International University of Health and Welfare, Mita Hospital, Tokyo, Japan
[4]Department of Radiology, Faculty of Medicine, The University of Tokyo, Tokyo, Japan
[5]Department of Radiology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

*Corresponding author, Phone: +81-3-3443-8111, Fax: + 81-3-5449-5746, E-mail: akrk-tky@umin.ac.jp

Received January 4, 2018 | Accepted April 4, 2018

necrosis, which results in irregular ring-like enhancement patterns on contrast-enhanced T_1-weighted images. In contrast, PCNSL arising in immunocompetent patients typically appears as a homogeneously enhancing tumor on contrast-enhanced T_1-weighted images.[7,8] However, these well-known imaging patterns are not reliable in some rare occasions where homogeneously enhancing glioblastoma without visible necrosis or PCNSL with visible necrosis is present.

Machine learning comprises a broad class of statistical analysis algorithms that iteratively improve in response to training data to build models for autonomous predictions.[9] Rapidly developing computer vision technology uses machine learning in object detection, localization, and classification on a large amount of digital images.[10] Machine learning can be applied to medical images as well, attracting increasing interest in the field of radiology. Since such machines have very recently demonstrated the capability to learn and even master tasks that were thought to be too complex for machines, machine learning is now considered to be a potentially useful component of computer-aided diagnosis and decision support systems.[10,11]

A few previous studies reported accuracy of support vector machine (SVM) classification between glioblastoma and PCNSL, based on textural features of MR images.[12–14] However, these studies performed cross-validation of a classifier model within a single training dataset but lacked external validation or testing with an independent dataset. Beyond the standard practice of internal validation, one method of which is cross-validation, testing a model with a dataset independent from a training dataset would be beneficial to assess generalizability of the model.[15] The purpose of the present study was to evaluate the feasibility of machine learning-based differentiation between glioblastoma and PCNSL, using contrast-enhanced T_1-weighted image data, with independent training and test datasets on a potential clinical scenario.

Materials and Methods

Ethics statement

This retrospective study was approved by the local ethics committee and informed consent was waived.

Study population

The study population consisted of two independent patient groups: the training group and the test group. As described below, the training image data were obtained from the training group and the test image data were obtained from the test group respectively. All patients were enrolled at one single institution.

The training group in the present study was comprised of the patients studied in our previous study,[16] and here we briefly describe how to enroll patients in the training group. We first performed a text search on our radiology report database up to August 2015. The search criteria were: 1) Untreated (i.e., not recurrent) case of pathologically proven glioblastoma or PCNSL, and 2) contrast-enhanced brain MR imaging performed on a 3T unit with the brain tumor imaging protocol at our institution before surgical resection or biopsy (Fig. 1). Fifty-one consecutive patients with glioblastoma and 16 consecutive patients with PCNSL, from December 2006 through August 2015, fulfilled these inclusion criteria. All of the PCNSL were diffuse large

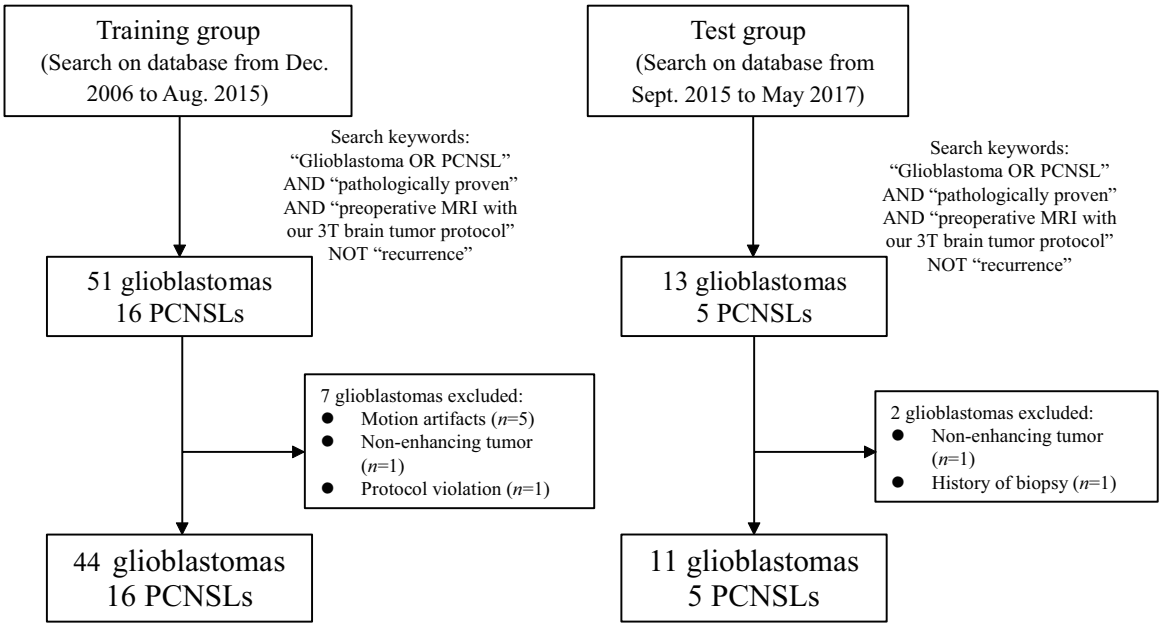

Fig. 1 Flow chart of subject enrollment for the training and the test groups. PCNSL, primary central nervous system lymphoma.

Table 1 Summary of the study population

Characteristic	Total number of patients (n = 76)			
	Training group (n = 60)		Test group (n = 16)	
	Glioblastoma	PCNSL	Glioblastoma	PCNSL
No. of patients	44	16	11	5
Women	14	3	5	3
Men	30	13	6	2
Mean age (years)	61.5	60.6	58.7	74.8
Age range (years)	26–81	42–75	38–75	61–81

PCNSL, primary central nervous system lymphoma.

B-cell subtype that developed in immunocompetent patients. Neither adult immunodeficiency syndrome-related nor Epstein–Barr virus-related lymphoma was included in our cases.[17] Histological diagnosis was made by experienced pathologists at our institution and was obtained from the medical records. Seven of the 51 patients with glioblastoma were excluded due to presence of motion artifacts (n = 5), non-enhancing tumor (n = 1), and protocol violation (n = 1) on MR imaging. Finally, 44 glioblastoma patients (30 men and 14 women; mean age: 61.5 years, range 26–81) and 16 PCNSL patients (13 men and 3 women, mean age: 60.6 years, range 42–75) were designated as the training group (Table 1).

In the present study, we additionally enrolled patients who presented to our hospital more recently than the patients in the training group. The second text search on our radiology report database was performed with the same search criteria but on a more recent search period (from September 2015 through May 2017) such that the test image data obtained from the test group could be "unseen" data to an established image classifier (Fig. 1). This scenario can realistically happen in clinical practice where newly incoming data are evaluated with already existing decision criteria. Thirteen consecutive patients with glioblastoma and 5 consecutive patients with PCNSL newly diagnosed in that period fulfilled the criteria, but two glioblastoma patients were excluded because of the presence of a non-enhancing tumor (n = 1) and a history of a previous biopsy at the referring hospital (n = 1). Finally, 11 glioblastoma patients (6 men and 5 women; mean age: 58.7 years, range 38–75) and 5 PCNSL patients (2 men and 3 women: mean age: 74.8 years, range 61–81) were designated as the test group (Table 1). The ratio of the number of patients in the training group over the test group resulted in approximately four to one.

MR imaging

Using the picture archiving and communication system of our institution, post-contrast trans-axial spin-echo T_1-weighted images were retrieved from MR images obtained with the brain tumor protocol for 3T. All MR images were obtained on 3T scanners (Singa HDx and HDxt after system upgrade, GE Healthcare, WI, USA, and Magnetom Skyra, Siemens, Erlangen, Germany) with the following imaging parameters: TR/TE (ms) = 400/9-12 (Signa HDx and HDxt) or 406-466/9 (Skyra), FOV = 21.0 cm, matrix = 256 × 256, slice thickness = 5 mm, interslice gap = 1 mm, number of excitations = 1. All patients received intravenous injection of a gadolinium-based contrast agent at the rate of 0.1 mmol/kg of body weight.

Feature extraction and development of a classifier with the training data

For image post-processing and statistical analyses, we used image processing software (ImageJ 1.48, National Institute of Health, Bethesda, MD, USA, http://imagej.nih.gov/ij), numerical analysis software (Matlab 2017a, The MathWorks, Natick, MA, USA), and statistical computing and graphics software (R 3.1.2, The R Foundation for Statistical Computing, Vienna, Austria, https://cran.r-project.org/). We used a computer containing two Intel Xeon E5-2609 2.4-GHz processors (Intel, Santa Clara, CA, USA) as the central processing unit (CPU), 2 TB of hard disk space and 32 GB of random access memory. Most of calculations in the following analyses took less than several minutes with the CPU.

A series of post-processing procedures to build a classifier on the training image data included: 1) Image feature extraction, 2) image feature selection, and 3) training and cross-validation (Fig. 2). We had already performed the first (image feature extraction) and the second (feature selection) processes in our previous study,[16] and therefore, we delivered the results of our previous study to the third process (training and cross-validation) in the present study. Here we provide a brief description on the entire process.

First, a post-contrast T_1-weighted image that harbored the largest contrast-enhancing lesion was selected for each case and the digital imaging and communication in medicine images were converted to 8-bit tagged image file format images with ImageJ, along with linearly scaling from minimum to maximum intensity between 0 and 255. Subsequently, a rectangular ROI was carefully placed within the enhancing lesion as large as possible but not to include surrounding brain tissue outside the enhancing lesion (Fig. 3), by two independent radiologists (both with more than 20 years of experience in neuroradiology), and all voxel values within the ROI were recorded as a matrix-style variable using the Matlab platform.

A total of 67 radiomic texture features, provided with the "radiomics" package of the R software,[18] were extracted from the matrices and Z-score transformation was performed per feature. These features included first-order texture features and second- or higher order features calculated with gray-level co-occurrence matrix, gray-level run length matrix (GLRLM), gray-level size zone matrix, and multiple gray-level size zone matrix.

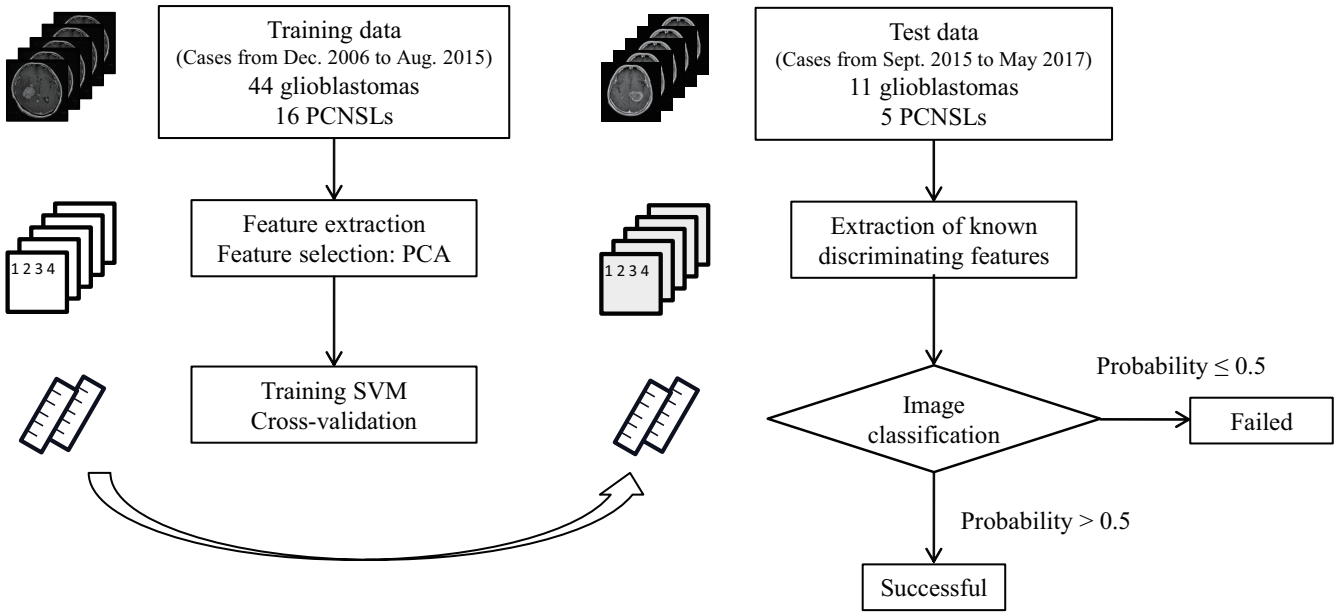

Fig. 2 Flow chart of classifier development with the training data and subsequent classification on the test data. PCA, principal component analysis; PCNSL, primary central nervous system lymphoma; SVM, support vector machine.

Feature selection was the second step. Interobserver reproducibility on the extracted features was assessed using the intraclass correlation coefficient of the two-way random model with the "psych" package of the R software, after the Kolmogrov–Smirnov test for normality.[19] An intraclass correlation coefficient of greater than 0.70 was considered reliable according to the literature,[20,21] and 28 features with coefficients above 0.70 were included after averaging. Principal component analysis, a standard method for feature dimension reduction,[22] was subsequently performed with the Matlab platform for feature selection and principal components were selected such that cumulative percentage contribution exceeded 80%. According to our previous study,[16] four features were selected: first-order entropy, first-order median, GLRLM run length non-uniformity, and GLRLM run percentage. The rationale for feature dimension reduction was to avoid overtraining. In the present study, the training data with 60 samples and four features satisfied the provoked sample-per-feature ratio of at least 5 to 10.[23]

The third step was training and cross-validation of a classifier, which was implemented in the Matlab platform. Prior to the present study, we preliminary tested classification performance of four common classification algorithms (*k*-nearest neighbor, decision tree, discriminant analysis, and SVM)[11] on the training data with sixfold cross-validation, using the classification learner application implemented in the Matlab. The sixfold split was deduced from the Sturges' formula.[24] We found that SVM showed the highest performance among the four algorithms and we developed two typical SVM classifier models, with the above-mentioned four selected texture features as the input variables, based on linear (classifier-L) or Gaussian (classifier-G) kernels, respectively.[25] The training dataset fulfilled the minimal sample size reported in the literature.[26] The sequential minimal optimization algorithm, a standard reported method,[25] was used in training. Leave-one-out cross-validation, which is widely used in machine learning studies and can provide a uniquely determined value for cross-validation, was performed to estimate the accuracy of each classifier model.[27,28] It is well known that the leave-one-out cross-validation provides an almost unbiased estimate of the accuracy of a classifier model.[29]

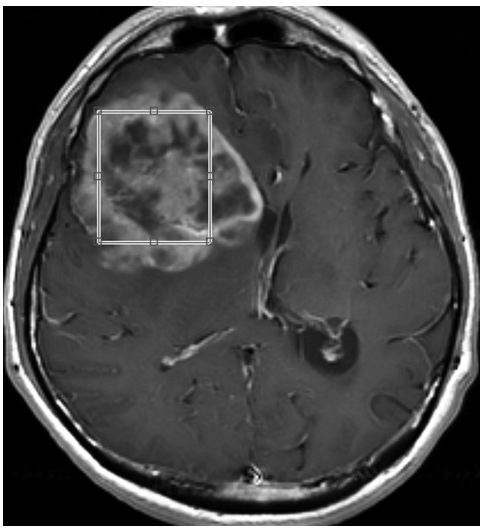

Fig. 3 Example of region of interest placement on the contrast-enhanced T_1-weighted image of a 43-year-old man subsequently diagnosed with glioblastoma. A rectangular region of interest is placed within the enhancing tumor region.

Statistical analysis

Statistical analysis was performed on the diagnostic performance of the classifiers across kernels. Four features (first-order entropy, first-order median, GLRLM run length non-uniformity, and GLRLM run percentage) were used as the explanatory (or input) variables of the classifier models and the responses were recorded on the training data. The performance metric was an area under the curve (AUC) of the receiver operating characteristic (ROC) analysis. The value of AUC with 95% confidence interval (CI) was obtained on each classifier using the "pROC" package of the R software.[30]

Classification of the test data

Feature extraction was similarly performed on the test image data as was done in the training process. *Z*-scores were calculated using the mean and the standard deviation values obtained in the training process for each of the four features. As described above, this followed a realistic scenario in clinical practice where an established classifier is applied to the next incoming data. The capability of binary classification of the test image data was estimated on each of the two classifiers supposing the four texture features as input variables, and the results were expressed as the posterior probability of the predicted classification for glioblastoma or PCNSL (Fig. 2). The posterior probability means the probability with which an observation (i.e. a given image) represents a particular disease (glioblastoma or PCNSL) in the context of the present study.

Results

Cross-validation and ROC curve analysis of the classifiers

The results are summarized in Table 2. Leave-one-out cross-validation demonstrated overall accuracies of 0.70 for the classifier-L and 0.80 for the classifier-G, respectively. When subdivided into glioblastoma and PCNSL, accuracies of 0.66 and 0.82 were calculated for glioblastoma, and accuracies of 0.81 and 0.75 for PCNSL, with the classifiers-L and -G, respectively. In Fig. 4, the classifiers-G and -L had the AUC values (95% CI) of 0.99 (0.96–1.00) and 0.87 (0.77–0.95), respectively.

Classification of the test data

Figure 5 demonstrates the posterior probability in assigning the feature vector data calculated from a given test image into the correct pathological diagnosis (i.e., putting a label of glioblastoma on an unseen glioblastoma image or a label of PCNSL on an unseen PCNSL image) by each of the classifier models. The median posterior probability (1st–3rd quartile) was 0.82 (0.59–0.96) for the classifier-L and 0.87 (0.50–0.96) for the classifier-G. When using a standard decision criterion in which the posterior probability above 0.50 indicated correct or successful classification in our binary classification scheme, both of the classifiers yielded prediction accuracy of 75% (12/16) of the test data. As for glioblastoma or PCNSL, glioblastoma was correctly assigned in 10 out of 11 cases by the classifier-G and 9 out of 11 cases by the classifier-L. PCNSL was correctly assigned in 3 out of 5 cases by the classifier-L and 2 out of 5 cases by the classifier-G (Figs. 5 and 6).

Discussion

Even today, in the midst of imaging technology advancement, differential diagnosis between glioblastoma and PCNSL by conventional MR findings remains challenging in some cases. In the present study, we evaluated the feasibility of SVM-based classification approaches, commonly used in machine learning, for differentiation of glioblastoma from PCNSL. Our results demonstrated that standard SVM-based classifiers, trained with contrast-enhanced T_1-weighted images, gave an accuracy of up to 0.80 in the cross-validation and an AUC value of up to 0.99 in the ROC curve analysis.

A strength of our study is that we used contrast-enhanced T_1-weighted images alone, which is generally included in conventional MR examination protocols for brain tumor evaluation. Recent studies focused on the usefulness of advanced MR imaging for differentiation between glioblastoma and PCNSL, including quantitative diffusion-weighted, perfusion-weighted, and susceptibility-weighted imaging, or multiparametric combinations.[6,31–33] However, advanced MR imaging techniques may not be necessarily available in every institution depending on the capability of MR scanners or to avoid a long or prolonged total scanning time. In contrast, contrast-enhanced T_1-weighted imaging is often performed for imaging diagnosis of cerebral mass lesions.

Table 2 Results of cross-validation and ROC curve analysis of SVM-based classifiers

	Cross-validation			ROC curve analysis
	Accuracy			AUC (95% CI)
	Glioblastoma	PCNSL	Overall	
Classifier-L	0.66	0.81	0.70	0.87 (0.77–0.95)
Classifier-G	0.82	0.75	0.80	0.99 (0.96–1.00)

AUC, area under the curve; CI, confidence interval; Classifier-L, the classifier with linear kernel; Classifier-G, the classifier with Gaussian kernel; PCNSL, primary central nervous system lymphoma; ROC, receiver operating characteristic; SVM, support vector machine.

In this study, the classifiers obtained with a typical SVM algorithm demonstrated an overall accuracy of 0.80 in leave-one-out cross-validation. A classifier model with high accuracy is expected to be applicable to predicting data in the future. In contrast, the AUC is commonly used as a summary measure of the diagnostic performance of a classifier.[34] A recent study by Rodriguez Gutierrez et al.[35] on SVM-based classification of three posterior fossa tumors reported averaged classification accuracy of 0.76–0.79, using contrast-enhanced T_1-weighted image histogram features. Overall accuracies of the classifier-G in our study reached 0.80, and our results appear to be similar with their findings of classifier performance. In addition, our ROC curve analysis revealed AUC values of 0.99 in classifier-G and of 0.87 in classifier-L. AUC values around 0.80 were reported in a recent publication by Tiwari et al.[36] for differentiation between radionecrosis and recurrent brain tumor using MR image-derived texture features and SVMs. Another SVM-based MR image classification study conducted by Alcaide-Leon et al.[12] reported the AUC value of 0.88 for differentiation between enhancing glioma and PCNSL. These previous studies reinforce our results suggesting that a SVM-based classification scheme may be feasible for differentiating glioblastoma from PCNSL.

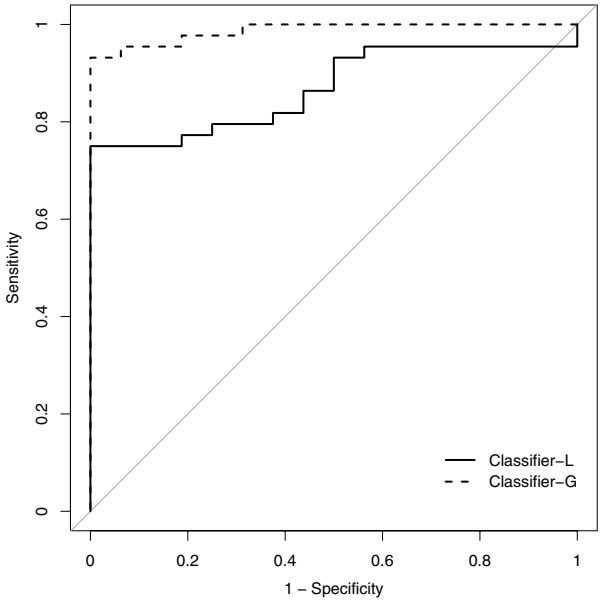

Fig. 4 The receiver operating characteristic (ROC) curves for the diagnostic performance of support vector machine (SVM) classifiers on the training data. The area under the curve (AUC) values (95% confidence interval [CI]) were 0.87 (0.77–0.95) and 0.99 (0.96–1.00) for the classifiers with linear and Gaussian kernels, respectively. Classifier-L, the classifier with linear kernel; Classifier-G, the classifier with Gaussian kernel.

Machine learning studies in the field of medicine often report prediction accuracy estimated from cross-validation but less often report prediction results with "real-world" testing.[11] Cross-validation likely yields optimistic results,[37] and it is encouraged to perform a validation test, after cross-validation, using an independent dataset. The testing process is conducted to check a bias, and high performance on the test process suggests that a developed model can be generalizable. However, no testing process was conducted in the previous studies.[12–14] In addition, we separated subjects into the training data and the test data in a time-course to allow a realistic clinical scenario where newly incoming data should be evaluated with a classifier obtained beforehand. Although, there may be concern about a historical control bias when patient groups are separated by the examination time, our study enrolled consecutive patients who had been studied with the uniformed MR imaging protocols on 3T scanners at the particular institution; therefore, we consider that data instability or bias was minimized.

We presented the results of the prediction on the test data using the posterior probability (Fig. 5). When we conventionally assumed that the class of input data should be assigned to the class with the largest posterior probability (i.e. the posterior probability above 0.50 in the binary classification), the classifier models in the present study excellently identified unseen images in glioblastoma cases, but

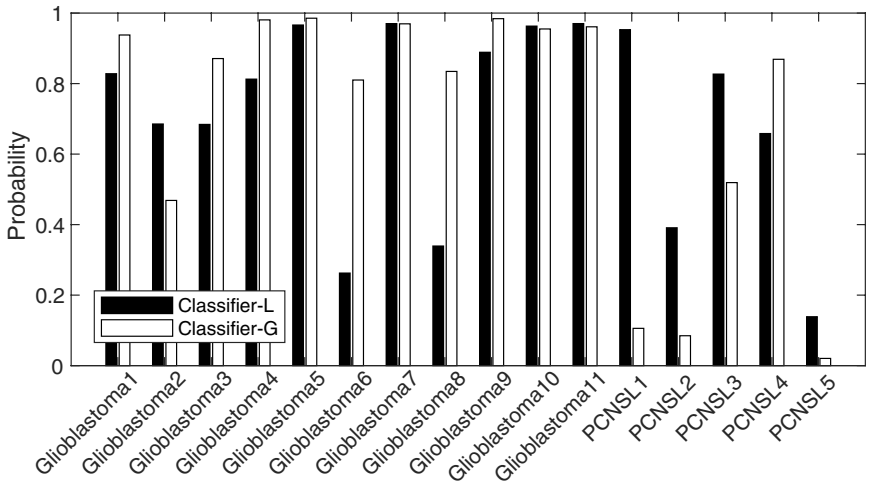

Fig. 5 Bar graphs representing the posterior probability of classification (prediction) on the test data. Results are presented per case and per classifier. Classifier-L, the classifier with linear kernel; Classifier-G, the classifier with Gaussian kernel; PCNSL, primary central nervous system lymphoma.

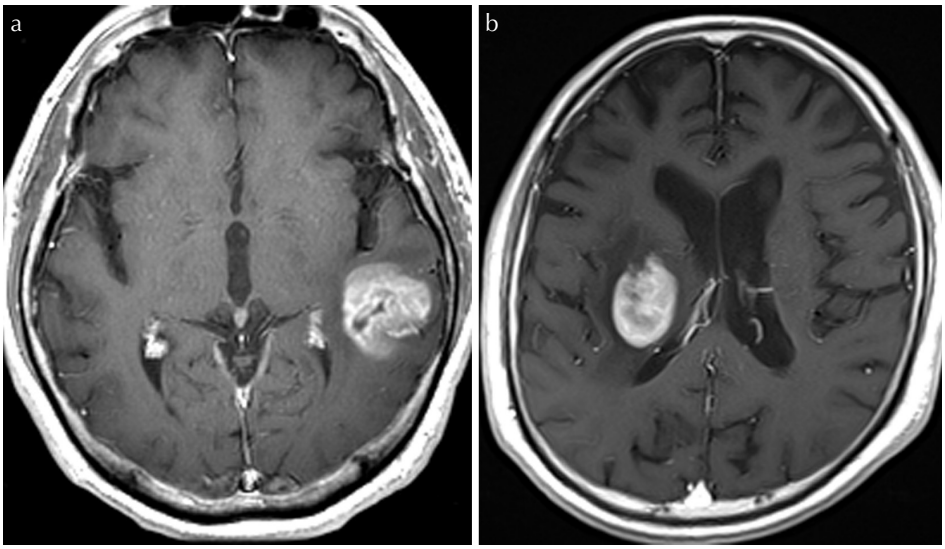

Fig. 6 Representative cases of image classification. (**a**) Contrast-enhanced T_1-weighted image of a 65-year-old man with glioblastoma in the left temporal lobe (glioblastoma3 in Fig. 5). Classifiers-G and -L assigned this image into glioblastoma. (**b**) Contrast-enhanced T_1-weighted image of a 81-year-old man with primary central nervous system lymphoma (PCNSL) affecting the right basal ganglia and deep white matter (PCNSL4 in Fig. 5). This image was classified as PCNSL by both classifiers. Classifier-L, the classifier with linear kernel; Classifier-G, the classifier with Gaussian kernel.

moderately in PCNSL cases. This may be attributable to the low and unbalanced number of PCNSL cases compared with glioblastoma cases in the training data. To cope with a class imbalance that can adversely impact the performance of a classifier, oversampling, including resampling, is advocated as a potential solution,[38] but some controversies exist on the effects of oversampling.[34] Moreover, the SVM-based classifiers in our study demonstrated accuracies of 0.66–0.82 for glioblastoma and 0.75–0.81 for PCNSL in cross-validation of the training data; the performance of the classifiers seemed to be similar regardless of the tumor type (Table 2). It is unclear whether the addition of oversampling in our SVM-classification scheme can improve the prediction performance with newly incoming or independent data, as it is beyond the scope of the present study.

Image classification tasks with supervised machine learning typically include feature extraction and classification (or output). In the present study, texture analysis was used to extract image features and SVM was used to classify given image features. Image features with known image classes (i.e. the training data) were fed to classifiers, and the classifiers were trained with SVM so as to be applicable to the test data. We used four texture features in the present study (first-order entropy, first-order median, GLRLM run length non-uniformity, and GLRLM run percentage). The former two are histogram features and the latter two are GLRLM features. GLRLM features can evaluate the coarseness of a texture in a predetermined direction within an image or within a ROI.[22] On the other hand, SVM is a learning system that uses a hypothesis space of linear functions in a high dimensional feature space. In SVM, classifiers are trained with a learning algorithm from optimization theory that implements minimization of a learning bias.[25] Support vector machine is a traditional but sophisticated technique that is still in use in machine learning studies.[28]

However, very recently, deep learning has started to be used in image recognition.[10] Deep learning, also known as deep neural network learning, has the benefit of not requiring image feature extraction as a first step, rather, features are autonomously identified by computer during the learning process.[11] Application of deep learning to the field of medical imaging will be rapidly encouraged. Recently, Yasaka et al.[39] reported that the median accuracy of classification of liver masses were 0.84 and the median AUC value for differentiating malignant from non-malignant categories was 0.92 using a deep convolutional neural network on liver CT images.[39] However, deep learning typically requires tens of thousands of images and this can be one of the biggest issues to utilize it in clinical practice where the number of patients with a rare disease is rather small. In contrast, SVM can be applied to a dataset even with a relatively small sample size.[26,35,36]

Our study has several limitations. First, this was a retrospective feasibility study with a small number of patients at a single institution, which may have led to bias in patient selection. We enrolled 55 glioblastoma patients and 21 PCNSL patients in the present study, and the numbers of each tumor seemed to follow the naturally-observed, higher incidence of glioblastoma over PCNSL.[2] In addition, we enrolled consecutive patients studied with the uniformed MR imaging protocols on 3T scanners; therefore, data instability or bias derived from MR scanners and protocols could be minimized. Second, we did not make a head-to-head comparison between results with SVM classifiers and diagnosis with radiologists. However, one previous study suggests that non-inferiority of our SVM-based classification accuracy to that of radiologists, as comparable AUC values were observed in the present study.[12] It remains undetermined whether machine learning, especially deep learning, can build a classifier with better diagnostic performance than that of experienced radiologists. Future work is needed to explore the added value of machine learning-based classification in clinical practice. Third, a considerable amount of our pathological diagnoses was made based on

conventional histology before the World Health Organization classification was revised, adopting the integrated molecular-histological diagnosis in 2016.[1] Forth, we focused on the classification between glioblastoma and PCNSL, and therefore, the classifiers developed in the present study should not be used in the classification between other tumors. In addition, we used accuracy to estimate prediction performance of the test data, but it is known that accuracy depends on the incidence of a disease (i.e. pretest probability).[11] Moreover, because of the small sample size of test patients, the prediction accuracy in the present study is vulnerable to a selection bias. A further study with a larger sample size is needed to conclude the generalizable accuracy. Finally, we did not use other MR images than contrast-enhanced T_1-weighted images in the present study. It remains unclear whether machine learning works best with contrast-enhanced T_1-weighted images. The rationale for the use of contrast-enhanced T_1-weighted images was clear delineation of tumor boundaries and high visibility of necrosis. Necrosis is a pathological hallmark of glioblastoma,[40] compared with PCNSL. Combined use of T_2-weighted or other MR sequences may enable better classification performance; however, this was beyond the scope of the present study.

Conclusion

In this feasibility study, we evaluated the performance of SVM-based classification, by using radiomic texture features computed from routine contrast-enhanced T_1-weighted images, on a training group and an independent test group from a single institution. Now that more powerful deep learning has started to be used in medical image recognition, machine learning is expected to facilitate laborious radiology practice. Although further improvement is needed, our results suggest that machine learning may provide complementary diagnostic information on routine MRI.

Disclosure Statement

A. Kunimatsu has received scholarship/donation from Daiichi-Sankyo Co. Ltd.; all of the other authors have no conflicts of interest.

References

1. International Agency for Research on Cancer. WHO classification of tumours of the central nervous system (World health organization classification of tumours) Revised 4th ed. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK eds. Lyon, France: IARC Publications; 2016.
2. Ostrom QT, Gittleman H, Fulop J, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008-2012. Neuro Oncol 2015; 17 Suppl 4: iv1–iv62.
3. Batchelor T, Loeffler JS. Primary CNS lymphoma. J Clin Oncol 2006; 24:1281–1288.
4. Hochberg FH, Baehring JM, Hochberg EP. Primary CNS lymphoma. Nat Clin Pract Neurol 2007; 3:24–35.
5. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 2005; 352:987–996.
6. Kickingereder P, Wiestler B, Sahm F, et al. Primary central nervous system lymphoma and atypical glioblastoma: multiparametric differentiation by using diffusion-, perfusion-, and susceptibility-weighted MR imaging. Radiology 2014; 272:843–850.
7. Koeller KK, Smirniotopoulos JG, Jones RV. Primary central nervous system lymphoma: radiologic-pathologic correlation. Radiographics 1997; 17:1497–1526.
8. Rees JH, Smirniotopoulos JG, Jones RV, Wong K. Glioblastoma multiforme: radiologic–pathologic correlation. Radiographics 1996; 16:1413–1438; quiz 1462–1463.
9. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. AJR Am J Roentgenol 2017; 208:754–760.
10. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015; 115:211–252.
11. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. Radiographics 2017; 37: 505–515.
12. Alcaide-Leon P, Dufort P, Geraldo AF, et al. Differentiation of enhancing glioma and primary central nervous system lymphoma by texture-based machine learning. AJNR Am J Neuroradiol 2017; 38:1145–1150.
13. Yamasaki T, Chen T, Hirai T, Murakami R. Classification of cerebral lymphomas and glioblastomas featuring luminance distribution analysis. Comput Math Methods Med 2013; 2013:619658.
14. Liu YH, Muftah M, Das T, Bai L, Robson K, Auer D. Classification of MR tumor images based on gabor wavelet analysis. Journal of Medical and Biological Engineering 2012; 32:22–28.
15. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Informatics 2006; 2:59–77.
16. Kunimatsu A, Kunimatsu N, Kamiya K, Watadani T, Mori H, Abe O. Comparison between glioblastoma and primary central nervous system lymphoma using MR image-based texture analysis. Magn Reson Med Sci 2018; 17:50–57.
17. Lee HY, Kim HS, Park JW, Baek HJ, Kim SJ, Choi CG. Atypical imaging features of Epstein-Barr virus-positive primary central nervous system lymphomas in patients without AIDS. AJNR Am J Neuroradiol 2013; 34: 1562–1567.
18. Parmar C, Rios Velazquez E, Leijenaar R, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. PLoS One 2014; 9:e102107.
19. Bartko JJ. On various intraclass correlation reliability coefficients. Psychol Bull 1976; 83:762–765.
20. Anvari A, Halpern EF, Samir AE. Statistics 101 for radiologists. Radiographics 2015; 35:1789–1801.

21. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol 2008; 31:466–475.
22. Kassner A, Thornhill RE. Texture analysis: a review of neurologic MR imaging applications. AJNR Am J Neuroradiol 2010; 31:809–816.
23. Somorjai RL, Dolenko B, Baumgartner R. Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. Bioinformatics 2003; 19:1484–1491.
24. Sturges HA. The choice of a class interval. J Am Stat Assoc 1926; 21:65–66.
25. Cristianini N, Shawe-Taylor J. An introduction to support vector machines: and other kernel-based learning methods; Cambridge University Press: Cambridge, New York, 2000; 189.
26. Mukherjee S, Tamayo P, Rogers S, et al. Estimating dataset size requirements for classifying DNA microarray data. J Comput Biol 2003; 10:119–142.
27. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Stat Surv 2010; 4:40–79.
28. Zhang X, Yan LF, Hu YC, et al. Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features. Oncotarget 2017; 8:47816–47830.
29. Ancona N, Maglietta R, Piepoli A, et al. On the statistical assessment of classifiers using DNA microarray data. BMC Bioinform 2006; 7:387.
30. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinform 2011; 12:77.
31. Doskaliyev A, Yamasaki F, Ohtaki M, et al. Lymphomas and glioblastomas: differences in the apparent diffusion coefficient evaluated with high b-value diffusion-weighted magnetic resonance imaging at 3T. Eur J Radiol 2012; 81:339–344.
32. Toh CH, Wei KC, Chang CN, Ng SH, Wong HF. Differentiation of primary central nervous system lymphomas and glioblastomas: comparisons of diagnostic performance of dynamic susceptibility contrast-enhanced perfusion MR imaging without and with contrast-leakage correction. AJNR Am J Neuroradiol 2013; 34:1145–1149.
33. Wang S, Kim S, Chawla S, et al. Differentiation between glioblastomas, solitary brain metastases, and primary cerebral lymphomas using diffusion tensor and dynamic susceptibility contrast-enhanced MR imaging. AJNR Am J Neuroradiol 2011; 32:507–514.
34. Wei L, Yang Y, Nishikawa RM, Jiang Y. A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. IEEE Trans Med Imaging 2005; 24:371–380.
35. Rodriguez Gutierrez D, Awwad A, Meijer L, et al. Metrics and textural features of MRI diffusion to improve classification of pediatric posterior fossa tumors. AJNR Am J Neuroradiol 2014; 35:1009–1015.
36. Tiwari P, Prasanna P, Wolansky L, et al. Computer-extracted texture features to distinguish cerebral radionecrosis from recurrent brain tumors on multiparametric MRI: a feasibility study. AJNR Am J Neuroradiol 2016; 37:2231–2236.
37. Waljee AK, Higgins PDR, Singal AG. A primer on predictive models. Clin Transl Gastroenterol 2014; 5:e44.
38. Chawla NV, Japkowicz N, Kotcz A. Editorial: special issue on learning from imbalanced data sets. SIGKDD Explor 2004; 6:1–6.
39. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. Radiology 2018; 286:887–896.
40. Raza SM, Lang FF, Aggarwal BB, Fuller GN, Wildrick DM, Sawaya R. Necrosis and glioblastoma: a friend or a foe? A review and a hypothesis. Neurosurgery 2002; 51:2–12; discussion 12–13.