From dose reduction to contrast maximization: can deep learning amplify the impact of contrast media on brain MR image quality? A reader study

Alexandre Bône, Samy Ammari, Yves Menu, Corinne Balleyguier, Eric Moulton, Emilie Chouzenoux, Andreas Volk, Gabriel C.T.E. Garcia, François Nicolas, Philippe Robert, et al.

Title

From dose reduction to contrast maximization: can deep learning amplify the impact of contrast media on brain MR image quality? A reader study

Manuscript type: original article.

Authors

- Alexandre Bône[1, *], PhD
- Samy Ammari[2, 3, *], MD, MSc
- Yves Menu[2], MD, PhD
- Corinne Balleyguier[2, 3], MD, PhD
- Eric Moulton[1], PhD
- Émilie Chouzenoux[4], PhD
- Andreas Volk[2, 3], PhD
- Gabriel C.T.E. Garcia[2], MD, MSc
- François Nicolas[1], PhD
- Philippe Robert[1], PhD
- Marc-Michel Rohé[1], PhD
- Nathalie Lassau[2, 3], MD, PhD

(: equal contributions)*

Institutions

1. Guerbet Research, Villepinte, France
2. Imaging Department, Gustave Roussy Cancer Campus, Université Paris-Saclay, Villejuif, France
3. BioMaps (UMR1281), Université Paris-Saclay, CNRS, Inserm, CEA, Villejuif, France
4. Center for Visual Computing, CentraleSupélec, Inria, Université Paris-Saclay, Gif-sur-Yvette, France

Funding information

Research study funded by Guerbet. Among its portfolio of solutions for medical imaging, Guerbet commercializes Dotarem (gadoterate meglumine), a gadolinium-based contrast agent.

Abbreviations

- GBCA = gadolinium-based contrast agent,
- ADC = apparent diffusion coefficient map,
- T1c = contrast-enhanced gradient echo T1,
- T1c-vfa = contrast-enhanced turbo spin echo T1 with variable flip angles,
- T1c-low = low-dose contrast-enhanced gradient echo T1,
- T1c-pro = processed contrast-enhanced gradient echo T1.

Abstract

Objectives

The aim of this study is to evaluate a deep learning method designed to increase the contrast-to-noise ratio in contrast-enhanced gradient echo T1-weighted brain MRI acquisitions. The processed images are quantitatively evaluated in terms of lesion detection performance.

Materials and methods

A total of 250 multiparametric brain MRIs, acquired between November 2019 and March 2021 at Gustave Roussy Cancer Campus (Villejuif, France), were considered for inclusion in this retrospective monocentric study. Independent training (107 cases, age 55y±14, 58 women) and test (79 cases, age 59y±14, 41 women) samples were defined. Patients had glioma, brain metastasis, meningioma, or no enhancing lesion. Gradient echo and turbo spin echo with variable flip angles postcontrast T1 sequences were acquired in all cases. For the cases that formed the training sample, “low-dose” postcontrast gradient echo T1 images using 0.025mmol/kg injections of contrast agent were also acquired. A deep neural network was trained to synthetically enhance the low-dose T1 acquisitions, taking standard-dose T1 MRI as reference. Once trained, the contrast enhancement network was used to process the test gradient echo T1 images. A read was then performed by two experienced neuroradiologists to evaluate the original and processed T1 MRI sequences in terms of contrast enhancement and lesion detection performance, taking the turbo spin echo sequences as reference.

Results

The processed images were superior to the original gradient echo and reference turbo spin echo T1 sequences in terms of contrast-to-noise ratio (44.5 versus 9.1 and 16.8, p<.001), lesion-to-brain ratio (1.66 versus 1.31 and 1.44, p<.001) and contrast enhancement percentage (112.4% versus 85.6% and 92.2%, p<.001) for cases with enhancing lesions. The overall image quality of processed T1 was preferred by both readers (graded 3.4/4 on average versus 2.7/4, p<.001). Finally, the proposed processing improved the average sensitivity of gradient echo T1 MRI from 88% to 96% for lesions larger than 10mm (p=.008*), whereas no difference was found in terms of the false detection rate (0.02/case in both cases, p>.99). The same effect was observed when considering all lesions larger than 5mm: sensitivity increased from 70% to 85% (p<.001*) while false detection rates remained similar (0.04/case versus 0.06/ case, p=.48). With all lesions included regardless of their size, sensitivities were 59% and 75% for original and processed T1 images respectively (p<.001*), and the corresponding false detection rates were 0.05/ case and 0.14/ case (p=.06).

Conclusion

The proposed deep learning method successfully amplified the beneficial effects of contrast agent injection on gradient echo T1 image quality, contrast level, and lesion detection performance. In particular, the sensitivity of the MRI sequence was improved by up to 16%, whereas the false detection rate remained similar.

Key words

Diagnostic imaging, multiparametric MRI, deep learning, image enhancement, contrast media, brain neoplasms, lesion detection.

Introduction

Magnetic resonance imaging (MRI) is the method of choice to detect and assess brain neoplasms (1,2). Over the last decades, the ever-increasing sensitivity of MRI allowed the detection of smaller tumors and thus earlier treatment onsets, ultimately improving patient survival and quality of life (3). Further quality enhancement of MRI is expected to produce similar positive effects and forms an active research field. Among other innovations, intravenous gadolinium-based contrast agents (GBCA) were found particularly powerful to increase MRI sensitivity and are now very commonly administered. The recent evidence of gadolinium deposition in various tissues (4,5) however prevents from simply increasing GBCA doses to further improve MRI diagnosis performance (6–8), as some studies have suggested (9–12).

To improve contrast-enhanced MRI sensitivity without increasing GBCA doses, three research avenues can be identified in the recent literature. First, new GBCAs with improved chemical properties could improve the image contrast without increasing the injected dose of gadolinium (13). Second, novel MRI sequences may replace or complement routine contrast-enhanced T1 sequences. In particular, the turbo spin echo T1 sequence with variable flip angles is more sensitive than its gradient echo counterpart (14–17) and is now recommended for brain tumor imaging (18), although some qualitative limitations were also identified (19). Similarly, a recent proof-of-concept study introduced the T1 relaxation-enhanced steady-state sequence which doubles the tumor-to-brain contrast with respect to spoiled GRE T1 sequence (20). Third, image processing methods may be leveraged to correct artifacts, reduce noise, and improve lesion conspicuity. A growing corpus of deep learning approaches (21,22) is offering varied algorithms ranging from generic MRI noise reduction (23,24) to arterial spin labelling improvement (25) or PET/MR attenuation correction (26). Regarding contrast-enhanced T1 MRI in particular, maximum intensity projection was shown to increase the sensitivity of turbo spin echo sequences (27). A bulk of deep learning methods have also been proposed for low-dose acquisitions, with the aim of converting them into surrogate images for standard-dose sequences (28–32) and reducing the dose of GBCA in neuro-oncology. While small-size lesions are still challenging (30), this stream of studies demonstrates the potential of deep learning for automatic enhancement of MRI contrast. To the best of our knowledge, this potential has never been evaluated for routine full-dose sequences – switching the perspective from GBCA dose reduction to MRI contrast maximization.

This study introduces a deep learning method that amplifies the beneficial effects of GBCA injection on gradient echo brain T1 MRI quality, contrast level, and lesion conspicuity. The image enhancement algorithm is calibrated by training a neural network to predict T1 sequences acquired after standard 0.1mmol/kg injections of GBCA, from the analysis of pre-contrast and 0.025mmol/kg-dose multimodal MRI. Once calibrated, the method is used to process routine contrast-enhanced T1 MRI. The processed images can be understood as a novel synthetic MRI sequence, which are qualitatively and quantitively evaluated in terms of general appearance, contrast enhancement, and lesion detection performance – taking contrast-enhanced turbo spin echo T1 with variable flip angles as reference.

Materials and methods

Data acquisition

Patient inclusion

A total of 250 multiparametric brain MRIs, consecutively acquired between November 2019 and March 2021 at Gustave Roussy Cancer Campus (Villejuif, France) to explore for brain neoplasms, were considered for inclusion in this retrospective monocentric study. Approval for data access was granted by the institutional review board (registration number 2021-18), in accordance with GDPR provisions. Training and test sets were defined using an arbitrary 60/40% split, ensuring that no patient could belong to both sets.

MRI Protocol

Three pre-contrast MRI sequences were acquired from all patients: T1, T2-Flair, and Diffusion Weighted Imaging (DWI) from which ADC maps were generated. Two post-contrast sequences were then acquired, after a 0.1mmol/kg dose of gadoterate meglumine (Dotarem, Guerbet, Villepinte, France) was injected intravenously: gradient echo T1 and turbo spin echo T1 with variable flip angles. These imaging sequences will be respectively noted T1, T2-Flair, ADC, T1c and T1c-vfa, in order of appearance. All were obtained from one of the two machines (General Electric, Milwaukee, US) whose imaging parameters are detailed in Table 1. For all patients in the training sample, the administration of GBCA was split into two successive injections with 0.025mmol/kg and 0.075mmol/kg doses, with an additional "low-dose" gradient echo T1 sequence acquired in between – hereinafter noted T1c-low. Note that two-injection MRI protocols have been recently included in consensus guidelines for glioma imaging, the first injection playing the role of preload bolus for a possible later perfusion sequence (33). To minimize the total gadolinium exposure, a 0.075mmol/kg dose was however chosen for the second injection instead of the suggested 0.1mmol/kg.

Machine	Modality	Sequence	TR	TE	Slice thickness
Optima MR450w 1.5T Installed in 2016, 70cm tunnel, 32 channels, 50cm z-axis FOV, gradients 40mT/m SR 200T/m/s.	T1	3D rapid gradient echo (flip angle 12°)	9ms	4.2ms	1mm
	T2-Flair	Turbo spin echo	7002ms	138ms	1.4mm
	ADC	EPI (two b-values: 0 and 1000 mm^2/s)	3349ms	79ms	4mm
	T1c	3D rapid gradient echo (flip angle 12°)	6.1ms	1.2ms	1mm
	T1c-low				
	T1c-vfa	3D turbo spin echo with variable flip angles	575ms	10ms	1.2mm
Discovery MR750w 3T Installed in 2012, 70cm tunnel, 32 channels, 50cm z-axis FOV, gradients 44mT/m SR 200T/m/s.	T1	3D rapid gradient echo (flip angle 12°)	5.9ms	2.1ms	1mm
	T2-Flair	Turbo spin echo	7002ms	118ms	1mm
	ADC	EPI (two b-values: 0 and 1000 mm^2/s)	5375ms	62.6ms	3mm
	T1c	3D rapid gradient echo (flip angle 12°)	6.1ms	2.1ms	1mm
	T1c-low				
	T1c-vfa	3D turbo spin echo with variable flip angles	575ms	10ms	1.2mm

Table 1: MRI parameters.

Deep learning method

Preprocessing

Figure 1 depicts how the different MRI sequences were leveraged to train and evaluate the deep learning algorithm. The T1c-vfa and T1c-low sequences were not required for the training and test cases, respectively. All other sequences were required for inclusion. Follow-up cases were excluded from the training sample to avoid skewing the calibration of the neural network. The remaining brain images were registered to the ICBM 2009a nonlinear symmetric atlas using the affine deformations of FSL FLIRT v6.0, resampled to the isotropic 1x1x1mm^3 resolution using spline-based interpolation, skull-stripped using HD-BET and center-cropped to the 160x192x160 image size. Brain signals were standardized using z-scores, clipped to the global train-set first and last percentiles, and mapped to the [0, 1] interval. An exhaustive quality control was finally performed to remove the sequences with important artifacts or failed alignment.

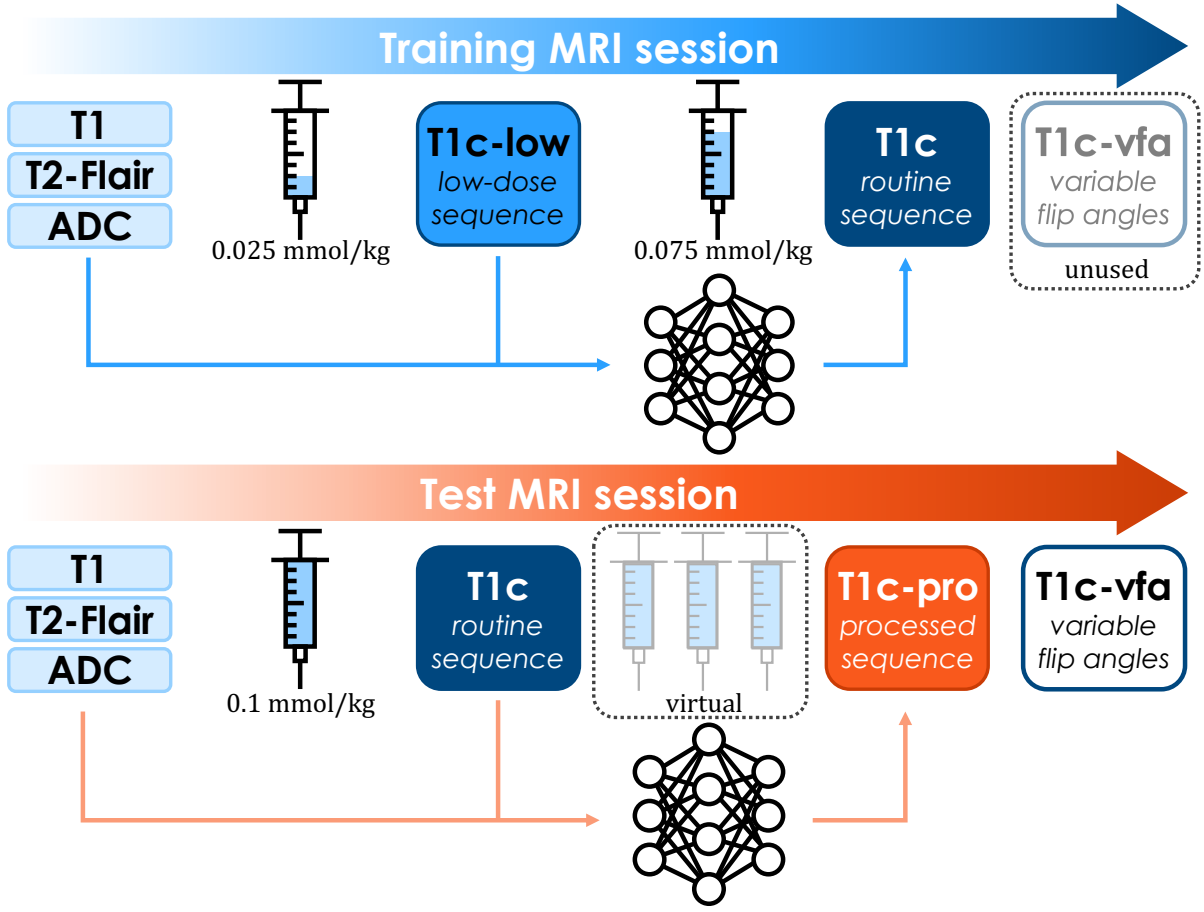

Figure 1: Proposed deep learning paradigm. A deep network is trained to predict standard contrast-enhanced gradient echo T1c sequences from multi-modal pre-contrast and low-dose MRI. Zero-dose inputs are T1, T2-Flair, and apparent diffusion coefficient (ADC) maps computed from diffusion MRI. Low-dose (T1c-low) MRI is acquired after the injection of a 0.025mmol/kg dose of gadolinium-based contrast agent (GBCA). Once trained, the enhancing model is leveraged to amplify the impact of full-dose GBCA injection on routine contrast-enhanced (T1c) MRI by processing them into high-contrast (T1c-pro) images. Their sensitivity for lesion detection is quantitatively evaluated, with respect to reference contrast-enhanced turbo spin echo with variable flip angles (T1c-vfa) sequences.

Deep network architecture

Figure 2 details the deep network architecture, introduced in (34) and validated in (30) for contrast dose minimization. This three-dimensional UNet architecture is composed of an encoding path that alternates residual and downsampling modules, followed by a decoding path that symmetrically alternates residual and upsampling modules (35). Both short and long skip connections are leveraged to facilitate gradient backpropagation through the hierarchical architecture (36); they are respectively based on addition and concatenation operators. Downsampling and upsampling convolution blocks rely on 2x2x2 kernels, all other kernels are 3x3x3 at the exception of the final 1x1x1 convolution. All activation functions but the last sigmoid are 0.2-LeakyReLU.

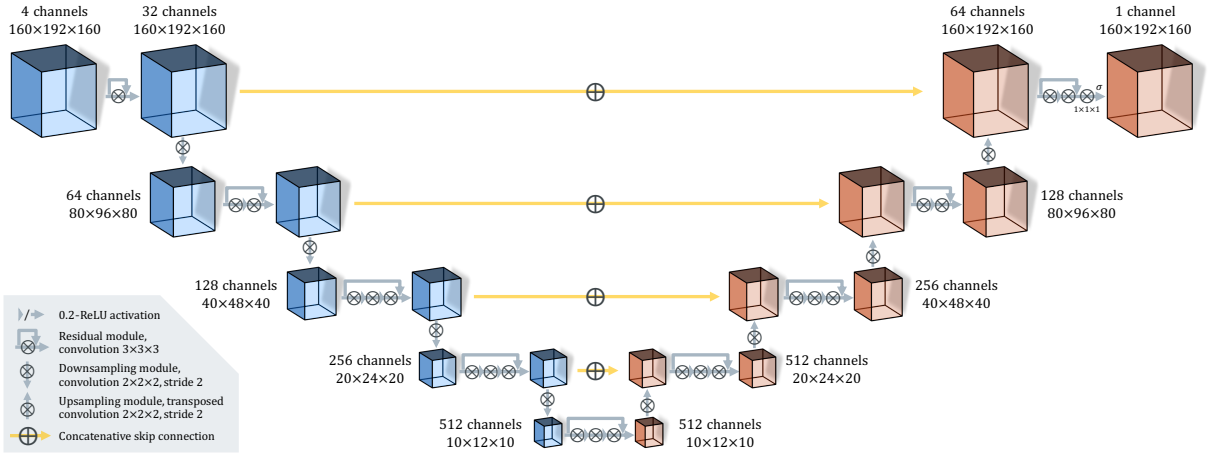

Figure 2: Architecture of the deep network, trained to amplify the impact of gadolinium injection on MRI contrast. The model takes T1, T2-Flair and ADC maps as pre-contrast inputs, plus a fourth contrast-enhanced sequence that is filtered and returned as an output.

Training and evaluating the model

Five independent neural networks were calibrated using imaging data from the training sample, repeatedly split in a five-fold fashion. The T1, T2-Flair, ADC and T1c-low sequences were given as input information. Using 300 Adam epochs with a learning rate of 1/1000 and a mini-batch size of 1, the models were trained to minimize the mean square error between their predictions and the routine T1c sequences. Random mirroring was used for all three axes to provide data augmentation during training.

At test-time, the trained networks processed the T1, T2-Flair, ADC and T1c sequences. Note that the fourth input provided was the routine T1c image, unlike during training where the low-dose T1c-low was supplied (see Figure 1). Hypothesizing that the deep networks primarily learned to amplify the difference of contrast between their pre-contrast and post-contrast inputs, replacing T1c-low by T1c images was expected to trigger the synthesis of approximate quadruple-dose sequences. Exhaustive 8-way mirroring was used at test-time for all five models. For each test case, the forty resulting predictions were averaged to produce the final processed image, that will be noted T1c-pro in the rest of this article.

All computations were performed using a single Azure NC6s workstation, equipped with a Nvidia V100 GPU. Training the deep learning models took approximately 14 hours and 30 minutes for each of the five folds. Predicting the 79 test T1c+ images took 20 minutes and 12 seconds in total, which

corresponds to a mean prediction time of 15 seconds per case. The Python code is publicly available at https://dev.azure.com/alexandrebone/_git/virtual-gadolinium.

Evaluation

Reader study protocol

The original T1c, processed T1c-pro, and reference T1c-vfa images were qualitatively and quantitatively evaluated thanks to a read performed by G.C.G. and S.A., neuroradiologists with 7 and 10 years of experience, respectively. In three successive phases respectively dedicated to the T1c, T1c-pro and T1c-vfa sequences, the readers independently graded the general image quality on 4-point scales and marked the major and minor axes of all visible enhancing lesions. For T1c-vfa sequences, lesion marking was only performed once by S.A. to define a single set of reference annotations. A minimal delay of 10 days was enforced between each reading phase to prevent recall bias. For all reads, the complementary T2-Flair sequences were also available to support the identification of brain lesions. Similarly, the routine T1c sequences were available when reading T1c-pro and T1c-vfa images. Cases with more than 20 lesions were not annotated and were excluded from the downstream statistical analysis. Reading times were automatically recorded by the annotation tool – which was specifically developed for this study to minimize the annotation burden.

Performance metrics

The contrast-to-noise ratio (CNR), lesion-to-brain ratio (LBR), and contrast enhancement percentage (CEP) metrics were computed for each case with at least one reference lesion, according to the formulae:

$$CNR = \frac{SI_{lesion} - SI_{brain}}{SD_{brain}} \quad ; \quad LBR = \frac{SI_{lesion}}{SI_{brain}} \quad ; \quad CEP = \frac{SI_{lesion} - SI_{lesion}^{pre}}{SI_{lesion}^{pre}}$$

where SI_{roi} and SD_{roi} respectively denote the average signal intensity (SI) and its standard deviation (SD) in a given region of interest (ROI). Lesion ROIs were automatically determined as the parallelograms built from the reference long and short axes defined on the T1c-vfa sequences. If a case contained more than one reference lesion, signal averages were computed on all lesion ROIs pooled together. The healthy brain ROIs were reviewed by S.A. after being manually defined by A.B., a computer scientist with 6 years of experience in the field of neuroimaging. Finally, the notation SI_{lesion}^{pre} corresponds to the average lesion intensity in the pre-contrast T1 sequence.

Taking the lesions identified by S.A. on the T1c-vfa sequences as reference, sensitivity (SE) and false detection rate (FDR), were computed for both readers for the original T1c and post-processed T1c+ sequences. The positive predictive value (PPV) and the F1-score (F1), i.e. the harmonic mean between SE and PPV, were also derived as secondary performance metrics. Lesion tracking was automatically computed by pairing lesions whose centers were closer than half the average of their long and short axes. The lesion detection metrics were stratified with respect to three minimum size thresholds: 10mm, 5mm, and 0mm – all lesions being included in the latter case. Lesion size was defined as the length of the long axis.

Statistical analysis

Average image quality grades, contrast enhancement metrics, and reading times were compared across readers and post-contrast MRI sequences using t-tests. The differences in SE were compared using McNemar mid-p test (37) as appropriate for moderate sample sizes of binary matched-pairs

data, and the differences in FDR using the Wilcoxon signed-rank test. Corresponding confidence intervals were determined for SE and FDR using the Clopper-Pearson and Student methods, respectively. All analyses were conducted using SciPy 1.5.2. All tests were two-tailed, and the 5% level was used for confidence intervals and statistical significance.

Results

Training and test samples

As detailed in Figure 3, 107 and 79 cases were finally included in the training and test samples respectively. Table 2 details their demographic and clinical characteristics.

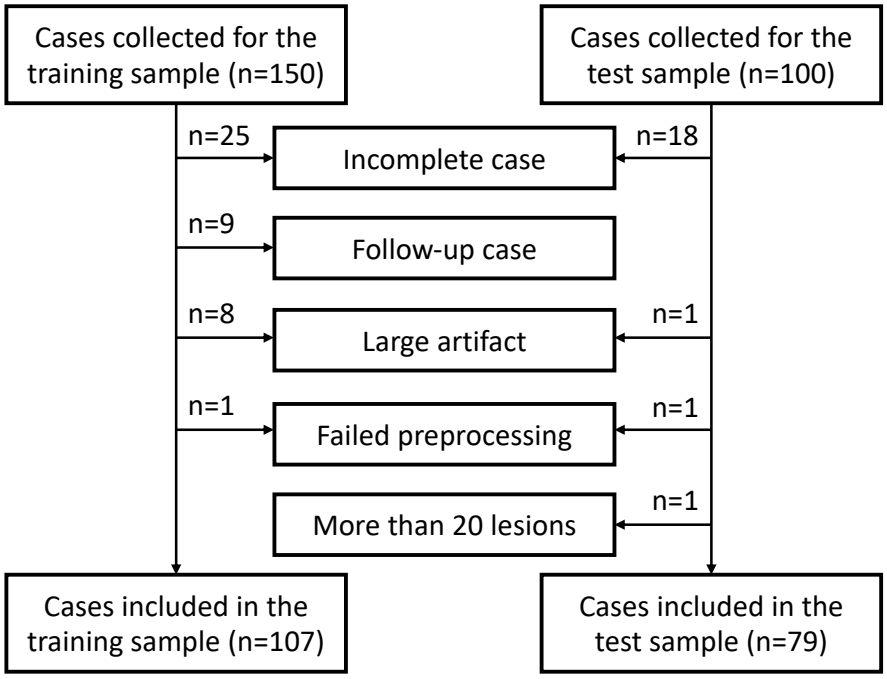

Figure 3: Exclusion flowchart for the training and test samples.

Variable	Training sample	Test sample	*P-value*
Nb. of included cases	107	79	
Women	54% (58/107)	52% (41/79)	*P=.77*
Age	55y (±14)	59y (±14)	*P=.07*
Weight	70kg (±18)	74kg (±16)	*P=.06*
Indication			*P=.18*
Glioma	21% (22/107)	10% (8/79)	
Brain metastases	49% (52/107)	56% (44/79)	
Meningioma	1% (1/107)	0% (0/79)	
No enhancing lesion	29% (32/107)	34% (27/79)	

Table 2: Demographics and baseline characteristics for included cases in training and test samples. Statistical significance of differences between training and test samples was estimated using the Fisher exact test for gender and indication, and the Wilcoxon-Mann-Whitney test for age and weight.

Qualitative inspection

Figure 4 displays the T2-Flair, T1c, T1c-pro and T1c-vfa sequences for four example cases from the test sample, either with brain metastases or glioma, plus one additional case which was excluded from this sample and the downstream statistical analysis because of large imaging artifacts.

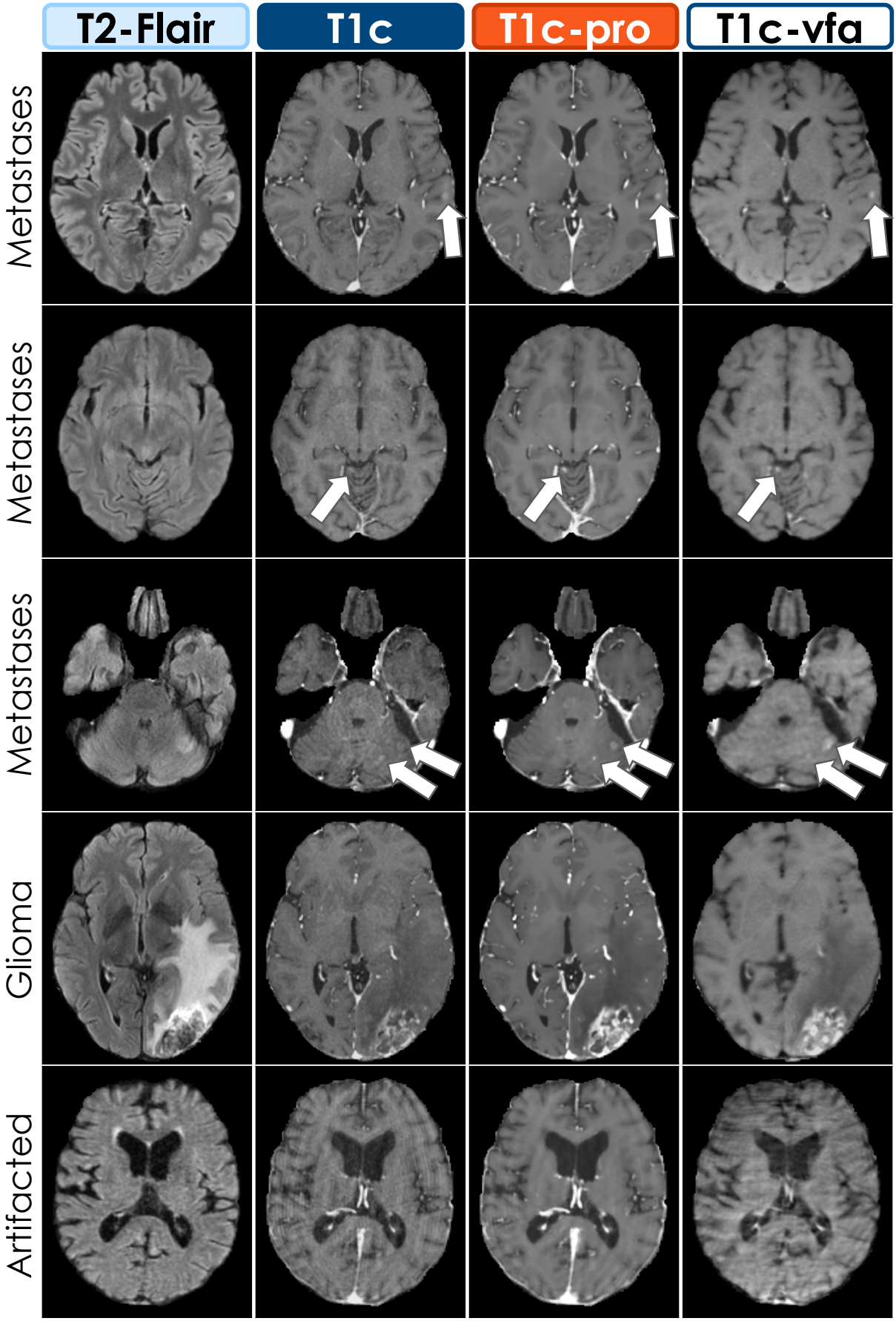
T2-Flair
T1c
T1c-pro
T1c-vfa
Metastases
Metastases
Metastases
Glioma
Artifacted

Figure 4: Axial slices of T2-Flair, T1c, T1c-pro and T1c-vfa images (in columns, from left to right) for five example cases (in rows). The three first rows correspond to patients with brain metastases. These metastases appeared brighter in the T1c-pro images when compared to the T1c sequences from which they are predicted. Brain parenchyma also appeared smoother. These favorable characteristics for lesion detection were shared with the T1c-vfa images. The fourth row corresponds to a patient with glioma, whose internal morphology appeared clearer on the post-processed T1c-pro image than the T1c-vfa sequence. Finally, the last row corresponds to a case which was excluded from the test sample because of its large ring-shaped artifacts. The post-processed T1c-pro image presented artifacts as well, but of a lesser magnitude in comparison with T1c and T1c-vfa sequences.

Image quality grades

As detailed in Table 3, the processed T1c-pro images were preferred by both readers for their general image quality, whereas no difference in quality was found between T1c and T1c-vfa sequences. On average between readers, T1c and T1c-vfa were graded 2.7/4 (average to good), whereas T1c-pro images were graded 3.4/4 (good to excellent).

	IQ ±std (grades on a 4-point scale)			***T1c** vs. **T1c-pro***	***T1c** vs. **T1c-vfa***	***T1c-pro** vs. **T1c-vfa***
	T1c	**T1c-pro**	**T1c-vfa**			
Reader Average	2.7 (±0.5)	**3.4** (±0.3)	2.7 (±0.6)	*P<.001**	*P=.62*	*P<.001**
Reader 1	2.8 (±0.5)	**2.9** (±0.4)	2.7 (±0.6)	*P=.02**	*P=.36*	*P=.007**
Reader 2	2.6 (±0.7)	**3.9** (±0.4)	2.7 (±0.8)	*P<.001**	*P=.14*	*P<.001**
R1 vs. R2	*P=.007**	*P<.001**	*P=.89*			

Table 3: Average image quality (IQ) scores for the 79 cases of the test sample. Grades are expressed on a 4-point Likert scale ranging from 1 (poor) to 4 (excellent). Standard deviations are given between parentheses. Differences across readers and post-contrast MRI sequences are compared using two-tailed t-tests, and p-values are reported. Best metrics are shown in bold when the 5% significance threshold is met.

Contrast enhancement

Table 4 reports the average CNR, LBR and CEP performance metrics for the T1c, T1c-pro and T1c-vfa sequences from the test sample. With an average CNR of 44.5, LBR of 1.66 and CEP of 112.4%, the processed T1c-pro images outperformed both T1c and T1c-vfa sequences for all considered metrics. Similarly, and in alignment with the literature (14), T1c-vfa images reached higher scores than T1c acquisitions, although the difference was not significant for CEP.

	T1c	**T1c-pro**	**T1c-vfa**	***T1c** vs. **T1c-pro***	***T1c** vs. **T1c-vfa***	***T1c-pro** vs. **T1c-vfa***
CNR	9.1 (±5.8)	**44.5** (±25.7)	16.8 (±9.7)	*P<.001**	*P<.001**	*P<.001**
LBR	1.31 (±0.18)	**1.66** (±0.27)	1.44 (±0.24)	*P<.001**	*P<.001**	*P<.001**
CEP	85.6 (±35.5)	**112.4** (±36.5)	92.2 (±38.1)	*P<.001**	*P=.0.11*	*P<.001**

Table 4: Average contrast-to-noise ratio (CNR), lesion-to-brain ratio (LBR), and contrast enhancement percentage (CEP) for the 52 cases of the test sample with at least one reference lesion. Standard deviations are given between parentheses. Differences across post-contrast MRI sequences are compared using two-tailed t-tests, and p-values are reported. Best metrics are shown in bold when the 5% significance threshold is met.

		Reader 1		Reader 2		R1 vs. R2		Reader Average		
		T1c *(& T2-Flair)*	**T1c-pro** *(& T2-Flair, T1c)*	**T1c** *(& T2-Flair)*	**T1c-pro** *(& T2-Flair, T1c)*	**T1c**	**T1c-pro**	**T1c** *(& T2-Flair)*	**T1c-pro** *(& T2-Flair, T1c)*	**T1c** vs. **T1c-pro**
≥0 mm	**SE**	57% (106/187)	74% (138/187)	59% (111/187)	76% (142/187)	*P=.15*	*P=.41*	59% [51; 65]	**75%** [68; 81]	*P<.001**
	FDR	0.04 (3/79)	0.09 (7/79)	0.06 (5/79)	0.19 (15/79)	*P=.32*	*P=.09*	0.05 [±0.04]	0.14 [±0.09]	*P=.06*
	PPV	97% (106/109)	95% (138/145)	96% (111/116)	90% (142/157)	*n/a*		96%	93%	*n/a*
	F1	72%	83%	73%	83%			72%	83%	
≥5 mm	**SE**	69% (100/145)	86% (124/145)	71% (103/145)	85% (123/145)	*P=.34*	*P=.79*	70% [62; 77]	**85%** [78; 91]	*P<.001**
	FDR	0.03 (2/79)	0.04 (3/79)	0.06 (5/79)	0.09 (7/79)	*P=.18*	*P=.10*	0.04 [±0.04]	0.06 [±0.05]	*P=.48*
	PPV	98% (100/102)	98% (124/127)	95% (103/108)	95% (123/130)	*n/a*		97%	96%	*n/a*
	F1	81%	91%	81%	89%			81%	90%	
≥10 mm	**SE**	87% (77/89)	96% (85/89)	89% (79/89)	96% (85/89)	*P=.25*	*P>.99*	88% [79; 94]	**96%** [89; 99]	*P=.008**
	FDR	0.00 (0/79)	0.01 (1/79)	0.04 (3/79)	0.03 (2/79)	*P=.08*	*P=.56*	0.02 [±0.02]	0.02 [±0.02]	*P>.99*
	PPV	100% (77/77)	99% (85/86)	96% (79/82)	98% (85/87)	*n/a*		98%	98%	*n/a*
	F1	93%	97%	92%	97%			93%	97%	

Table 5: Sensitivity (SE, expressed in percentages of the total number of reference lesions), false detection rate (FDR, expressed in average number per case), positive predictive value (PPV, expressed in percentages of the total number of detected lesions), and F1-score (F1, harmonic mean between SE and PPV), in nested evaluation configurations where only lesions larger than a given threshold (of the long axis) are considered. Two-tailed McNemar mid-p and Wilcoxon signed-rank tests were performed for SE and FDR respectively, and p-values are reported. Best metrics are shown in bold when the 5% significance level is met. Clopper-Pearson and Student's t-distribution 95% confidence intervals are also reported for SE and FDR respectively, between brackets.

Lesion detection performance

Table 5 summarizes and compares the average lesion detection performance reached when reading the T1c sequence in addition to when also jointly reading the associated post-processed T1c-pro. In both reading scenarios, the pre-contrast T2-Flair sequence was available to readers as well. A read with access to T2-Flair, T1c and T1c-vfa sequences identified 187 reference lesions with a median long axis length of 9.2mm (inter-quartile range of 10.7mm). Three nested evaluation configurations were considered, depending on the minimum included lesion size: 10mm, 5mm or 0mm – all lesions being considered in the latter case.

The access to T1c-pro images increased the sensitivity for lesion detection for both readers in all evaluation configurations. On average across readers, the overall SE was 75% when T1c-pro images were available (59% with T1c sequences only, P<.001*), 85% for lesions larger than 5mm (70% with T1c, P<.001*), and 96% for lesions larger than 10mm (88% with T1c, P=.008*). No difference was found in terms of FDR, which remained below 0.19/case across readers, reading scenarios, and evaluation configurations. No difference was found between readers for either SE or FDR.

PPV remained higher than 90% in all configurations. On average across readers, quantitative differences in PPV were at most 3% between the two reading scenarios. F1 was higher than 90% across readers and readings for lesions larger than 10mm, higher than 80% for lesions larger than 5mm, and higher 70% when all lesions were included. On average across readers, F1 was higher when T1c-pro was available to readers by a margin that varied between 4% and 11% according to the considered range of lesion sizes.

Reading times

Table 6 details the recorded times spent by readers when identifying and marking lesions on the T1c, T1c-pro, and T1c-vfa sequences. Note that only a single reader (reader 2, S.A.) determined the reference lesions on T1c-vfa acquisitions. No significant difference was found between readers for the T1c and T1c-pro images. The readings dedicated to T1c images were executed faster than the following readings where T1c-pro images were jointly analyzed. On average across readers, 1'48'' were required in the former case, versus 2'46'' in the latter. The reference read based on T1c-vfa acquisitions took longer, 3'03'' in total. Normalizing these reading times by the average number of lesions marked in each sequence (see Table 5), we found that 0.96s/lesion was required for T1c, 1.10s/lesion for T1c-pro, and 0.98s/lesion for T1c-vfa. In other words, the additional burden of reading T1c-pro (respectively T1-vfa) sequences translated to an extra 140 (respectively 20) milliseconds spent on each lesion.

	RT ±std (minutes' and seconds'')			***T1c*** *vs.* ***T1c-pro***	***T1c*** *vs.* ***T1c-vfa***	***T1c-pro*** *vs.* ***T1c-vfa***
	T1c *(& T2-Flair)*	**T1c-pro** *(& T2-Flair, T1c)*	**T1c-vfa** *(& T2-Flair, T1c)*			
Reader Average	**1'48''** (±1'25'')	2'46'' (±2'01'')	n/a	*P<.001**	*n/a*	*n/a*
Reader 1	**1'32''** (±1'35'')	2'53'' (±2'09'')	n/a	*P<.001**	*n/a*	*n/a*
Reader 2	**2'05''** (±2'03'')	2'39'' (±2'43'')	3'03'' (±2'31'')	*P=.15*	*P=.01**	*P=.36*
R1 vs. R2	*P=.07*	*P=.44*				

Table 6: Average reading times (RT) for the 79 cases from the test sample. Delays are expressed in minutes (') and seconds (''). Standard deviations are given between parentheses. Differences across

readers and post-contrast MRI sequences are compared using two-tailed t-tests, and p-values are reported. Best metrics shown in bold if the 5% significance threshold is met.

Discussion

Focusing on neuro-oncology and contrast-enhanced MRI, our study proposed a deep learning method that amplifies the impact of gadolinium injection on MRI contrast, lesion conspicuity, and sensitivity for tumor detection. Taking standard contrast-enhanced T1c acquisitions as training targets, the deep network learned to enhance low-dose T1c-low sequences. On a separate test sample, the enhancing model was then leveraged to process clinical-routine T1c acquisitions. The processed T1c-pro images were qualitatively and quantitatively evaluated, in a reading study, against the original T1c and complementary turbo spin echo with variable flip angles T1c-vfa sequences.

The processed T1c-pro images were preferred to both the original T1c acquisitions and the reference T1c-vfa sequences by both readers. On average across cases with at least one reference lesion, the original T1c contrast-to-noise (CNR) ratio was 9.1. This average CNR was increased to 44.5 for T1c-pro images with our method. For reference, the CNR of T1c-vfa was 16.8. Similar improvements were also found in terms of lesion-to-brain ratio (LBR) and contrast enhancement percentage (CEP), which were respectively raised from 1.31 to 1.66 and from 85.6% to 112.4%, demonstrating the contrast amplification effect of the proposed algorithm. For the reference T1c-vfa sequences, LBR was 1.44 and CEP was 92.2% on average.

Beyond their refined aspect and augmented contrast, the processed T1c-pro images were found to significantly improve the sensitivity for lesion detection for T1c acquisitions, which increased from 59% to 75% on average across readers. This sensitivity improved from 70% to 85% when considering lesions larger than 5mm, and from 88% to 96% for lesions larger than 10mm. While the false detection rate (FDR) was slightly more elevated with T1c-pro images, reaching 0.14/case versus 0.05/case with T1c sequences, this difference was not statistically significant, and the total amount of false positives remained modest. Lesions identified with T1c-pro images had a positive predictive value (PPV) of 93% on average across readers, and the corresponding F1-score was 83%.

This substantial gain in lesion detection performance came however at the cost of longer reading times. Reading T1c-pro images in addition increased the mean duration of lesion marking by 81 seconds for reader 1, 34 seconds for reader 2, corresponding to a significant increase of 58 seconds from 1’48’’ to 2’46’’ on average across readers. Considering that more lesions were found thanks to T1c-pro images, we computed that this difference in total reading times translated to an extra 140 milliseconds spent on each lesion. We may hypothesize that this increase in reading time is partly explained by the lack of experience of readers with the non-standard T1c-pro images, whose general smooth aspect was found qualitatively quite different from the T1c and T1c-vfa sequences.

From a methodological point of view, deep learning was exploited in an original manner to build the proposed contrast enhancement algorithm. Trained on a low-to-normal contrast prediction task, the deep network became a normal-to-high contrast enhancement method at inference time. We believe that this novel approach, where the learned model is deflected from its training configuration to fill a more creative purpose than the mere reproduction of some reference targets, could be useful to tackle other medical imaging challenges such as super-resolution or denoising, and forms a key contribution of this paper. As an exemplar future research lead, we may evaluate whether this

contrast amplification method could be advantageously applied to the same base image not only once, but repeatedly.

In the absence of self-declared reference images, validating the deep network predictions required to resort to a third-party MRI sequence: the T1c-vfa acquisitions from which the reference lesions were defined. This choice was motivated by the prolonged injection-to-imaging delay and the high sensitivity of T1c-vfa MRI for tumor detection, demonstrated in recent studies (14–18). Yet this reference remains somewhat arbitrary. It could be argued that higher-quality reference lesions would be based on histological samples or, more pragmatically, longitudinal follow-ups. We claim nonetheless that the chosen reference was sufficient to demonstrate the superior sensibility of T1c-pro images with respect to standard T1c MRI in this proof-of-concept study.

Choosing T1c-vfa MRI as a reference for lesion detection naturally raises the question: should this sequence supersede classical T1c MRI and all its post-processed offspring altogether? This question is far from being naive, as recent works have indeed suggested that T1c-vfa MRI may soon become a new standard for neuroimaging (18). However, three main lines of answer can be opposed. First, this T1c-vfa sequence typically requires a longer acquisition time with respect to T1c MRI, leading to higher costs and more frequent motion artifacts – explaining its limited adoption in most medical institutions. Second, several publications have pointed out that T1c and T1c-vfa sequences have complementary imaging qualities and should not be systematically opposed (19). In particular, the grey/white matter interface has been consistently described as more clearly visible in T1c MRI, which is of critical importance for surgery or radiotherapy planning (38). Finally, we argue that future work may straightforwardly adapt our proof-of-concept study to develop a post-processing algorithm dedicated to T1c-vfa acquisitions.

More generally, future work will evaluate the potential of our methodology for other contrast-enhanced MRI sequences, in neuroimaging or beyond. Image enhancement algorithms could be similarly developed for CT imaging, radiation level playing the role of gadolinium dose. Finally, although our monocentric retrospective study showed that the proposed deep learning method successfully amplified the beneficial effects of contrast agent injection on gradient echo T1 MRI, improving its sensitivity by up to 16%, a prospective multicentric follow-up study will be necessary to confirm the results.

References

1. Weller M, Wick W, Aldape K, et al. Glioma. Nat Rev Dis Prim. 2015;1(July). doi: 10.1038/nrdp.2015.17.

2. Achrol AS, Rennert RC, Anders C, et al. Brain metastases. Nat Rev Dis Prim. Springer US; 2019;5(1). doi: 10.1038/s41572-018-0055-y.

3. Arnold M, Rutherford MJ, Bardot A, et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. Lancet Oncol. Elsevier; 2019;20(11):1493–1505. doi: 10.1016/S1470-2045(19)30456-5.

4. Bjørnerud A, Vatnehol SAS, Larsson C, Due-Tønnessen P, Hol PK, Groote IR. Signal enhancement of the dentate nucleus at unenhanced MR imaging after very high cumulative doses of the macrocyclic gadolinium-based contrast agent gadobutrol: An observational study. Radiology.

Radiology; 2017;285(2):434–444. doi: 10.1148/radiol.2017170391.

5. Ramalho J, Semelka RC, Ramalho M, Nunes RH, AlObaidy M, Castillo M. Gadolinium-based contrast agent accumulation and toxicity: An update. Am J Neuroradiol. American Journal of Neuroradiology; 2016;37(7):1192–1198. doi: 10.3174/ajnr.A4615.

6. Fraum TJ, Ludwig DR, Bashir MR, Fowler KJ. Gadolinium-based contrast agents: A comprehensive risk assessment. J Magn Reson Imaging. John Wiley & Sons, Ltd; 2017;46(2):338–353. doi: 10.1002/jmri.25625.

7. Medicines Agency E. PRAC confirms restrictions on the use of linear gadolinium agents. Eur. Med. Agency. 2017. p. Document Reference: EMA/625317/2017. www.ema.europa.eu/contact.

8. United States Food and Drug Administration. FDA warns that gadolinium-based contrast agents (GBCAs) are retained in the body; requires new class warnings [Press release]. Fda. 2018. p. Document Reference: 109825. https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-fda-warns-gadolinium-based-contrast-agents-gbcas-are-retained-body. Accessed February 25, 2021.

9. Yuh WTC, Engelken JD, Muhonen MG, Mayr NA, Fisher DJ, Ehrhardt JC. Experience with high-dose gadolinium MR imaging in the evaluation of brain metastases. Am J Neuroradiol. American Society of Neuroradiology; 1992;13(1):335–345. /pmc/articles/PMC8331800/?report=abstract. Accessed August 25, 2021.

10. Åkeson P, Larsson EM, Kristoffersen DT, Jonsson E, Holtås S. Brain metastases — comparison of gadodiamide injection-enhanced MR imaging at standard and high dose, contrast-enhanced CT and non-contrast-enhanced MR imaging. Acta radiol. SAGE PublicationsSage UK: London, England; 1995;36(3):300–306. doi: 10.1177/028418519503600318.

11. Essig M, Anzalone N, Combs SE, et al. MR imaging of neoplastic central nervous system lesions: Review and recommendations for current practice. Am J Neuroradiol. American Society of Neuroradiology; 2012;33(5):803–817. doi: 10.3174/ajnr.A2640.

12. Kim ES, Chang JH, Choi HS, Kim J, Lee SK. Diagnostic yield of double-dose gadobutrol in the detection of brain metastasis: Intraindividual comparison with double-dose gadopentetate dimeglumine. Am J Neuroradiol. AJNR Am J Neuroradiol; 2010;31(6):1055–1058. doi: 10.3174/ajnr.A2010.

13. Bendszus M, Roberts D, Kolumban B, et al. Dose Finding Study of Gadopiclenol, a New Macrocyclic Contrast Agent, in MRI of Central Nervous System. Invest Radiol. Lippincott Williams and Wilkins; 2020;55(3):129–137. doi: 10.1097/RLI.0000000000000624.

14. Park J, Kim J, Yoo E, Lee H, Chang JH, Kim EY. Detection of small metastatic brain tumors: Comparison of 3D contrast-enhanced whole-brain black-blood imaging and MP-RAGE imaging. Invest Radiol. 2012;47(2):136–141. doi: 10.1097/RLI.0b013e3182319704.

15. Majigsuren M, Abe T, Kageji T, et al. Comparison of brain tumor contrast-enhancement on t1-cube and 3d-spgr images. Magn Reson Med Sci. Magn Reson Med Sci; 2016;15(1):34–40. doi: 10.2463/mrms.2014-0129.

16. Oh J, Choi SH, Lee E, et al. Application of 3D fast spin-echo T1 black-blood imaging in the diagnosis and prognostic prediction of patients with leptomeningeal carcinomatosis. Am J Neuroradiol. AJNR Am J Neuroradiol; 2018;39(8):1453–1459. doi: 10.3174/ajnr.A5721.

17. Suh CH, Jung SC, Kim KW, Pyo J. The detectability of brain metastases using contrast-enhanced spin-echo or gradient-echo images: a systematic review and meta-analysis. J Neurooncol. J Neurooncol; 2016;129(2):363–371. doi: 10.1007/s11060-016-2185-y.

18. Kaufmann TJ, Smits M, Boxerman J, et al. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases. Neuro Oncol. Neuro Oncol; 2020;22(6):757–772. doi: 10.1093/neuonc/noaa030.

19. Bapst B, Amegnizin JL, Vignaud A, et al. Post-contrast 3D T1-weighted TSE MR sequences (SPACE, CUBE, VISTA/BRAINVIEW, isoFSE, 3D MVOX): Technical aspects and clinical applications. J Neuroradiol. Elsevier Masson SAS; 2020;47(5):358–368. doi: 10.1016/j.neurad.2020.01.085.

20. Edelman R, Leloudas N, Pang J, Bailes J, Merrell R, Koktzoglou I. Twofold improved tumor-to-brain contrast using a novel T1 relaxation-enhanced steady-state (T1RESS) MRI technique. Sci Adv. American Association for the Advancement of Science; 2020;6(44):1635. doi: 10.1126/sciadv.abd1635.

21. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. Jpn J Radiol. Jpn J Radiol; 2019;37(1):73–80. doi: 10.1007/s11604-018-0796-2.

22. Zhu G, Jiang B, Tong L, Xie Y, Zaharchuk G, Wintermark M. Applications of deep learning to neuro-imaging techniques. Front Neurol. Frontiers Media SA; 2019;10(AUG):869. doi: 10.3389/fneur.2019.00869.

23. Jiang D, Dou W, Vosters L, Xu X, Sun Y, Tan T. Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network. Jpn J Radiol. Jpn J Radiol; 2018;36(9):566–574. doi: 10.1007/s11604-018-0758-8.

24. Gondara L. Medical Image Denoising Using Convolutional Denoising Autoencoders. IEEE Int Conf Data Min Work ICDMW. IEEE Computer Society; 2016;0:241–246. doi: 10.1109/ICDMW.2016.0041.

25. Kim KH, Choi SH, Park SH. Improving arterial spin labeling by using deep learning. Radiology. Radiology; 2018;287(2):658–666. doi: 10.1148/radiol.2017171154.

26. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep learning MR imaging-based attenuation correction for PET/MR imaging. Radiology. Radiology; 2018;286(2):676–684. doi: 10.1148/radiol.2017170700.

27. Yoon BC, Saad AF, Rezaii P, Wintermark M, Zaharchuk G, Iv M. Evaluation of thick-slab overlapping MIP images of contrast-enhanced 3D T1-weighted CUBE for detection of intracranial metastases: A pilot study for comparison of lesion detection, interpretation time, and sensitivity with nonoverlapping CUBE MIP, CUBE, a. Am J Neuroradiol. American Journal of Neuroradiology; 2018;39(9):1635–1642. doi: 10.3174/ajnr.A5747.

28. Pasumarthi S, Tamir JI, Christensen S, Zaharchuk G, Zhang T, Gong E. A generic deep learning model for reduced gadolinium dose in contrast-enhanced brain MRI. Magn Reson Med. Wiley; 2021;86(3):1687–1700. doi: 10.1002/mrm.28808.

29. Luo H, Zhang T, Gong NJ, et al. Deep learning–based methods may minimize GBCA dosage in brain MRI. Eur Radiol. Springer Science and Business Media Deutschland GmbH; 2021;31(9):6419–6428. doi: 10.1007/s00330-021-07848-3.

30. Ammari S, Bône A, Balleyguier C, et al. Can Deep Learning Replace Gadolinium in Neuro-Oncology? Invest Radiol. 2021;Publish Ah(00):1–9. doi: 10.1097/rli.0000000000000811.

31. Kleesiek J, Morshuis JN, Isensee F, et al. Can Virtual Contrast Enhancement in Brain MRI Replace Gadolinium?: A Feasibility Study. Invest Radiol. LWW; 2019;54(10):653–660. doi: 10.1097/RLI.0000000000000583.

32. Gong E, Pauly JM, Wintermark M, Zaharchuk G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J Magn Reson Imaging. 2018;48(2):330–340. doi: 10.1002/jmri.25970.

33. Thust SC, Heiland S, Falini A, et al. Glioma imaging in Europe: A survey of 220 centres and recommendations for best clinical practice. Eur Radiol. Springer; 2018;28(8):3306–3317. doi: 10.1007/s00330-018-5314-5.

34. Bone A, Ammari S, Lamarque JP, et al. Contrast-enhanced brain MRI synthesis with deep learning: Key input modalities and asymptotic performance. Proc - Int Symp Biomed Imaging. 2021;2021-April:1159–1163. doi: 10.1109/ISBI48211.2021.9434029.

35. Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. Proc - 2016 4th Int Conf 3D Vision, 3DV 2016. 2016. p. 565–571. doi: 10.1109/3DV.2016.79.

36. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). Springer, Cham; 2016;10008 LNCS:179–187. doi: 10.1007/978-3-319-46976-8_19.

37. Fagerland MW, Lydersen S, Laake P. The McNemar test for binary matched-pairs data: Mid-p and asymptotic are better than exact conditional. BMC Med Res Methodol. BioMed Central; 2013;13(1):1–8. doi: 10.1186/1471-2288-13-91.

38. Scoccianti S, Detti B, Gadda D, et al. Organs at risk in the brain and their dose-constraints in adults and in children: A radiation oncologist's guide for delineation in everyday practice. Radiother Oncol. Radiother Oncol; 2015;114(2):230–238. doi: 10.1016/j.radonc.2015.01.016.