



Machine Learning and Deep Learning CT-Based Models for Predicting the Primary Central Nervous System Lymphoma and Glioma Types: A Multicenter Retrospective Study

Guang Lu¹, Yuxin Zhang², Wenjia Wang³, Lixin Miao^{4*} and Weiwei Mou^{5*}

¹ Department of Hematology, Shengli Oilfield Central Hospital, Dongying, China, ² Department of Neurosurgery, Guangrao County People's Hospital, Dongying, China, ³ GE Healthcare, Shanghai, China, ⁴ Department of Medical Imaging Center, Shengli Oilfield Central Hospital, Dongying, China, ⁵ Department of Pediatrics, Shengli Oilfield Central Hospital, Dongying, China

OPEN ACCESS

Edited by:

Hyunjin Park,
Sungkyunkwan University,
South Korea

Reviewed by:

Francesco Grussu,
Vall d'Hebron Institute of Oncology
(VHIO), Spain
Alissa A. Thomas,
University of Vermont, United States

*Correspondence:

Lixin Miao
miaolixin656@163.com
Weiwei Mou
MVVmouweiwei@126.com

Specialty section:

This article was submitted to
Neuro-Oncology and Neurosurgical
Oncology,
a section of the journal
Frontiers in Neurology

Received: 26 March 2022

Accepted: 23 June 2022

Published: 30 August 2022

Citation:

Lu G, Zhang Y, Wang W, Miao L and
Mou W (2022) Machine Learning and
Deep Learning CT-Based Models for
Predicting the Primary Central
Nervous System Lymphoma and
Glioma Types: A Multicenter
Retrospective Study.
Front. Neurol. 13:905227.
doi: 10.3389/fneur.2022.905227

Purpose and Background: Distinguishing primary central nervous system lymphoma (PCNSL) and glioma on computed tomography (CT) is an important task since treatment options differ vastly from the two diseases. This study aims to explore various machine learning and deep learning methods based on radiomic features extracted from CT scans and end-to-end convolutional neural network (CNN) model to predict PCNSL and glioma types and compare the performance of different models.

Methods: A total of 101 patients from five Chinese medical centers with pathologically confirmed PCNSL and glioma were analyzed retrospectively, including 50 PCNSL and 51 glioma. After manual segmentation of the region of interest (ROI) on CT scans, 293 radiomic features of each patient were extracted. The radiomic features were used as input, and then, we established six machine learning models and one deep learning model and three readers to identify the two types of tumors. We also established a 2D CNN model using raw CT scans as input. The area under the receiver operating characteristic curve (AUC) and accuracy (ACC) were used to evaluate different models.

Results: The cohort was split into a training (70, 70% patients) and validation cohort (31, 30% patients) according to the stratified sampling strategy. Among all models, the MLP performed best, with an accuracy of 0.886 and 0.903, sensitivity of 0.914 and 0.867, specificity of 0.857 and 0.937, and AUC of 0.957 and 0.908 in the training and validation cohorts, respectively, which was significantly higher than the three primary physician's diagnoses (ACCs ranged from 0.710 to 0.742, $p < 0.001$ for all) and comparable with the senior radiologist (ACC 0.839, $p = 0.988$). Among all the machine learning models, the AUC ranged from 0.605 to 0.821 in the validation cohort. The end-to-end CNN model achieved an AUC of 0.839 and an ACC of 0.840 in the validation cohort, which had no significant difference in accuracy compared to the MLP model ($p = 0.472$) and the senior radiologist ($p = 0.470$).

Conclusion: The established PCNSL and glioma prediction model based on deep neural network methods from CT scans or radiomic features are feasible and provided high performance, which shows the potential to assist clinical decision-making.

Keywords: computed tomography, glioma, primary central nervous system lymphoma, machine learning (ML), deep neural networks

INTRODUCTION

Primary central nervous system lymphoma (PCNSL) and glioma are two common kinds of malignant primary tumors (1). PCNSL consists of about 2% of brain tumors, with an increasing incidence over the past decades (2). PCNSL is an aggressive type of extra nodal lymphoma without coexisting systemic disease at diagnosis (3). Glioma is the most aggressive primary malignant brain tumor in adults, accounting for about 15% of brain tumors (4). The two brain tumors can be identified, localized, and characterized using conventional techniques, such as computed tomography (CT) and magnetic resonance imaging (MRI). It is important to differentiate PCNSL from glioma since treatment options are vastly different for the two diseases. Patients with PCNSL usually have a good response to noninvasive treatments, such as chemotherapy, target therapies, and whole brain radiation treatment (5). Resection provides no therapeutic benefit and is reserved only for rare cases of neurologic deterioration due to brain herniation. But for patients with glioma, the standard treatment is the invasive maximum safe surgical resection followed by concurrent chemo-radiotherapy (6). Therefore, preoperative differentiation of PCNSL and glioma is clinically critical to guide neurosurgical treatment strategies, avoid unnecessary and potentially harmful surgery, and thus optimize patient outcomes, quality of care, and cost-effectiveness (7).

Imaging has a central role in the differentiation task of PCNSL and glioma, especially for patients whose lesion pathology cannot be obtained by puncture. However, the two diseases are challenging to differentiate based on radiology alone since they share overlapping imaging characteristics. A simple and accurate method is needed to identify PCNSL and glioma. There are multiple previous studies attempted to distinguish the two tumors. One way is to use advanced imaging techniques, such as different types of MR perfusion (8, 9), diffusion-tensor imaging (DTI) (10), and dynamic CT perfusion (11). These advanced methods have been assessed with modest success but require additional expense, time, and radiation and may not be performed routinely for every patient. Most recently, machine learning (ML) and deep learning (DL) have been applied in correctly diagnosing PCNSL and glioma. Radiomics is one successful method that extracts high-dimensional quantitative features from medical images using data-characterization algorithms and provides the information that represents the underlying pathophysiology that is difficult to acquire by visual inspection (12, 13). Suh et al. demonstrated that the diagnostic performance of MR radiomics-based machine-learning algorithm for differentiating PCNSL from atypical

glioma yielded a better diagnostic performance than human radiologists and ADC values on a single medical center (14). Kunimatsu et al. developed an ML-based image classifier to differentiate between GBM and PCNSL using texture features from contrast-enhanced T1-weighted images although the prediction accuracy was only 75% on the test data (15). Yun et al. revealed that a combination of radiomic features and multilayer perceptron (MLP) network classifier served a high-performing and generalizable model for the two tumors classification tasks on a small MR dataset (16). Bathla et al. compared the predictive performance of various ML techniques to differentiate between PCNSL and glioma using a combination of various feature selection and ML algorithms on several MR sequences (17). Xia et al. (18) investigated the use of CNN model to differentiate between PCNSL and glioma without delineation from 289 MRI scans and proved that the proposed model was comparable to the radiomic models and radiologist.

Prior studies have shown the MR radiomics-based or MRI scan-based methods could successfully differentiate between PCNSL and glioma. However, some patients are not suitable for MR checks, such as critically ill patients with impaired consciousness and patients with implantable medical devices, including pacemakers. For these patients, CT check is routinely performed. CT is also a cost-effective check, and the scan time is short. To our knowledge, the effectiveness of ML or DL CT-based models for predicting the two tumors has been rarely explored. On the other hand, the amount, type, completeness, and diversity of data determine the performance of classification models. The results from multicenter trials are more representative for future clinical practice. Therefore, the aim of our study was to explore various machine learning and deep learning methods based on radiomic features extracted from CT scans or raw CT scans to predict PCNSL and glioma types and compare the performance of different models on multicenter data. For the end-to-end CNN model, we also employed visualization techniques to superimpose heatmaps which explained the important regions for making decision.

MATERIALS AND METHODS

Study Population

The retrospective study complied with the Declaration of Helsinki (2000) and was approved by the Independent Ethics Committees of the Shengli Oilfield Central Hospital, the Affiliated Hospital of Qingdao University, Yantai Yuhuangding Hospital, the Affiliated Hospital of Weifang Medical University, and the Affiliated Hospital of Binzhou Medical University. The requirement for informed patient consent was waived. Patients

who underwent brain CT and pathologically proven diagnosis of PCNSL and glioma between 1 August 2015 and 31 December 2020 were recruited from the electronic medical records. All 51 identified PCNSL cases and 56 randomly selected glioma cases were assessed for the exclusion and inclusion criteria. The exclusion criteria included the following: (1) previous brain biopsy or surgery before CT (3 glioma cases), (2) more than one lesion (1 glioma case), (3) the absence of available index CT scan (1 glioma case), and (4) unsuccessful feature extraction (1 PCNSL case). Finally, the cohort consisted of 50 PCNSL and 51 glioma who were successfully assessed (**Figure 1**).

Image Acquisition

All examinations were conducted on four CT scanners: two dual-source CT scanners with ≥ 64 -detector rows (Somatom Definition Flash/Force; Siemens, Forchheim, Germany) and two multi-detector rows (Optima 660, GE Healthcare, Milwaukee, WI, USA; Philips iCT 256, Philips Healthcare, Amsterdam, Nederland), following the guidelines of head CT. The tube voltage of the CT scan was 120 kV and the tube current was 200–250 mAs. Axial images were reconstructed with as slice thickness of 5 mm, spacing of 5 mm, and the matrix of 512×512 mm. The CT images were reconstructed with a standard kernel. The CT images were transferred to an external workstation (Syngo MMWP VE 36 A; Siemens Healthcare, Forchheim, Germany) for further postprocessing.

Image Preprocessing

The overview of the study workflow is provided in **Figure 2**. A radiologist (5 years of experience in radiology) from Shengli Oilfield Central Hospital reviewed the CT images and manually delineated the brain tumors at the axial site using MITK software version 2018.04.2 (www.mitk.org). The marked regions of interest (ROIs) were confirmed by another senior neuroradiologist (10 years of experience in radiology) who were blinded to the assessment. For the end-to-end CNN method, we converted 3D CT scans to 2D slices to overcome the small number of data points.

Feature Extraction

Radiomic features were extracted using an open-source python package, Pyradiomics 2.2.0 (<https://pyradiomics.readthedocs.io/en/latest/index.html>) (19). The binWidth was set to 25 and the interpolator was “sitkBSpline.” The resampledSpacing was (1, 1, 1). In total, 293 radiomic features of each patient were extracted, which included 18 first-order histogram features, 24 gray-level co-occurrence matrix features, 14 3D-shape features, 14 gray-level dependency matrix features, 16 gray-level size zone matrix features, 16 gray-level run length matrix features, 5 neighboring gray tone difference matrix features, and 186 Laplacian of Gaussian ($\text{LoG}_{\sigma=2.0/3.0}$) features.

Feature Selection

We preprocessed the data and normalized the extracted features. When the data value exceeded the range of mean and standard deviation, the median of specific variance vector was used to

replace the outliers. In addition, we standardized the data in a specific interval.

Due to the high dimension of the possible feature sets to be used relative to the sample size and highly correlated variables, feature selection is generally considered a critical piece of the model building process and could reduce overfitting in further differentiation model. A total of three feature selection methods were considered. The analysis of variance (ANOVA) threshold of 1.0 was applied first and 57 features remained. Then, we used filter model correlation analysis with a threshold of 0.7 and obtained 21 features. Finally, an embedded model gradient boosting decision tree (GBDT) was implemented and 7 features were selected. The feature selection process was performed using Python (version 3.5.6).

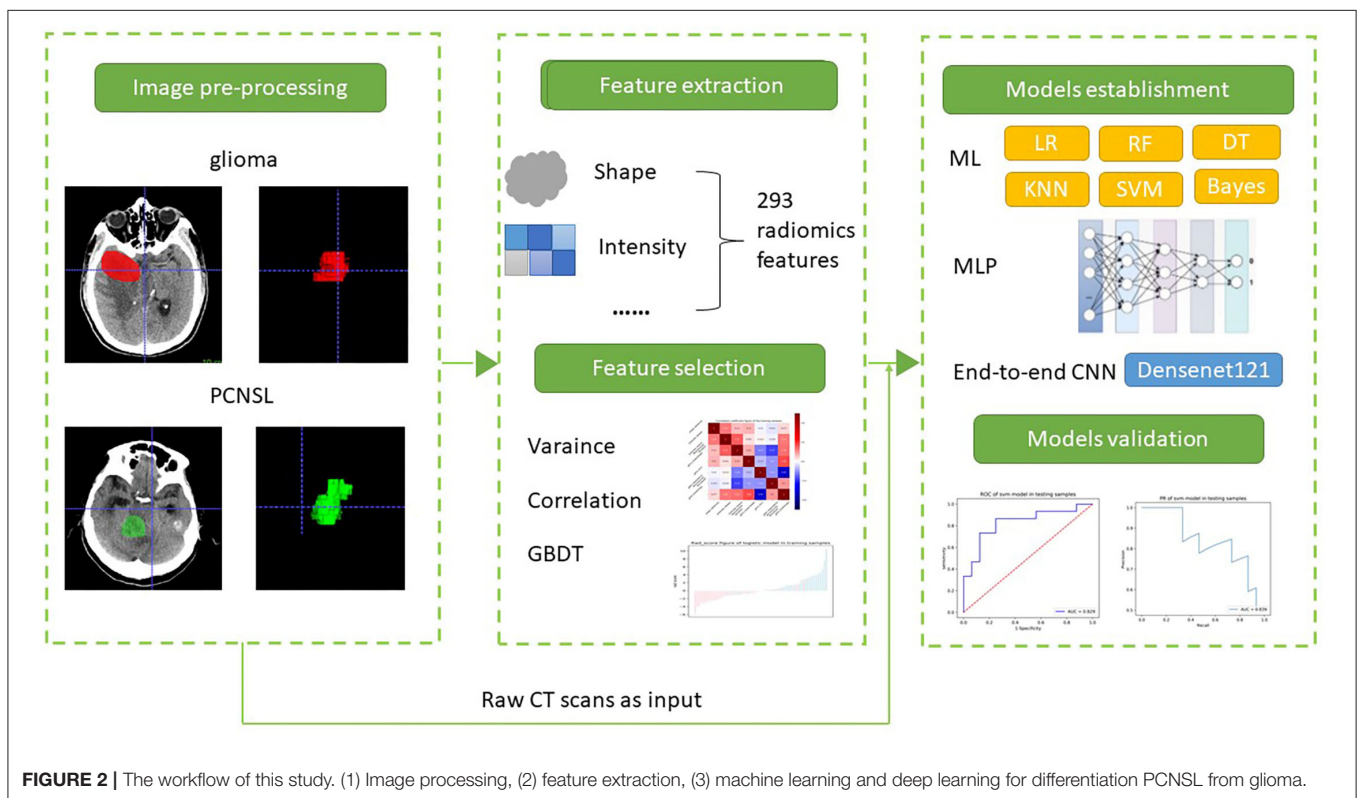
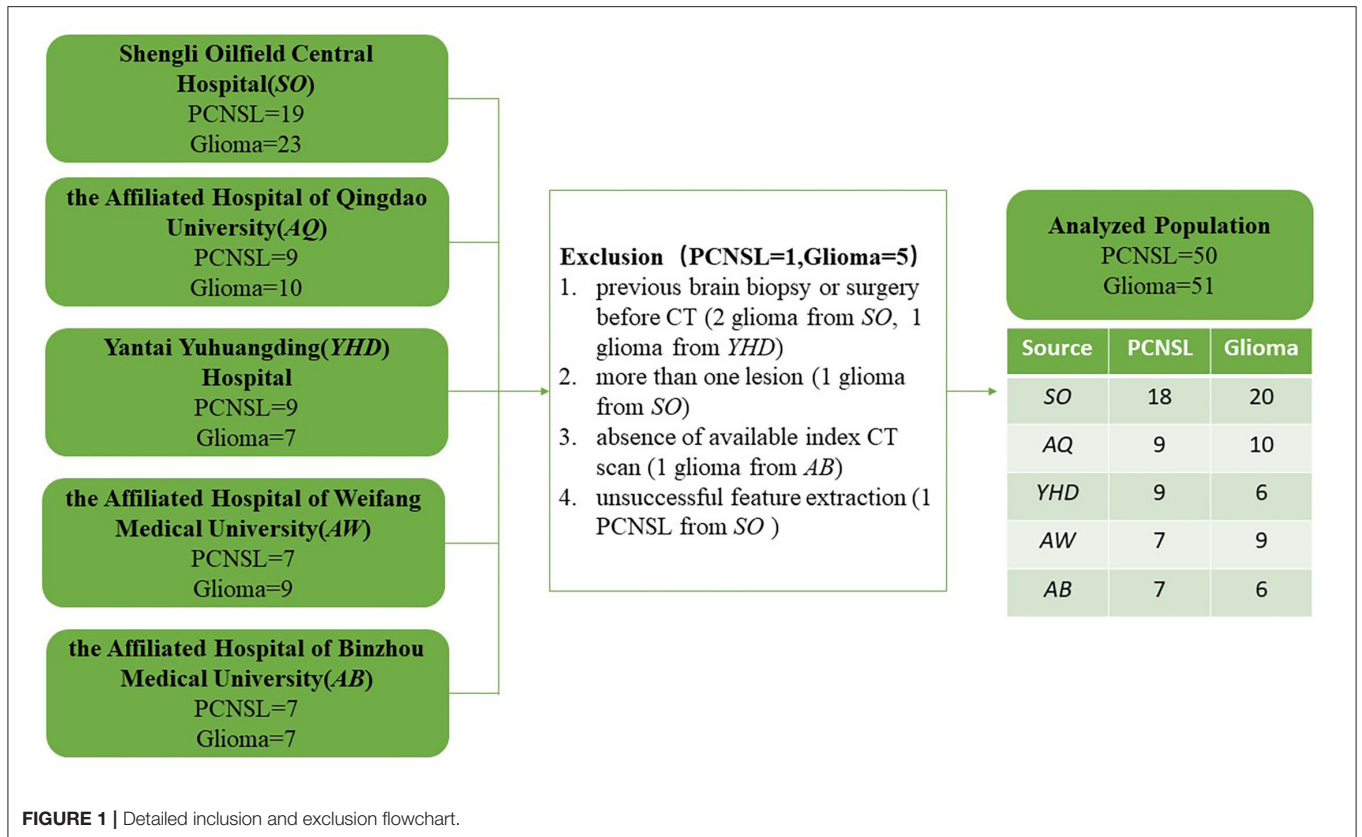
Machine Learning Models Establishment and Validation

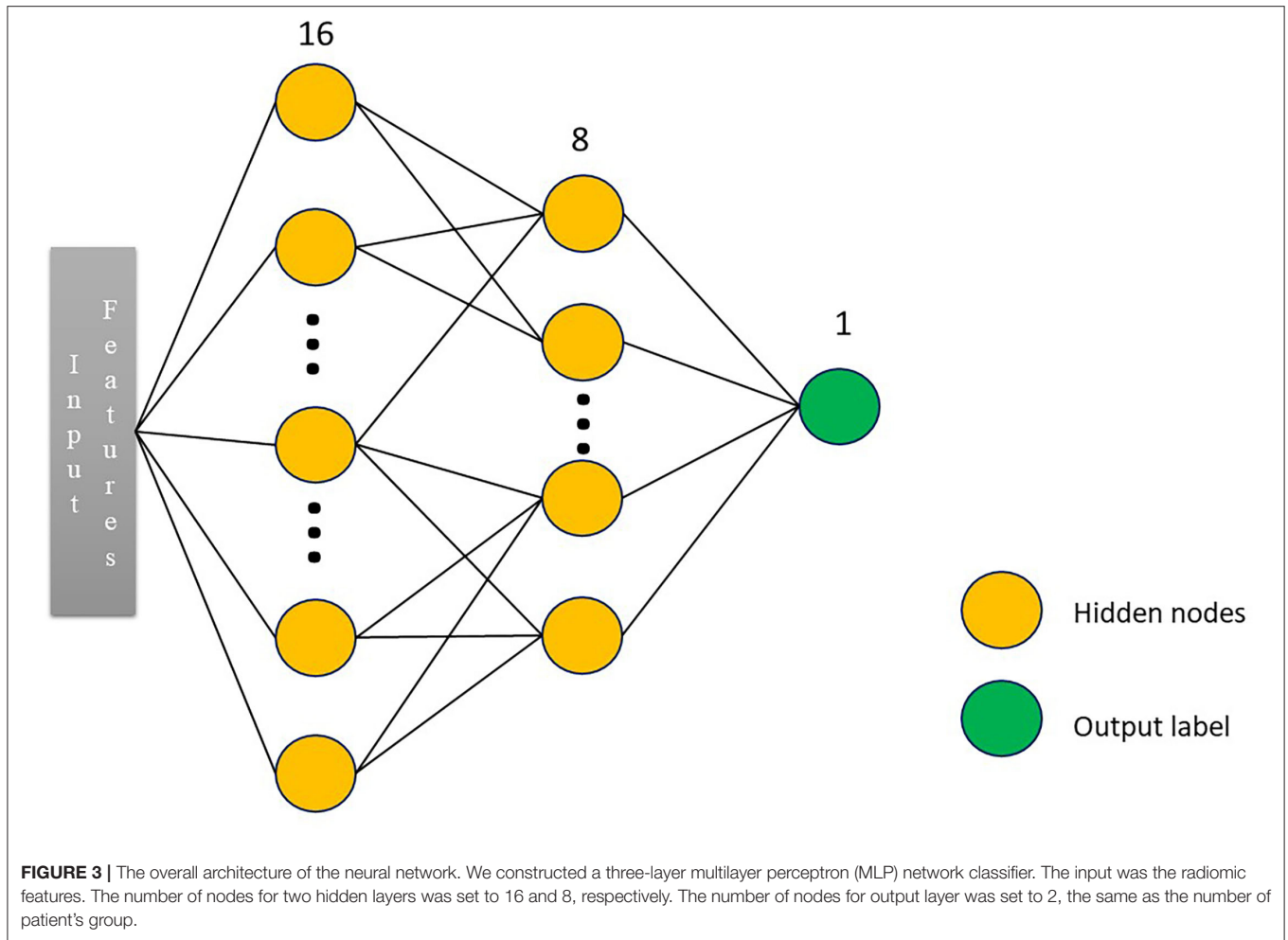
We randomly divided the patients into the training ($n = 70$, 35 PCNSL and 35 glioma) and validation ($n = 31$, 15 PCNSL and 16 glioma) sets by a ratio of 7:3. Then, we established six different radiomic models from the established optimal feature subsets of the training dataset using logistic regression (LR), random forest (RF), decision tree (DT), k-nearest neighboring (KNN), support vector machine (SVM), and naïve Bayes. The hyper-parameters of the RF, DT, KNN, and SVM were automatically selected by search method.

Models were trained with the training set using the 5-fold cross-validation method, and estimation performance was evaluated within the validation set. The performance of different models was assessed using the area under the receiver operating characteristic curve (AUC), accuracy (ACC), sensitivity, specificity, a positive predictive value (PPV), and a negative value (NPV). All machine learning methods were performed using Python (version 3.5.6).

Radiomic-Based Deep Learning Models

In this study, a four-layer convolutional neural network model, namely, multilayer perceptron (MLP) model, was constructed with the selected seven radiomic features, such as original_shape_Sphericity, original_firstorder_Median, original_firstorder_RobustMeanAbsoluteDeviation, original_glcm_ClusterShade, original_glcm_Imc1, original_gldm_DependenceNonUniformityNormalized, and original_glrlm_RunEntropy as input. MLP is an artificial neural network that has also performed well in the previous studies (20, 21). The overall architecture of MLP is shown in **Figure 3**. Random search was used to select the best configuration of hidden layer number nodes. The number of nodes for two hidden layers was set to 16 and 8, respectively. The number of nodes for output layer was set to 2, the same as the number of patient's group. The learning rate was 0.0001. The optimizer was stochastic gradient descent, and the loss function was the cross-entropy. Approximately 5-fold cross-validation was used in the training process. We derived this radiomics-based deep learning model using Pytorch 1.7.1 (<https://pytorch.org>), and we stopped training when the model converged. The model was trained with 150 epochs.





End-to-end CNN Model and Visualization

We trained an 2D end-to-end CNN model that integrated an automatic feature extraction and a discriminative classifier into one model. The raw CT scans were the input and the classification label was the direct output. There were 1,138 glioma and 1,218 PCNSL slices in the training set. Transfer learning was used, and a pretrained Densenet-121 (22) model on ImageNet (<http://www.image-net.org/>) was fine-tuned to perform classification on our dataset. We converted our 2D slices into color images and resized them to the Densenet input size (224×224). Slices for each patient were classified by our trained model, and the final decision was performed by majority voting. The loss function was cross-entropy, and the optimizer was Adam. The learning rate was set to 0.0001 without weight decay. The batch size was 8. We applied 5-fold cross-validation and trained 500 epochs.

For the growing importance of interpretability in deep learning, we applied Grad-Cam++ (23) technique to draw coarse localization heatmaps, which highlighted the important regions for CNN model to make diagnosis decisions. The model paid more attention to the deeper red region.

Physician's Diagnosis

In addition, all images were read by four independent radiologists, and three were primary radiologists (with 3, 4, and 4 years of experience in neuroradiology). One was a senior neuroradiologist (17 years of experience). The four independent radiologists were blinded to the initial diagnosis report, the pathological and clinical information, but were aware of that the segmented tumors were either PCNSL or glioma. The consistency of diagnosis from three readers was evaluated using intra- and interclass correlation coefficients (ICCs). An ICC of <0.4 indicated poor correlation; an ICC of 0.4 to 0.75 indicated fair to good correlation; and an ICC of more than 0.75 indicated excellent correlation (24).

Statistical Analysis

Area under the receiver operating characteristic curve (AUC) and accuracy (ACC) were used to evaluate the performance of the models. All statistical analyses were performed using Python (version 3.5.6). Categorical variables were presented as absolute numbers and counts with percentages. Summary statistics are presented as means \pm standard deviation (SD) for normally distributed continuous variables, or as median [interquartile

TABLE 1 | Demographic and clinical characteristics of all patients.

Characteristic	PCNSL	glioma	P value
Gender			0.913
Male	27 (54.0%)	28 (54.9%)	
Female	23 (46.0%)	23 (45.1%)	
Age, mean \pm SD (years)	61.1 \pm 12.1	56.4 \pm 13.0	0.017
Tumor Location			0.066
Telencephalon	40	49	
Thalamus	5	2	
Brainstem	2	0	
Cerebellum	3	0	
History of malignancy			0.624
No	49 (98.0%)	48 (94.2%)	
Yes	1 (2.0%)	3 (5.8%)	

range (IQR)] for non-normally distributed continuous variables. Mann–Whitney *U* test was performed to compare continuous variables, whereas chi-squared test was used for categorical variables between groups. DeLong tests was performed to compare the AUC of different models. A two-tailed $p < 0.05$ indicated statistical significance. ICC was used to evaluate the consistency between different readers.

The training and validation of the MLP and CNN models were implemented using Pytorch 1.7.1 (<https://pytorch.org>) with Tesla V100 GPU support.

RESULTS

Study Population

In this study, the PCNSL cohort comprised of 50 patients (27 men and 23 women) with a mean age of 61.1 ± 12.1 years whereas the glioma cohort comprised of 51 patients (28 men and 23 women) with a mean age of 56.4 ± 13.0 years. The detailed patient demographics of the dataset are listed in **Table 1**. There were no significant differences between the PCNSL and glioma sets in terms of gender ($p = 0.913$), history of malignancy ($p = 0.624$), and tumor location ($p = 0.066$).

Performance of Different Models

The training and validation performances of six machine learning models, MLP model, end-to-end CNN model, and four radiologists are presented in **Table 2**. **Figure 4** demonstrated the area under the receiver operating characteristic curves (AUCs) of all models. Among all the machine learning models, RF had the best performance (AUC = 0.998, ACC = 0.957), followed by SVM (AUC = 0.930, ACC = 0.829), DT (AUC = 0.923, ACC = 0.900), LR (AUC = 0.885, ACC = 0.814), KNN (AUC = 0.852, ACC = 0.771), and naive Bayes (AUC = 0.796, ACC = 0.714) in the training set. In the validation set, the performance of SVM (AUC = 0.829, ACC = 0.742) was the best among the six RMs, followed by LR (AUC = 0.821, ACC = 0.774), KNN (AUC = 0.819, ACC = 0.774), RF (AUC = 0.740, ACC = 0.710), Bayes (AUC = 0.644, ACC = 0.581), and DT (AUC = 0.605, ACC = 0.608). There was obvious overfitting in DT and RF models.

In comparing diagnostic performances, the ACCs of the three primary radiologists were 0.742, 0.710, and 0.710 for readers 1, 2 and 3, respectively. The ICC of the three radiologists was 0.651 ($p < 0.001$). The senior radiologist achieved an accuracy of 0.839, which was higher than the three primary radiologists.

The MLP performed best than other models, with an accuracy of 0.886 and 0.903, sensitivity of 0.914 and 0.867, specificity of 0.857 and 0.937, and AUC of 0.957 and 0.908 in the training and validation cohorts, respectively. There was no significant difference compared to the senior radiologist in accuracy (0.903 vs. 0.839, $p = 0.988$ using chi-squared test). The end-to-end CNN model achieved an accuracy of 0.957 and 0.839, sensitivity of 0.971 and 0.867, specificity of 0.943 and 0.813, and AUC of 0.957 and 0.840 in the training and validation cohorts, respectively, which had no significant difference in validation accuracy compared to the MLP model ($p = 0.472$) and the senior radiologist ($p = 0.470$).

We added **Figure 5** which includes six cases Grad-Cam++ heatmaps that were obtained from the last convolutional layer in the network. The heatmaps might provide the most important regions that influenced the diagnosis decision. The deep red regions overlapped with the tumor area.

DISCUSSION

Differentiation between PCNSL and glioma is an important but challenging task. The question of “PCNSL” or “glioma” may become “resection or nonsurgical treatment” bypassing the steps of biopsy, histological evaluation, and postsurgical patient recovery. This would undoubtedly help expert neurosurgeons to optimize both patient outcomes and the cost-effectiveness of neurosurgical care (25). In this review, we explored various machine learning and deep learning methods based on radiomic features extracted from CT scans to predict PCNSL and glioma types and compare the performance of different models on 101 patients from five Chinese medical centers. Among all the models, the MLP model performed best, with an accuracy of 0.886 and 0.903, sensitivity of 0.914 and 0.867, specificity of 0.857 and 0.937, and AUC of 0.957 and 0.908 in the training and validation cohorts, respectively, and was significantly higher than the three primary physician’s diagnoses (ACCs ranged from 0.710 to 0.742, $p < 0.001$ for all) and comparable with the senior radiologist (ACC 0.839, $p = 0.988$). The end-to-end CNN model achieved an AUC of 0.839 and an ACC of 0.840 in the validation cohort, which had no significant difference in accuracy compared to the MLP model ($p = 0.472$) and the senior radiologist ($p = 0.470$). The CNN model used raw CT scans as input. Feature extraction and selection were automated and implemented. The coarse localization heatmaps help to give an intuitive understanding of the mechanisms of the model.

Patients with PCNSL and glioma shared many similar clinical symptoms and overlapping imaging characteristics. In our study, the three primary physician’s diagnoses results ranged from 0.710 to 0.742 and the experienced radiologist achieved 0.839. The fine-grained features of tumor regions were difficult to identify with the naked eyes. The low accuracy indicated the difficulties

TABLE 2 | Performance of different models in training set and validation set.

Method	Training						Validation					
	AUC (95%CI)	ACC (p)	Sensitivity	Specificity	PPV	NPV	AUC (95%CI)	ACC (p)	Sensitivity	Specificity	PPV	NPV
LR	0.885 (0.816–0.942)	0.814 (0.810)	0.800	0.829	0.824	0.806	0.821 (0.711–0.939)	0.774 (0.470)	0.800	0.750	0.750	0.800
RF	0.998 (0.995,1.0)	0.957 (0.810)	0.943	0.971	0.971	0.944	0.740 (0.577,0.880)	0.710 (0.149)	0.800	0.625	0.667	0.769
SVM	0.930 (0.877–0.974)	0.829 (0.231)	0.857	0.800	0.811	0.848	0.829 (0.709–0.946)	0.742 (0.071)	0.867	0.625	0.684	0.833
DT	0.923 (0.874–0.967)	0.900 (0.151)	0.971	0.829	0.850	0.967	0.605 (0.492–0.727)	0.581 (0.470)	0.600	0.562	0.562	0.600
Naive Bayes	0.796 (0.705–0.870)	0.714 (0.632)	0.771	0.657	0.692	0.742	0.644 (0.475–0.803)	0.516 (0.988)	0.467	0.562	0.500	0.529
KNN	0.852 (0.777–0.914)	0.771 (0.031)	0.629	0.914	0.880	0.711	0.819 (0.699–0.932)	0.774 (0.988)	0.733	0.812	0.786	0.765
Radiologist1	-	0.743 (0.810)	0.714	0.771	0.758	0.730	-	0.742 (0.718)	0.733	0.750	0.733	0.750
Radiologist2	-	0.714 (0.810)	0.686	0.743	0.727	0.703	-	0.710 (0.470)	0.733	0.688	0.688	0.733
Radiologist3	-	0.729 (0.339)	0.771	0.686	0.711	0.750	-	0.710 (0.470)	0.600	0.813	0.750	0.684
Radiologist4	-	0.843 (1.000)	0.829	0.857	0.853	0.833	-	0.839 (1.000)	0.800	0.875	0.857	0.824
MLP	0.957 (0.923,0.980)	0.886 (0.473)	0.914	0.857	0.865	0.910	0.908 (0.885,0.941)	0.903 (0.988)	0.867	0.937	0.928	0.882
CNN	0.957 (0.928–0.979)	0.957 (0.632)	0.971	0.943	0.944	0.971	0.840 (0.797–0.900)	0.839 (0.470)	0.867	0.813	0.813	0.867

AUC, area under curve; ACC, accuracy; PPV, positive predictive value; NPV, negative predictive value.

Radiologist4 is a senior neuroradiologist with 17 years' experience while the other three are senior neuroradiologists with less than 5 years' experience.

P is significant difference in accuracy compared to the senior radiologist4 using chi-squared test.

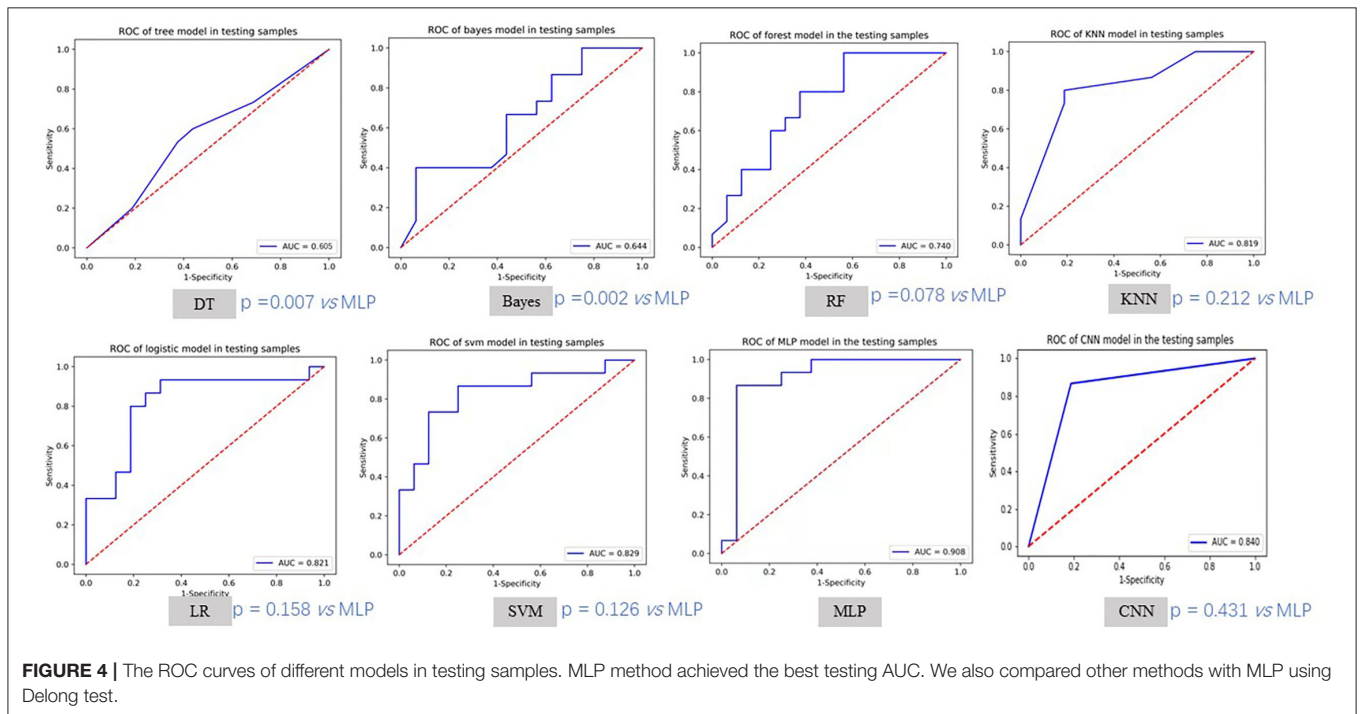
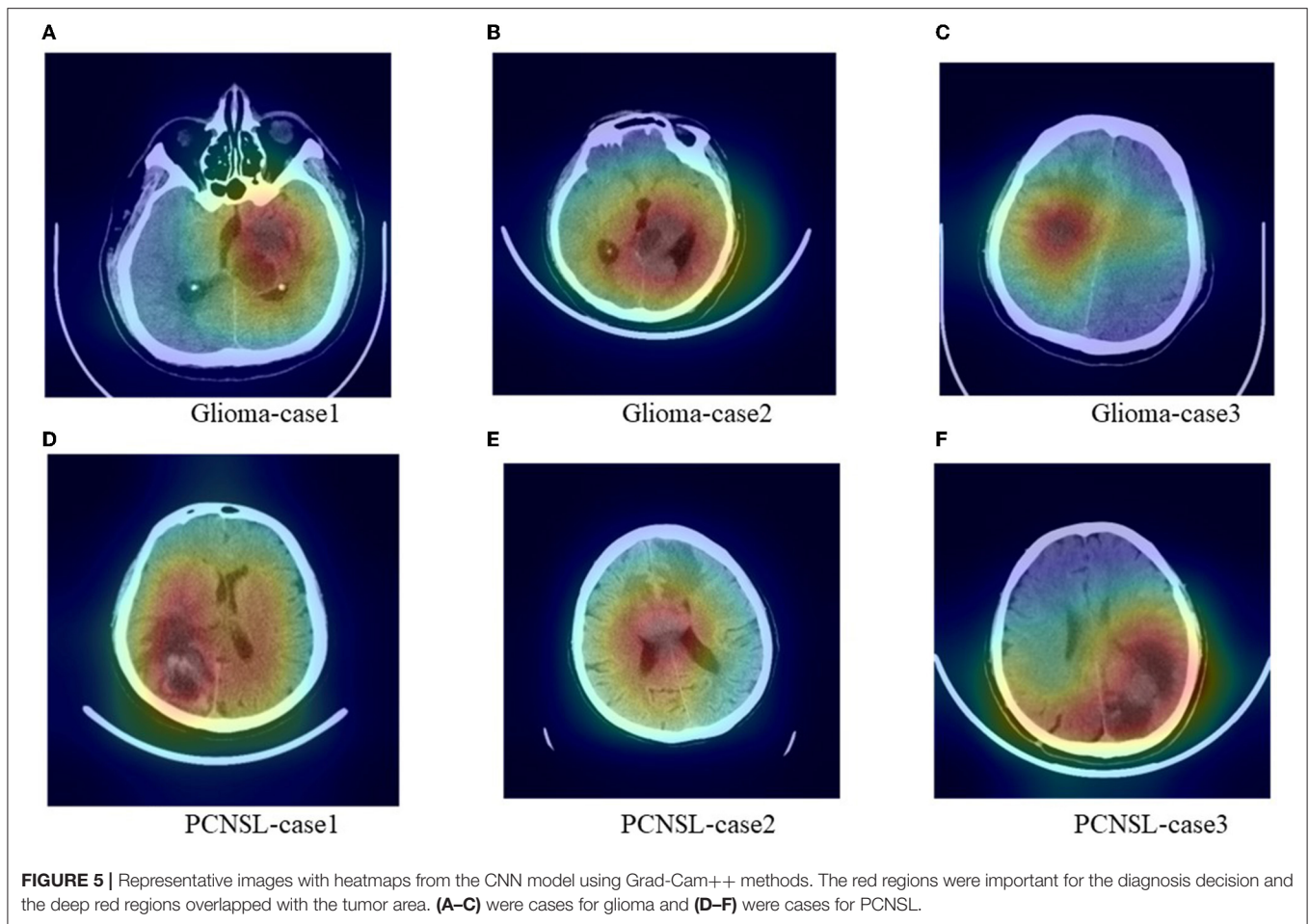


FIGURE 4 | The ROC curves of different models in testing samples. MLP method achieved the best testing AUC. We also compared other methods with MLP using Delong test.

in identifying the two brain tumors with the gross visual. The studies reported that even using high-resolution MR sequences, visual differentiation of the two tumors was still challenging

(26, 27). Radiomics enabled the conversion of original images into high-dimensional feature spaces that allowed an improved performance in PCNSL and glioma. Radiomic ML algorithms



performed as well as or better than radiologists in several studies (14, 28, 29). Bathla et al. (17) compared the predictive performance of various ML techniques to differentiate between PCNSL and glioma using a combination of various feature selection and ML algorithms on 94 patients, and several models achieved comparable performance.

The MLP model is a type of feed-forward artificial neural network model that has the input and output layers connected by a hidden layer. It helps to evaluate nonlinear relationships in classification tasks (30) and improve efficiency (31). Yun et al. (16) established an MRI radiomic MLP model to distinguish PCNSL from glioma and found high performance, which was even better than unsupervised convolutional neural networks. In our study, among all of our machine learning methods, the nonlinear classifier SVM ranked first, which indicated that a more complex model was needed for the input features. Therefore, we established an MLP classifier that had the capability of producing a higher level and more abstracted feature selection algorithm and it performed the best.

Convolutional neural network model is an automatic model for differentiating between PCNSL and glioma. It was built on raw CT slices and required no segmentation of tumor region. Training from scratch was usually data hungry, so transfer learning was applied. Natural images shared some underlying

features with medical images and the pretrained model helped to improve the convergence rate. The validation results showed that CNN model had no significant difference in accuracy compared to the senior radiologists. According to the heatmaps, the red important regions overlapped with the tumor regions, which showed that the diagnosis decision of CNN mode was based on the tumor rather than background regions. The effectiveness and reasonability have been proved.

Although the MRI is the recommended check for brain tumor diseases diagnose and it has a better resolution in brain tissue, cost-effective CT is routinely performed for patients in clinical practice. In our cases, there are final pathologically confirmed patients with PCNSL unable to accept prepuncture due to the tumor location and mistakenly underwent surgery. CT has its own advantages, such as short scan time and low price. Also, CT is performed for some patients who cannot have MR check, such as people with implantable medical devices, including pacemakers and cochlear implants. The ML or DL algorithms have achieved great success in differentiating between PCNSL and glioma on MR scans in previous studies and inspired us to verify distinguish on CT scans. To our knowledge, the effectiveness of CT-based ML or DL models for predicting the two tumors has been rarely explored before. Therefore, we collected data from multicenter and established different models

to explore the performance on the two tumors classification tasks. We followed the standard radiomic process: ROI segmentation, feature selection, and model establishment and validation. An effective feature selection is a crucial step because radiomic features are multiple collinear and correlated predictors that could produce unstable estimates and might overfit predictions. In our models, 7 features remained after feature selection, and using fewer features was a good strategy against overfitting. In this study, MLP model was significantly higher than the three primary physician's diagnoses and the six machine learning methods. CNN model made comparable prediction result with MLP method ($p = 0.431$ in AUC) and senior radiologist ($p = 0.470$ in ACC). CNN was also recommended since the whole process rarely required human intervention. The multicenter study validated the robustness and generalizability of our model, and it could provide a convenient and accurate tool for radiologists to identify PCNSL and glioma types on CT scans.

Besides the retrospective nature, our study also had several limitations. First, we used a relatively small number of patients. The small sample size could limit the effectiveness of supervised machine learning and deep learning methods and caused overfitting. It was quite possible that the model performance might vary with additional training data. We collected data from five Chinese medical centers and divided all data into training and validation sets by a ratio of 7:3. Unlike glioma, there was no free available PCNSL dataset which could be readily used. The absence of an external validation set could not help us to determine the generalizability of our findings. Nevertheless, our methods for CT radiomic feature analysis showed the potential to be reproducible with other datasets, although our classifiers might exhibit limited value. Second, the proposed models were used especially for the discrimination of PCNSL and glioma, whereas in clinical practice, the single brain metastasis could show similar appearance on CT which might cause diagnostic difficulties. Further study with collecting

more data to validate the generalizability of the developed models and a more advanced multiclassification model needs to be explored. Our classifiers could serve as a base model for the discrimination of PCNSL and glioma and might have the potential to be an effective aided diagnostic tool for clinical practice to some extent. If more kinds of brain tumor data could be collected, a baseline model could be conducted quickly.

In conclusion, we established machine learning and deep learning models from CT scans to help to differentiate PCNSL and glioma and verified that the models were feasible and provided high-performance -like senior physician's diagnoses, which showed the potential to assist clinical decision-making.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of privacy and ethical concerns. Requests to access the datasets should be directed to mvmouweiwei@126.com.

ETHICS STATEMENT

Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Guarantor of integrity of the entire study and literature research: WM. Definition of intellectual content: LM. Experimental studies: GL and WM. Data acquisition: GL and YZ. Data analysis and statistical analysis: WW and GL. Paper writing: All authors. All authors contributed to the article and approved the submitted version.

REFERENCES

- Ricard D, Ahmed I, François D, Marion L, Hoang-Xuan K, Jean-Yves D. Primary brain tumours in adults. *Lancet*. (2012) 379:1984–96. doi: 10.1016/S0140-6736(11)61346-9
- Ostrom QT, Haley G, Jordonna F, Max L, Rachel B, Courtney K, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008–2012. *Neuro Oncol*. (2015) 17:iv1–iv62. doi: 10.1093/neuonc/nov189
- Fine HA, Mayer RJ. Primary central nervous system lymphoma. *Ann Intern Med*. (1993) 119:1093–104. doi: 10.7326/0003-4819-119-11-199312010-00007
- Bondy ML, Scheurer ME, Malmer B, Barnholtz-Sloan JS, Davis FG, Il'Yasova D, et al. Brain tumor epidemiology: consensus from the Brain Tumor Epidemiology Consortium. *Cancer*. (2008) 113:1953–68. doi: 10.1002/cncr.23741
- Grommes C, Lisa MD. Primary CNS lymphoma. *J Clin Oncol*. (2017) 35:2410. doi: 10.1200/JCO.2017.72.7602
- Stupp R, Warren PM, Martin JV, Michael W, Barbara F, Martin JBT, et al. "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. (2005) 325:987–96. doi: 10.1056/NEJMoa043330
- Zusman EE, Deborah LB. The continuum of neurosurgical care: increasing the neurosurgeon's role and responsibility. *Neurosurgery*. (2017) 80:S34–41. doi: 10.1093/neuros/nyw151
- Furtner J, Schöpf V, Preusser M, Asenbaum U, Woitek R, Wöhrer A, 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. doi: 10.1016/j.ejrad.2014.01.017
- Choi YS, Lee HJ, Ahn SS, Chang JH, Kang SG, Kim EH, et al. Primary central nervous system lymphoma and atypical glioblastoma: differentiation using the initial area under the curve derived from dynamic contrast-enhanced MR and the apparent diffusion coefficient. *Eur Radiol*. (2017) 27:1344–51. doi: 10.1007/s00330-016-4484-2
- Doskalyev A, Yamasaki F, Ohtaki M, Kajiwara Y, Takeshima Y, Watanabe Y, et al. Lymphomas and glioblastomas: differences in the apparent diffusion coefficient evaluated with high b-value diffusion-weighted magnetic resonance imaging at 3 T. *Eur J Radiol*. (2012) 81:339–44. doi: 10.1016/j.ejrad.2010.11.005
- Schramm P, Xyda A, Klotz E, Tronnier V, Knauth M, Hartmann M. Dynamic CT perfusion imaging of intra-axial brain tumours: differentiation of high-grade gliomas from primary CNS lymphomas. *Eur Radiol*. (2010) 20:2482–90. doi: 10.1007/s00330-010-1817-4

12. Kickingereder P, Burth S, Wick A, Götz M, Eidel O, Schlemmer HP, et al. Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology*. (2016) 280:880–97. doi: 10.1148/radiol.2016160845
13. Prasanna P, Patel J, Partovi S, Madabhushi A, Tiwari P. Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. *Eur Radiol*. (2017) 27:4188–97. doi: 10.1007/s00330-016-4637-3
14. Suh HB, Choi YS, Bae S, Ahn SS, Chang JH, Kang SG, et al. Primary central nervous system lymphoma and atypical glioblastoma: differentiation using radiomics approach. *Eur Radiol*. (2018) 28:3832–9. doi: 10.1007/s00330-018-0568-4
15. 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*. (2018) 17:50. doi: 10.2463/mrms.mp.2017-0044
16. Yun J, Park JE, Lee H, Ham S, Kim N, Kim HS. Radiomic features and multilayer perceptron network classifier: a robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. *Sci Rep*. (2019) 9:1–10. doi: 10.1038/s41598-019-42276-w
17. Bathla G, Priya S, Liu Y, Ward C, Le NH, Soni N, et al. Radiomics-based differentiation between glioblastoma and primary central nervous system lymphoma: a comparison of diagnostic performance across different MRI sequences and machine learning techniques. *Eur Radiol*. (2021) 31:8703–13. doi: 10.1007/s00330-021-07845-6
18. Xia W, Hu B, Li H, Shi W, Tang Y, Yu Y, et al. Deep learning for automatic differential diagnosis of primary central nervous system lymphoma and glioblastoma: multi-parametric magnetic resonance imaging based convolutional neural network model. *J Magnet Res Imag*. (2021) 54:880–7. doi: 10.1002/jmri.27592
19. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
20. Bou Assi E, Gagliano L, Rihana S, Nguyen DK, Sawan M. Bispectrum features and multilayer perceptron classifier to enhance seizure prediction. *Sci Rep*. (2018) 8:1–8. doi: 10.1038/s41598-018-33969-9
21. Kwon K, Kim D, Park H. A parallel MR imaging method using multilayer perceptron. *Med Phys*. (2017) 44:6209–24. doi: 10.1002/mp.12600
22. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2017). p. 4700–4708. doi: 10.1109/CVPR.2017.243
23. Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE winter conference on applications of computer vision (WACV). [I]IEEE[I/]. (2018) 839–47. doi: 10.1109/WACV.2018.00097
24. Rosner B. Fundamentals of biostatistics. Boston, MA, USA: Cengage learning. (2015).
25. Nguyen AV, Blears EE, Ross E, Lall RR, Ortega-Barnett J. Machine learning applications for the differentiation of primary central nervous system lymphoma from glioblastoma on imaging: a systematic review and meta-analysis. *Neurosurg Focus*. (2018) 45:E5. doi: 10.3171/2018.8.FOCUS18325
26. Koeller KK, Smirniotopoulos JG, Jones RV. Primary central nervous system lymphoma: radiologic-pathologic correlation. *Radiographics*. (1997) 17:1497–526. doi: 10.1148/radiographics.17.6.9397461
27. Al-Okaili RN, Krejza J, Woo JH, Wolf RL, O'Rourke DM, Judy KD, et al. Intraaxial brain masses: MR imaging-based diagnostic strategy—initial experience. *Radiology*. (2007) 243:539–50. doi: 10.1148/radiol.2432060493
28. Alcaide-Leon P, Dufort P, Geraldo AF, Alshafai L, Maralani PJ, Spears J, et al. Differentiation of enhancing glioma and primary central nervous system lymphoma by texture-based machine learning. *Am J Neuroradiol*. (2017) 38:1145–50. doi: 10.3174/ajnr.A5173
29. Kang D, Park JE, Kim YH, Kim JH, Oh JY, Kim J, et al. Diffusion radiomics as a diagnostic model for atypical manifestation of primary central nervous system lymphoma: development and multicenter external validation. *Neuro Oncol*. (2018) 20:1251–61. doi: 10.1093/neuonc/noy021
30. Hamerla G, Meyer HJ, Schob S, Ginat DT, Altman A, Lim T, et al. Comparison of machine learning classifiers for differentiation of grade 1 from higher gradings in meningioma: A multicenter radiomics study. *Magn Reson Imaging*. (2019) 63:244–9. doi: 10.1016/j.mri.2019.08.011
31. Bibault JE, Giraud P, Housset M, Durdux C, Taieb J, Berger A, et al. Deep learning and radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Sci Rep*. (2018) 8:1–8. doi: 10.1038/s41598-018-30657-6

Conflict of Interest: Author WW was employed by GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lu, Zhang, Wang, Miao and Mou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.