AJNR

Differentiation of Enhancing Glioma and Primary Central Nervous System Lymphoma by Texture-Based Machine Learning

P. Alcaide-Leon, P. Dufort, A.F. Geraldo, L. Alshafai, P.J. Maralani, J. Spears and A. Bharatha

Differentiation of Enhancing Glioma and Primary Central Nervous System Lymphoma by Texture-Based Machine Learning

P. Alcaide-Leon, P. Dufort, A.F. Geraldo, L. Alshafai, P.J. Maralani, J. Spears, and A. Bharatha

ABSTRACT

BACKGROUND AND PURPOSE: Accurate preoperative differentiation of primary central nervous system lymphoma and enhancing glioma is essential to avoid unnecessary neurosurgical resection in patients with primary central nervous system lymphoma. The purpose of the study was to evaluate the diagnostic performance of a machine-learning algorithm by using texture analysis of contrast-enhanced T1-weighted images for differentiation of primary central nervous system lymphoma and enhancing glioma.

MATERIALS AND METHODS: Seventy-one adult patients with enhancing gliomas and 35 adult patients with primary central nervous system lymphomas were included. The tumors were manually contoured on contrast-enhanced T1WI, and the resulting volumes of interest were mined for textural features and subjected to a support vector machine–based machine-learning protocol. Three readers classified the tumors independently on contrast-enhanced T1WI. Areas under the receiver operating characteristic curves were estimated for each reader and for the support vector machine classifier. A noninferiority test for diagnostic accuracy based on paired areas under the receiver operating characteristic curve was performed with a noninferiority margin of 0.15.

RESULTS: The mean areas under the receiver operating characteristic curve were 0.877 (95% CI, 0.798–0.955) for the support vector machine classifier; 0.878 (95% CI, 0.807–0.949) for reader 1; 0.899 (95% CI, 0.833–0.966) for reader 2; and 0.845 (95% CI, 0.757–0.933) for reader 3. The mean area under the receiver operating characteristic curve of the support vector machine classifier was significantly noninferior to the mean area under the curve of reader 1 ($P = .021$), reader 2 ($P = .035$), and reader 3 ($P = .007$).

CONCLUSIONS: Support vector machine classification based on textural features of contrast-enhanced T1WI is noninferior to expert human evaluation in the differentiation of primary central nervous system lymphoma and enhancing glioma.

ABBREVIATIONS: AUC = area under the receiver operating characteristic curve; PCNSL = primary central nervous system lymphoma; SVM = support vector machine

G liomas and primary central nervous system lymphoma (PCNSL) represent the 2 most common primary malignant brain tumors.[1] Treatment of PCNSL consists of chemotherapy and/or radiation.[2] Because resection of PCNSL confers no survival benefit for patients,[3] stereotactic brain biopsy sampling is the standard procedure for obtaining a pathologic diagnosis.[4] In

high-grade gliomas, on the contrary, extensive resections have been shown to improve survival.[5,6] Accurate preoperative diagnosis is also important to avoid administration of steroids before biopsy in PCNSL because this medication can cause false-negative results of histologic examinations.[7]

Differentiation between enhancing glial tumors and PCNSL by conventional MR imaging can be challenging. Multiple imaging techniques have been used to solve this problem, including different types of MR perfusion,[8-10] ADC quantification,[10,11] SWI,[12] DTI,[13] and [18F]-fluorodeoxyglucose positron-emission tomography.[14] Texture analysis has also been used to differentiate high-grade gliomas and PCNSL,[15,16] and only 1 study[15] has combined this approach with machine learning to improve the diagnostic accuracy of textural features on conventional MR images. To our knowledge, no prior studies on the differentiation between glioma and lymphoma have adequately compared the accuracy of a machine-learning algorithm and neuroradiologists.

PCNSL typically demonstrates intense homogeneous en-

hancement as opposed to more heterogeneous enhancement of glial tumors. We hypothesized that the extraction of textural features of tumors and posterior input of these features in a machine-learning algorithm could provide a model for accurate and robust tumor classification. In machine learning, support vector machines (SVMs) are supervised learning algorithms that analyze data used for classification. From a set of training examples, each of them belonging to one of the categories, the SVM can build a model that classifies new data in the different categories. The purpose of this study was 3-fold: 1) to develop a classification model by using texture analysis and a machine-learning algorithm to differentiate PCNSL and enhancing glial tumors; 2) to compare the diagnostic accuracy of the SVM classifier with that of neuroradiologists; and 3) to examine whether the SVM classifier and the radiologists tend to misclassify the same cases.

MATERIALS AND METHODS
Study Design
A noninferiority statistical design with a noninferiority margin of 0.15 was adopted for this study. The study entailed comparisons of diagnostic accuracy between the radiologists and the SVM classifier in the differentiation of enhancing glioma and PCNSL. The area under the receiver operating characteristic curve (AUC) was the primary outcome measure. The sample size for the comparison of diagnostic accuracies between the radiologists and the SVM classifier was estimated by using 1-sided calculations with an α of .05 and a power of 80% based on a noninferiority margin[17] of -15%. The selection of this noninferiority margin was based on the goal of this technique not substituting for the radiologist's judgment but assisting in the diagnosis; therefore a noninferiority margin of -15% seems clinically acceptable. A priori sample size calculation was based on prior reported accuracies of 99.1% for texture analysis in a machine-learning algorithm[16] and 88.9% for radiologists.[16] The total sample size required was 22 (11 gliomas and 11 PCNSLs) according to the formula described by Blackwelder[18]: $n = f(\alpha, \beta) \times (\pi s \times [100 - \pi s] + \pi e \times [100 - \pi e]) / (\pi s - \pi e - d)^2$, where πs and πe are the true percentage "success" in the standard and experimental treatment group, respectively, and $f(\alpha, \beta) = [\Phi\text{-}1(\alpha) + \Phi - 1(\beta)]^2$, with Φ-1 being the cumulative distribution function of a standardized normal deviate. We opted for a more conservative approach with a larger sample size because our sample of tumors was more heterogeneous compared with other studies and accuracies may differ substantially.

Subjects
Institutional review board approval was obtained and informed consent was waived for this Health Insurance Portability and Accountability Act–compliant retrospective study. Inclusion criteria consisted of consecutive adult patients (older than 18 years of age) with a pathologic diagnosis of PCNSL or enhancing glial tumor and preoperative MR imaging performed at St. Michael's Hospital, including contrast-enhanced T1WI, between January 2005 and December 2015. An exclusion criterion was poor image quality due to motion or other artifacts. A random sample of 10% of patients with enhancing gliomas and 20% of patients with PCNSLs was selected. Two patients with enhancing glial tumors were excluded due to motion artifacts degrading the images. One

hundred six patients were included (71 patients with enhancing glial tumors and 35 patients with PCNSLs). Surgery and histologic evaluation were performed within a month interval after imaging.

Image Acquisition
Thirty-two patients (20 with gliomas and 12 with PCNSLs) were scanned in a 3T magnet (Magnetom Skyra; Siemens, Erlangen, Germany) equipped with a 20-channel head-neck coil. A T1WI FLASH sequence (TR/TE, 250/2.49 ms; flip angle, 70°; section thickness, 5 mm; in-plane voxel size, 0.6 × 0.6 mm; FOV, 200 mm; gap, 0.5 mm; NEX, 1) was performed after administration of 10 mL of gadobenate dimeglumine. The total duration of the sequence was 1:1 minutes. Seventy-four patients (51 with gliomas and 23 with PCNSLs) were scanned in a 1.5T magnet (Intera; Philips Healthcare, Best, the Netherlands) equipped with a 6-channel head coil. A T1WI spin-echo sequence was acquired in the axial plane after administration of 10 mL of gadobenate dimeglumine (TR/TE, 400/8 ms; flip angle, 90°; section thickness, 5 mm; in-plane voxel size, 0.83 × 0.83 mm; FOV, 200 mm; gap, 1 mm; NEX, 2). The total duration of the sequence was 4:19 minutes.

Reading of Radiologists
Three neuroradiologists (L.A., A.F.G., and P.J.M. with 3, 2, and 4 years of experience in neuroradiology after residency) classified 106 tumors as gliomas or PCNSLs independently and blinded to clinical information and pathology reports. The 3 readers evaluated the contrast-enhanced T1WI of 106 patients and recorded their diagnoses and degrees of confidence by using a 4-point scale: 1, definite glioma; 2, likely glioma; 3, likely PCNSL; and 4, definite PCNSL. The readers were selected from other hospitals to ensure lack of prior exposure to the cases, and they were not informed of the number of cases in each category. The readers spent between 1 and 2 hours reviewing the images.

Texture Metrics
A neuroradiologist (P.A.-L.) with 6 years of experience in neuroradiology created tumor volumes of interest by contouring the outer margin of the enhancing component of the tumors in all sections on the contrast-enhanced T1WI sequence. In cases of multiple enhancing lesions, only the 2 largest lesions were contoured. The process of manual VOI generation took around 10 hours.

The generation of the texture features was accomplished by using a customized code written by one of the authors (P.D.) and took on the order of a few seconds for each study. The calculation of most texture features involves 2 steps: The first is the accumulation of histograms, and the second is the evaluation of nonlinear functions that take the histograms as input. The first-order texture metrics require 1D histograms that count the number of times image voxels of each possible value occur in the VOI. The functions that take these histograms as input can evaluate percentiles of the distribution or other measures of its shape such as means, variances, skewness, and kurtosis. The second-order metrics are based on 2D histograms that count the number of times voxels of one value are found spatially adjacent to voxels of another value over the entire VOI. Many nonlinear functions take

these histograms as input to produce second-order texture metrics such as entropy, correlation, contrast, and the angular second moment.

A set of 11 first-order and 142 second-order texture metrics was generated from each VOI. The first-order metrics consisted of the 11 image-intensity percentiles from each VOI, ranging from 0% (minimum value) to 100% (the maximum value) with 9 steps of 10% between them. These metrics provide a characterization of the 1D image-intensity histogram shape.

Before we computed the 142 second-order texture metrics, the intensities within each VOI were binned into 32 equal-sized bins spanning the range of image intensities between the first percentile at the bottom and the 99th percentile at the top. The binning is a standard technique for minimizing histogram noise when computing second-order texture metrics, while the use of image intensities between the first and 99th percentiles serves to minimize the effect of outliers on the bin layout. The second-order texture features consisted of metrics from 4 classes computed from multidimensional histograms: 1) the mean and range of the 13 Haralick features computed from the gray-scale co-occurrence matrix[19] taken over all 13 neighbor orientations[20]; 2) 5 features based on the neighborhood gray tone difference matrix[21]; 3) 10 features from the gray-level run-length matrix[22]; and 4) the same 10 features from the gray-level size zone matrix.[23] A detailed, illustrated description of these metrics has been previously published.[20] The result of this computation is a set of 153 texture features that are then fed into the machine-learning algorithm as predictors.

Machine Learning

The goal of the machine learning was to train a classifier to predict whether each tumor was a glioma or lymphoma based on the texture features extracted from the VOIs. All machine learning was performed by using the SVM algorithm with a radial basis function kernel. The Matlab (MathWorks, Natick, Massachusetts) interface to the LibSVM software library (http://www.csie.ntu.edu.tw/~cjlin/libsvm/)[24] was used to apply the SVM training algorithm to the data. The SVM[25] was selected over other machine-learning methods such as deep learning (eg, convolutional neural networks) for 2 reasons: first, because deep learning in general and convolutional neural networks in particular requires very large datasets for training; second, because the tumors investigated in this study have very predictable internal structures and whatever exploitable regularity may be present in tumors has so far been shown to be primarily statistical in nature, a category of pattern that is much better quantified by using texture metrics than convolutional kernels. For each SVM training run, it was necessary to tune 3 hyperparameters governing the behavior of the classifier. The first hyperparameter pertained to feature selection. An *F*-statistic approach[26] was used to rank the 153 input texture features in the order of their association with the response classification. A tunable hyperparameter representing the fraction of the most highly associated features to keep was then applied to select the features that were used. The second hyperparameter was the standard cost parameter common to all types of SVM, while the third was the width of the Gaussian that makes up the radial basis function kernel.

A nested cross-validation scheme was used to tune the 3 hyperparameters while keeping the assessment of accuracy completely independent. In each of 100 iterations of the outer loop, 10-fold cross-validation was used to hold out 10% of the data for testing, while the remaining 90% was passed to the inner loop. Within the inner loop, a further 10-fold cross-validation protocol was used for each point in a 3D grid covering a range of fractions of the best features to retain, values of the SVM cost parameter, and values of the radial basis function width. The inner loop cross-validation result was recorded for each grid point searched, and at the conclusion of the inner loop, the best performing triple of the hyperparameters was used to train a classifier by using all of the inner loop data. This classifier was then applied to classify the held-out data from the outer loop. A SVM classifier does not produce a dichotomous binary classification as its output, but rather a single, continuous number on the real line. Only when a threshold is applied, is it transformed into a classification. Repeating the outer loop of the nested cross-validation protocol 100 times yields 100 such numbers for each tumor. Each of the 100 numbers for a particular tumor represents an instance in which it was held out during cross-validation with a different 10% of the data. The percentage of trials in which each case was classified as a PCNSL was recorded.

The training of the classifier took a few days of computer time to complete, and the estimation of the accuracy of the classifier took 3 weeks. After the classifier has been produced, its application to each new case in a production environment would take only a small fraction of a second.

Statistical Analysis

Receiver operating characteristic curves were constructed for each reader and for the SVM classifier by using SPSS, Version 21 (IBM, Armonk, New York). For the receiver operating characteristic curve and AUC calculation, glioma was considered "negative" and PCNSL was considered "positive." The AUCs were estimated in each case by nonparametric methods. The noninferiority test for diagnostic accuracy based on the paired AUCs described in Zhou et al[27] was performed to compare each radiologist with the SVM classifier. The standard error of the difference between AUCs was calculated by taking into account the correlation derived from the paired nature of the data as described by Hanley and McNeil.[28]

To assess whether the radiologists and the SVM classifier tended to misclassify the same cases, we estimated interrater agreement among the 3 readers, and the SVM classifier was estimated by a linearly weighted κ.[29] The results from the SVM classifier were simplified to 4 categories so that they could be compared with the radiologists' readings. These categories were defined by the percentage of trials in which each case classified as PCNSL: 0%–25%, definite glioma; 26%–50%, likely glioma; 51%–75%, likely PCNSL; and 76%–100%, definite PCNSL.

RESULTS

In the glioma group ($n = 71$), there were 23 women (mean age, 59.5 years; range, 33–88 years) and 48 men (mean age, 54.5 years; range, 19–84 years). Two gliomas were grade III, and 69 were grade IV. In the PCNSL group ($n = 35$), there were 14 women
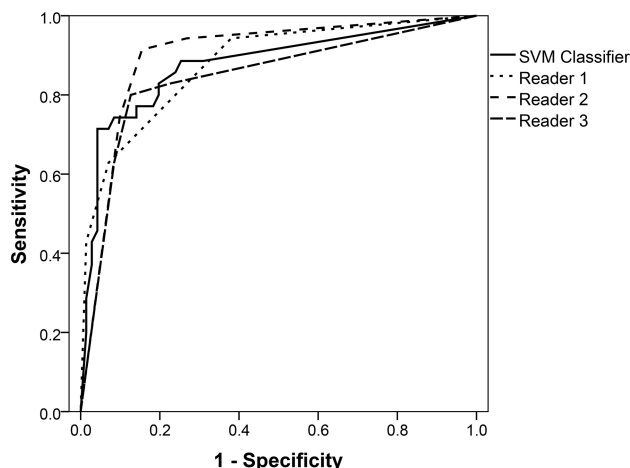
FIG 1. Receiver operating characteristic curves for discrimination of primary central nervous system lymphoma (positive) and glioblastoma (negative) of the support vector machine classifier (*continuous line*) and the 3 readers (*dashed lines*). The mean areas under the curve estimated under the nonparametric assumption were 0.877 (95% confidence interval, 0.798−0.955) for the SVM classifier; 0.878 (95% confidence interval, 0.807−0.949) for reader 1; 0.899 (95% confidence interval, 0.833−0.966) for reader 2; and 0.845 (95% confidence interval, 0.757−0.933) for reader 3.

Table 1: Differences in mean AUC between the SVM classifier and the neuroradiologists

Comparison	Difference	95% CI for Difference	P Value[a]
SVM classifier vs reader 1	−0.001	−0.096, 0.094	.021
SVM classifier vs reader 2	−0.022	−0.106, 0.062	.035
SVM classifier vs reader 3	0.032	−0.074, 0.138	.007

[a] P value for the test of noninferiority.

(mean age, 55.7 years; range, 41–71 years) and 21 men (mean age, 58.9 years; range, 39−83 years). Thirty-four cases of PCNSL corresponded to diffuse large B-cell lymphomas, and 1 was a T-cell lymphoma. Thirty-three cases of PCNSL occurred in immunocompetent patients, 1 in a patient with HIV, and 1 corresponded to an Epstein-Barr virus–driven lymphoma in a patient with a kidney transplant.

Diagnostic Accuracy

The mean AUCs were 0.877 (95% CI, 0.798−0.955) for the SVM classifier; 0.878 (95% CI, 0.807−0.949) for reader 1; 0.899 (95% CI, 0.833−0.966) for reader 2; and 0.845 (95% CI, 0.757−0.933) for reader 3. Receiver operating characteristic curves are shown in Fig 1. The mean AUC of the SVM classifier was significantly noninferior to the radiologists' mean AUCs. Differences in the AUCs between the SVM classifier and each of the readers are detailed in Table 1 and featured in Fig 2.

Agreement

Table 2 shows the linearly weighted Cohen κ coefficients for each pair of readers or reader-SVM classifier. Agreement was slightly higher among radiologists than between the SVM classifier and the radiologists.

Figure 3 shows the percentage of correctly classified trials by the SVM classifier in the order of decreasing accuracy on a case-by-case basis. The number of radiologists who classified each tumor correctly is also represented.
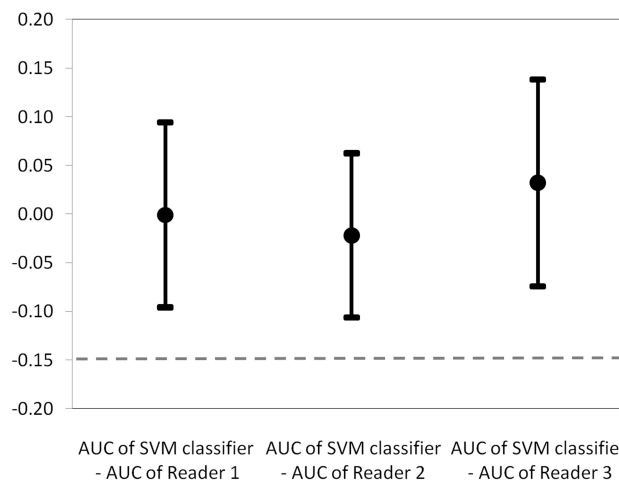
FIG 2. The chart shows the mean differences in area under the curve (95% confidence interval between the support vector machine classifier and reader 1 = −0.001 (95% CI, −0.096−0.094); between the SVM classifier and reader 2 = −0.022 (95% CI, −0.106−0.062); and between SVM classifier and reader 3 = 0.032 (95% CI, −0.074−0.138). All the confidence intervals sit wholly above the −0.15 limit (*dashed line*) representing the noninferiority margin.

Table 2: Linearly weighted κ coefficients representing the agreement between the neuroradiologists and the SVM classifier

|	Reader 1	Reader 2	Reader 3	SVM Classifier
Reader 1	−	0.58	0.56	0.40
Reader 2		−	0.63	0.55
Reader 3			−	0.46

Figure 4 shows images from 2 cases in which there was agreement between the radiologists but a mismatch between the SVM classifier and the radiologists.

DISCUSSION

This article presents an SVM classification scheme for differentiating enhancing glioma and PCNSL noninferior to human evaluation. Prior studies with smaller samples have used texture analysis for differentiation of PCNSL and glioblastoma with[16] and without [15] machine learning. Yamasaki et al,[15] in a study including 40 patients, reported an accuracy of 91%. Their higher accuracy can be explained by lack of grade III glial tumors in their sample, which was limited to grade IV glial tumors. Grade III tumors typically lack necrosis, making the differential diagnosis with PCNSL more challenging. This study also lacks details regarding enrollment and a comparison with the accuracy of radiologists. The work by Liu et al,[16] also based on texture analysis, incorporates machine learning. They included only 18 patients and excluded not only non-grade IV glial tumors but also immunocompromised patients with PCNSLs. PCNSLs in immunocompromised patients commonly show atypical features (necrosis and hemorrhage), mimicking high-grade glial tumors and metastases. These exclusion criteria may explain the high accuracy of the machine learning algorithm (99.1%) in the work by Liu et al,[16] which was reported to be higher than that of the radiologists (88.9%) despite lack of statistical analysis for this comparison. In summary, prior studies on the topic lack representative samples and direct comparison with the diagnostic performance of radiologists. Our study on a random sample of
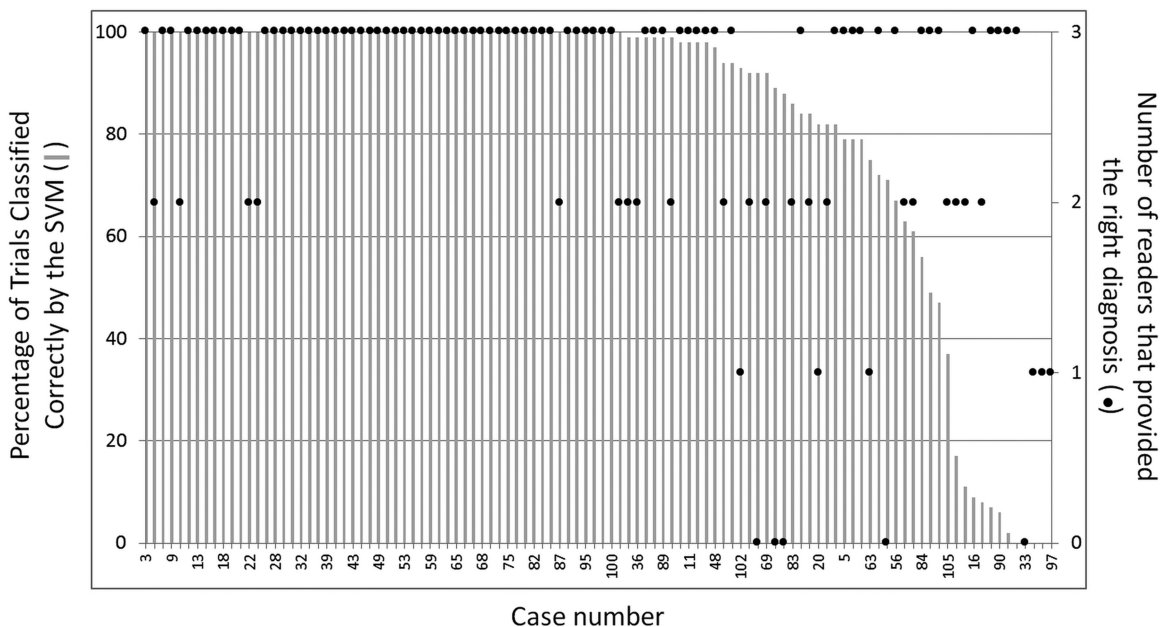
FIG 3. Comparison between the accuracy of the radiologists and the support vector machine classifier for each of the 106 cases. The horizontal axis shows the different cases sorted in order of decreasing SVM classifier accuracy. The left vertical axis shows the percentage of correctly classified trials by the SVM across 100 nested cross-validation trials. The right vertical axis shows the number of radiologists that classified the tumor correctly. For this graph, the results of the radiologists were simplified to 2 categories "glioma" and "lymphoma" without taking into account the degree of certainty. Although agreement is slightly better among radiologists than between radiologists and the SVM classifier, the cases in which the SVM provides different results for different trials (midright area of the graph) correspond to cases with more disagreements among the radiologists.

FIG 4. *A*, Axial contrast-enhanced T1-weighted image of a 51-year-old woman with a grade IV glioma. All 3 radiologists incorrectly classified the tumor as PCNSL, whereas the SVM classified it correctly in 92% of the trials. *B* and *C*, Axial contrast-enhanced T1WI of a 47-year-old woman with a grade IV glioma. All 3 radiologists incorrectly classified the tumor as PCNSL, whereas the SVM classifier provided the right diagnosis in 88% of the trials.

consecutive patients including 106 subjects is more likely to encompass the whole imaging spectrum of enhancing gliomas and PCNSLs, providing more realistic estimates of diagnostic accuracy than prior work.

The radiologists tended to agree slightly more among themselves than with the SVM classifier. It is interesting to analyze the disagreements, particularly those cases in which the SVM provided the right diagnosis and the radiologists failed. Figure 4 shows 2 such cases. In the case illustrated in Fig 4A, the tumor has a very heterogeneous appearance, more typical of gliomas; however, the radiologists classified it as a lymphoma, likely due to the periventricular location. The SVM classifier, on the contrary, only

uses textural information and classified the case correctly as a glioma. One of the sources of disagreements between radiologists and the SVM classifier may be that radiologists take other tumor features into account such as the location and the presence of nonenhancing infiltrative components. Another possible source of disagreement is that the SVM classifier had only textural information from the 2 largest enhancing lesions, whereas the radiologists analyzed the whole brain. In the future, SVM and other types of machine-learning algorithms will be able to analyze the full dataset of images, combine it with the clinical information, and provide more reliable results. Adequately trained SVMs may support preoperative tumor diagnosis, especially in centers with-

out experienced neuroradiologists. This support will help avoid unnecessary neurosurgical resections in patients with PCNSL.

Our study has a number of limitations. First, the evaluation of contrast-enhanced T1WI in isolation from other valuable sequences such as ADC, perfusion, and T2 gradient-echo is not representative of the real clinical scenario. Second, the requirement of VOI tracing from an expert makes our approach semiautomatic and therefore subject to intra- and interobserver variability. Third, only the 2 largest enhancing lesions were segmented and analyzed by the SVM in cases of multiple lesions.

CONCLUSIONS

Our results show that SVMs can be trained to distinguish PCNSL and enhancing gliomas on the basis of textural features of contrast-enhanced T1WI with an accuracy significantly noninferior to that of neuroradiologists. The testing of larger datasets including other MR images will not only provide better accuracy estimations but also further improve the performance of the classifier, because SVM classification systems benefit from more extensive training.

REFERENCES

1. Dolecek TA, Propp JM, Stroup NE, et al. **CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005–2009.** *Neuro Oncol* 2012;14(suppl 5):v1–49 CrossRef Medline
2. DeAngelis L. **Primary CNS lymphoma: treatment with combined chemotherapy and radiotherapy.** *J Neurooncol* 1999;43:249–57 CrossRef Medline
3. Bataille B, Delwail V, Menet E, et al. **Primary intracerebral malignant lymphoma: report of 248 cases.** *J Neurosurg* 2000;92:261–66 CrossRef Medline
4. Sherman ME, Erozan YS, Mann RB, et al. **Stereotactic brain biopsy in the diagnosis of malignant lymphoma.** *Am J Clin Pathol* 1991;95:878–83 CrossRef Medline
5. Sanai N, Polley MY, McDermott MW, et al. **An extent of resection threshold for newly diagnosed glioblastomas.** *J Neurosurg* 2011;115:3–8 CrossRef Medline
6. Li YM, Suki D, Hess K, et al. **The influence of maximum safe resection of glioblastoma on survival in 1229 patients: can we do better than gross-total resection?** *J Neurosurg* 2016;124:977–88 CrossRef Medline
7. Weller M. **Glucocorticoid treatment of primary CNS lymphoma.** *Neurooncol* 1999;43:237–39 CrossRef Medline
8. Kickingereder P, Sahm F, Wiestler B, et al. **Evaluation of microvascular permeability with dynamic contrast-enhanced MRI for the differentiation of primary CNS lymphoma and glioblastoma: radiologic-pathologic correlation.** *AJNR Am J Neuroradiol* 2014;35:1503–08 CrossRef Medline
9. Furtner J, Schöpf V, Preusser M, et al. **Non-invasive assessment of intratumoral vascularity using arterial spin labeling: a comparison to susceptibility-weighted imaging for the differentiation of primary cerebral lymphoma and glioblastoma.** *Eur J Radiol* 2014;83:806–10 CrossRef Medline
10. Shim WH, Kim HS, Choi CG, et al. **Comparison of apparent diffusion coefficient and intravoxel incoherent motion for differentiating among glioblastoma, metastasis, and lymphoma focusing on diffusion-related parameter.** *PLoS One* 2015;10:e0134761 CrossRef Medline
11. Ahn SJ, Shin HJ, Chang JH, et al. **Differentiation between primary cerebral lymphoma and glioblastoma using the apparent diffusion coefficient: comparison of three different ROI methods.** *PLoS One* 2014;9:e112948 CrossRef Medline
12. Radbruch A, Wiestler B, Kramp L, et al. **Differentiation of glioblastoma and primary CNS lymphomas using susceptibility weighted imaging.** *Eur J Radiol* 2013;82:552–56 CrossRef Medline
13. Toh CH, Castillo M, Wong AM, et al. **Primary cerebral lymphoma and glioblastoma multiforme: differences in diffusion characteristics evaluated with diffusion tensor imaging.** *AJNR Am J Neuroradiol* 2008;29:471–75 CrossRef Medline
14. Zhou W, Wen J, Hua F, et al. **FDG-PET in immunocompetent patients with primary central nervous system lymphoma: differentiation from GBM and correlation with DWI.** *J Nucl Med* 2016;57(suppl 2):1613
15. Yamasaki T, Chen T, Hirai T, et al. **Classification of cerebral lymphomas and glioblastomas featuring luminance distribution analysis.** *Comput Math Methods Med* 2013;2013:619658 CrossRef Medline
16. Liu Y, Muftah M, Das T, et al. **Classification of MR tumor images based on Gabor wavelet analysis.** *J Med Biol Eng* 2012;32:22–28 CrossRef
17. Ahn S, Park SH, Lee KH. **How to demonstrate similarity by using noninferiority and equivalence statistical testing in radiology research.** *Radiology* 2013;267:328–38 CrossRef Medline
18. Blackwelder WC. **"Proving the null hypothesis" in clinical trials.** *Control Clin Trials* 1982;3:345–53 CrossRef Medline
19. Haralick RM, Shanmugam K, Dinstein I. **Textural features for image classification.** *IEEE Trans Syst Man Cybern* 1973;3:610–21 CrossRef
20. Tixier F, Le Rest CC, Hatt M, et al. **Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer.** *J Nucl Med* 2011;52:369–78 CrossRef Medline
21. Amadasun M, King R. **Textural features corresponding to textural properties.** *IEEE Trans Syst Man Cybern* 1989;19:1264–74 CrossRef
22. Loh HH, Leu JG, Luo RC. **The analysis of natural textures using run length features.** *IEEE Trans Ind Electron* 1988;35:323–28 CrossRef
23. Thibault G, Fertil B, Navarro C, et al. **Shape and texture indexes application to cell nuclei classification.** *Intern J Pattern Recognit Artif Intell* 2013;27:1357002
24. Chang CC, Lin CJ. **LIBSVM: a library for support vector machines.** *ACM Trans Intell Syst Technol* 2011;2:1–27
25. Orrù G, Pettersson-Yeo W, Marquand AF, et al. **Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review.** *Neurosci Biobehav Rev* 2012;36:1140–52 CrossRef Medline
26. Chen YW, Lin CJ. **Combining SVMs with various feature selection strategies.** In: Guyon I, Nikravesh M, Gunn S, et al, eds. *Feature Extraction: Foundations and Applications.* Berlin, Heidelberg: Springer; 2006:315–324
27. Zhou XH, Obuchowski NA, McClish DK. **Comparing the accuracy of two diagnostic tests.** In: *Statistical Methods in Diagnostic Medicine.* Hoboken: John Wiley & Sons; 2011:165–92
28. Hanley JA, McNeil BJ. **A method of comparing the areas under receiver operating characteristic curves derived from the same cases.** *Radiology* 1983;148:839–43 CrossRef Medline
29. Fleiss JL, Cohen J. **The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability.** *Educ Psychol Meas* 1973;33:613–19