



Comparison of diagnostic performance of radiologist- and AI-based assessments of T2-FLAIR mismatch sign and quantitative assessment using synthetic MRI in the differential diagnosis between astrocytoma, IDH-mutant and oligodendroglioma, IDH-mutant and 1p/19q-codeleted

Kazufumi Kikuchi¹ · Osamu Togao¹ · Koji Yamashita² · Daichi Momosaka² · Yoshitomo Kikuchi² · Daisuke Kuga³ · Sangatsuda Yuhei³ · Yutaka Fujioka³ · Fumiya Narutomi⁴ · Makoto Obara⁵ · Koji Yoshimoto³ · Kousei Ishigami²

Received: 30 October 2023 / Accepted: 7 January 2024
© The Author(s) 2024

Abstract

Purpose This study aimed to compare assessments by radiologists, artificial intelligence (AI), and quantitative measurement using synthetic MRI (SyMRI) for differential diagnosis between astrocytoma, IDH-mutant and oligodendroglioma, and IDH-mutant and 1p/19q-codeleted and to identify the superior method.

Methods Thirty-three cases (men, 14; women, 19) comprising 19 astrocytomas and 14 oligodendrogliomas were evaluated. Four radiologists independently evaluated the presence of the T2-FLAIR mismatch sign. A 3D convolutional neural network (CNN) model was trained using 50 patients outside the test group (28 astrocytomas and 22 oligodendrogliomas) and transferred to evaluate the T2-FLAIR mismatch lesions in the test group. If the CNN labeled more than 50% of the T2-prolonged lesion area, the result was considered positive. The T1/T2-relaxation times and proton density (PD) derived from SyMRI were measured in both gliomas. Each quantitative parameter (T1, T2, and PD) was compared between gliomas using the Mann–Whitney *U*-test. Receiver-operating characteristic analysis was used to evaluate the diagnostic performance.

Results The mean sensitivity, specificity, and area under the curve (AUC) of radiologists vs. AI were 76.3% vs. 94.7%; 100% vs. 92.9%; and 0.880 vs. 0.938, respectively. The two types of diffuse gliomas could be differentiated using a cutoff value of 2290/128 ms for a combined 90th percentile of T1 and 10th percentile of T2 relaxation times with 94.4/100% sensitivity/specificity with an AUC of 0.981.

Conclusion Compared to the radiologists' assessment using the T2-FLAIR mismatch sign, the AI and the SyMRI assessments increased both sensitivity and objectivity, resulting in improved diagnostic performance in differentiating gliomas.

Keywords Artificial intelligence · Convolutional neural network · Synthetic MRI · T2-FLAIR mismatch sign

✉ Kazufumi Kikuchi
kikuchi.kazufumi.953@m.kyushu-u.ac.jp

¹ Department of Molecular Imaging and Diagnosis, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-Ku, Fukuoka 812-8582, Japan

² Department of Clinical Radiology, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-Ku, Fukuoka 812-8582, Japan

³ Department of Neurosurgery, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-Ku, Fukuoka 812-8582, Japan

⁴ Department of Anatomic Pathology, Pathological Sciences, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-Ku, Fukuoka 812-8582, Japan

⁵ Philips Japan Ltd., 2-13-37, Konan, Minato-Ku, Tokyo 108-8507, Japan

Introduction

Isocitrate dehydrogenase (IDH) enzymes play a key role in glioma tumorigenesis [1]. A previous study revealed that IDH mutations were more frequently observed in diffuse low-grade gliomas, including astrocytomas and oligodendrogliomas [2]. The two types of gliomas share the same IDH mutation status, but their prognoses differ [3]. Oligodendrogliomas, IDH-mutant and 1p/19q-codeleted have a better prognosis and respond better to chemotherapy or radiotherapy than astrocytomas, IDH-mutant [4], while astrocytomas require more intensive treatment. Therefore, an accurate diagnosis is essential for effective patient management [5]. In 2017, Patel et al. [3] reported that astrocytoma, IDH-mutant exhibited the T2-FLAIR mismatch sign. Subsequently, numerous studies on the T2-FLAIR mismatch sign have been published [5–8]. The T2-FLAIR mismatch sign has a high specificity of 100% but a low sensitivity ranging from 12 to 51% in the diagnosis of astrocytoma, IDH-mutant [3, 7]. This is due to the strict criteria used to maintain high specificity and the wide range of the interobserver agreement, which makes it dependent on observer subjectivity, leading to significant interobserver variability [6].

Potential solutions to this problem include artificial intelligence (AI) modalities such as deep learning and quantitative approaches based on relaxation time measurements. In recent years, there has been considerable research regarding AI as an adjunct to imaging diagnostics, with some studies suggesting that it can outperform radiologists in certain diagnostic tasks [9–12]. Moreover, there are reports that the combination of AI-detected lesions and human assessment can lead to an even higher diagnostic accuracy [13, 14]. By eliminating subjective judgments and highlighting areas of T2-FLAIR mismatch, it is proposed that AI could be a valuable asset in this context. To the best of our knowledge, no previous studies on AI assessments of the T2-FLAIR mismatch sign have been reported.

Another solution might be a quantitative method that offers inherent objectivity. In addition to ensuring consistency, the capacity to make numerical judgments eliminates much of the subjectivity that can occasionally result in biases or errors in the interpretation of data. Previous studies have reported that measurement of relaxation time can improve sensitivity in T2-FLAIR mismatch lesions [5, 15].

Therefore, this study aimed to compare assessments by radiologists, artificial intelligence (AI), and quantitative measurement using synthetic MRI (SyMRI) for differential diagnosis between astrocytoma, IDH-mutant and oligodendroglioma, and IDH-mutant and 1p/19q-codeleted and to identify the superior method.

Materials and methods

The institutional review board of our hospital approved this retrospective study, and the requirement for informed consent was waived. All methods were performed in accordance with the relevant guidelines and regulations.

Patients

From June 2019 to December 2021, all patients at our institution who received a glioma diagnosis in a timely manner were eligible for this study. Inclusion criteria were (1) a diagnosis of IDH-mutant and/or 1p/19q-codeleted glioma based on the WHO 2021 classifications [1], (2) MRI scans performed within 2 weeks preceding surgery, and (3) available SyMRI evaluations. The exclusion criterion was image distortion, such as motion artifacts or noise. A total of 33 patients (men, 14; women, 19; age, 29–79 [median, 44] years) including 19 astrocytomas, IDH-mutant (men, 6; women, 13; age, 29–60 [median, 43] years; Grade 2, $n = 12$; Grade 3, $n = 5$; Grade 4, $n = 2$), and 14 oligodendrogliomas, IDH-mutant and 1p/19q-codeleted (men, 8; women, 6; age, 37–63 [median, 48] years; Grade 2, $n = 10$; Grade 3, $n = 4$) were included in the study.

For the machine learning training data, 50 patients other than those included in the study, treated at our hospital from June 2002 to December 2014, were selected. These comprised 28 astrocytomas, IDH-mutant (men, 20; women, 8; age, 20–52 [median, 33] years; Grade 2, $n = 18$; Grade 3, $n = 10$), and 22 oligodendrogliomas, IDH-mutant and 1p/19q-codeleted (men, 10; women, 12; age, 20–73 [median, 46] years; Grade 2, $n = 13$; Grade 3, $n = 9$).

MR imaging

MR sequences used in this study were previously described [5]. All examinations were performed using a 3 T MR scanner (Ingenia 3.0 T CX; Philips Healthcare, Best, Netherlands) with a 15-channel head coil. Quantitative MRI was performed using the two-dimensional axial quantification of relaxation times and proton densities by the multi-echo acquisition of a saturation recovery using a turbo spin-echo readout (QRAPMASTER) pulse sequence with two echo times (TEs; 13 and 100 ms) and four delay times to generate eight real images and eight imaginary images [16]. The other parameters included repetition time (TR), 4831 ms; flip angle (FA), 90°; number of excitations (NEX), 1; sensitivity-encoding factor, 2.2; field of view (FOV), 230 × 189 (recon. 230 × 230) mm²; matrix, 512 × 512; echo-train length, 10; thickness/gap, 4.0/1.0 mm; 30 slices; and scan time 6 min 36 s. Quantification map acquisition was performed using SyMRI software (Version 19.0; SyMRI,

Linköping, Sweden, <https://syntheticmr.com/>) [16]. Standard MR sequences (T1-weighted imaging [T1WI], T2WI, FLAIR, and contrast-enhanced T1WI) were also obtained. The sequence parameters of the two-dimensional axial T2WI and FLAIR sequences were as follows: T2WI—TR/TE, 3000/80 ms; FA, 90°; NEX, 1; FOV, 230×230 mm²; matrix, 512×375 (recon. 512×512); echo-training length, 15; thickness/gap, 5.0/1.0 mm; 22 slices; and scan time, 2 min 36 s, and FLAIR—TR/TE/TI, 10,000/120/2700 ms; FA, 180°; NEX, 1; FOV, 230×207 (recon. 230×230) mm²; matrix, 320×228 (recon. 512×512); echo-training length, 27; thickness/gap, 5.0/1.0 mm; 22 slices; and scan time, 3 min.

Convolutional neural network model architecture

U-net and DeepMedic are widely used as convolutional neural networks (CNNs) in AI research. As for this study, DeepMedic was chosen because it has been reported to outperform U-net in intracranial atherosclerotic diseases [17]. We applied the DeepMedic network developed by Kamnitsas et al. [18], which is a multi-scale 3D CNN, to assess the T2-FLAIR mismatch lesion. In order to create a large receptive field for the final classification while retaining a low computational cost, this design comprises 11 layers and 2 parallel convolutional pathways that process the input at various scales. This architecture uses 3³ kernels, which are fewer than the usual 5³ kernels, to convolve quickly and minimize the weight. With these small kernels, deep network variants can be designed efficiently by reducing the number of multiplications and trainable parameters for each

element. To incorporate both local and more general contextual information into the 3D CNN, a down-sampled second pathway was introduced. In the first pathway, the structure's specific local appearance is recorded, whereas, in the second pathway, higher-level information like the structure's location in the brain is learned [18]. The identification of the T2-FLAIR mismatch lesion as ground truth was manually determined by a certificated radiologist (K.K., with 16 years of experience in diagnostic radiology) who knew the pathological information of patients with diffuse glioma. Figure 1 shows the preprocessing pipeline using IntelliSpace Discovery (version 3, Philips Healthcare, Best, Netherlands). Both images of T2WI and FLAIR were bias-field-corrected. Then, in the second step, the corrected FLAIR images were coregistered to the reference space defined by the T2WI. Third, a brain mask was computed and applied to obtain skull-stripped images. Finally, labels were drawn using the semi-automated contouring tool. For detecting the T2-FLAIR mismatch lesion, the deep learning model (DeepMedic) was applied to the preprocessed data.

T2-FLAIR mismatch evaluation and statistical analysis

Radiologist assessment

Four board-certified neuroradiologists (with 23, 21, 10, and 8 years of experience) were blinded to the patient information of the evaluated T2-FLAIR mismatch sign. The T2-FLAIR mismatch sign was defined by the presence

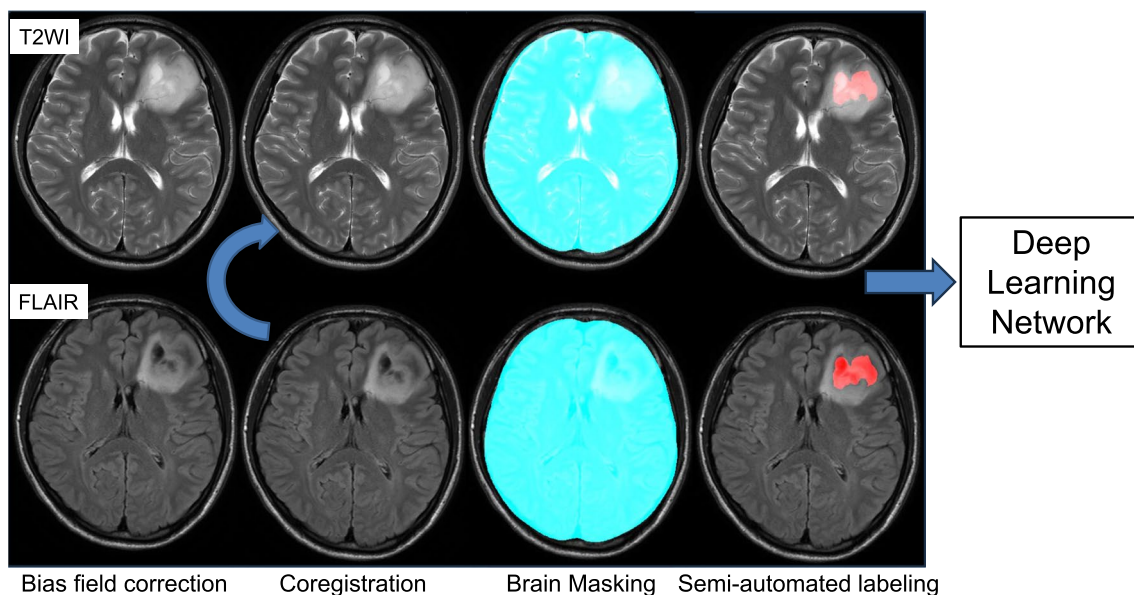


Fig. 1 Image-preprocessing pipeline. (1) Bias field correction was applied to T2WI and FLAIR images, (2) FLAIR images were coregistered with T2WI images, (3) brain masking was created on T2WI

images and propagated to the registered FLAIR images, and (4) labels were drawn using the semi-automated contouring tool

of two distinct MRI features as follows [3, 6]: (1) The tumor displayed a complete or nearly complete and nearly homogeneous hyperintense signal on T2WI and (2) the tumor displayed a relatively hypointense signal on the FLAIR sequence except for a hyperintense peripheral rim. Further, Jain et al. [6] introduced additional imaging features aiding in the accurate identification of the T2-FLAIR mismatch sign: (3) Necrotic cavities do not represent the T2-FLAIR mismatch sign; small cysts do not meet the criteria for the T2-FLAIR mismatch sign. (4) The T2-FLAIR mismatch lesion is typically accompanied by little or no contrast enhancement. (5) The degree of FLAIR signal suppression could be inhomogeneous within the tumor. (6) Common imaging correlates include homogeneous hypointensity on non-contrast T1WI, markedly elevated apparent diffusion coefficient values, low blood volume on perfusion maps, and diffuse hypodensity on CT. After independent data collection, the interreader agreement was calculated. Four radiologists read both the T2WI and FLAIR images based on whether the T2-FLAIR mismatch sign was present or absent. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were calculated. In the radiologist's evaluation, the interrater agreement for the T2-FLAIR mismatch sign among the four observers was evaluated using Fleiss's kappa coefficient [19]. The kappa value was interpreted as follows: almost perfect agreement, 1.00–0.81; substantial agreement, 0.80–0.61; moderate agreement, 0.60–0.41; fair agreement, 0.40–0.21; slight agreement, 0.20–0.01; and poor agreement, < 0 [20].

Artificial intelligence assessment based on the convolutional neural network

If the CNN labeled more than 50% of the T2-prolonged lesion area, it was considered positive, defining the presence of the T2-FLAIR mismatch sign. While there is a method using the Dice coefficient, this study is binary

in nature; therefore, we simply determined it based on appearance.

Quantitative assessment using synthetic MRI

The DICOM data of the T1 and T2 relaxation times and proton density map were extracted by SyMRI software (version 19.0; SyMRI, Linköping, Sweden, <https://syntheticmr.com/>) [16]. We used a single maximum section of each tumor for the regions of interest (ROI) analysis on the T2-prolonged region in the tumor using an ImageJ plugin (ImageJ/Fiji; version 2.0.0-rc-59/1.51 k, National Institutes of Health, Bethesda, MD). The maximum section of the tumor was visually determined as the largest orthogonal cross-product of the tumor on the axial T2WI/FLAIR [5]. Using the ROI manager tool of ImageJ/Fiji, the ROI mask from the T2-prolonged region on conventional T2WI scans was copied and placed on each parameter map (T1 and T2 relaxations and proton density maps) to obtain pixel-by-pixel values for the histogram analyses. The 10th, 25th, 50th, 75th, and 90th percentiles and the mean, skewness, and kurtosis of each parameter were recorded from the histograms. Each parameter (i.e., T1 and T2 relaxation times and proton density) was compared between astrocytomas, IDH-mutant and oligodendrogliomas, IDH-mutant and 1p/19q-codeleted using the Mann–Whitney *U*-test. The diagnostic performance of each parameter was evaluated using a receiver-operating characteristic curve analysis.

All statistical analyses were performed using commercial software programs (JMP, version 15.0.0; SAS Institute, Cary, NC, USA; Prism 7.0, GraphPad Software, La Jolla, CA, USA). $P < 0.05$ was considered statistically significant.

Results

Radiologist and artificial intelligence assessments

Table 1 shows the results from the four radiologists and AI. The mean sensitivity, specificity, accuracy, PPV, NPV, and

Table 1 Radiologist and artificial intelligence assessment of T2-FLAIR mismatch sign

	Reader 1	Reader 2	Reader 3	Reader 4	Average*	AI
Sensitivity (%)	68.4	73.7	84.2	79.0	76.3	94.7
Specificity (%)	100.0	100.0	100.0	100.0	100.0	92.9
Accuracy (%)	81.8	84.9	90.9	87.9	86.4	93.9
PPV (%)	100.0	100.0	100.0	100.0	100.0	94.7
NPV (%)	70.0	73.7	82.4	77.8	76.0	92.9
AUC	0.850	0.868	0.912	0.889	0.880	0.938

AI artificial intelligence, AUC area under the curve, NPV negative predictive value, PPV positive predictive value

*The kappa coefficient among four radiologists was 0.88

the area under the curve (AUC) of radiologists vs. AI were 76.3% vs. 94.7%, 100% vs. 92.9%, 86.4% vs. 93.9%, 100% vs. 94.7%, 76.0% vs. 92.9%, and 0.880 vs. 0.938, respectively. An almost perfect interrater agreement was observed (kappa coefficient = 0.88).

Quantitative assessment using synthetic MRI

Figure 2 and Supplementary Table S1 show the histograms of each parameter over all the pixels in the tumor ROIs. T1 and T2 relaxation times and proton densities from the astrocytomas all exhibited a slight rightward shift relative to those from the oligodendrogliomas. T1 and T2 relaxation times and proton densities were larger for astrocytomas than for oligodendrogliomas (median values, 95% confidence intervals, and p -values—2503 (1967–2891) vs. 1385 (1119–1783) ms, $p < 0.0001$ for mean T1 relaxation time; 259 (214–343) vs. 121 (94–143) ms, $p < 0.0001$ for mean T2 relaxation time; and 94.9% (89.2–96.7%) vs. 85.2% (76.8–88.3%), $p < 0.0001$ for mean proton density, respectively). There were also significant differences in the 10–90th percentiles for T1 and T2 relaxation times and proton densities (all $p < 0.05$). Table 2 and Supplementary

Table S2 show the diagnostic performance in differentiating the two glioma groups; the most useful values of each parameter are shown in Table 2. The two types of diffuse gliomas could be differentiated using a cutoff value of 2290/128 ms for a combined 90th percentile of T1 and 10th percentile of T2 relaxation times with 94.4% sensitivity, 100% specificity, 96.9% accuracy, 100% PPV, and 93.3% NPV, with an AUC of 0.981.

Figures 3 and 4 show representative images of patients with astrocytoma and oligodendroglioma, respectively.

Discussion

We found that both AI and SyMRI improved the sensitivity of T2-FLAIR mismatch lesions as well as the diagnostic performance in the differential diagnosis between astrocytoma, IDH-mutant and oligodendroglioma, and IDH-mutant and 1p/19q-codeleted compared to radiologists in this study. The determination of the T2-FLAIR mismatch sign is subjective, resulting in variability. This study showed that by eliminating subjectivity, sensitivity was improved. The use of AI offers a distinct advantage regarding versatility; once

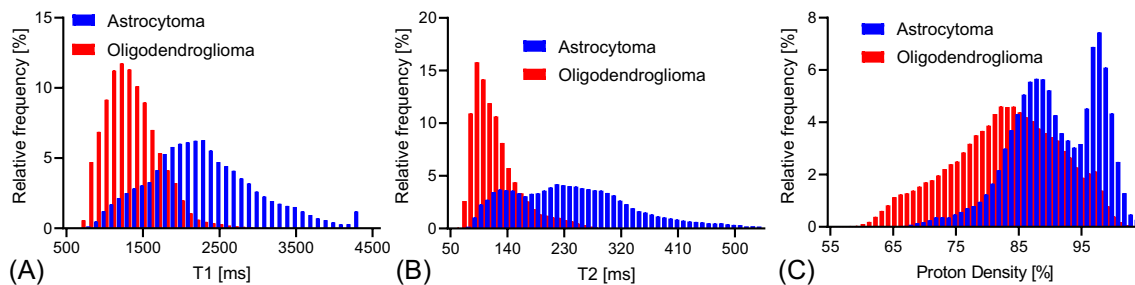


Fig. 2 Histograms of T1 and T2 relaxation times and proton density (PD) between astrocytomas, IDH-mutant and oligodendrogliomas, IDH-mutant and 1p/19q-codeleted. All parameters (T1 and T2 relaxa-

tion times and PD) in astrocytomas exhibit a slight rightward shift relative to those in oligodendrogliomas

Table 2 Diagnostic performance of parameters in differentiating between astrocytoma, IDH-mutant and oligodendroglioma, and IDH-mutant and 1p/19q-codeleted

Parameters	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	Cutoff	AUC
T1 [ms]							
75 th percentile	84.2	100.0	90.9	100.0	82.4	2107	0.966
90 th percentile	89.5	92.9	90.9	94.4	86.7	2290	0.966
T2 [ms]							
10 th percentile	100.0	85.7	93.9	90.5	100.0	99	0.977
PD [%]							
10 th percentile	79.0	100.0	87.9	100.0	77.8	83.4	0.917
Combined 90 th T1 and 10 th T2	94.4	100.0	96.9	100.0	93.3	2290/128	0.981

AUC area under the curve, IDH isocitrate dehydrogenase, NPV negative predictive value, PD proton density, PPV positive predictive value

Fig. 3 Images from a 39-year-old man with astrocytoma, IDH-mutant (WHO Grade 2). **a** T2WI shows a homogeneous T2-prolonged mass in the right insula (arrow). **b** FLAIR shows partial signal suppression, indicating a T2-FLAIR mismatch sign (arrowheads). **c** Our artificial intelligence correctly detects this T2-FLAIR mismatch lesion (arrow). T1 (**d**), T2 (**e**), and relaxation time and proton density (**f**) maps derived from SyMRI show T1 (2891 ms*) and T2 (375 ms*) relaxation time prolongations and increased PD (96.1%*) (arrows) in the tumor. Asterisks (*) beside the values in the caption indicates that each value is expressed as the mean

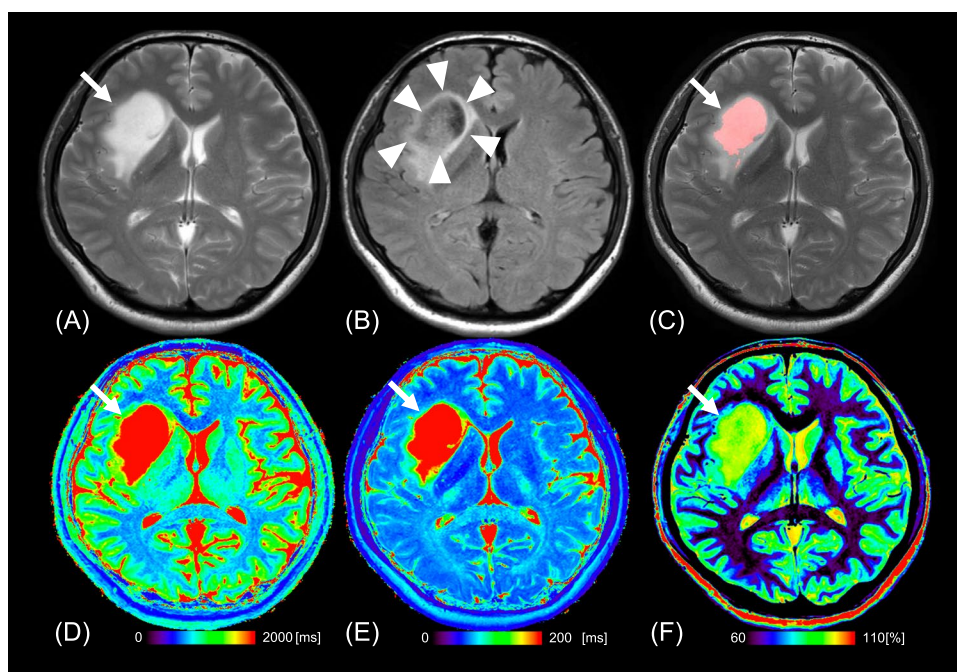
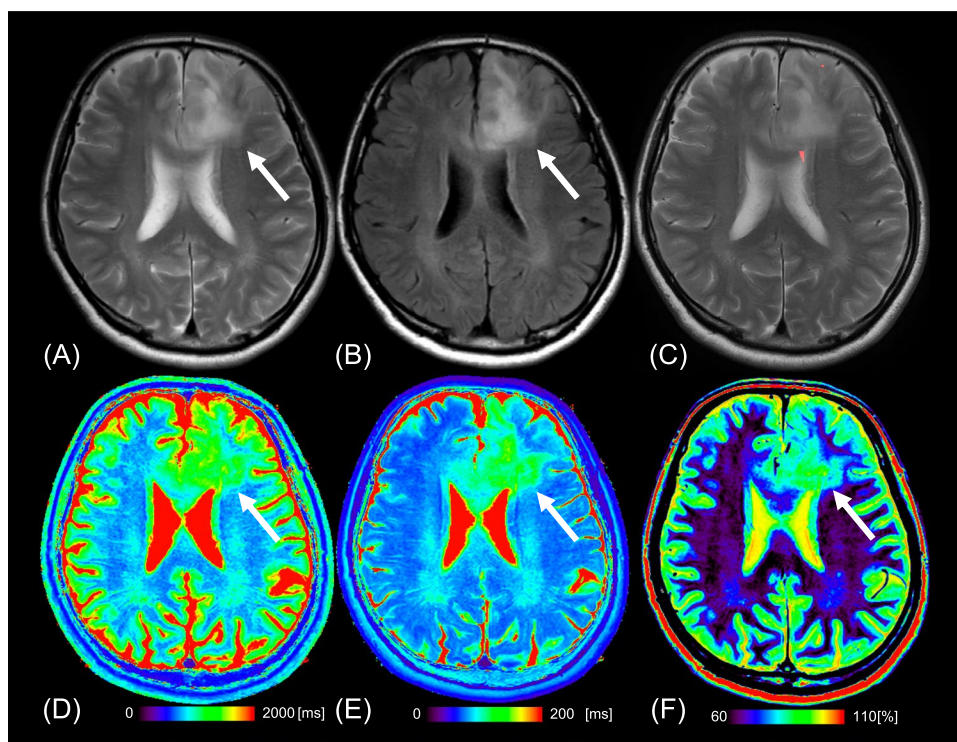


Fig. 4 Images from a 54-year-old man with oligodendroglioma, IDH-mutant and 1p/19q-codeleted (WHO Grade 2). **a** T2WI shows a heterogeneous, poorly circumscribed mass in the bilateral frontal lobes (arrow). **b** FLAIR shows no signal suppression in the tumor, indicating no T2-FLAIR mismatch sign (arrow). **c** Our artificial intelligence does not correctly show this T2-FLAIR mismatch lesion. T1 (**d**), T2 (**e**), and relaxation time and proton density (**f**) maps derived from SyMRI show mild T1 (1783 ms*) and T2 (121 ms*) relaxation time prolongations and mildly increased PD (81.7%*) (arrows) in the tumor. Asterisks (*) beside the values in the caption indicate that each value is expressed as the mean



a model is refined and completed, it can be deployed across different institutions or settings, ensuring widespread applicability. This universal adaptability is a compelling strength of AI-driven solutions. On the other hand, the quantitative assessment by SyMRI has unique advantages. The ability to

evaluate data numerically provides inherent objectivity. This ability to provide numerical assessments not only ensures consistency but also removes much of the human subjectivity that can sometimes lead to inconsistencies or biases in data interpretation. Therefore, while AI offers flexibility and

adaptability, tools like SyMRI offer rigorous, objective analysis. The present study method is expected to increase the accuracy of preoperative brain tumor diagnosis. Because the AI model can be used for transfer learning, PACS equipped with AI applications could find widespread use. On the other hand, relaxation time can also be measured with conventional MRI using the multi-echo method instead of SyMRI.

Previous studies have provided valuable insight into radiologists' assessment of the T2-FLAIR mismatch sign. Sensitivity in these studies has been found to range from 22 to 57% [3, 7]. The interrater agreement among radiologists has also shown a considerable range, with κ (kappa coefficient) values ranging from 0.38 to 0.88 [8]. Notably, Jain et al. emphasized the importance of applying strict criteria to maintain a high level of specificity, even though this is often at the expense of reduced sensitivity [6]. Our research findings are consistent with the trends observed in these earlier studies. The relatively low interrater agreement among radiologists is likely due to the binary scoring system that is commonly used. This system may not adequately capture the nuances of the T2-FLAIR mismatch sign, as subtle variations in imaging characteristics may lead to different interpretations by different readers [5, 8].

The high sensitivity of AI in detecting T2-FLAIR mismatch lesions is believed to be due to its retrospective learning approach, where it learns to identify these lesions after having access to pathology results. In other words, it operates in a "cheat mode" that enables it to achieve greater sensitivity. To our knowledge, there are no previous studies that have attempted to detect T2-FLAIR mismatch lesions using AI. We used the DeepMedic network proposed by Kamnitsas et al. [18], which is efficient in learning even with a small dataset. Because this model employs a multi-scale approach to capture information at different levels of detail, it can still extract useful features from the data even when the dataset is small, thereby optimizing the available information. Kikuchi et al. reported that they used DeepMedic trained on 50 patients with 165 lesions to detect brain metastases [9]. Although the amount of training data is smaller than in previous studies (number of training cases/lesions = 188–469/917–1149) [11, 12], DeepMedic demonstrates a detection sensitivity for brain metastases that is comparable with that of radiologists [9]. This suggests that DeepMedic can effectively learn from a limited number of cases.

SyMRI has been shown to be valuable in increasing the sensitivity for differential diagnosis between astrocytoma, IDH-mutant and oligodendroglioma, IDH-mutant and 1p/19q-codeleted. Previous studies have also highlighted the benefits of using quantitative relaxation time assessments in the context of T2-FLAIR mismatch lesions, which have consistently resulted in increased sensitivity [5, 15]. The results of the present study are consistent with these previous research findings and underscore the utility of such quantitative assessments. In this study, an increase in sensitivity

was observed in the measurement of relaxation time; however, the specificity of the single parameter decreased slightly. Previous research on the pathologic evaluation of the T2-FLAIR mismatch sign has shown that regions with T2-FLAIR mismatch have microcystic changes, leading to prolongation of relaxation time due to increased fluid components [21]. Differential diagnosis is difficult when astrocytoma, IDH-mutant presents without microcystic change because of the lack of prolonged relaxation time. Nevertheless, the combination with the relaxation parameters of T1 and T2 exhibited improved diagnostic performance. Based on this result, it can be concluded that the combination of T1 and T2 relaxation times provides a better understanding of the tissue structure within the T2-FLAIR mismatch lesions.

This study has several limitations. First, our study comprised postoperative cases and had a small sample size since IDH-mutant-type gliomas are relatively rare and there are limitations to collecting cases at a single center. Because an AI-based classification study requires a large sample size, this may raise concerns about the reliability of the results; therefore, further studies may require multicenter validation. Second, we did not assess IDH-wild-type astrocytomas in our investigation. A follow-up study with patients with IDH-wild-type astrocytomas would be useful. Third, we did not include the whole tumor volume for the histogram analysis in the quantitative evaluation. Instead, we used the maximum section of the tumor, with its boundary defined by the hyperintensity on T2WI. However, only the largest region of the tumor was used in the previous research on the T2-FLAIR mismatch sign; whole-volume histogram analysis was not conducted. Since the T2-FLAIR mismatch sign criteria are designed to retain high specificity rather than boost sensitivity, by using these tight criteria, a simple evaluation based on the maximum-sized slice of the tumor may be sufficient. Although a quantitative whole-tumor analysis would probably yield results that differ from those currently presented, it is likely that astrocytomas, IDH-mutant would have exhibited longer T1 and T2 values than oligodendrogliomas, IDH-mutant and 1p/19q-codeleted.

In conclusion, compared to radiologists' assessments using the T2-FLAIR mismatch sign, the AI and the SyMRI assessments increased both sensitivity and objectivity, resulting in improved diagnostic performance in differentiating astrocytomas, IDH-mutant from oligodendrogliomas, IDH-mutant and 1p/19q-codeleted.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00234-024-03288-0>.

Author contributions KK: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, and writing—original draft, visualization; OT: validation, investigation,

and writing—review and editing; KY: validation, investigation, and writing—review and editing; DM: validation, investigation, resources, data curation, and writing—review and editing; YK: validation, investigation, resources, data curation, and writing—review and editing; DK: validation, investigation, resources, data curation, and writing—review and editing; SY: validation, investigation, resources, data curation, and writing—review and editing; YF: validation, investigation, resources, data curation, and writing—review and editing; FN: validation, investigation, resources, and writing—review and editing; MO: validation, investigation, resources, and writing—review and editing; KY: validation, investigation, resources, data curation, and writing—review and editing; KI: conceptualization, methodology, validation, investigation, resources, writing—review and editing, supervision, and project administration.

Funding This work was supported by the GE Healthcare Pharma Educational Grant, Philips Japan, Ltd., and the Shin-Nihon Foundation of Advanced Medical Research.

Data, materials, and/or code availability The datasets generated during the current study are not available because of our institutional policy.

Declarations

Competing interests None of the authors have any conflict of competing interests to disclose.

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This retrospective study was approved by our Institutional Review Board (number 2020–228).

Consent Consent to participate was waived due to the retrospective study design.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- WHO Classification of Tumours Editorial Board (2022) Central nervous system tumours: WHO classification of tumours. WHO classification of tumours
- Ichimura K, Pearson DM, Kocialkowski S, Bäcklund LM, Chan R, Jones DT, Collins VP (2009) IDH1 mutations are present in the majority of common adult gliomas but rare in primary glioblastomas. *Neuro Oncol* 11:341–347. <https://doi.org/10.1215/15228517-2009-025>
- Patel SH, Poisson LM, Brat DJ, Zhou Y, Cooper L, Snuderl M, Thomas C, Franceschi AM, Griffith B, Flanders AE, Golfinos JG, Chi AS, Jain R (2017) T2-FLAIR mismatch, an imaging biomarker for IDH and 1p/19q status in lower-grade gliomas: a TCGA/TCIA project. *Clin Cancer Res* 23:6078–6085. <https://doi.org/10.1158/1078-0432.ccr-17-0560>
- Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, Pekmezci M, Rice T, Kosel ML, Smirnov IV, Sarkar G, Caron AA, Kollmeyer TM, Praska CE, Chada AR, Halder C, Hansen HM, McCoy LS, Bracci PM, Marshall R, Zheng S, Reis GF, Pico AR, O'Neill BP, Buckner JC, Giannini C, Huse JT, Perry A, Tihan T, Berger MS, Chang SM, Prados MD, Wiemels J, Wiencke JK, Wrensch MR, Jenkins RB (2015) Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N Engl J Med* 372:2499–2508. <https://doi.org/10.1056/nejmoa1407279>
- Kikuchi K, Togao O, Yamashita K, Momosaka D, Kikuchi Y, Kuga D, Hata N, Mizoguchi M, Yamamoto H, Iwaki T, Hiwatashi A, Ishigami K (2022) Quantitative relaxometry using synthetic MRI could be better than T2-FLAIR mismatch sign for differentiation of IDH-mutant gliomas: a pilot study. *Sci Rep* 12:9197. <https://doi.org/10.1038/s41598-022-13036-0>
- Jain R, Johnson DR, Patel SH, Castillo M, Smits M, van den Bent MJ, Chi AS, Cahill DP (2020) “Real world” use of a highly reliable imaging sign: “T2-FLAIR mismatch” for identification of IDH mutant astrocytomas. *Neuro Oncol* 22:936–943. <https://doi.org/10.1093/neuonc/noaa041>
- Broen MPG, Smits M, Wijnenga MMJ, Dubbink HJ, Anten MHME, Schijns OEMG, Beckervordersandforth J, Postma AA, van den Bent MJ (2018) The T2-FLAIR mismatch sign as an imaging marker for non-enhancing IDH-mutant, 1p/19q-intact lower-grade glioma: a validation study. *Neuro Oncol* 20:1393–1399. <https://doi.org/10.1093/neuonc/noy048>
- Bhandari AP, Liang R, Koppen J, Murthy SV, Lasocki A (2021) Noninvasive determination of IDH and 1p19q status of lower-grade gliomas using MRI radiomics: a systematic review. *AJNR Am J Neuroradiol* 42:94–101. <https://doi.org/10.3174/ajnr.a6875>
- Kikuchi Y, Togao O, Kikuchi K, Momosaka D, Obara M, Van Cauteren M, Fischer A, Ishigami K, Hiwatashi A (2022) A deep convolutional neural network-based automatic detection of brain metastases with and without blood vessel suppression. *Eur Radiol* 32:2998–3005. <https://doi.org/10.1007/s00330-021-08427-2>
- Kusunoki M, Nakayama T, Nishie A, Yamashita Y, Kikuchi K, Eto M, Oda Y, Ishigami K (2022) A deep learning-based approach for the diagnosis of adrenal adenoma: a new trial using CT. *Br J Radiol* 95:20211066. <https://doi.org/10.1259/bjr.20211066>
- Bousabarah K, Ruge M, Brand J-S, Hoevens M, Rueß D, Borggrefe J, Große Hokamp N, Visser-Vandewalle V, Maintz D, Treuer H, Kocher M (2020) Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data. *Radiat Oncol* 15:87. <https://doi.org/10.1186/s13014-020-01514-6>
- Park YW, Jun Y, Lee Y, Han K, An C, Ahn SS, Hwang D, Lee SK (2021) Robust performance of deep learning for automatic detection and segmentation of brain metastases using three-dimensional black-blood and three-dimensional gradient echo imaging. *Eur Radiol* 31:6686–6695. <https://doi.org/10.1007/s00330-021-07783-3>
- Chen W, Wu J, Wei R, Wu S, Xia C, Wang D, Liu D, Zheng L, Zou T, Li R, Qi X, Zhang X (2022) Improving the diagnosis of acute ischemic stroke on non-contrast CT using deep learning: a multicenter study. *Insights Imaging* 13:184. <https://doi.org/10.1186/s13244-022-01331-3>
- Toyohara Y, Sone K, Noda K, Yoshida K, Kurokawa R, Tanishima T, Kato S, Inui S, Nakai Y, Ishida M, Gonoji W, Tanimoto S, Takahashi Y, Inoue F, Kukita A, Kawata Y, Taguchi A, Furusawa A, Miyamoto Y, Tsukazaki T, Tanikawa M, Iriyama T, Mori-Uchino M, Tsuruga T, Oda K, Yasugi T, Takechi K, Abe O, Osuga Y (2022) Development of a deep learning method for improving

- diagnostic accuracy for uterine sarcoma cases. *Sci Rep* 12:19612. <https://doi.org/10.1038/s41598-022-23064-5>
15. Kinoshita M, Arita H, Takahashi M, Uda T, Fukai J, Ishibashi K, Kijima N, Hirayama R, Sakai M, Arisawa A, Takahashi H, Nakanishi K, Kagawa N, Ichimura K, Kanemura Y, Narita Y, Kishima H (2020) Impact of inversion time for FLAIR acquisition on the T2-FLAIR mismatch detectability for IDH-mutant, non-CODEL astrocytomas. *Front Oncol* 10:596448. <https://doi.org/10.3389/fonc.2020.596448>
 16. Warntjes JBM, Dahlqvist O, Lundberg P (2007) Novel method for rapid, simultaneous T1, T2*, and proton density quantification. *Magn Reson Med* 57:528–537. <https://doi.org/10.1002/mrm.21165>
 17. Patel TR, Pinter N, Sarayi SMMJ, Siddiqui AH, Tutino VM, Rajabzadeh-Oghaz H (2021) Automated cerebral vessel segmentation of magnetic resonance imaging in patients with intracranial atherosclerotic diseases. *Conf Proc IEEE Eng Med Biol Soc* 2021:3920–3923. <https://doi.org/10.1109/embc46164.2021.9630626>
 18. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B (2017) Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 36:61–78. <https://doi.org/10.1016/j.media.2016.10.004>
 19. Fleiss JL, Cohen J (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 33:613–619. <https://doi.org/10.1177/001316447303300309>
 20. Bartko JJ (1966) The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 19:3–11. <https://doi.org/10.2466/pr0.1966.19.1.3>
 21. Deguchi S, Oishi T, Mitsuya K, Kakuda Y, Endo M, Sugino T, Hayashi N (2020) Clinicopathological analysis of T2-FLAIR mismatch sign in lower-grade gliomas. *Sci Rep* 10:10113. <https://doi.org/10.1038/s41598-020-67244-7>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.