ARTICLE

Deep Learning-Assisted Quantitative Susceptibility Mapping as a Tool for Grading and Molecular Subtyping of Gliomas

Wenting Rui[1] · Shengjie Zhang[2,3] · Huidong Shi[1] · Yaru Sheng[1] · Fengping Zhu[4] · YiDi Yao[1] ·
Xiang Chen[2,3] · Haixia Cheng[5] · Yong Zhang[6] · Ababikere Aili[7] · Zhenwei Yao[1] · Xiao-Yong Zhang[2,3] ·
Yan Ren[1]

Abstract

This study aimed to explore the value of deep learning (DL)-assisted quantitative susceptibility mapping (QSM) in glioma grading and molecular subtyping. Forty-two patients with gliomas, who underwent preoperative T2 fluid-attenuated inversion recovery (T2 FLAIR), contrast-enhanced T1-weighted imaging (T1WI+C), and QSM scanning at 3.0T magnetic resonance imaging (MRI) were included in this study. Histopathology and immunohistochemistry staining were used to determine glioma grades, and *isocitrate dehydrogenase* (*IDH*) *1* and *alpha thalassemia/mental retardation syndrome X-linked gene* (*ATRX*) subtypes. Tumor segmentation was performed manually using Insight Toolkit-SNAP program (www.itksnap.org). An inception convolutional neural network (CNN) with a subsequent linear layer was employed as the training encoder to capture multi-scale features from MRI slices. Fivefold cross-validation was utilized as the training strategy (seven samples for each fold), and the ratio of sample size of the training, validation, and test dataset was 4:1:1. The performance was evaluated by the accuracy and area under the curve (AUC). With the inception CNN, single modal of QSM showed better performance in differentiating glioblastomas (GBM) and other grade gliomas (OGG, grade II–III), and predicting *IDH1* mutation and *ATRX* loss (accuracy: 0.80, 0.77, 0.60) than either T2 FLAIR (0.69, 0.57, 0.54) or T1WI+C (0.74, 0.57, 0.46). When combining three modalities, compared with any single modality, the best AUC/accuracy/F1-scores were reached in grading gliomas (OGG and GBM: 0.91/0.89/0.87, low-grade and high-grade gliomas: 0.83/0.86/0.81), predicting *IDH1* mutation (0.88/0.89/0.85), and predicting *ATRX* loss (0.78/0.71/0.67). As a supplement to conventional MRI, DL-assisted QSM is a promising molecular imaging method to evaluate glioma grades, *IDH1* mutation, and *ATRX* loss.

Keywords Quantitative susceptibility mapping · Glioma classification · *Isocitrate dehydrogenase* · *Alpha thalassemia/mental retardation syndrome X-linked gene* · Deep learning

Abbreviations

WHO	World Health Organization
HGG	High-grade gliomas
LGG	Low-grade gliomas
IDH	*Isocitrate dehydrogenase*
ATRX	*Alpha thalassemia/mental retardation syndrome X-linked gene*
MRI	Magnetic resonance imaging
T2 FLAIR	T2 fluid-attenuated inversion recovery
T1WI+C	Contrast-enhanced T1 weighted images
QSM	Quantitative susceptibility mapping
GRE	Gradient-echo
GBM	Glioblastomas
ML	Machine learning
CNN	Convolutional neural networks
DL	Deep learning
ROI	Region of interest
OGG	Other grade glioma
T2WI	T2-weighted image
TR	Repetition time
TE	Echo time
FOV	Field of view
PPV	Positive predictive value

Wenting Rui, Shengjie Zhang and Huidong Shi contributed equally to this work as co-first authors.

✉ Xiao-Yong Zhang
xiaoyong_zhang@fudan.edu.cn

✉ Yan Ren
yan_ren@139.com

Extended author information available on the last page of the article

NPV   Negative predictive value
ROC   Receiver operating characteristic curve
AUC   Area under the curve
t-SNE   T-distributed stochastic neighbor embedding
SVM   Support vector machine
GLM   Generalized linear model

Introduction

Gliomas are the most common primary intracranial tumor, accounting for 81% of all malignant brain tumors (Ostrom et al. 2014). According to the World Health Organization (WHO) criteria, gliomas are assigned grades I–IV indicating different degrees of malignancy (Weller et al. 2015). Patients with gliomas of different grades are treated with significantly different surgical plans, radiotherapy, and adjuvant chemotherapy strategies (Weller et al. 2014, 2017). Low-grade gliomas (LGG, grade I–II) are typically associated with a longer life expectancy than high-grade gliomas (HGG, grade III–IV) (Weller et al. 2015). Grade I gliomas are characterized by slow growth and high possibility of cure by surgical resection alone (Weller et al. 2015). Glioblastomas (GBM, grade IV) are the most aggressive tumor type, with the median survival time of about 12–15 months (Louis et al. 2016). In contrast to GBMs, the other grade gliomas (OGG, grade II–III) show more favorable outcomes and share similar histopathologic and genomic characteristics (Louis et al. 2016). According to the WHO tumor classification, the mutation status of *isocitrate dehydrogenase* (*IDH*) *1* codon 132 or *IDH2* codon 172 plays a major role in diagnosing and treating gliomas (Louis et al. 2016; Weller et al. 2017). The presence of *IDH* mutation (*IDH*(+)) distinguishes glioma entities with distinct biology. In the case of gliomas, patients carrying *IDH*(+) usually have a significant favorable response to treatment and outcomes compared to the patients with its wild type (*IDH*(−)) (Reifenberger et al. 2017; Weller et al. 2015). Loss of *nuclear alpha thalassemia/mental retardation syndrome X-linked gene* (*ATRX*) expression has never been found to be accompanied by *1p/19q* codeletion (Ikemura et al. 2016) and is characteristic in diagnosing diffuse astrocytomas rather than oligodendrogliomas (Louis et al. 2016). *ATRX* loss (*ATRX*(−)) is also associated with poor outcomes in LGG patients (Ogishima et al. 2017). Therefore, it is critical to distinguish LGG from HGG, GBM from OGG, and to identify *IDH* and *ATRX* subtypes. Currently, histopathology procedures, immunohistochemistry, or sequencing following biopsy or surgical resection are the main methods used for glioma grading and molecular subtyping; however, all of these methods are invasive (Ferris et al. 2017).

Preoperative noninvasive diagnosis of gliomas is made mainly based on conventional magnetic resonance imaging (MRI) such as T2 fluid-attenuated inversion recovery (T2 FLAIR) and contrast-enhanced T1-weighted images (T1WI + C), but with limited value in grading and genetic classification (Ly et al. 2020). Pathologically, HGG exhibit a higher rate of vascular proliferation, microhemorrhages, and small vessels than LGG (Ferris et al. 2017). Intratumoral calcification tends to be found in *1p/19* co-deleted oligodendrogliomas, which are associated with improved prognosis and responsiveness to therapy (Saito et al. 2016). Different forms of iron in blood products and calcification inside gliomas result in susceptibility variations, all of which can be detected by quantitative susceptibility mapping (QSM) (Wang et al. 2017a, b).

QSM has become a sensitive and reliable quantitative technique to determine the bulk magnetic susceptibility distribution caused by iron load of tissues (Langkammer et al. 2012). By identifying the magnetic field produced by tissue susceptibility and solving the field-to-magnetization (tissue susceptibility) inverse problem (Kee et al. 2017), QSM deconvolves the blooming artifacts in gradient-echo (GRE) phase data and shows much better contrast than R2* or T2 methods (Haacke et al. 2015). So far, QSM has been widely used in the quantitative study of brain iron content (Li et al. 2019). Previous studies have shown the application of QSM in distinguishing HGG with hemorrhage and less aggressive brain tumors with or without calcification (Bandt et al. 2019), and differentiation between blood deposits and calcification in GBMs (Deistung et al. 2013).

Recently, machine learning (ML) has been developed to capture complex patterns in imaging data that are beyond human perception and provide quantitative evaluation of radiographic features for data-driven prediction tasks (Hosny et al. 2018; Lotan et al. 2019). A large amount of ML-assisted research has been applied to determine histological glioma grade, molecular profiles, and prognosis (Lotan et al. 2019). With the huge success of convolutional neural networks (CNN) (especially ResNet and its variants etc.) in medical image classification tasks (Cheng et al. 2022), deep learning (DL) can build a pipeline for feature extraction and image classification, while reducing the subjective bias of manual feature extraction. Especially, inception module and pyramid module can benefit the multi-scale feature extraction, which is significant in medical image classification because the region of interest (ROI) may be small and discrete. However, a dataset with a small number of samples may suffer from an overfitting problem. Inspired by the semi-supervised learning strategy for training a consistency loss, we proposed an inception CNN encoder with consistency loss computed by output of adjacent slices for glioma diagnosis. The purpose of this study was to explore

Fig. 1 Flowchart of the study population. *IDH1*(+): *IDH1* mutation; *IDH1*(−): *IDH1* wildtype; *ATRX*(−): *ATRX* expression loss; *ATRX*(+): *ATRX* retention
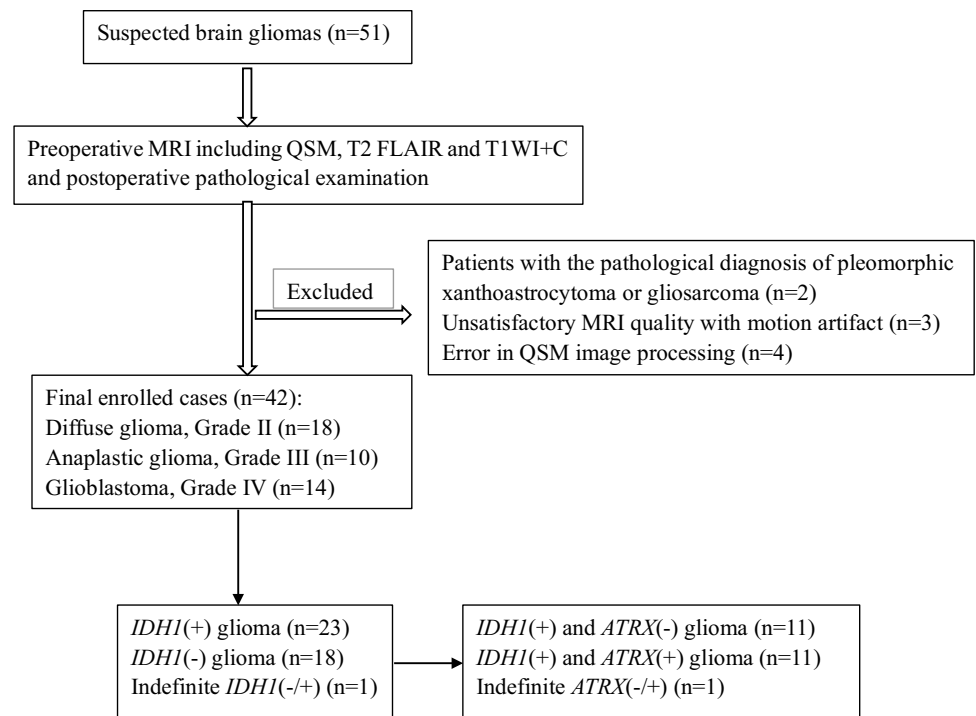
the value of DL-assisted QSM in the prediction of glioma grades, *IDH1*(+), and *ATRX*(−).

Materials and Methods

Study Population

This study was approved by the Ethics Review Board of our institution, and individual consent for this retrospective analysis was waived. A total of 51 patients with clinically suspected gliomas underwent a unified preoperative MR examination protocol. These patients met three main criteria: (1) clinically newly diagnosed primary brain gliomas without any pharmacotherapy or radiotherapy; (2) available histopathological diagnosis and molecular genetic characteristics including *IDH* and *ATRX*; (3) available preoperative MR images, including QSM, T2 FLAIR, and T1WI+C sequences. Nine patients were excluded for the following reasons: (1) uncommon pathological diagnosis of gliosarcoma ($n = 1$) and pleomorphic xanthoastrocytoma ($n = 1$); (2) image artifacts owing to patient movement ($n = 3$); and (3) error in QSM image processing ($n = 4$). Finally, 42 patients (18 female and 24 male, mean age: 47 years, age range: 26–75 years) were enrolled in this study (Fig. 1).

In the present study, three stratified detection tasks were designed: (1) the detection of glioma grades (LGG and HGG, OGG and GBM); (2) the detection of gliomas with

IDH1(+) or *IDH1*(−); (3) the detection of *IDH1* mutated glioma with *ATRX*(−) or *ATRX* retention (*ATRX*(+)).

Pathology Data Collection

The histopathologic grading and molecular subtyping data of gliomas were collected from an electronic database in the Neuropathology Department of our institution. Each resection specimen was sectioned and stained with hematoxylin and eosin. *IDH1* and *ATRX* molecular status were tested by immunohistochemical staining. According to the latest WHO classification, two experienced neuropathologists consistently diagnosed the glioma grades and performed molecular classification.

MRI Acquisition

MR images were acquired on a 3.0T MRI system (Discovery 750; GE Healthcare, Milwaukee, WI) with an eight-channel phased-array head coil (GE Medical Systems). The QSM was generated from a three-dimensional multi-echo GRE sequence. Each patient received the unified preoperative MR scan in the following order: axial T1WI, T2-weighted image (T2WI), T2 FLAIR, multi-echo GRE, and T1WI+C sequences. Specific parameters were as follows: T1WI with FLAIR technique: repetition time (TR)/echo time (TE) = 3195/24 ms, field of view (FOV) = 240 × 240 mm, matrix size = 256 × 256, slice thickness = 4 mm,

number of slices = 28; T2WI: TR/TE = 9185/108 ms, FOV = 240 × 240 mm, matrix size = 256 × 256, slice thickness = 4 mm, number of slices = 28; T2 FLAIR: TR/TE = 9491/140 ms, inversion time = 2200 ms, FOV = 240 × 240 mm, matrix size = 256 × 256, slice thickness = 4 mm, number of slices = 28; and multi-echo GRE: TR = 41.6 ms, number of echoes = 16, first TE = 3.2 ms, TE spacing = 2.4 ms, bandwidth = 62.50 kHz, flip angle = 12°, FOV = 256 × 256 mm, matrix size = 256 × 256, slice thickness = 1 mm, voxel size = 1 × 1 × 1 mm^3, number of slices = 140, acceleration factor = 2, acquisition time = 9 min. Array spatial sensitivity encoding technique was employed to accelerate the multi-echo GRE. Contrast-enhanced images were obtained immediately after administering a standard dose (0.1 mmol/kg body weight) of gadopentetate dimeglumine (Beilu, Beijing, China) at approximately 3–4 mL/s via the dorsal hand or elbow vein.

Image Reconstruction and Tumor Segmentation

QSM reconstructions were performed using susceptibility tensor imaging (STI) Suite software (Duke University) with reference to the previous studies (Li et al. 2011, 2015). First, the multi-echo phase images for each channel of the coil were collected and then averaged after subtracting the receiver phase of each channel. The phase was unwrapped with a Laplacian method. The unwrapped phase images were normalized by the corresponding echo times and averaged to determine the frequency shift. Second, the background phase was removed by the sophisticated harmonic artifact reduction for phase data method, and the filter radius was set as eight (Schweser et al. 2011). Third, the susceptibility map of the brain tissue was obtained from the frequency map by an improved least-squares method, and the regulatory threshold of Laplace filtering was set to 0.04 (Li et al. 2011, 2015).

Tumor segmentation was performed on axial T2 FLAIR, T1WI + C, and QSM images, respectively, using the Insight Toolkit-SNAP program (University of Pennsylvania, www.itksnap.org). With reference to T1WI, T2WI, T2 FLAIR, and T1WI + C, one neuroradiologist (WTR), who was blinded to the histopathologic and molecular information, delineated the tumor ROI guided by an experienced neuroradiologist (ZWY) with 20 years' experience in neuroradiology using the same criteria: at the image section with maximum diameter of solid tumor in each sequence, an arbitrarily shaped ROI was delineated around the area of tumor but avoiding peritumoral edema as much as possible (Rui et al. 2018).

Deep Learning

To avoid the potential data gap between training, validation and test dataset, intensity normalization across each individual modality was conducted. Then, MRI slices were center cropped, minimizing the impact of background. There was a binary mask over the specific slice, which could be a weak annotation to help guide the classification task. The binary mask was concatenated in the channel dimension for the guidance. As shown in Fig. 2, there are two paths in our proposed inception CNN—labeled path for annotated MRI slices and label-free path for unlabeled data. To automatically extract the efficient feature from the MRI slices, an inception module was introduced to capture potential biomarkers for glioma classification. For each path, three inception layers were employed to enhance the representation ability, followed by a global average pooling layer and a linear layer. For the multi-modality data, image slices were concatenated in the channel dimension (e.g., T1WI + C, T2 FLAIR, QSM for three modalities). To alleviate the influence of dataset split bias, fivefold cross-validation was leveraged and the ratio of training and validation and test dataset

Fig. 2 Pipeline of proposed inception convolutional neural network. For the ensemble of multi-modality, image slices of the same subject are concatenated in the channel dimension. Only one slice is annotated with a tumor mask, which is fed to the inception CNN for supervision and the adjacent slices are fed to the shared-weight encoder to achieve the similar output
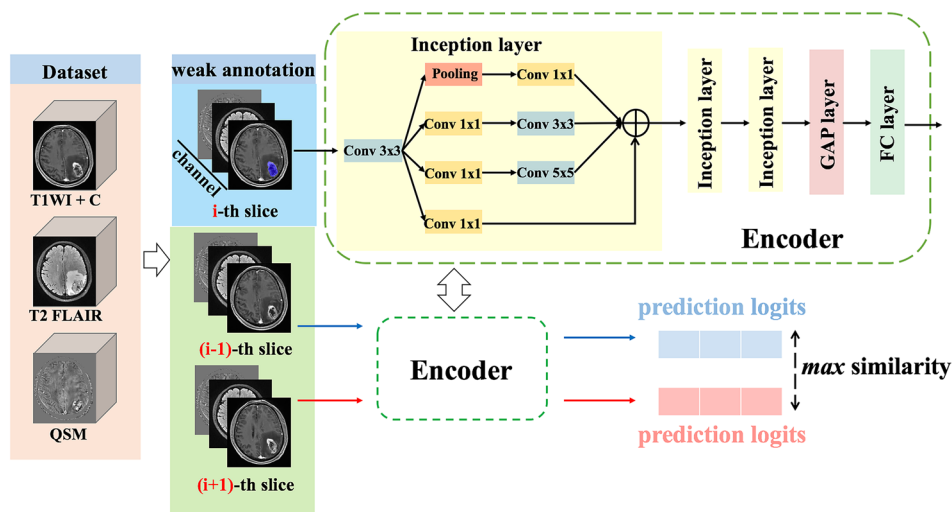
was 4:1:1. The classification algorithms aim to classify specific models via targeted metrics. For the grading task, we denoted the true labels with numbers 1, 2 and 3 corresponding to grade II, III and IV. For the prediction of *IDH1* mutation, the labels were divided into two classes: + and −. For the prediction of *ATRX* loss, − and + were the binary classification labels. For the proposed inception CNN and algorithms for comparison, each fold of dataset split was the same. The deep learning methods including inception CNN and standard CNN were trained through loss back propagation. For the corresponding details, refer to Supplementary Material 1.

The model performance (recorded as mean ± standard deviation) was evaluated by computing the total prediction accuracy, sensitivity/recall, specificity, positive predictive value (PPV)/precision, negative predictive value (NPV), and F1 score (defined as the harmonic mean of sensitivity and PPV) using the confusion matrix. A receiver operating characteristic (ROC) curve was also drawn and the area under the curve (AUC) was calculated to assess the discriminative ability of the model in the test dataset. To present the advantage of QSM in glioma classification, t-distributed stochastic neighbor embedding (t-SNE) and Shapley value analysis were employed to visualize the feature distribution of different modalities and the contribution to the model output of each modality. The standard pipeline of t-SNE was from manifold package in the scikit-learn, and the code of Shapley value analysis followed the official demo. To illustrate the effectiveness of proposed inception CNN, support vector machine (SVM), standard CNN (three-layer), and generalized linear model (GLM) were used for comparisons in the three-modality fusion classification tasks.
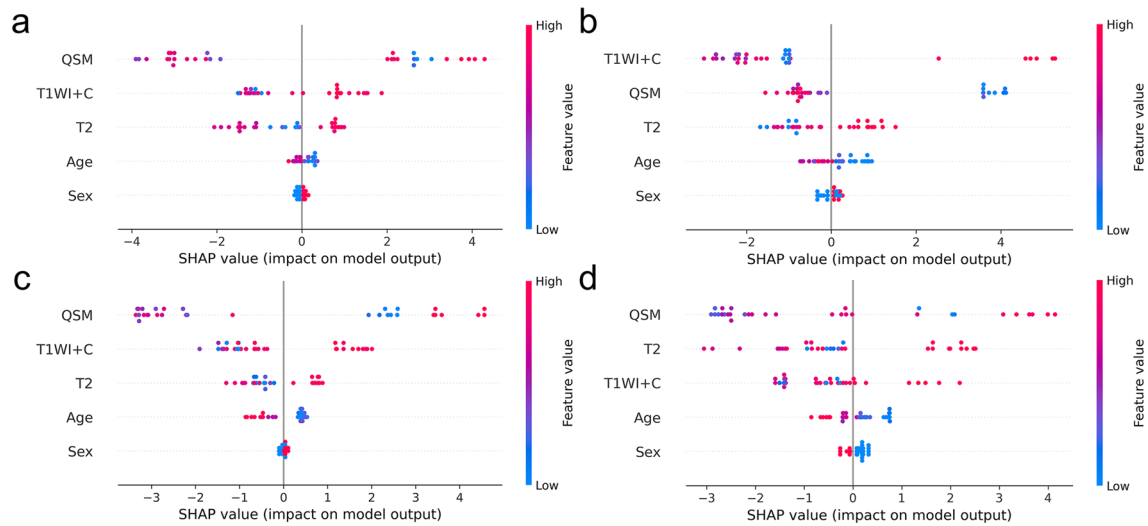
Results

Independent Efficiency of T2 FLAIR, T1WI + C, and QSM by DL in Glioma Tasks

The diagnostic performance of each MRI modality by DL for predicting glioma grades and the molecular subtype is listed in Table 1. For the grading task, QSM modality showed higher diagnostic accuracy of 0.80 and F1 score of 0.75 in differentiating OGG from GBM than T2 FLAIR (0.69, 0.60) and T1WI + C (0.74, 0.70); however, T1WI + C modality (0.74, 0.72) performed better in distinguishing LGG from HGG than QSM (0.69, 0.64) and T2 FLAIR (0.69, 0.61). For the *IDH1* task, QSM modality (accuracy: 0.77, F1 score: 0.70) was superior to T2 FLAIR (0.57, 0.53) and T1WI + C (0.57, 0.52) in predicting *IDH1*(+). For the *ATRX* task, QSM modality (accuracy: 0.60, F1 score: 0.52) showed a better performance than T2 FLAIR (0.54, 0.46) and T1WI + C (0.46, 0.42) in diagnosing *ATRX*(−). The figures of exemplary gliomas of different grades and phenotypes can be found in the Supplementary Figs. 1–3.

Importance of Different MRI Modalities and Clinical Features in Glioma Classification

Feature importance explanations showed that QSM modality feature was the most important variable for the classification of OGG/HGG, *IDH1*(+) or *IDH1*(−), *ATRX*(−), or *ATRX*(+) (Fig. 3). The spread of the Shapley values reflects the corresponding impacts on the model output of four classification tasks. t-SNE plot showed QSM modality feature extracted by inception network could partially identify

Table 1 Comparisons of the performance for prediction of glioma grades and molecular subtypes on different MR modalities by fivefold cross-validation of deep learning

Tasks	Modality	Accuracy	Sensitivity	Specificity	PPV	NPV	F1-score
OGG/GBM	T2 FLAIR	0.69 ± 0.11	0.75 ± 0.22	0.60 ± 0.25	0.75 ± 0.14	0.65 ± 0.18	0.60 ± 0.12
	T1WI + C	0.74 ± 0.11	0.75 ± 0.22	0.73 ± 0.25	0.83 ± 0.15	0.64 ± 0.08	0.70 ± 0.14
	QSM	**0.80 ± 0.07**	**0.86 ± 0.12**	**0.73 ± 0.25**	**0.84 ± 0.13**	**0.75 ± 0.13**	**0.75 ± 0.14**
LGG/HGG	T2 FLAIR	0.69 ± 0.11	**0.80 ± 0.19**	0.53 ± 0.16	0.70 ± 0.07	**0.67 ± 0.18**	0.61 ± 0.13
	T1WI + C	**0.74 ± 0.14**	0.76 ± 0.16	**0.73 ± 0.25**	**0.81 ± 0.17**	0.62 ± 0.10	**0.72 ± 0.17**
	QSM	0.69 ± 0.11	0.70 ± 0.19	0.67 ± 0.21	0.76 ± 0.14	0.59 ± 0.08	0.64 ± 0.12
IDH1(+)/(−)	T2 FLAIR	0.57 ± 0.20	0.65 ± 0.26	0.47 ± 0.16	0.60 ± 0.17	0.48 ± 0.17	0.53 ± 0.19
	T1WI + C	0.57 ± 0.20	0.60 ± 0.26	0.53 ± 0.27	0.63 ± 0.22	0.57 ± 0.28	0.52 ± 0.19
	QSM	**0.77 ± 0.11**	**0.86 ± 0.12**	**0.67 ± 0.30**	**0.81 ± 0.17**	**0.72 ± 0.16**	**0.70 ± 0.18**
ATRX(−)/(+)	T2 FLAIR	0.54 ± 0.17	0.60 ± 0.26	0.47 ± 0.27	0.62 ± 0.22	0.54 ± 0.26	0.46 ± 0.13
	T1WI + C	0.46 ± 0.11	0.50 ± 0.16	0.40 ± 0.13	0.52 ± 0.11	0.38 ± 0.10	0.42 ± 0.09
	QSM	**0.60 ± 0.11**	**0.65 ± 0.20**	**0.53 ± 0.27**	**0.69 ± 0.17**	**0.59 ± 0.22**	**0.52 ± 0.09**

OGG other grade glioma (grade II + III), *GBM* glioblastoma multiforme (grade IV), *LGG* low-grade glioma (grade II), *HGG* high-grade glioma (grade III + IV), *IDH1(+)* IDH1 mutation, *IDH1(−)* IDH1 wild, *ATRX(−)* ATRX expression loss, *ATRX(+)* ATRX retention, *PPV* positive predictive value, *NPV* negative predictive value

The best results are marked in bold

Fig. 3 Shapley values for analyzing importance scores of image features extracted by inception network, when the extracted features are combined with the non-imaging features (sex, age) for glioma classification (**a** for OGG/GBM, **b** for LGG/HGG, **c** for *IDH1* (±), and **d** for *ATRX* (±))
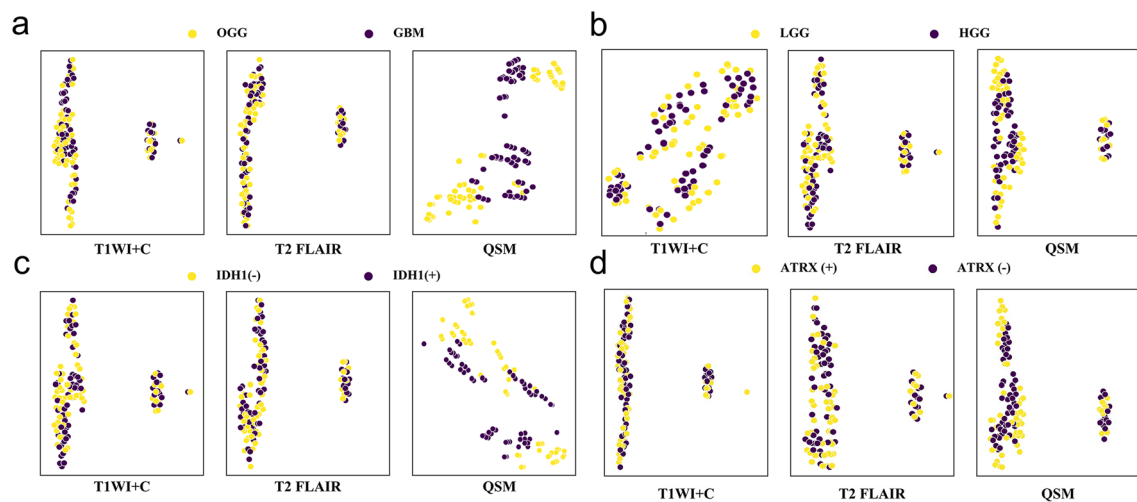
Fig. 4 t-distributed stochastic neighbor embedding (t-SNE) visualization of modality (T1WI, T2 FLAIR, and QSM) feature distribution of four glioma classification tasks (**a** for OGG/GBM, **b** for LGG/HGG, **c** for *IDH1* (±) and **d** for *ATRX* (±))

GBM patients from OGG patients and distinguish *IDH1*(+) or *IDH1*(−), while T1WI + C and T2 FLAIR modality features could not (Fig. 4).

Added Value of QSM in Contrast to Conventional MRI

Performance metrics of multi-modal MRI in glioma tasks by DL are presented in Table 2. For the OGG and GBM classification, T1WI + C plus T2 FLAIR showed a diagnostic accuracy of 0.80 and F1 score of 0.78, and improved diagnostic efficiency could be achieved by adding the QSM modality (accuracy: 0.89, F1-score: 0.87, sensitivity or recall: 0.90, specificity: 0.87, PPV or precision: 0.91, NPV: 0.82). Similarly, better diagnostic performance (accuracy: 0.86, F1-score: 0.81) was acquired with QSM modality added than only routine T2 FLAIR plus T1WI + C (0.77, 0.76) in discriminating LGG from HGG. For the *IDH1* task, diagnostic efficacy of T1WI + C plus QSM (accuracy: 0.80, F1 score: 0.74) or T2 FLAIR plus QSM (0.80, 0.75) was superior to T1WI + C plus T2 FLAIR modalities (0.69, 0.60). By combining the three modalities, satisfactory diagnostic efficiency was obtained (accuracy: 0.89, F1-score: 0.85, sensitivity: 0.81, specificity: 1.00, PPV: 1.00, NPV:

Table 2 Performance metrics of the deep learning model on multi-modal MRI for glioma grading and molecular subtyping

Tasks	Modality	Accuracy	Sensitivity	Specificity	PPV	NPV	F1-score
OGG/GBM	A+B	0.80±0.07	0.76±0.16	0.87±0.16	0.91±0.11	0.67±0.05	0.78±0.09
	A+C	0.83±0.11	0.90±0.12	0.73±0.25	0.82±0.15	0.82±0.15	0.77±0.17
	B+C	0.77±0.07	0.80±0.19	0.73±0.25	0.84±0.13	0.74±0.14	0.71±0.12
	A+B+C	**0.89±0.11**	**0.90±0.12**	**0.87±0.16**	**0.91±0.11**	**0.82±0.15**	**0.87±0.11**
LGG/HGG	A+B	0.77±0.11	0.71±0.19	**0.87±0.16**	0.89±0.14	0.64±0.08	0.76±0.12
	A+C	0.74±0.14	0.76±0.16	0.73±0.25	0.81±0.17	0.62±0.10	0.72±0.17
	B+C	0.83±0.11	0.86±0.20	0.80±0.27	0.89±0.14	0.79±0.18	0.77±0.17
	A+B+C	**0.86±0.09**	**0.91±0.11**	0.80±0.27	**0.89±0.14**	**0.82±0.15**	**0.81±0.17**
IDH1(+)/(−)	A+B	0.69±0.11	0.75±0.22	0.60±0.25	0.75±0.14	0.65±0.18	0.60±0.12
	A+C	0.80±0.07	**0.86±0.12**	0.73±0.25	0.84±0.14	0.75±0.13	0.75±0.14
	B+C	0.80±0.07	0.81±0.19	0.80±0.27	0.84±0.13	0.74±0.14	0.74±0.14
	A+B+C	**0.89±0.11**	0.81±0.19	**1.00±0.00**	**1.00±0.00**	0.80±0.17	**0.85±0.14**
ATRX(−)/(+)	A+B	0.51±0.15	0.55±0.19	0.47±0.16	0.57±0.14	0.45±0.15	0.48±0.14
	A+C	0.66±0.15	0.65±0.20	0.67±0.30	0.72±0.16	**0.62±0.22**	0.60±0.16
	B+C	0.63±0.15	0.60±0.12	0.67±0.30	0.75±0.21	0.54±0.14	0.59±0.16
	A+B+C	**0.71±0.13**	**0.70±0.19**	**0.73±0.25**	**0.81±0.17**	0.60±0.10	**0.67±0.15**

A T2 FLAIR, *B* T1WI+C, *C* QSM, *OGG* other grade glioma (grade II+III), *GBM* glioblastoma multiforme (grade IV), *LGG* low-grade glioma (grade II), *HGG* high-grade glioma (grade III+IV), *IDH1*(+) *IDH1* mutation, *IDH1*(−) *IDH1* wildtype, *ATRX*(−) *ATRX* expression loss, *ATRX*(+) *ATRX* retention, *PPV* positive predictive value, *NPV* negative predictive value

The best results are marked in bold
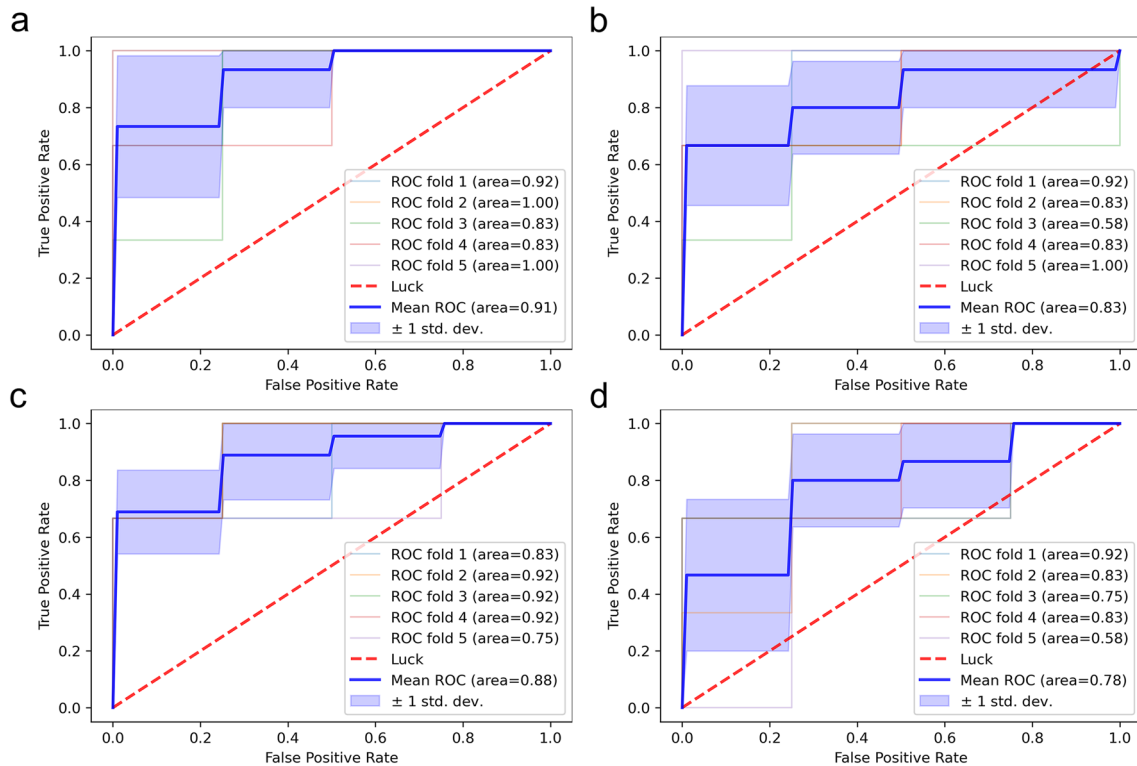
Fig. 5 Fivefold ROC curves of the proposed method combining three MRI modalities for four glioma classification tasks (**a** for OGG/GBM, **b** for LGG/HGG, **c** for *IDH1* (±), and **d** for *ATRX* (±)). The shadow range represents the standard deviation of fivefold

0.80) in determining *IDH1*(+). For the *ATRX* task, diagnostic performance of T1WI + C plus T2 FLAIR modalities was unsatisfactory (accuracy: 0.51, F1 score: 0.48), and the efficacy was slightly better when using T1WI + C plus QSM (0.63, 0.59) or T2 FLAIR plus QSM modalities (0.66, 0.60). When the three modalities participated together, moderate diagnostic ability (accuracy: 0.71, F1-score: 0.67, sensitivity: 0.70, specificity: 0.73, PPV: 0.81, NPV: 0.60) could be achieved in predicting *ATRX*(−). ROC curves in diagnosing gliomas based on three MRI modalities by DL are shown in Fig. 5. The mean AUC was 0.83 in differentiating LGG from HGG, 0.91 in discriminating OGG from GBM, and 0.88 and 0.78 in predicting *IDH1*(+) and *ATRX*(−), respectively. The comparisons of glioma classification results based on GLM, SVM, CNN, and our method are shown in Fig. 6. Performance comparisons of simpler histogram analysis and inception CNN model on QSM modality in glioma classification tasks are listed in Supplementary Table 1.

Discussion

QSM has shown a high degree of reproducibility on one scanner, and across different vendors, field strengths, and sites (Deh et al. 2015; Wang et al. 2017a, b), although its accuracy can be affected by spatial resolution, echo time, tissue orientation and other factors (Karsa et al. 2019; Lancione et al. 2017, 2019; Li et al. 2012). QSM has been widely used in brain diseases involving susceptibility variations (Bandt et al. 2019; Wang et al. 2017a, b). Radiomics or ML-based study using QSM has been applied for the diagnosis of Parkinson's disease and early Alzheimer's disease (Kim et al. 2020; Li et al. 2019). To the best of our knowledge, no ML-assisted QSM studies on glioma grading and molecular subtyping have been reported till now. Recently, DL-based methods have been widely used in medical image classification. Generally, consistency loss was utilized in the semi-supervised learning. In our research, although the binary masks over tumor regions could not directly provide the result of tumor classification, it can be used as an approximate guidance for classification. Thus, the inception module with consistency loss could enhance the ability of efficient feature extraction.

In this study, we extracted features of all ROIs from MR images automatically and assigned them to the final linear classifiers for glioma grading, *IDH1*(+), and *ATRX*(−) prediction. DL-assisted T1WI + C and T2 FLAIR exhibited the diagnostic accuracy of < 80% in glioma grading, which is similar to the published computer-aided performance and not ideal for clinical practice (Hsieh et al. 2017). DL-assisted
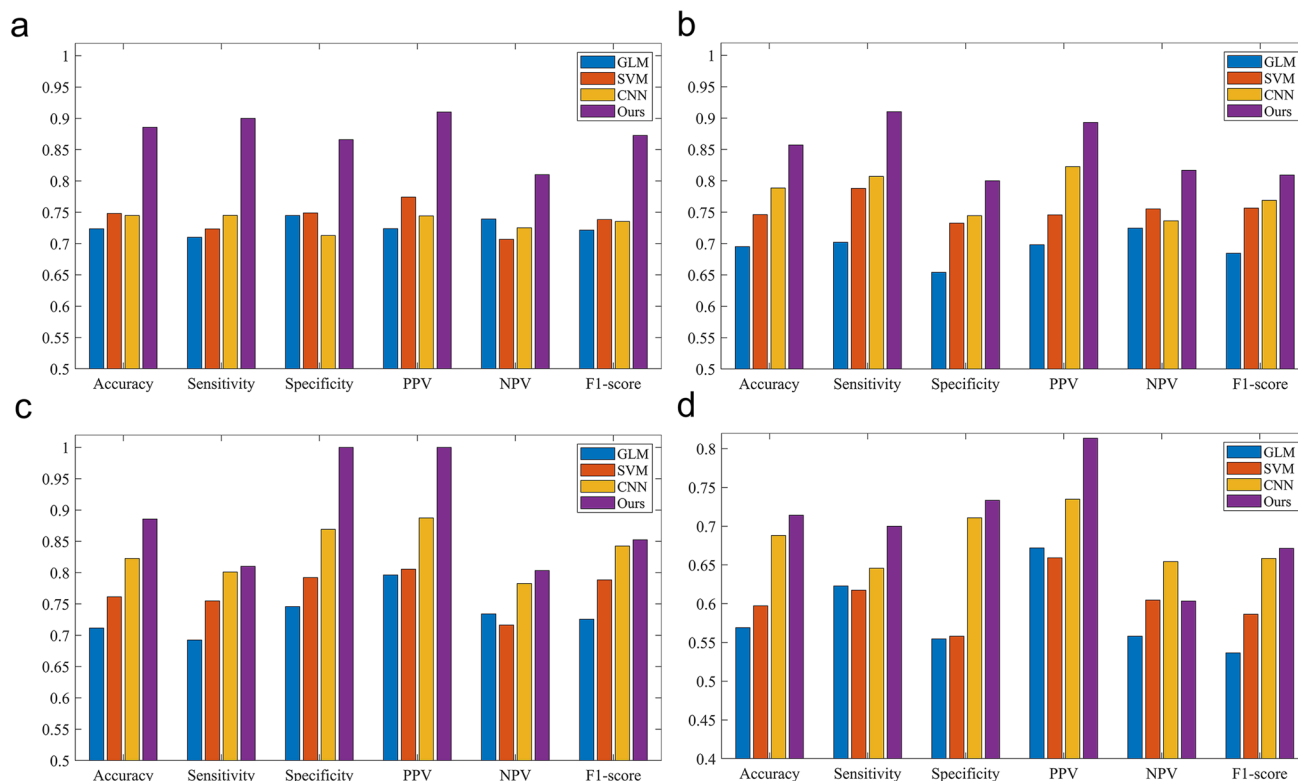
Fig. 6 Glioma classification results based on different algorithms (GLM, SVM, and CNN (three-layer convolution)) and our method using three modalities concatenated in the channel dimension (**a** for OGG/GBM, **b** for LGG/HGG, **c** for *IDH1* (±), and **d** for *ATRX* (±))

QSM was superior to conventional MRI modalities in distinguishing OGG from GBM, but not as good as T1WI + C in differentiating LGG from HGG. This is likely because GBM is characterized by microvascular proliferation and necrosis and is more prone to hemorrhage than OGG (Ferris et al. 2017), the induced susceptibility changes of which can be easily detected by QSM (Zhang et al. 2019). In addition, grade II and III oligodendrogliomas both feature extensive calcifications and a branching network of delicate capillaries in the appearance of "chicken wire" (Wesseling et al. 2015), which increases the confusion of QSM in differentiating between grade II and III gliomas. After adding QSM to conventional MRI modalities, the accuracy of glioma grading could be improved to 85%–90%, and the AUC was 0.91 in identifying OGG from GBM. As OGGs share many similar histologic and genomic characteristics and therapeutic strategies, which are quite different from GBMs (Reifenberger et al. 2017), distinguishing OGG from GBM is very critical, and DL-assisted QSM shows great clinical significance.

Single-modal T1WI + C and T2 FLAIR showed limited value in determining the molecular subtypes of gliomas. A previous study has shown that *IDH*(+) gliomas have lower levels of hypoxia-inducible-factor 1-alpha and decreased angiogenesis and vasculogenesis in comparison with *IDH*(−) ones through the 2-hydroxyglutarate-mediated prolyl hydroxylase enzymes (EGLN, also called PHD) inhibition (Kickingereder et al. 2015). The resulting magnetic susceptibility differences between gliomas with *IDH*(+) and *IDH*(−) make it possible for QSM to distinguish between molecular subtypes of *IDH1*. When conventional T1WI + C and T2 FLAIR were combined with QSM modality during DL, satisfactory efficiency with AUC of 0.88 and accuracy of 0.89 could be achieved in predicting *IDH1*(+). DL-assisted QSM showed greater potential for both the *IDH1* and *ATRX* task than conventional MRI. *ATRX*(−) is associated with alternate telomere lengthening, which can promote cellular immortality (Venneti and Huse 2015). In diagnosing *ATRX*(−), three MRI modalities together displayed a moderate accuracy with AUC of 0.78. In the four glioma classification tasks, QSM modality features were superior to demographic characteristics (age and sex), and QSM features were the most important variables for the classification of OGG/HGG, *IDH1*(+) or *IDH1*(−), *ATRX*(−) or *ATRX*(+). Compared with GLM, SVM, and standard CNN, our proposed inception CNN performed best in the three-modality fusion classification tasks.

Our previous study has shown the potential of multiparametric MR radiomic features in predicting *IDH1* and *ATRX* subtypes in LGGs (Ren et al. 2019). In recent years, many ML-based studies have been reported to be effective in determining glioma grades and *IDH* and *1p/19q* status by multi-modal MRI (Lu et al. 2018; Sengupta et al. 2019), but the contribution of intratumoral hemorrhage and calcification has not yet been studied. One study showed the potential of deep learning based on susceptibility-weighted imaging in detecting cerebral microbleeds (Liu et al. 2019). Considering the susceptibility variations of gliomas, the present study explored the highly reproducible QSM technique in combination with an inception CNN network, which revealed a great potential of DL-assisted QSM in glioma grading and molecular subtyping.

This study had some limitations. First, due to the small sample size, *IDH1* prediction task was not performed in OGGs and GBMs separately, and the results of DL need to be externally verified in larger study samples. Second, patients with HGG who were unable to tolerate long MR scans were underrepresented during the course of patient enrollment. Third, the volume information was not considered when delineating the tumor ROIs. In future studies, we plan to include multi-center QSM and perfusion MRI data to predict the molecular subtypes of OGG and GBM, respectively.

Conclusion

In conclusion, compared to conventional MRI modalities, DL-assisted QSM shows great advantages in distinguishing OGG from GBM, and predicting *IDH1* and *ATRX* subtypes.

Declarations

References

Bandt SK, de Rochefort L, Chen W, Dimov AV, Spincemaille P, Kopell BH, Gupta A, Wang Y (2019) Clinical integration of quantitative susceptibility mapping magnetic resonance imaging into neurosurgical practice. World Neurosurg 122:e10–e19. https://doi.org/10.1016/j.wneu.2018.08.213

Cheng J, Tian S, Yu L, Gao C, Kang X, Ma X, Wu W, Liu S, Lu H (2022) ResGANet: Residual group attention network for medical image classification and segmentation. Med Image Anal 76:102313. https://doi.org/10.1016/j.media.2021.102313

Deh K, Nguyen TD, Eskreis-Winkler S, Prince MR, Spincemaille P, Gauthier S, Kovanlikaya I, Zhang Y, Wang Y (2015) Reproducibility of quantitative susceptibility mapping in the brain at two field strengths from two vendors. J Magn Reson Imaging 42:1592–1600. https://doi.org/10.1002/jmri.24943

Deistung A, Schweser F, Wiestler B, Abello M, Roethke M, Sahm F, Wick W, Nagel AM, Heiland S, Schlemmer HP, Bendszus M, Reichenbach JR, Radbruch A (2013) Quantitative susceptibility mapping differentiates between blood depositions and calcifications in patients with glioblastoma. PLoS ONE 8:e57924. https://doi.org/10.1371/journal.pone.0057924

Ferris SP, Hofmann JW, Solomon DA, Perry A (2017) Characterization of gliomas: from morphology to molecules. Virchows Arch 471:257–269. https://doi.org/10.1007/s00428-017-2181-4

Haacke EM, Liu S, Buch S, Zheng W, Wu D, Ye Y (2015) Quantitative susceptibility mapping: current status and future directions. Magn Reson Imaging 33:1–25. https://doi.org/10.1016/j.mri.2014.09.004

Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H (2018) Artificial intelligence in radiology. Nat Rev Cancer 18:500–510. https://doi.org/10.1038/s41568-018-0016-5

Hsieh KL, Lo CM, Hsiao CJ (2017) Computer-aided grading of gliomas based on local and global MRI features. Comput Methods Programs Biomed 139:31–38. https://doi.org/10.1016/j.cmpb.2016.10.021

Ikemura M, Shibahara J, Mukasa A, Takayanagi S, Aihara K, Saito N, Aburatani H, Fukayama M (2016) Utility of ATRX immunohistochemistry in diagnosis of adult diffuse gliomas. Histopathology 69:260–267. https://doi.org/10.1111/his.12927

Karsa A, Punwani S, Shmueli K (2019) The effect of low resolution and coverage on the accuracy of susceptibility mapping. Magn Reson Med 81:1833–1848. https://doi.org/10.1002/mrm.27542

Kee Y, Liu Z, Zhou L, Dimov A, Cho J, de Rochefort L, Seo JK, Wang Y (2017) Quantitative susceptibility mapping (QSM) algorithms: mathematical rationale and computational implementations. IEEE Trans Biomed Eng 64:2531–2545. https://doi.org/10.1109/tbme.2017.2749298

Kickingereder P, Sahm F, Radbruch A, Wick W, Heiland S, Deimling A, Bendszus M, Wiestler B (2015) IDH mutation status is associated with a distinct hypoxia/angiogenesis transcriptome signature which is non-invasively predictable with rCBV imaging in human glioma. Sci Rep 5:16238. https://doi.org/10.1038/srep16238

Kim HG, Park S, Rhee HY, Lee KM, Ryu CW, Lee SY, Kim EJ, Wang Y, Jahng GH (2020) Evaluation and prediction of early alzheimer's disease using a machine learning-based optimized combination-feature set on gray matter volume and quantitative susceptibility mapping. Curr Alzheimer Res 17:428–437. https://doi.org/10.2174/1567205017666200624204427

Lancione M, Tosetti M, Donatelli G, Cosottini M, Costagli M (2017) The impact of white matter fiber orientation in single-acquisition quantitative susceptibility mapping. NMR Biomed. https://doi.org/10.1002/nbm.3798

Lancione M, Donatelli G, Cecchi P, Cosottini M, Tosetti M, Costagli M (2019) Echo-time dependency of quantitative susceptibility mapping reproducibility at different magnetic field strengths. Neuroimage 197:557–564. https://doi.org/10.1016/j.neuroimage.2019.05.004

Langkammer C, Schweser F, Krebs N, Deistung A, Goessler W, Scheurer E, Sommer K, Reishofer G, Yen K, Fazekas F, Ropele S, Reichenbach JR (2012) Quantitative susceptibility mapping (QSM) as a means to measure brain iron? A post mortem validation study. Neuroimage 62:1593–1599. https://doi.org/10.1016/j.neuroimage.2012.05.049

Li W, Wu B, Liu C (2011) Quantitative susceptibility mapping of human brain reflects spatial variation in tissue composition. Neuroimage 55:1645–1656. https://doi.org/10.1016/j.neuroimage.2010.11.088

Li W, Wu B, Avram AV, Liu C (2012) Magnetic susceptibility anisotropy of human brain in vivo and its molecular underpinnings. Neuroimage 59:2088–2097. https://doi.org/10.1016/j.neuroimage.2011.10.038

Li W, Wang N, Yu F, Han H, Cao W, Romero R, Tantiwongkosi B, Duong TQ, Liu C (2015) A method for estimating and removing streaking artifacts in quantitative susceptibility mapping. Neuroimage 108:111–122. https://doi.org/10.1016/j.neuroimage.2014.12.043

Li G, Zhai G, Zhao X, An H, Spincemaille P, Gillen KM, Ku Y, Wang Y, Huang D, Li J (2019) 3D texture analyses within the substantia nigra of Parkinson's disease patients on quantitative susceptibility maps and R2(∗) maps. Neuroimage 188:465–472. https://doi.org/10.1016/j.neuroimage.2018.12.041

Liu S, Utriainen D, Chai C, Chen Y, Wang L, Sethi SK, Xia S, Haacke EM (2019) Cerebral microbleed detection using Susceptibility Weighted Imaging and deep learning. Neuroimage 198:271–282. https://doi.org/10.1016/j.neuroimage.2019.05.046

Lotan E, Jain R, Razavian N, Fatterpekar GM, Lui YW (2019) State of the art: machine learning applications in glioma imaging. AJR Am J Roentgenol 212:26–37. https://doi.org/10.2214/ajr.18.20218

Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW (2016) The 2016 world health organization classification of tumors of the central nervous system: a summary. Acta Neuropathol 131:803–820. https://doi.org/10.1007/s00401-016-1545-1

Lu CF, Hsu FT, Hsieh KL, Kao YJ, Cheng SJ, Hsu JB, Tsai PH, Chen RJ, Huang CC, Yen Y, Chen CY (2018) Machine learning-based radiomics for molecular subtyping of gliomas. Clin Cancer Res 24:4429–4436. https://doi.org/10.1158/1078-0432.ccr-17-3445

Ly KI, Wen PY, Huang RY (2020) Imaging of central nervous system tumors based on the 2016 World Health Organization Classification. Neurol Clin 38:95–113. https://doi.org/10.1016/j.ncl.2019.08.004

Ogishima T, Tamura K, Kobayashi D, Inaji M, Hayashi S, Tamura R, Nariai T, Ishii K, Maehara T (2017) ATRX status correlates with 11 C-methionine uptake in WHO grade II and III gliomas with IDH1 mutations. Brain Tumor Pathol 34:20–27. https://doi.org/10.1007/s10014-017-0280-1

Ostrom QT, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE, Pekmezci M, Schwartzbaum JA, Turner MC, Walsh KM, Wrensch MR, Barnholtz-Sloan JS (2014) The epidemiology of glioma in adults: a "state of the science" review. Neuro Oncol 16:896–913. https://doi.org/10.1093/neuonc/nou087

Reifenberger G, Wirsching HG, Knobbe-Thomsen CB, Weller M (2017) Advances in the molecular genetics of gliomas - implications for classification and therapy. Nat Rev Clin Oncol 14:434–452. https://doi.org/10.1038/nrclinonc.2016.204

Ren Y, Zhang X, Rui W, Pang H, Qiu T, Wang J, Xie Q, Jin T, Zhang H, Chen H, Zhang Y, Lu H, Yao Z, Zhang J, Feng X (2019) Non-invasive prediction of IDH1 mutation and ATRX expression loss in low-grade gliomas using multiparametric MR radiomic

features. J Magn Reson Imaging 49:808–817. https://doi.org/10.1002/jmri.26240

Rui W, Ren Y, Wang Y, Gao X, Xu X, Yao Z (2018) MR textural analysis on T2 FLAIR images for the prediction of true oligodendroglioma by the 2016 WHO genetic classification. J Magn Reson Imaging 48:74–83. https://doi.org/10.1002/jmri.25896

Saito T, Muragaki Y, Maruyama T, Komori T, Tamura M, Nitta M, Tsuzuki S, Kawamata T (2016) Calcification on CT is a simple and valuable preoperative indicator of 1p/19q loss of heterozygosity in supratentorial brain tumors that are suspected grade II and III gliomas. Brain Tumor Pathol 33:175–182. https://doi.org/10.1007/s10014-016-0249-5

Schweser F, Deistung A, Lehr BW, Reichenbach JR (2011) Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: an approach to in vivo brain iron metabolism? Neuroimage 54:2789–2807. https://doi.org/10.1016/j.neuroimage.2010.10.070

Sengupta A, Ramaniharan AK, Gupta RK, Agarwal S, Singh A (2019) Glioma grading using a machine-learning framework based on optimized features obtained from T(1) perfusion MRI and volumes of tumor components. J Magn Reson Imaging 50:1295–1306. https://doi.org/10.1002/jmri.26704

Venneti S, Huse JT (2015) The evolving molecular genetics of low-grade glioma. Adv Anat Pathol 22:94–101. https://doi.org/10.1097/pap.0000000000000049

Wang R, Xie G, Zhai M, Zhang Z, Wu B, Zheng D, Hong N, Jiang T, Wen B, Cheng J (2017a) Stability of R2* and quantitative susceptibility mapping of the brain tissue in a large scale multi-center study. Sci Rep 7:45261. https://doi.org/10.1038/srep45261

Wang Y, Spincemaille P, Liu Z, Dimov A, Deh K, Li J, Zhang Y, Yao Y, Gillen KM, Wilman AH, Gupta A, Tsiouris AJ, Kovanlikaya I, Chiang GC, Weinsaft JW, Tanenbaum L, Chen W, Zhu W, Chang S, Lou M, Kopell BH, Kaplitt MG, Devos D, Hirai T, Huang X, Korogi Y, Shtilbans A, Jahng GH, Pelletier D, Gauthier SA, Pitt D, Bush AI, Brittenham GM, Prince MR (2017b) Clinical quantitative susceptibility mapping (QSM): biometal imaging and its emerging roles in patient care. J Magn Reson Imaging 46:951–971. https://doi.org/10.1002/jmri.25693

Weller M, van den Bent M, Hopkins K, Tonn JC, Stupp R, Falini A, Cohen-Jonathan-Moyal E, Frappaz D, Henriksson R, Balana C, Chinot O, Ram Z, Reifenberger G, Soffietti R, Wick W (2014) EANO guideline for the diagnosis and treatment of anaplastic gliomas and glioblastoma. Lancet Oncol 15:e395-403. https://doi.org/10.1016/s1470-2045(14)70011-7

Weller M, Wick W, Aldape K, Brada M, Berger M, Pfister SM, Nishikawa R, Rosenthal M, Wen PY, Stupp R, Reifenberger G (2015) Glioma. Nat Rev Dis Primers 1:15017. https://doi.org/10.1038/nrdp.2015.17

Weller M, van den Bent M, Tonn JC, Stupp R, Preusser M, Cohen-Jonathan-Moyal E, Henriksson R, Le Rhun E, Balana C, Chinot O, Bendszus M, Reijneveld JC, Dhermain F, French P, Marosi C, Watts C, Oberg I, Pilkington G, Baumert BG, Taphoorn MJB, Hegi M, Westphal M, Reifenberger G, Soffietti R, Wick W (2017) European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. Lancet Oncol 18:e315–e329. https://doi.org/10.1016/s1470-2045(17)30194-8

Wesseling P, van den Bent M, Perry A (2015) Oligodendroglioma: pathology, molecular mechanisms and markers. Acta Neuropathol 129:809–827. https://doi.org/10.1007/s00401-015-1424-1

Zhang S, Liu Z, Nguyen TD, Yao Y, Gillen KM, Spincemaille P, Kovanlikaya I, Gupta A, Wang Y (2019) Clinical feasibility of brain quantitative susceptibility mapping. Magn Reson Imaging 60:44–51. https://doi.org/10.1016/j.mri.2019.04.003

Authors and Affiliations

Wenting Rui[1] · **Shengjie Zhang[2,3]** · **Huidong Shi[1]** · **Yaru Sheng[1]** · **Fengping Zhu[4]** · **YiDi Yao[1]** · **Xiang Chen[2,3]** · **Haixia Cheng[5]** · **Yong Zhang[6]** · **Ababikere Aili[7]** · **Zhenwei Yao[1]** · **Xiao-Yong Zhang[2,3]** · **Yan Ren[1]**

Wenting Rui
wennyrui@126.com

Shengjie Zhang
zsjxll@gmail.com

Huidong Shi
396958945@qq.com

Yaru Sheng
446396841@qq.com

Fengping Zhu
zhugeny@163.com

YiDi Yao
983414709@qq.com

Xiang Chen
20110850014@fudan.edu.cn

Haixia Cheng
Chenghaixia99@126.com

Yong Zhang
dryongzhang@126.com

Ababikere Aili
3351724032@qq.com

Zhenwei Yao
zwyao@fudan.edu.cn

1   Department of Radiology, Huashan Hospital, Fudan University, Mid 12 Wulumuqi Road, Shanghai 200040, China

2   Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China

3   MOE Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China

4   Department of Neurosurgery, Huashan Hospital, Fudan University, Shanghai 200040, China

[5]   Department of Neuropathology, Huashan Hospital, Fudan
      University, Shanghai 200040, China

[6]   GE Healthcare, MR Research, Huatuo Road,
      Shanghai 201203, China

[7]   Department of Radiology, Kuqa County People's Hospital,
      Xinjiang 842000, China