



## OPEN ACCESS

### EDITED BY

Ying Xie,  
Kennesaw State University, United States

### REVIEWED BY

Mahdi Bashiri Bawil,  
Sahand University of Technology, Iran  
Dinesh Chowdary Attota,  
Kennesaw State University—Marietta  
Campus, United States

### \*CORRESPONDENCE

Yunyan Zhang  
✉ yunyzhan@ucalgary.ca

RECEIVED 08 December 2025

REVISED 21 February 2026

ACCEPTED 24 February 2026

PUBLISHED 12 March 2026

### CITATION

Shahrokhi S, Oladosu O, Tariq R and  
Zhang Y (2026) CycleGAN models show  
consistent brain MRI synthesis across  
datasets supporting downstream tissue  
characterization in multiple sclerosis.  
*Front. Neuroinform.* 20:1762794.  
doi: 10.3389/fninf.2026.1762794

### COPYRIGHT

© 2026 Shahrokhi, Oladosu, Tariq and  
Zhang. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance  
with accepted academic practice. No  
use, distribution or reproduction is  
permitted which does not comply with  
these terms.

# CycleGAN models show consistent brain MRI synthesis across datasets supporting downstream tissue characterization in multiple sclerosis

Shayan Shahrokhi<sup>1,2</sup>, Olayinka Oladosu<sup>2,3</sup>, Rehman Tariq<sup>4</sup> and  
Yunyan Zhang<sup>2,3,5\*</sup>

<sup>1</sup>Neuroscience Graduate Program, Faculty of Graduate Studies, University of Calgary, Calgary, AB, Canada, <sup>2</sup>Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada, <sup>3</sup>Department of Radiology, University of Calgary, Calgary, AB, Canada, <sup>4</sup>Biomedical Engineering Graduate Program, Faculty of Graduate Studies, University of Calgary, Calgary, AB, Canada, <sup>5</sup>Department of Clinical Neurosciences, University of Calgary, Calgary, AB, Canada

**Background:** Secondary quantitative analysis of brain magnetic resonance imaging (MRI) can provide valuable information for many neurological diseases, including multiple sclerosis (MS), but it demands complete datasets that are often unavailable clinically. We investigated how image synthesis via deep learning using cycle-consistent generative adversarial networks (CycleGANs) compared with Pix2Pix as a related method, based on T1-weighted and T2-weighted brain MRI in MS, following verification on two streamlined datasets. The synthesized images were also evaluated against the source data.

**Methods:** The streamlined datasets involved 1,113 healthy participants from the Human Connectome Project (HCP) and 318 participants from the Parkinson's Progression Markers Initiative (PPMI). The MS cohort in this study included 105 participants scanned with different protocols. Image synthesis was bidirectional between T1- and T2-weighted MRI using CycleGAN with and without spectral normalization, as well as Pix2Pix. Utility testing focused on T1-weighted MRI that was most often unavailable in MS, and that involved lesion detection, brain volumetry, and lesion texture analysis.

**Results:** All CycleGAN models performed competitively, while Pix2Pix performed better, mostly with streamlined datasets ( $p < 0.001$ ). The average peak signal-to-noise ratio ranged from 24.860–28.570 versus 28.520–31.100, and the structural similarity index ranged from 0.838–0.901 versus 0.924–0.943. With spectral normalization, CycleGAN improved in PPMI but not in HCP and generally not in MS ( $p < 0.001$ ). Furthermore, the synthesized images showed high similarity to the source data in utility tests, although Pix2Pix T1 images appeared more heterogeneous in lesion texture than source T1 images.

**Conclusion:** CycleGAN without spectral normalization appeared feasible for synthesizing common clinical brain MRI, including T1-weighted images usable for subsequent quantitative analysis in MS.

### KEYWORDS

brain MRI, CycleGAN, image synthesis, multiple sclerosis, Pix2Pix, spectral normalization

## 1 Introduction

Magnetic resonance imaging (MRI) is an essential tool for the evaluation and management of neurological diseases such as multiple sclerosis (MS) (Thompson et al., 2018). Clinical MS imaging, including T1-weighted (T1w) and T2-weighted (T2w) brain MRI, provides valuable diagnostic information, such as the number and activity of focal lesions (McGinley et al., 2021). However, these features reflect only part of MS pathology (Fisniku et al., 2008). Secondary analysis of clinical MRI using computational approaches has the potential to characterize both visible and invisible tissue changes, thereby generating competitive data-driven hypotheses (Vázquez-Marrufo et al., 2023). Nonetheless, such approaches demand complete datasets, which are often limited in a clinical setting due to constraints in time or cost. There is a clear need to investigate new methods, such as deep learning techniques, that can synthesize missing imaging sequences from available data.

One common deep learning method for image synthesis is the generative adversarial network (GAN) (Goodfellow et al., 2020; Xu et al., 2022). GANs work by combining two neural networks, namely a generator and a discriminator, which create new images and determine their validity, respectively. These methods have shown promise in different applications, including synthesizing brain MRI to support segmentation of brain tumors (Nema et al., 2020) and MS lesions (Zhang et al., 2018). However, base models of GANs are not without limitations, as demonstrated by training instability in their discriminators due to inefficient learning, especially in cross-domain image translation (Lee and Choi, 2020). Developments are underway to overcome these challenges, including applying weight clipping (Arjovsky et al., 2017) or spectral normalization (weight control) (Lin et al., 2021) to the discriminators, and developing new architectures such as conditional GAN.

Two recognized conditional GANs are Pix2Pix and cycle-consistent GAN (CycleGAN). Pix2Pix targets paired image synthesis and is trained unidirectionally (Isola et al., 2017). CycleGAN, on the other hand, can handle unpaired images, enabled by the inclusion of two pairs of generators and discriminators (Zhu et al., 2017). Various studies demonstrate the utility of these models. With brain MRI of MS, Pix2Pix has assisted in image sequence translation and enhancement for cortical (Finck et al., 2022) and white matter lesion segmentation (La Rosa et al., 2021). CycleGAN has demonstrated competence in creating pseudo-healthy and lesion images (Basaran et al., 2022), and in adapting MRI between centers for segmenting white matter MS lesions (Julián Alberto et al., 2020). Furthermore, spectral normalization has demonstrated value in stabilizing CycleGAN in cross-modality image translation from computed tomography (CT) to MRI (Xu et al., 2020). However, current studies focus mainly on synthesizing images for tissue segmentation rather than structural analysis. Additionally, few studies have investigated synthesizing frequently non-acquired sequences, such as T1w brain MRI in MS, using routine clinical care data. Routine MRI is typically more heterogeneous than streamlined datasets in both scanner type and acquisition parameters (Eche et al., 2021), requiring extra attention.

The purposes of this study were threefold. The first was to investigate how CycleGAN models trained using large, streamlined datasets compared with our heterogeneous clinical MS data in image synthesis between T1w and T2w MRI. The second was to evaluate whether and how spectral normalization impacted CycleGAN performance. The third was to explore the utility of synthesized images versus source

images in different analysis tasks, focusing on T1w MRI in MS. In contemporary MS protocols, T1w MRI is not always acquired in routine diagnostic or follow-up examinations (Wattjes et al., 2021); however, many image processing pipelines require a T1w sequence, as seen in co-registration and brain volume measurement (Jenkinson et al., 2012; Fischl, 2012). T2w brain MRI is commonly available in MS protocols (Wattjes et al., 2021), but the acquisition parameters often vary across scanners. Throughout the process, a Pix2Pix model was studied for comparison.

## 2 Methods

### 2.1 Data characteristics

Two publicly available datasets were included, with each acquired using streamlined protocols. One involved healthy brain MRIs from 1,113 participants (aged 22–36 years, 606 women) in the Human Connectome Project (HCP) (Van Essen et al., 2012). The other comprised brain MRIs from 318 participants (aged 60–70 years, 125 women) in the Parkinson's Progression Markers Initiative (PPMI) (Marek et al., 2011). Data utilization followed all required practices.

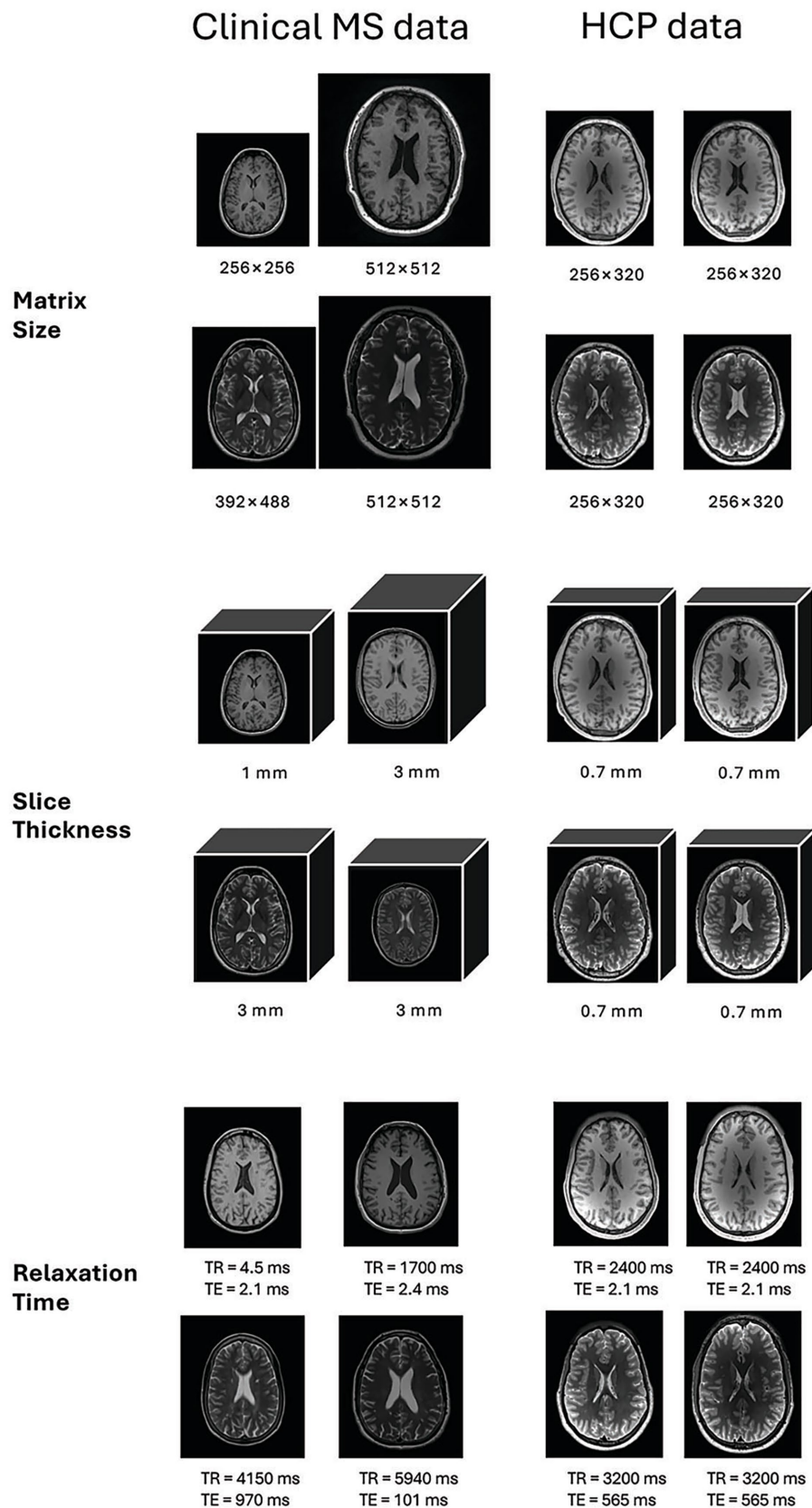
Our local dataset involved clinical brain MRI from a sample of 169 participants with relapsing–remitting MS, who were enrolled in an ongoing cohort study known as Clinical Impact of Multiple Sclerosis (Ethics ID: REB14-1926). Of these, 104 participants (73 women) who had both T1w and T2w sequences available were selected. The mean  $\pm$  standard deviation age was  $36.38 \pm 8.89$  years, disease duration was  $4.81 \pm 6.88$  years, and the median Expanded Disability Status Scale (EDSS) was 2.0. This study was approved by the Institutional Ethics Review Board, with written informed consent obtained from each participant.

### 2.2 Imaging protocol

The HCP scans were acquired with a 3T scanner (Magnetom Skyra, Siemens Healthineers, Erlangen, Germany). T1w imaging used a magnetization-prepared rapid gradient-echo (MPRAGE) sequence with a repetition/echo time (TR/TE) of 2,400 ms/2.14 ms, slice thickness of 0.7 mm, field of view (FOV) of 224 mm, and a matrix size of  $256 \times 320$ . T2w MRI used the Sampling Perfection with Application optimized Contrasts using different flip-angle evolutions (SPACE) sequence with TR/TE of 3,200 ms/565 ms, slice thickness of 0.7 mm, FOV of 224 mm, and the same matrix size.

The PPMI images were acquired using 1.5T and 3T Siemens scanners, where only baseline scans were used. T1w MRI used an MPRAGE sequence with a TR/TE of 3,000 ms/2.98 ms, slice thickness of 1 mm, FOV of 256 mm, and a matrix size of  $240 \times 256$ . T2w MRI used a turbo spin-echo sequence with a TR/TE of 3,000 ms/101 ms, slice thickness of 3 mm, FOV of 256 mm, and a matrix size of  $240 \times 256$ .

Our MS imaging involved both Siemens and GE scanners at 1.5T or 3T. T1w MRI used either an MPRAGE or fast spoiled gradient-echo sequence, with TR of 4.3–2,490 ms, TE of 2.08–33.96 ms, slice thickness of 1–3 mm, and matrix size of  $192 \times 256$  to  $512 \times 512$ . T2w MRI applied mainly a spin-echo sequence, with TR of 2,970–11,470 ms, TE of 83–122.90 ms, slice thickness of 3 mm, and matrix size of  $256 \times 256$  to  $512 \times 512$  (Figure 1).



**FIGURE 1**  
 Example brain MRI protocols of high and low heterogeneity from clinically standard MS (left 2 columns) and streamlined HCP datasets (right 2 columns), respectively. Shown are T1-weighted (top rows) and T2-weighted (bottom rows) images associated with three common protocol settings: matrix size, slice thickness, and repetition and echo times (TR & TE) that can impact image size, resolution, and anatomical contrast. Under relaxation time, the T2-weighted MRIs of clinical MS also show different matrix sizes of 336 × 384 and 392 × 488, while the HCP images have the same image size.

## 2.3 Image preprocessing

Four steps were performed to optimize image quality. These included nonuniformity correction using the N4ITK algorithm of Advanced Normalization Tools (ANTs) (Tustison et al., 2010) and brain extraction using FSL (FMRIB Software Library, Oxford, UK). All scans were then linearly co-registered to the standard MNI152 1-mm T1 template using ANTs. Finally, image intensity was normalized to the range 0–1 for subsequent modeling.

## 2.4 Image synthesis using CycleGAN

### 2.4.1 Model development for Vanilla CycleGAN

We started from a recognized two-dimensional (2D) CycleGAN architecture (Figure 2) (Zhu et al., 2017). Refinement was performed on hyperparameters that often largely impact model stability and output, including the number of down-/up-sampling blocks and the number of residual blocks in the generator, the down-sampling blocks in the discriminator, as well as the initial learning rate and the decay schedule. Candidate settings were chosen based on measures of validation image quality by peak signal-to-noise-ratio (PSNR) and structural similarity index measure (SSIM) across training epochs (Supplementary Figure 1). Accordingly, each model was trained for 200 epochs, where the metrics plateaued, with a batch size of 1 as commonly done in CycleGAN. The initial learning rate was set at 0.0002, which was reduced by a factor of 10 at the 100th epoch. The Adam Optimizer (Kingma and Ba, 2014) was used to minimize loss functions associated with the generators and discriminators. Model training started with individual images at a dimension of  $256 \times 256 \times 1$ , which was encoded to  $16 \times 16 \times 512$  and then gradually decreased to the initial dimension with the generators. The discriminators started with this dimension and decreased to  $16 \times 16 \times 1$  eventually. These settings were used in the analysis of all three datasets investigated in this study to ensure consistency. For each study cohort, the prepared data were randomly split at 0.8/0.1/0.1 for training, validation, and testing, respectively, with splitting randomness controlled by a random seed of 42. All training procedures were conducted using an Nvidia A100 80 GB GPU unit with the TensorFlow backend (Abadi et al., 2016).

### 2.4.2 Model development for CycleGAN with spectral normalization

These models were developed using the same procedures as vanilla CycleGAN, except for the addition of spectral normalization. The latter was added to the convolutional layers of the discriminators by scaling the weight matrices by their maximum singular values (the spectral norms) (Jenkinson et al., 2012). Weight matrices normalized this way were expected to control disproportionate feature amplifications for consistent backpropagation of gradients in the discriminators. To further understand the impact of spectral normalization, imaging contrast variations with or without its addition were also explored based on the healthy HCP data.

## 2.5 Image synthesis with Pix2Pix

The Pix2Pix model was implemented based on a published article (Isola et al., 2016). In contrast to CycleGAN, this method

was composed of one pair of generator and discriminator, with the hyperparameters decided based on established practices for GANs and empirical tuning. Specifically, the model was trained for 200 epochs, with a batch size of 1, and hinge loss optimized by Adam. The initial learning rate was 0.0001. Data splits for training, validation, and testing were conducted in the same way as for CycleGAN with each dataset.

## 2.6 Model evaluation

Based on the held-out testing datasets, CycleGAN models with or without spectral normalization were evaluated using three common metrics: PSNR, SSIM, and mean absolute error (MAE). These metrics were calculated per image slice and averaged per person per dataset for subsequent analyses.

## 2.7 Utility evaluation of the synthesized images

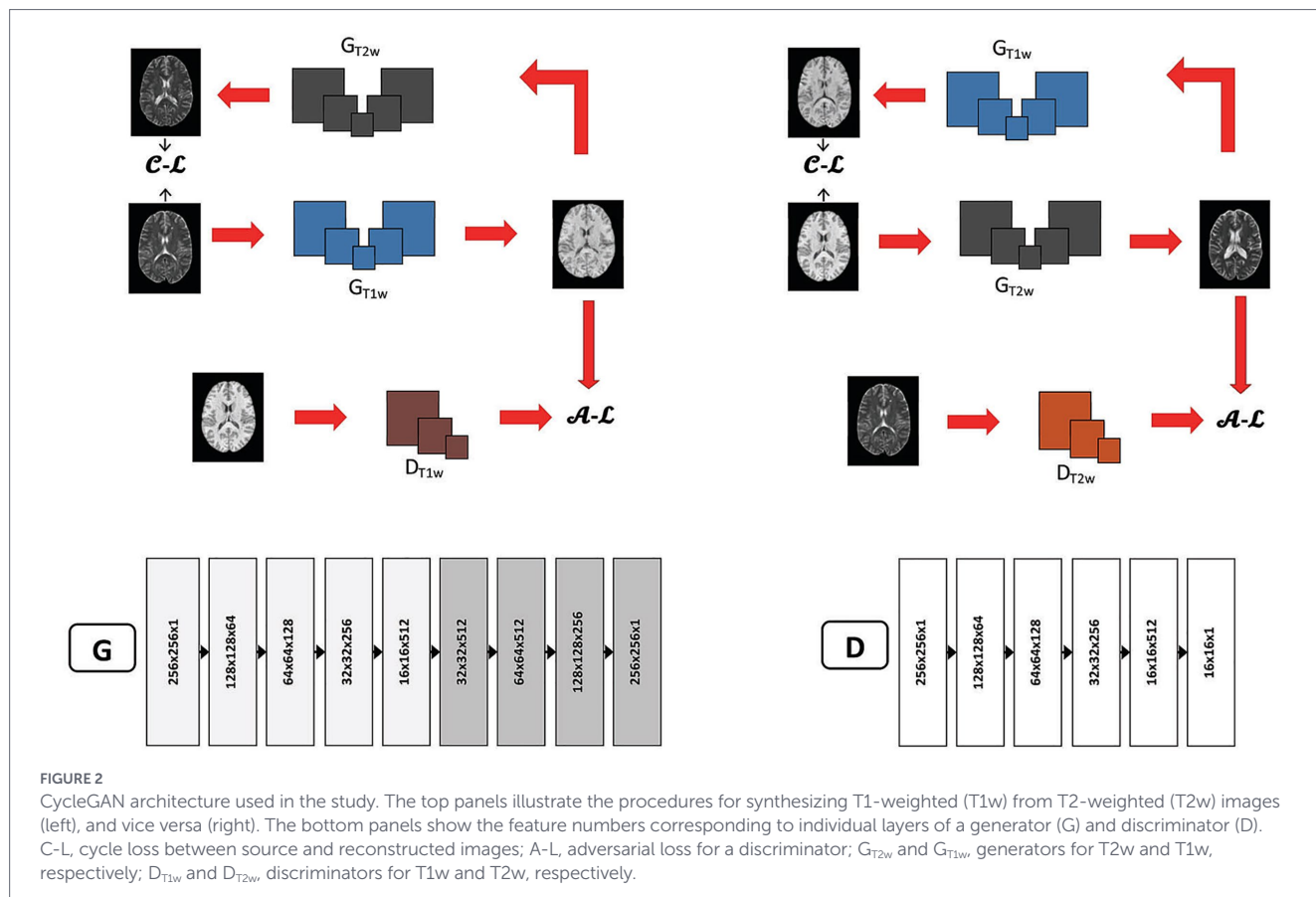
Based on the testing T1w brain MRI of MS, utility exploration involved three tasks. The first was lesion detection. Using the TrUENet method in FSL (Strain et al., 2024), MS lesions were segmented using T1w and the corresponding FLAIR images, where lesion probabilities below 0.3 were excluded. Segmentation accuracy was assessed using the Dice coefficient and 95th-percentile Hausdorff distance (HD95) for voxel-level overlap of lesions. Lesion count accuracy was measured using the lesion-wise F1 score and the lesion count agreement score. Furthermore, lesion size distribution and detection failure modes were presented for additional understanding of image synthesis methods. The latter included both the failure lesion number and the lesion count variant ratio between synthesized and source images.

The second task was brain volume measurement using T1w MRI following lesion filling as a standard practice, done using the SynthSeg method from FreeSurfer 8.0 (Harvard Medical School, USA). This assessment considered the following four measures: total intracranial volume (ICV), volumes of cortical gray matter, white matter, and cerebrospinal fluid (CSF).

The third was texture analysis around MS lesions using an optimized statistical approach, gray level co-occurrence matrix (GLCM) (Hosseinpour et al., 2022). Using its Scikit implementation in Python, two tested parameters were computed: texture contrast and texture dissimilarity. Both metrics detected subtle structural changes in MS using histology-informed brain MRI (Hosseinpour et al., 2022). Lesion texture was then averaged per subject.

## 2.8 Statistical analysis

Comparing results across imaging sources was performed using one-way ANOVA when normal distributions were satisfied, followed by Tukey correction for multiple comparisons. Friedman's test was used otherwise. Statistical analyses used the GraphPad Prism package (version 10.4.1, Boston, Massachusetts, USA) and the SciPy package in Python, with  $p \leq 0.05$  set as significance. Comparison between CycleGAN modeling results across datasets was not performed, given their inherent differences.



**FIGURE 2** CycleGAN architecture used in the study. The top panels illustrate the procedures for synthesizing T1-weighted (T1w) from T2-weighted (T2w) images (left), and vice versa (right). The bottom panels show the feature numbers corresponding to individual layers of a generator (G) and discriminator (D). C-L, cycle loss between source and reconstructed images; A-L, adversarial loss for a discriminator;  $G_{T2w}$  and  $G_{T1w}$ , generators for T2w and T1w, respectively;  $D_{T1w}$  and  $D_{T2w}$ , discriminators for T1w and T2w, respectively.

### 3 Results

#### 3.1 Comparable outcomes between synthesized and source images

Qualitatively (Figure 3), anatomical integrity was well preserved in generated images by both CycleGAN-based and Pix2Pix models, including ventricular morphology and contrast. Lesion size distribution was largely similar based on synthesized images from all models to source images, with a minor exception for relatively small lesions (Supplementary Figure 2). Spectral normalization in CycleGAN appeared to decrease imaging sharpness based on experiments with the testing cohort of HCP, as indicated by lower Laplacian energy and higher Gradient MAE (Supplementary Table 1). Quantitative results supported the overall fidelity of synthesized image from each model, with Pix2Pix performing slightly better than CycleGAN models in certain outputs.

##### 3.1.1 HCP dataset

In synthesizing either T1w or T2w images, CycleGAN with or without spectral normalization achieved high PSNR and SSIM, and low MAE, which improved in Pix2Pix, though the effect sizes varied (Tables 1, 2; Figure 4). Specifically, the average values of CycleGAN images ranged from 26.190 to 28.290 versus 28.520 and 31.100 of Pix2Pix for PSNR, 0.879–0.901 versus 0.933 and 0.938 for SSIM, and 0.014–0.029 versus 0.010 and 0.024 for MAE. In addition, CycleGAN without spectral normalization performed better than within both image synthesis directions, with PSNR and SSIM increased by 2.64–3.19% and 1.63–2.09%, and MAE decreased by 4.76–8.30% (all  $p \leq 0.001$ ).

##### 3.1.2 PPMI dataset

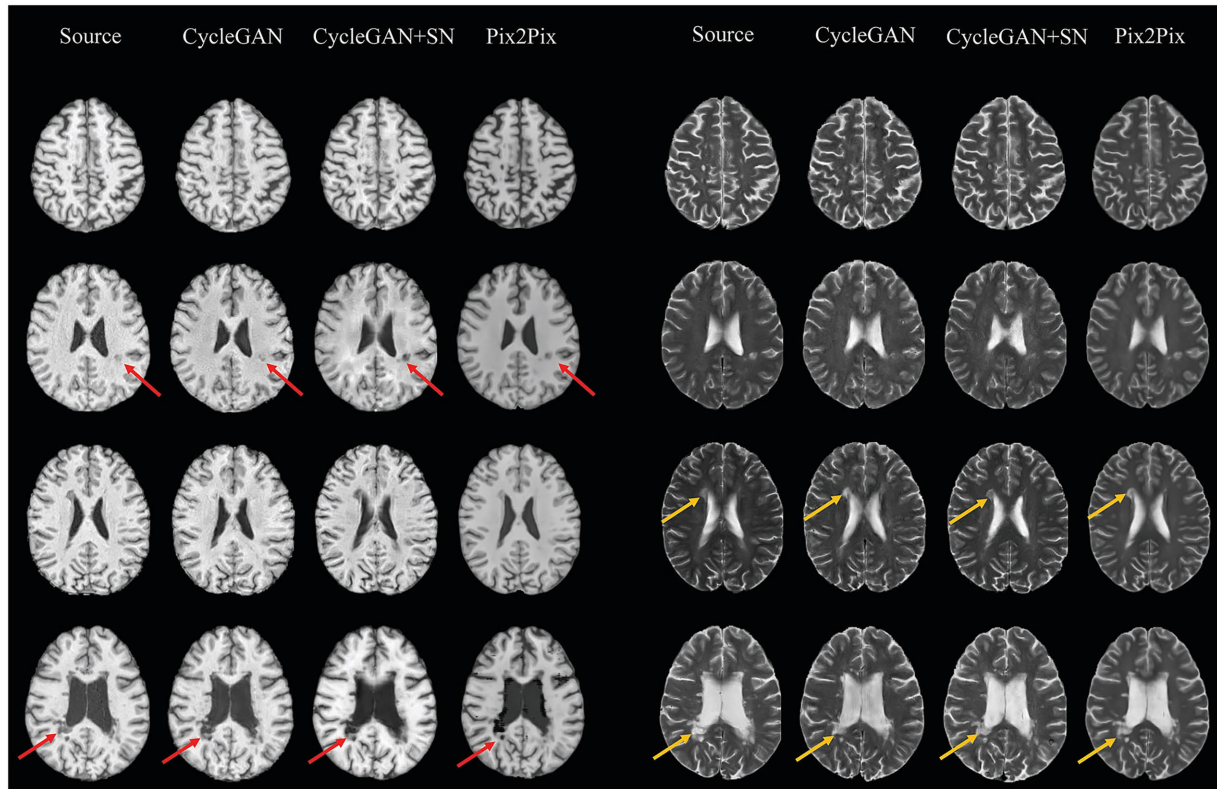
All CycleGAN images showed competitively high PSNR and SSIM, and low MAE, which also showed improvement in Pix2Pix with different effect sizes (Tables 1, 2; Figure 4). The average values of CycleGANs ranged from 24.860 to 27.130 versus 29.200 and 31.280 of Pix2Pix in PSNR, 0.838–0.888 versus 0.924 and 0.943 for SSIM, and 0.018–0.029 versus 0.011 and 0.019 for MAE. Between models, CycleGAN with spectral normalization showed increased PSNR by 1.89–3.57% and SSIM by 2.56–3.63%, and decreased MAE by 11.02% compared with vanilla CycleGAN, with all  $p \leq 0.001$  except MAE, where  $p \leq 0.05$ .

##### 3.1.3 Local MS dataset

Similarly, both PSNR and SSIM were high, and MAE was low with CycleGAN models, with Pix2Pix improved in select comparisons (Tables 1, 2; Figure 4). The average values of CycleGAN images ranged from 26.510–28.570 versus 28.260 and 28.010 of Pix2Pix for PSNR, 0.873–0.899 versus 0.917 and 0.914 for SSIM, and 0.019–0.029 versus 0.017 and 0.025 for MAE. Adding spectral normalization showed no positive impact on CycleGAN outputs, except for a mild improvement in PSNR at T1w to T2w translation.

#### 3.2 Equivalent utility between CycleGAN-synthesized and source T1w MRI in MS

In lesion detection, all CycleGAN and Pix2Pix T1w images achieved high Dice and low HD95 scores, with no significant differences between these models (mean Dice > 0.7 and mean HD95 < 10;



**FIGURE 3**  
Representative ground truth and synthesized brain MRI of MS based on different methods. CycleGAN models with or without spectral normalization (SN) both replicated lesion areas well in T1-weighted (hypointensity, red arrows) and T2-weighted (hyperintensity, yellow arrows) images, similar to ground truth. The Pix2Pix-generated images show similar findings with slightly higher contrast than the source in a few lesion areas.

**TABLE 1** Mean (95% confidence interval) testing metrics for the CycleGAN-based and Pix2Pix models on reciprocal synthesis between T1- and T2-weighted brain MRIs across datasets.

Dataset	Translation	Metric	CycleGAN	CycleGAN + SN	Pix2Pix
HCP (N = 112)	T2w to T1w	PSNR	26.880 [26.40, 27.37]	26.190 [25.75, 26.63]	28.520 [27.85, 29.19]
		SSIM	0.901 [0.897, 0.904]	0.886 [0.883, 0.890]	0.933 [0.930, 0.937]
		MAE	0.028 [0.025, 0.030]	0.029 [0.027, 0.031]	0.024 [0.021, 0.026]
	T1w to T2w	PSNR	28.290 [28.13, 28.44]	27.410 [27.26, 27.56]	31.100 [30.19, 31.29]
		SSIM	0.898 [0.895, 0.900]	0.879 [0.877, 0.882]	0.938 [0.936, 0.940]
		MAE	0.014 [0.014, 0.015]	0.016 [0.015, 0.016]	0.010 [0.010, 0.011]
PPMI (N = 32)	T2w to T1w	PSNR	24.860 [24.37, 25.35]	25.330 [24.80, 25.85]	29.220 [28.41, 30.30]
		SSIM	0.838 [0.832, 0.845]	0.860 [0.852, 0.868]	0.924 [0.919, 0.930]
		MAE	0.028 [0.027, 0.030]	0.029 [0.027, 0.031]	0.019 [0.016, 0.021]
	T1w to T2w	PSNR	26.190 [25.77, 26.61]	27.130 [26.71, 27.54]	31.280 [30.58, 31.98]
		SSIM	0.866 [0.860, 0.873]	0.888 [0.881, 0.896]	0.943 [0.937, 0.948]
		MAE	0.020 [0.019, 0.022]	0.018 [0.017, 0.019]	0.011 [0.010, 0.012]
MS (N = 11)	T2w to T1w	PSNR	28.570 [26.69, 30.46]	27.230 [26.00, 28.46]	28.260 [26.99, 29.53]
		SSIM	0.899 [0.882, 0.915]	0.892 [0.881, 0.903]	0.917 [0.904, 0.929]
		MAE	0.025 [0.020, 0.030]	0.029 [0.023, 0.030]	0.025 [0.021, 0.029]
	T1w to T2w	PSNR	26.99 [26.49, 27.49]	26.510 [25.92, 27.11]	28.010 [27.13, 28.89]
		SSIM	0.884 [0.878, 0.890]	0.873 [0.859, 0.886]	0.914 [0.906, 0.921]
		MAE	0.019 [0.017, 0.020]	0.019 [0.018, 0.021]	0.017 [0.014, 0.022]

SN, spectral normalization; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure; MAE, mean absolute error.

TABLE 2 Comparison of testing metrics between the two CycleGAN and Pix2Pix models in reciprocal synthesis between T1- and T2-weighted brain MRI across datasets.

Dataset	Modality	Metric	$F$	$p_{ANOVA}$	Cohen's $d$ (GAN-SN)	Cohen's $d$ (SN-Pix2Pix)	Cohen's $d$ (GAN-Pix2Pix)
HCP	T1w	PSNR	19.36	<0.001	0.79	-1.32	-0.92
		SSIM	187.42	<0.001	2.37	-4.2	-3.3
		MAE	6.30	0.0021	-0.35	0.84	0.6
	T2w	PSNR	534.53	<0.001	2.72	5.35	5.13
		SSIM	631.56	<0.001	3.73	6.97	6.62
		MAE	241.31	<0.001	-1.69	-3.83	-3.5
PPMI	T1w	PSNR	61.18	<0.001	-0.63	-1.97	-2.39
		SSIM	181.04	<0.001	-2.75	-2.98	-4.76
		MAE	32.77	<0.001	-0.11	1.65	1.76
	T2w	PSNR	109.67	<0.001	-1.47	2.66	2.88
		SSIM	148.35	<0.001	-2.68	3.36	4.7
		MAE	79.89	<0.001	1.09	-2.7	-2.99
MS	T1w	PSNR	1.10	0.35	0.67	-0.57	0.15
		SSIM	4.48	0.02	0.4	-2.03	-1.09
		MAE	1.11	0.34	-0.51	0.47	-0.04
	T2w	PSNR	6.26	0.0054	1.02	1.77	1.41
		SSIM	24.19	<0.001	0.81	3.15	5.13
		MAE	0.58	0.57	-0.56	-0.41	-0.21
		MAE	79.89	<0.001	1.09	-2.7	-2.99

GAN, CycleGAN; SN, spectral normalization (along with CycleGAN); PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure; MAE, mean absolute error.

Figure 5). In lesion count, the lesion-wise F1 ranged 0.638–0.734, and lesion count agreement scores ranged 0.664–0.756 for CycleGAN models and Pix2Pix, respectively (Table 3). The lesion failure rate was low and similar across models, with all lesions detected and the count variant ratio close to or greater than one, indicating high fidelity.

In brain volume measurement, CycleGAN T1w from either version was source-equivalent in cortical gray matter and CSF volumes ( $p > 0.05$ ), but lower than source in ICV and white matter ( $|\Delta| \approx 1.0$ – $1.5\%$ ;  $p < 0.01$ ). Pix2Pix T1w volumes showed a similar trend, except for a larger CSF volume than the source ( $p \leq 0.01$ ; Figure 6).

In texture analysis, lesion texture contrast was similar between the T1w of all models under investigation compared with source T1w. Lesion texture dissimilarity was higher in Pix2Pix T1w than in source images, while both CycleGAN models showed similar outputs to source results ( $p > 0.05$ ; Figure 7).

## 4 Discussion

Based on investigations using both streamlined and heterogeneous datasets, this study demonstrated the consistency of conditional GAN models in synthesizing T1w and T2w brain MRI. While pending confirmation, the vanilla CycleGAN appeared to be sufficient in synthesizing clinically standard images of MS, and Pix2Pix was mostly superior to CycleGANs in synthesizing images based on streamlined datasets. Utility tests supported the capacity of synthesized T1w images compared with source MRI for structural characterizations in MS.

Deep learning methods have emerged as promising candidates to address image synthesis questions in medical imaging, especially conditional GANs. The capability of CycleGAN was largely attributed to its cycle-consistent loss being minimized, allowing effective learning between the reversible imaging domains under study. Other deep learning methods, such as autoencoders, have also shown potential to synthesize medical images, but such relatively simple networks were mainly used as components of advanced models, including GANs (Wang et al., 2021). Pix2Pix facilitated paired supervision that could directly optimize correspondence between input and target images. Prior work also showed the promise of CycleGAN in advancing brain MRI in MS (La Rosa et al., 2021). However, much of the effort was based on streamlined datasets acquired with well-controlled protocols.

To compare, this study began with two high-quality, consistent datasets. With the healthy HCP data, all of our CycleGAN models achieved competitive results (PSNR > 26; SSIM > 0.8). These values appeared higher than the literature norm, where the PSNR was 22.9–23.2, and the SSIM was 0.78 in T1w to T2w, and 23.9–24.2 and 0.79, respectively, in T2w to T1w synthesis (Bui et al., 2020; Welander et al., 2018). Experiments using the PPMI dataset yielded equivalent results, but they provided an opportunity to further evaluate the potential of the models in handling pathological images based on a highly regarded dataset. Our findings with PPMI were also similar to the literature (PSNR = 19–30; SSIM = 0.65–0.91) (Zhang et al., 2022; Yang et al., 2020; Li et al., 2019). Comparably, the Pix2Pix metrics were mostly greater than those of CycleGANs in our experiments. This was not unexpected because of the benefit of paired supervision used in Pix2Pix. However, paired training may not always be possible in a

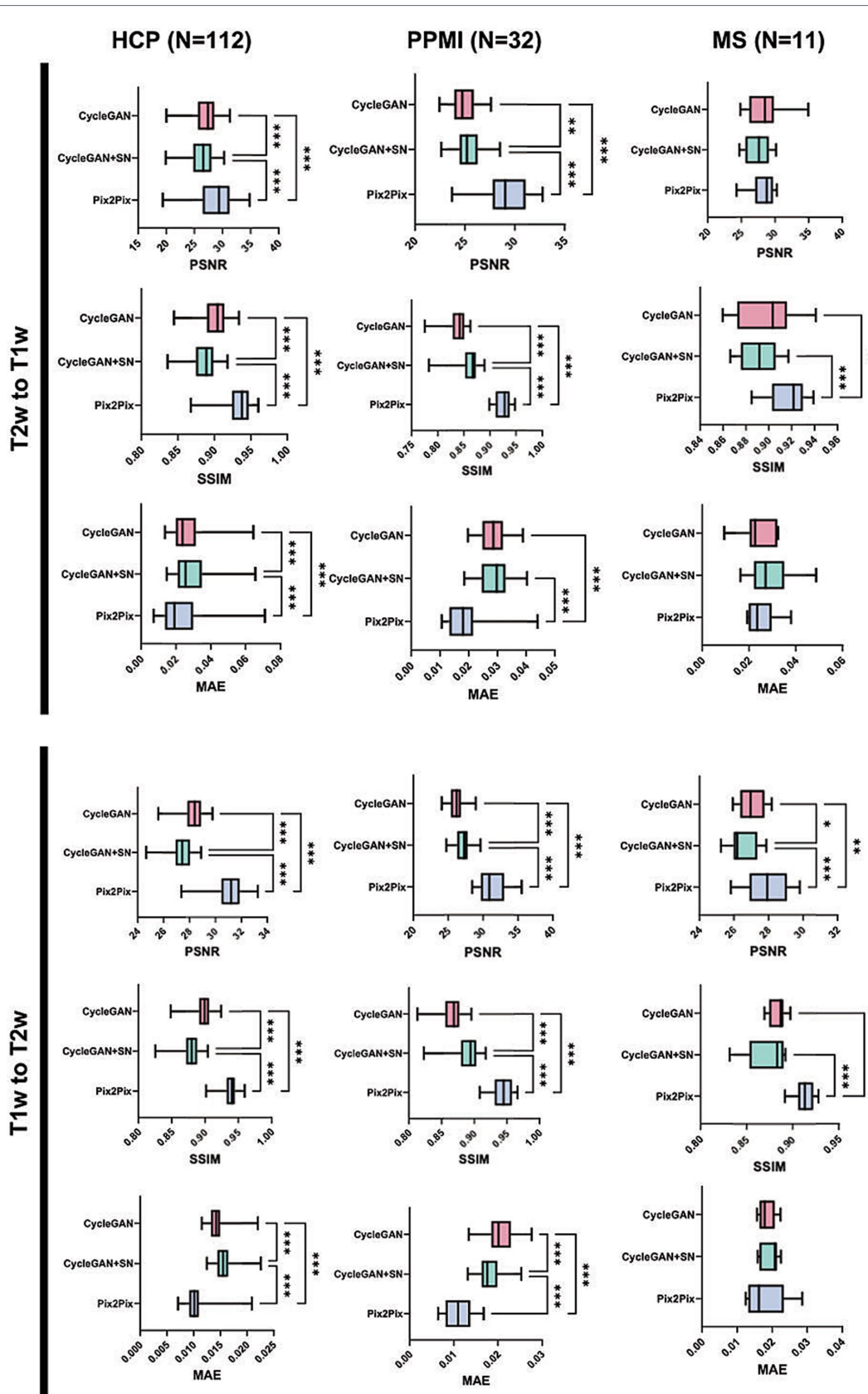


FIGURE 4 Comparison of CycleGAN models and Pix2Pix by evaluation metric and image synthesis direction. Error bars represent standard deviation. The stars represent significance values: \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ . PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure; MAE, mean absolute error; SN, spectral normalization.

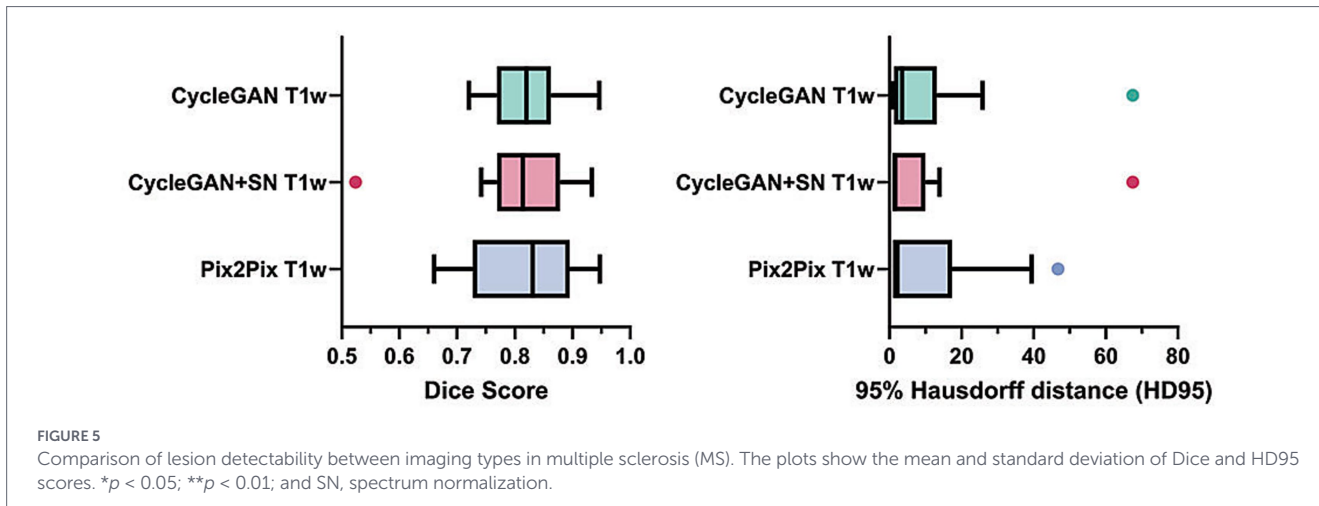
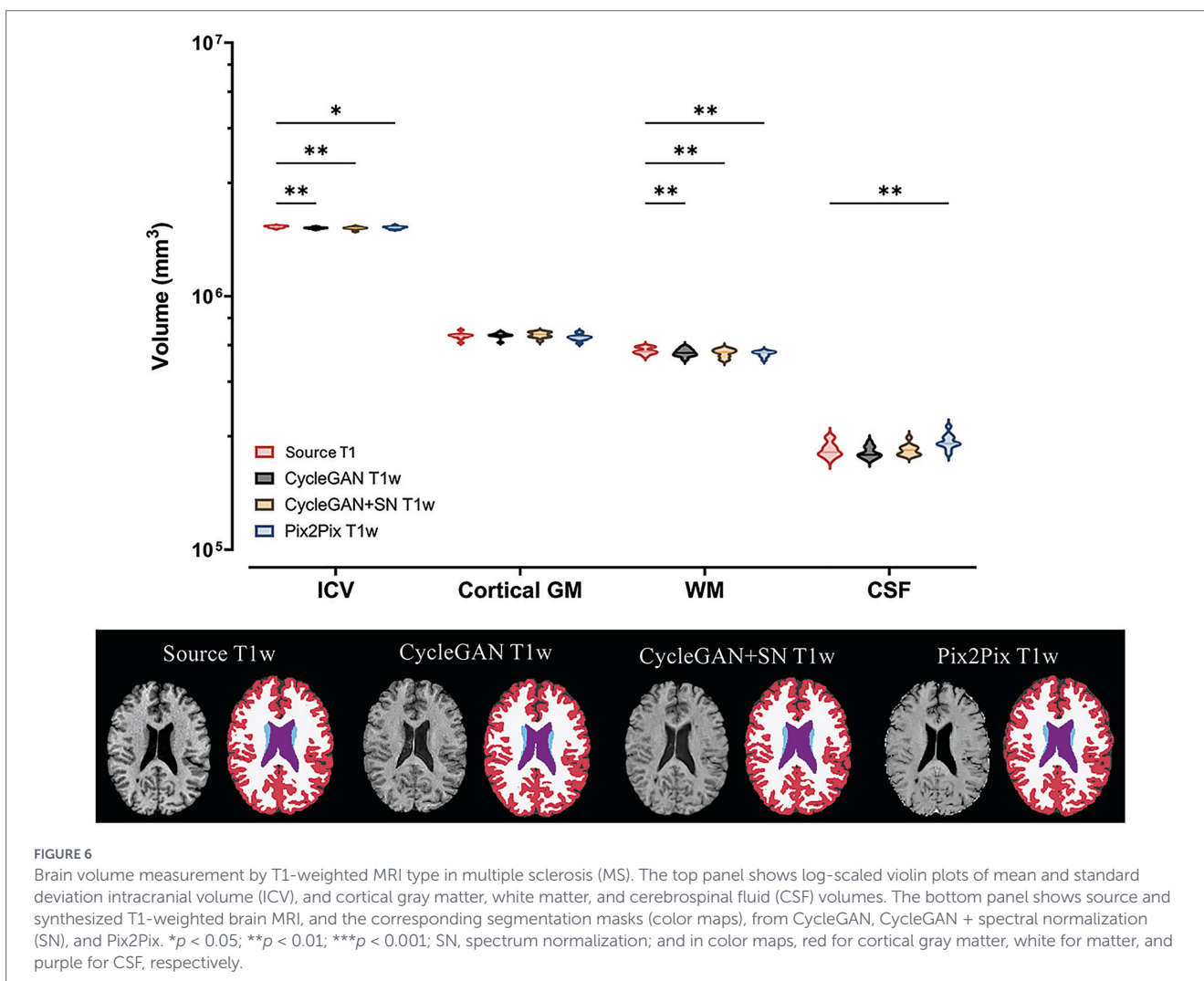
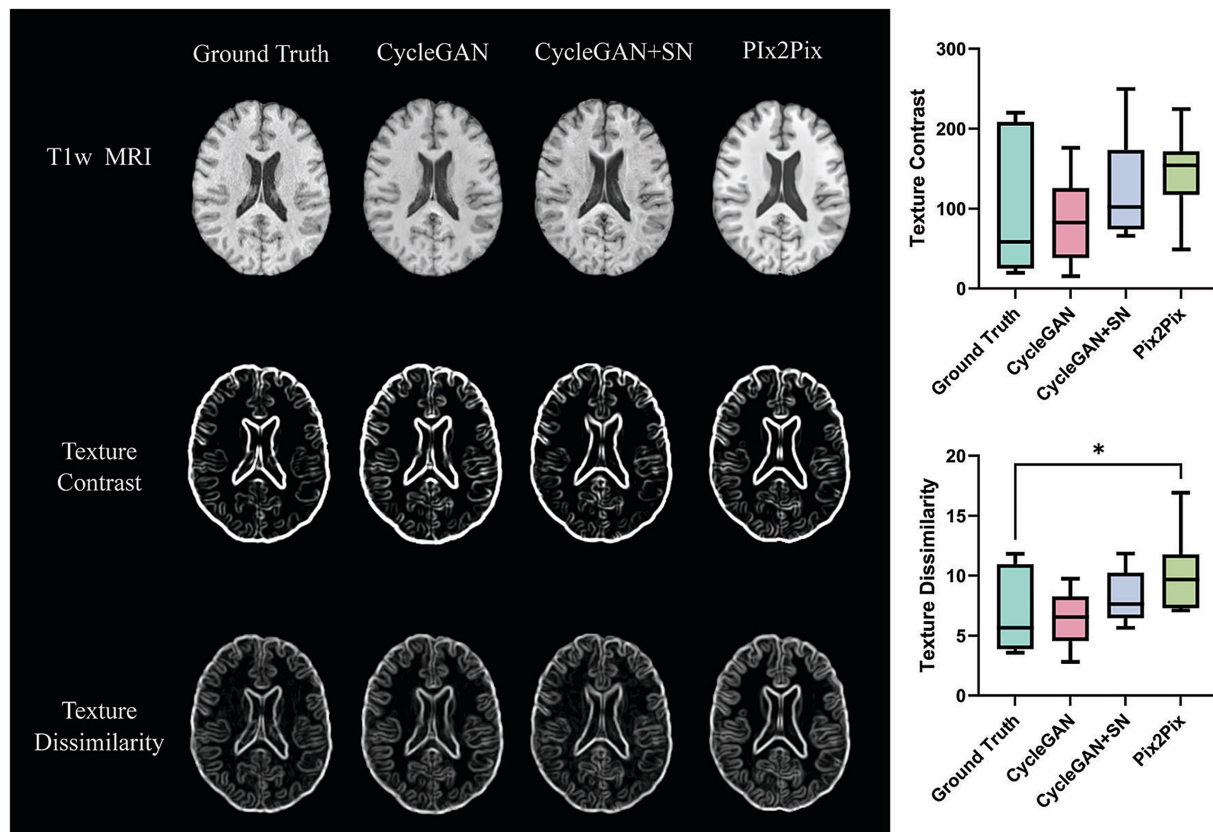


TABLE 3 Lesion count accuracy and failure rates.

Metric	CycleGAN	CycleGAN + SN	Pix2Pix
Lesion-wise F1	0.698	0.638	0.734
Lesion count agreement score	0.664	0.743	0.756
Failure lesion rate ( $n/N$ )	0/11	0/11	0/11
Count variance ratio	1.24	1.65	0.76

$n$  refers to the number of lesions failed to detect, and  $N$  refers to the total number of lesions that exist in the source images.





**FIGURE 7** Example images and corresponding texture analysis results by T1-weighted imaging type in multiple sclerosis (MS). Based on the gray level co-occurrence matrix method, the middle and bottom rows represent texture images from the source and synthesized T1 by CycleGAN, CycleGAN with spectral normalization (SN), and Pix2Pix. Box plots on the right show lesion texture computed from the above T1 types from 11 testing MS participants.

clinical setting due to the common phenomenon of missing data and heterogeneity of clinical scans.

In real-world applications, routine scans are more representative than streamlined ones. However, the former also present additional challenges due to inconsistency, complicating *post hoc* quantitative analyses. Image preprocessing may help address general quality-assurance issues but is typically not ideal for handling unavailable data. In the present study, although based on a relatively small sample size, our findings support the potential of vanilla CycleGAN models for generating potentially unavailable brain MRI scans, especially T1w in MS. The associated PSNR (>28) and SSIM (~0.9) were similar to or greater than values from the streamlined datasets and the literature. Pix2Pix could also help when paired images are available.

The results of this study with CycleGAN also suggested the dependence of spectral normalization on datasets. In HCP, including spectral normalization showed a decrease in model performance. In contrast, the same process in PPMI seemed to increase performance, which, however, was not replicated in the MS data. These findings could be due to several reasons, including differences in data characteristics, sample size, and imaging contrast variations. Our pilot experiments with the HCP test data suggested that including spectral normalization in CycleGAN decreased imaging sharpness or contrast. As the HCP scans were from healthy subjects that were already relatively 'uniform,' this added-smoothing action from spectral normalization might have contributed to worse performance than vanilla CycleGAN. On the other hand, the pathologic PPMI scans would feature increased diversity in imaging contrast and therefore likely

benefited from the over-smoothing process. Accordingly, while speculative, we hypothesized that images with greater heterogeneity would benefit more from spectral normalization in translations with CycleGAN. The same theory did not repeat in our clinical MS dataset, which could be due to the small sample size, requiring further confirmation. In line with our findings, prior studies also showed that spectral normalization might increase (Xu et al., 2022; Xu et al., 2020) or decrease (Zhao et al., 2021) CycleGAN capacity.

Our additional utility experiments in MS also supported the robustness of synthesized images by each of the conditional GAN models investigated in this study. In lesion detection, all models yielded high Dice and low HD95, agreeing with the literature (Rokuss et al., 2025). Likewise, lesion-specific metrics including number count and lesion size distribution appeared to be highly similar to source results, with no considerable differences between models. These findings suggest that a missing T1w sequence could be replaced by synthesized images from these models with little penalty to lesion burden or shape measures.

Brain volume measures also favored the synthesized T1w images. They reproduced source tissue volumes with about 1% deviation in assessing cortical gray matter and CSF, the important metrics in characterizing neurodegeneration as a critical pathology of MS (Rocca et al., 2021). Notably, the Pix2Pix images yielded a higher CSF volume than the CycleGAN models and the source data. This could be due to the hinge loss function used in this model, which made extra emphasis on edge detection.

To understand the utility of synthesized images in characterizing subtle structural patterns, texture analysis was also performed. The

lack of difference in lesion texture between CycleGAN-based and source T1w images would further suggest the fidelity of these methods, with vanilla CycleGAN-based values seemingly the closest to source results. The higher lesion texture dissimilarity from Pix2Pix T1w likely reflected greater image heterogeneity than other imaging types, although further confirmation is required. Collectively, although there is a lack of literature for direct comparisons, the concordant results from lesion, volume, and texture analyses support the robustness of our synthesized MS images across scales. The T1w MRI created this way could also serve as a reference in many image preprocessing tasks including co-registration (Smith et al., 2004). By itself, T1w MRI had also been a useful channel in different deep learning applications including disease classification (Zhang et al., 2021).

The study has some limitations. The sample size of our MS dataset was relatively small, limiting generalizability. However, as dramatically different acquisition protocols were involved, the dataset served as a good representation of real-world data. In future studies, our estimation based on a moderate effect size indicated that obtaining 80% power with 0.05 type I error would require 34 subjects per group in a paired design. Additionally, this study focused only on conditional GANs as example deep learning methods, with comparisons only to Pix2Pix. While this might have limited the scope of interpretation, these methods were well-recognized in synthesizing MR images, and our investigation using different types of datasets seemed to support their capacity. Further, our utility evaluation studies were rather exploratory. Nonetheless, our consistent results based on an important yet potentially missing T1w sequence should have provided proof-of-concept evidence. In the future, we plan to confirm the present findings using larger datasets, investigate new methods such as other GAN variants or diffusion models (Koetzier et al., 2024), and assess the role of synthesized images in different applications, including deep learning.

## 5 Conclusion

Our overall findings suggest the consistency and potential of conditional GAN models in synthesizing brain MRI, such as T1w and T2w sequences, across datasets of both low and high heterogeneity. Spectral normalization appeared not mandatory in most of the experiments, including those for synthesizing real-world clinical MS images, deserving further confirmation. Furthermore, CycleGAN-synthesized images appeared feasible for both gross and fine pathological analysis, as demonstrated with T1w brain MRI in MS, supporting further validation.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving humans were approved by the Institutional Research Information Services Solution (IRISS), University of Calgary.

The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

SS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. OO: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – review & editing. RT: Data curation, Investigation, Visualization, Writing – review & editing. YZ: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This study is funded in part by the Accelerating Innovations into CarE (AICE)—Concepts Program, Alberta Innovates, Canada; MS Canada; Canadian Institutes of Health Research; and Natural Sciences and Engineering Research Council of Canada. A part of the data was obtained from the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. Another part of the data was obtained on 5 June 2023 from the Parkinson's Progression Markers Initiative (PPMI) database ([www.ppmi-info.org/access-data%20specimens/download-data](http://www.ppmi-info.org/access-data%20specimens/download-data)), RRID: SCR\_006431. PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including 4D Pharma, AbbVie, AcureX, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's, AskBio, Avid Radiopharmaceuticals, BIAL, BioArctic, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol-Myers Squibb, Calico Labs, Capsida Biotherapeutics, Celgene, Cerevel Therapeutics, Coave Therapeutics, DaCapo BrainScience, Denali, Edmond J. Safra Foundation, Eli Lilly, Gain Therapeutics, GE HealthCare, Genentech, GSK, Golub Capital, Handl Therapeutics, Insiteo, Jazz Pharmaceuticals, Johnson & Johnson Innovative Medicine, Lundbeck, Merck, Meso Scale Discovery, Mission Therapeutics, Neurocrine Biosciences, Neuron23, Neuropore, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi, Servier, Sun Pharma Advanced Research Company, Takeda, Teva, UCB, Vanqua Bio, Verily, Voyager Therapeutics, the Weston Family Foundation, and Yumanity Therapeutics.

## Acknowledgments

The authors thank the participants and the local clinical research team involved in the research. The author Shayan Shahrokhi would

also like to thank his wife, Maryam, for her support throughout the working process of this manuscript.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy,

including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

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

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fninf.2026.1762794/full#supplementary-material>

## References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). TensorFlow: a system for large-scale machine learning. *arXiv*. (csDC) 16, 265–283. doi: 10.48550/arXiv.1605.08695
- Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein generative adversarial networks. In: D. Precup and Y. W. Teh, editors. *Proceedings of the 34th international conference on machine learning: PMLR*; (2017). p. 214–223.
- eds. B. D. Basaran, M. Qiao, P. M. Matthews and W. Bai (2022). "Subject-specific lesion generation and Pseudo-healthy synthesis for multiple sclerosis brain images," in *Simulation and Synthesis in Medical Imaging*. (Cham: Springer International Publishing).
- Bui, T. D., Nguyen, M., Le, N., and Luu, K. Flow-based deformation guidance for unpaired multi-contrast MRI image-to-image translation. *Medical image computing and computer assisted intervention (MICCAI)* (2020). p. 728–737.
- Eche, T., Schwartz, L. H., Mokrane, F. Z., and Derclé, L. (2021). Toward generalizability in the deployment of artificial intelligence in radiology: role of computation stress testing to overcome underspecification. *Radiol. Artif. Intell.* 3:e210097. doi: 10.1148/ryai.2021210097
- Finck, T., Li, H., Schlaeger, S., Grundl, L., Sollmann, N., Bender, B., et al. (2022). Uncertainty-aware and lesion-specific image synthesis in multiple sclerosis magnetic resonance imaging: a multicentric validation study. *Front. Neurosci.* 16:889808. doi: 10.3389/fnins.2022.889808
- Fischl, B. (2012). Freesurfer. *NeuroImage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021
- Fisniku, L. K., Brex, P. A., Altmann, D. R., Miszkiel, K. A., Benton, C. E., Lanyon, R., et al. (2008). Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain* 131, 808–817. doi: 10.1093/brain/awm329
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2020). Generative adversarial networks. *Commun. ACM* 63, 139–144. doi: 10.1145/3422622
- Hosseinpour, Z., Jonkman, L., Oladoso, O., Pridham, G., Pike, G. B., Inglese, M., et al. (2022). Texture analysis in brain T2 and diffusion MRI differentiates histology-verified grey and white matter pathology types in multiple sclerosis. *J. Neurosci. Methods* 379:109671. doi: 10.1016/j.jneumeth.2022.109671
- Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A. (2016). Image-to-image translation with conditional adversarial networks. *arXiv:161107004v3*. (csCV) doi: 10.1109/CVPR.2017.632
- Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A. Image-to-image translation with conditional adversarial networks. *2017 IEEE conference on computer vision and pattern recognition (CVPR)*: IEEE; (2017). p. 1125–1134.
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., and Woolrich, M. W. (2012). FSL. *NeuroImage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- P. Julián Alberto, S. Diego Fernandez and F. Enzo, editors. Unsupervised domain adaptation via CycleGAN for white matter hyperintensity segmentation in multicenter MR images. *ProcSPIE*; (2020).
- Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv*. (csLG) doi: 10.48550/arXiv.1412.6980
- Koetzier, L. R., Wu, J., Mastrodicasa, D., Lutz, A., Chung, M., Koszek, W. A., et al. (2024). Generating synthetic data for medical imaging. *Radiology* 312:e232471. doi: 10.1148/radiol.232471
- La Rosa, F., Yu, T., Barquero, G., Thiran, J. P., Granziera, C., and Bach Cuadra, M. (2021). MPRAGE to MP2RAGE UNI translation via generative adversarial network improves the automatic tissue and lesion segmentation in multiple sclerosis patients. *Comput. Biol. Med.* 132:104297. doi: 10.1016/j.combiomed.2021.104297
- Lee, J.-Y., and Choi, S.-I. (2020). Improvement of learning stability of generative adversarial network using variational learning. *Appl. Sci.* 10:4528. doi: 10.3390/app10134528
- Li, H., Paetzold, J. C., Sekuboyina, A., Kofler, F., Zhang, J., Kirschke, J. S., et al. (2019). "DiamondGAN: unified multi-modal generative adversarial networks for MRI sequences synthesis," in *Lecture Notes in Computer Science*, eds. D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, et al. (Cham: Springer International Publishing), 795–803.
- Lin, Z., Sekar, V., and Fanti, G. (2021). Why spectral normalization stabilizes GANs: analysis and improvements. *Adv. Neural Inf. Process. Syst.* 34, 9625–9638. doi: 10.48550/arXiv.2009.02773
- Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., et al. (2011). The Parkinson progression marker initiative (PPMI). *Prog. Neurobiol.* 95, 629–635. doi: 10.1016/j.pneurobio.2011.09.005
- McGinley, M. P., Goldschmidt, C. H., and Rae-Grant, A. D. (2021). Diagnosis and treatment of multiple sclerosis. *JAMA* 325:765. doi: 10.1001/jama.2020.26858
- Nema, S., Dudhane, A., Murala, S., and Naidu, S. R. (2020). An unpaired GAN for brain tumor segmentation. *Biomed. Signal Process. Control* 55:101641. doi: 10.1016/j.bspc.2019.101641
- Rocca, M., Valsasina, P., Meani, A., Gobbi, C., Zecca, C., Rovira, A., et al. (2021). Association of gray matter atrophy patterns with clinical phenotype and progression in multiple sclerosis. *Neurology* 96, e1561–e1573. doi: 10.1212/WNL.00000000000011494
- eds. M. R. Rokuss, Y. Kirchhoff, S. Roy, B. Kovacs, C. Ulrich and T. Waldet al. (2025). "Longitudinal segmentation of MS lesions via temporal difference weighting," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2024 Workshops*, (Cham: Springer Nature Switzerland).
- Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., et al. (2004). Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage* 23, S208–S219. doi: 10.1016/j.neuroimage.2004.07.051
- Strain, J. F., Rahmani, M., Dierker, D., Owen, C., Jafri, H., Vlassenko, A. G., et al. (2024). Accuracy of TrUE-net in comparison to established white matter hyperintensity segmentation methods: an independent validation study. *NeuroImage* 285:120494. doi: 10.1016/j.neuroimage.2023.120494
- Thompson, A. J., Baranzini, S. E., Geurts, J., Hemmer, B., and Ciccarelli, O. (2018). Multiple sclerosis. *Lancet* 391, 1622–1636. doi: 10.1016/S0140-6736(18)30481-1
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010). N4ITK: improved N3 Bias correction. *IEEE Trans. Med. Imaging* 29, 1310–1320. doi: 10.1109/TMI.2010.2046908

- Van Essen, D. C., Ugurbil, K., and Auerbach, E. (2012). The future of the human connectome. *NeuroImage* 62, 2222–2231. doi: 10.1016/j.neuroimage.2012.01.032
- Vázquez-Marrufo, M., Sarrias-Arrabal, E., García-Torres, M., Martín-Clemente, R., and Izquierdo, G. (2023). A systematic review of the application of machine-learning algorithms in multiple sclerosis. *Neurologia* 38, 577–590. doi: 10.1016/j.nrleng.2020.10.013
- Wang, T., Lei, Y., Fu, Y., Wynne, J. F., Curran, W. J., Liu, T., et al. (2021). A review on medical imaging synthesis using deep learning and its clinical applications. *J. Appl. Clin. Med. Phys.* 22, 11–36. doi: 10.1002/acm2.13121
- Wattjes, M. P., Ciccarelli, O., Reich, D. S., Banwell, B., De Stefano, N., Enzinger, C., et al. (2021). 2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. *Lancet Neurol.* 20, 653–670. doi: 10.1016/S1474-4422(21)00095-8
- Welander, P., Karlsson, S., and Eklund, A. (2018). Generative adversarial networks for image-to-image translation on multi-contrast mr images—a comparison of CycleGAN and unit. *arXiv:180607777*. doi: 10.48550/arXiv.1806.07777
- Xu, L., Zeng, X., Zhang, H., Li, W., Lei, J., and Huang, Z. (2020). BPGAN: bidirectional CT-to-MRI prediction using multi-generative multi-adversarial nets with spectral normalization and localization. *Neural Netw.* 128, 82–96. doi: 10.1016/j.neunet.2020.05.001
- Xu, L., Zhang, H., Song, L., and Lei, Y. (2022). Bi-MGAN: bidirectional T1-to-T2 MRI images prediction using multi-generative multi-adversarial nets. *Biomed. Signal Process. Control* 78:103994. doi: 10.1016/j.bspc.2022.103994
- Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E. I. C., and Xu, Y. (2020). MRI cross-modality image-to-image translation. *Sci. Rep.* 10:3753. doi: 10.1038/s41598-020-60520-6
- Zhang, Y., Hong, D., McClement, D., Oladosu, O., Pridham, G., and Slaney, G. (2021). Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. *J. Neurosci. Methods* 353:109098. doi: 10.1016/j.jneumeth.2021.109098
- Zhang, H., Li, H., Dillman, J. R., Parikh, N. A., and He, L. (2022). Multi-contrast MRI image synthesis using switchable cycle-consistent generative adversarial networks. *Diagnostics (Basel)* 12:816. doi: 10.3390/diagnostics12040816
- Zhang, C., Song, Y., Liu, S., Lill, S., Wang, C., Tang, Z., et al. (2018). “MS-GAN: GAN-based semantic segmentation of multiple sclerosis lesions in brain magnetic resonance imaging,” in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, (Piscataway, NJ: IEEE).
- Zhao, P., Pan, H., and Xia, S. MRI-trans-GAN: 3D MRI cross-modality translation. (2021) *2021 40th Chinese Control Conference (CCC): IEEE*.
- Zhu, J. -Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *2017 IEEE international conference on computer vision (ICCV): IEEE*.