

**Deep learning-based generation of synthetic multiphasic MRI in
hepatocellular carcinoma and cirrhosis**

Sara A. Abosabie, Salma A. S. Abosabie, Weicheng Dai, Junlin Yang, Moritz Gross,
Jeffrey Weinreb, Margarita V. Revzin, Gaurav Parmar, Chenyu You, MingDe Lin, Olivia
Gaddum, Bernhard Gebauer, Lynn Jeanette Savic, David Craig Madoff, James S.
Duncan, Julius Chapiro

Table of contents

Supplementary methods.....2

Supplementary figures.....5

Supplementary table.....8

Supplementary references.....9

SUPPLEMENTAL METHODS

MRI Preprocessing: MRI Coregistration

The study used three-dimensional liver contrast-enhanced multiphasic magnetic resonance imaging (CE-mpMRI), including T1-weighted and T2-weighted precontrast phases and T1-weighted postcontrast phases: arterial, portal venous, delayed, and hepatobiliary phase (20 s, 70 s, 180 s, 20 minutes). The exported NIFTI files of the MRI phases were de-identified and then coregistered using BioImage Suite (v3.5) (1). Within the software, the precontrast T1-weighted, precontrast T2-weighted phases, as well as postcontrast T1-weighted phases (arterial, delayed, and hepatobiliary) were nonlinearly coregistered to the postcontrast T1-weighted portal-venous phase using an intensity-based approach.

MRI Preprocessing: MRI Bias Field Correction and Intensity Normalization

After coregistration, automated N4 bias field correction algorithm (N4ITK) implemented in the SimpleITK library was used to correct local image intensity inhomogeneities (2). MRI images were also processed by clipping intensity values below the 0.5th percentile and above the 99.5th percentile, which helps minimize the influence of outlier voxels and enhances robustness. After clipping, min-max normalization was performed, scaling all voxel intensities to a standardized range (0 to 1).

Data Augmentation

During training, data augmentation of the input phase arrays included random rotations 0–3 times by 90° along a random spatial axis, random flipping along another random axis, and random cropping to generate patches with size 64x64x64 voxels for further augmentation.

Liver and tumor segmentations

For the 20% test set, liver and tumor segmentation masks were created for all ground-truth and synthetic CE-mpMRI exams from the target model across the four postcontrast phases (arterial, portal venous, delayed, hepatobiliary phase). Automated deep learning–based liver segmentation was performed for the HCC and cirrhosis cohorts, using a previously validated and published method (3). Target HCC lesions were manually segmented for the HCC cohort using 3D Slicer (version 5.4.0, www.slicer.org) (4). A board-certified radiologist (J.C.) with over 15 years of experience, who was not a study reader,

reviewed the manual tumor and automated liver segmentations while blinded to the ground-truth versus synthetic nature of the MRI exams, and necessary manual corrections were made accordingly.

Deep Cycle-consistent Generative Adversarial Network Architecture

A fully three-dimensional cycle-consistent generative adversarial network (CycleGAN) deep learning model adapted from Zhu and Park et al. (5) was implemented using Python version 3.11.9. The model comprised two 3D U-Net-based generators adapted from Wang et Al. (6) and two 3D PatchGAN-style discriminators adapted from Isola et al (7). All cycleGAN models were trained using $64 \times 64 \times 64$ -voxel patches from preprocessed images and a batch size of 8. 80% of the MRI dataset was used for the training set, and 20% of the MRI dataset was used for the completely independent test set. Within the 80% training set, a 10-fold cross-validation was used. For all models, early stopping was triggered after 30 consecutive epochs without improvement in validation loss to monitor performance and prevent overfitting.

Single and multi-phase generation for separate and combined cohorts

To assess whether the target model offered optimal performance and to ensure it did not underperform compared to alternative cycleGAN models that generate a single postcontrast phase or multiple postcontrast phases for either separate (HCC only or cirrhosis only) or combined cohorts, fourteen additional models were trained, each for the first cross-validation fold as shown in supplemental Figure E2. Model performance was assessed by calculating the structural similarity index and normalized root mean square error metrics in the liver between ground-truth and corresponding synthetic MRI exams. In only the first cross-validation fold, the target model was statistically compared to that of the fourteen additional models using the two-tailed Wilcoxon signed-rank test for paired data, as shown in supplemental Figure E2.

The fourteen additional models were as follows:

For the HCC and cirrhosis combined cohort, four additional "single-phase separate cohort" models were each trained to generate only one specific postcontrast phase (arterial, portal venous, delayed, and hepatobiliary). For the HCC cohort, five additional models were trained: four "single-phase separate cohort" models were trained to generate only one specific postcontrast phase (arterial, portal venous, delayed, and hepatobiliary), and one "multi-phase separate cohort" model was

trained to generate all four postcontrast phases simultaneously. Likewise, the same five models were trained independently for the cirrhosis cohort only.

Qualitative Analysis: Multi-reader Study

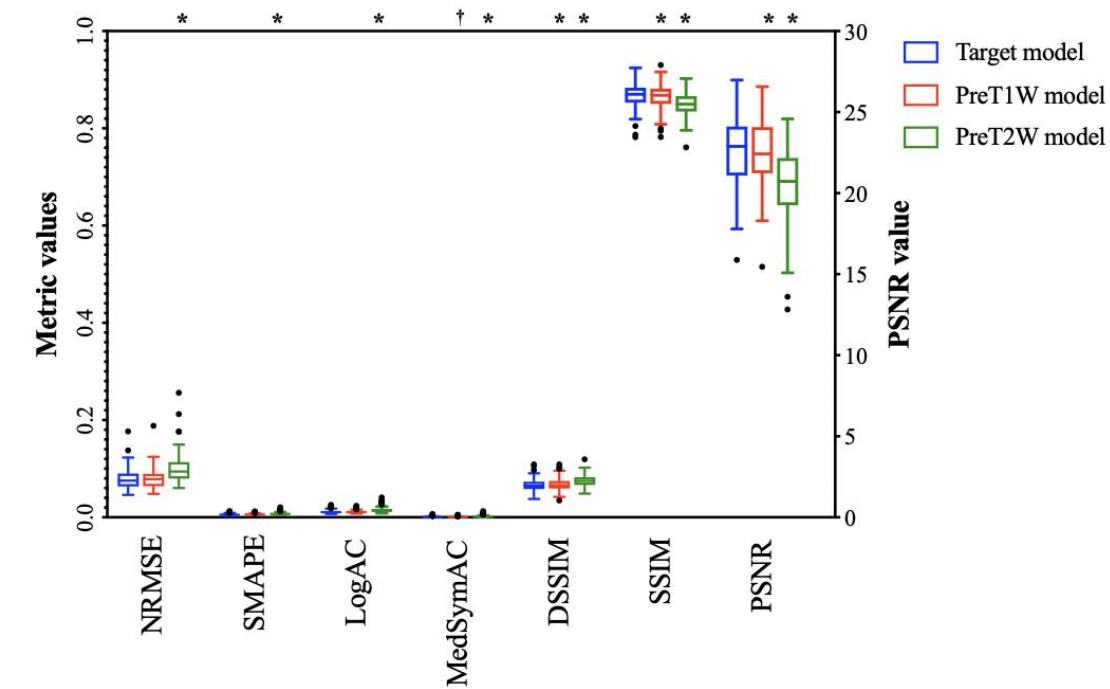
A multi-reader study was conducted by three blinded board-certified radiologist readers (J.W.), (M.V.R.), and (G.P.) with 11, 22, and 40 years of experience in abdominal imaging, respectively, including liver MRI interpretation. Readers were intentionally selected to represent a range of experience levels to evaluate inter-reader agreement across different expertise levels and to assess the generalizability, feasibility, and potential benefit of the synthetic MRI approach across varying clinical backgrounds. Each reader completed a detailed evaluation questionnaire while blinded to the ground-truth or synthetic nature of the CE-mpMRI exams as well as to the patient's clinical history. Their assessments were thus based solely on independent review during the reading sessions.

The ground-truth and synthetic liver CE-mpMRI exams (generated by the target model) of the 20% test dataset of the combined cohort were read across two reading sessions, with the second session after at least a 4-week memory washout period to minimize recall bias. Each CE-mpMRI exam consisted of either four ground-truth or four synthetic postcontrast phases (arterial, portal venous, delayed, and hepatobiliary), displayed across multiple viewing windows.

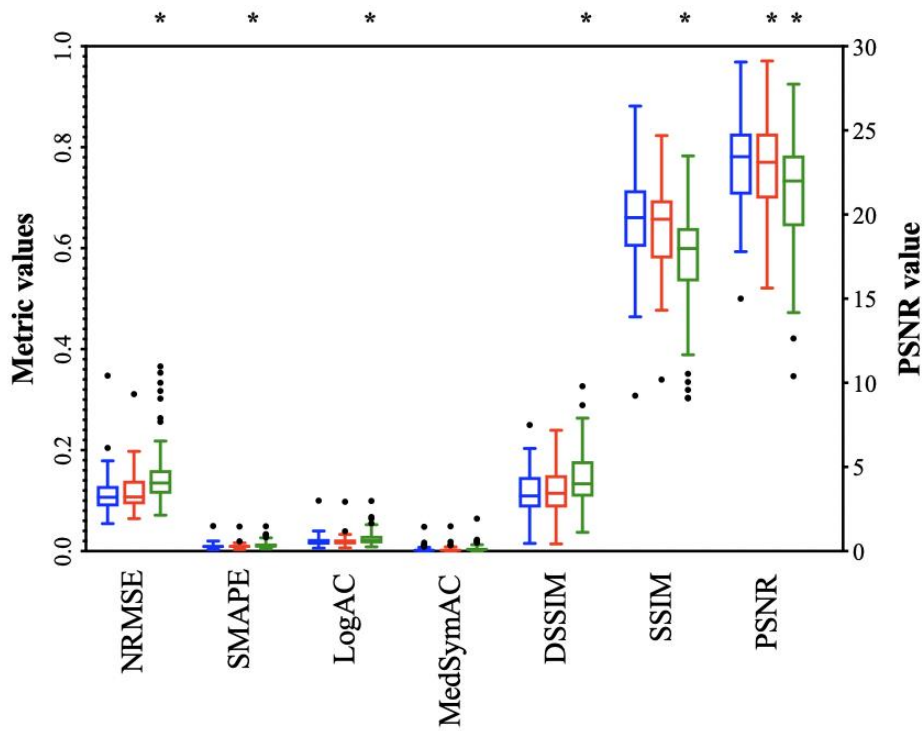
A comprehensive six-task questionnaire was developed to thoroughly assess the key aspects of the MRI exams based on a detailed review of established literature (8-13) and modified as necessary to fit the specific aims of this study. The final questionnaire was reviewed and refined by a board-certified radiologist (J.C.) with over 15 years of experience, who was not involved as a study reader. In questionnaire tasks 1 – 6, for the four postcontrast phases of each ground-truth or synthetic MRI exam, the readers assessed tasks 1) visual Turing test (ground-truth vs synthetic nature), 2) image quality, 3) radiologic diagnosability of disease (HCC or cirrhosis), 4) liver anatomic accuracy, 5) severity of image artifacts, and 6) presence of Liver Imaging Reporting and Data System (LI-RADS) features (14) within the target HCC lesions. In task 2, image quality scores were: 5 = excellent, 4 = very good, 3 = good, 2 = poor, or 1 = unacceptable. MRI exams with quality scores of 3–5 or 4–5 were later categorized as "diagnostic quality" or "high diagnostic quality", respectively. In task 5, artifact scores were: 0 = no artifacts, 1 = slight artifacts, 2 = moderate artifacts, or 3 = severe artifacts. MRI exams with artifact scores of 0–1 were later categorized as "minimal artifacts".

SUPPLEMENTAL FIGURES

A. Liver



B. Tumor



Supplemental Figure E1 shows (A) liver and (B) tumor similarity and error metrics between ground-truth and corresponding synthetic MRI exams for the target "multi-phase combined cohort" model and two other models that each

use only the precontrast T1-weighted phase or only the precontrast T2-weighted phase as input to synthesize the four postcontrast phases (arterial, portal venous, delayed, and hepatobiliary). Statistical comparisons were performed using the two-tailed Wilcoxon signed-rank test. The tests were run based on average results of the combined four postcontrast phases (arterial, portal venous, delayed, and hepatobiliary), i.e.: average per full CE-mpMRI exam), from the first cross-validation fold of each model.

Note: The left Y-axis shows values of all metrics except PSNR. The right Y-axis shows the PSNR values in decibels (dB). Higher values in similarity metrics (SSIM, PSNR) and lower values in error metrics (NRMSE, SMAPE, LogAC, MedSymAC, DSSIM) indicate better performance.

* indicates a model was significantly outperformed by the target model (two-tailed Wilcoxon signed-rank test: all $p < 0.001$, except for target model vs PreT1W model in tumor PSNR: $p = 0.002$)

† indicates a model that significantly outperformed the target model (two-tailed Wilcoxon signed-rank test: all $p < 0.001$).

NRMSE: normalized root mean square error, SMAPE: symmetric mean absolute percent error, LogAC: log accuracy ratio, MedSymAC: median symmetric accuracy, DSSIM: structural dissimilarity index, SSIM: structural similarity index, PSNR: peak signal-to-noise ratio, PreT1W: precontrast T1-weighted phase, PreT2W: precontrast T2-weighted phase, CE-mpMRI: contrast-enhanced multiphasic Magnetic Resonance Imaging.

SUPPLEMENTAL TABLES

Supplemental Table E1 shows the quantitative analysis in the form of similarity and error metrics for the target model between the ground-truth and corresponding synthetic CE-mpMRI exams in the liver in the combined HCC and cirrhosis cohort and in the tumor in the HCC cohort.

	SSIM	PSNR/dB	NRMSE	SMAPE (%)	LogAC	MedSymAC	DSSIM
Liver	0.862 ± 0.030	22.394 ± 2.988	0.081 ± 0.030	0.625 ± 0.229	0.013 ± 0.005	0.001 ± 0.001	0.069 ± 0.015
Tumor	0.758 ± 0.107	19.436 ± 4.196	0.120 ± 0.063	1.043 ± 0.733	0.021 ± 0.015	0.004 ± 0.008	0.121 ± 0.054

Note – Results are shown as mean ± SD

Note: Higher values in similarity metrics (SSIM, PSNR) and lower values in error metrics (NRMSE, SMAPE, LogAC, MedSymAC, DSSIM) indicate better performance.

HCC: Hepatocellular Carcinoma, CE-mpMRI: contrast-enhanced multiphasic Magnetic Resonance Imaging, SSIM: structural similarity index, PSNR: peak signal-to-noise ratio, NRMSE: normalized root mean square error, SMAPE: symmetric mean absolute percent error, LogAC: log accuracy ratio, MedSymAC: median symmetric accuracy, DSSIM: structural dissimilarity index

SUPPLEMENTAL REFERENCES

Author names in bold designate shared co-first authorship.

1. Papademetris X, Jackowski MP, Rajeevan N, et al. BioImage Suite: An integrated medical image analysis suite: An update. *Insight J.* 2006;2006:209.

2. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging.* 2010;29(6):1310-20.

3. Gross M, Huber S, Arora S, et al. Automated MRI liver segmentation for anatomical segmentation, liver volumetry, and the extraction of radiomics. *Eur Radiol.* 2024;34(8):5056-65.

4. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging.* 2012;30(9):1323-41.

5. **Zhu JY, Park T**, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *IEEE Int Conf Comput Vis*. 2017.
6. Wang G, Luo X, Gu R, et al. PyMIC: A deep learning toolkit for annotation-efficient medical image segmentation. *Comput Methods Programs Biomed*. 2023;231:107398.
7. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. *arXiv*. 2017;arXiv:1611.07004.
8. Chung M, Calabrese E, Mongan J, et al. Deep Learning to Simulate Contrast-enhanced Breast MRI of Invasive Breast Cancer. *Radiology*. 2023;306(3):e213199.
9. Calabrese E, Rudie JD, Rauschecker AM, et al. Feasibility of Simulated Postcontrast MRI of Glioblastomas and Lower-Grade Gliomas by Using Three-dimensional Fully Convolutional Neural Networks. *Radiol Artif Intell*. 2021;3(5):e200276.
10. Chen KT, Gong E, de Carvalho Macruz FB, et al. Ultra-Low-Dose (18)F-Florbetaben Amyloid PET Imaging Using Deep Learning with Multi-Contrast MRI Inputs. *Radiology*. 2019;290(3):649-56.
11. Muller-Franzes G, Huck L, Tayebi Arasteh S, et al. Using Machine Learning to Reduce the Need for Contrast Agents in Breast MRI through Synthetic Images. *Radiology*. 2023;307(3):e222211.
12. Tanenbaum LN, Tsiouris AJ, Johnson AN, et al. Synthetic MRI for Clinical Neuroimaging: Results of the Magnetic Resonance Image Compilation (MAGiC) Prospective, Multicenter, Multireader Trial. *AJNR Am J Neuroradiol*. 2017;38(6):1103-10.
13. Bluethgen C, Chambon P, Delbrouck JB, et al. A vision-language foundation model for the generation of realistic chest X-ray images. *Nat Biomed Eng*. 2024.
14. Chernyak V, Fowler KJ, Kamaya A, et al. Liver Imaging Reporting and Data System (LI-RADS) Version 2018: Imaging of Hepatocellular Carcinoma in At-Risk Patients. *Radiology*. 2018;289(3):816-30.