

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

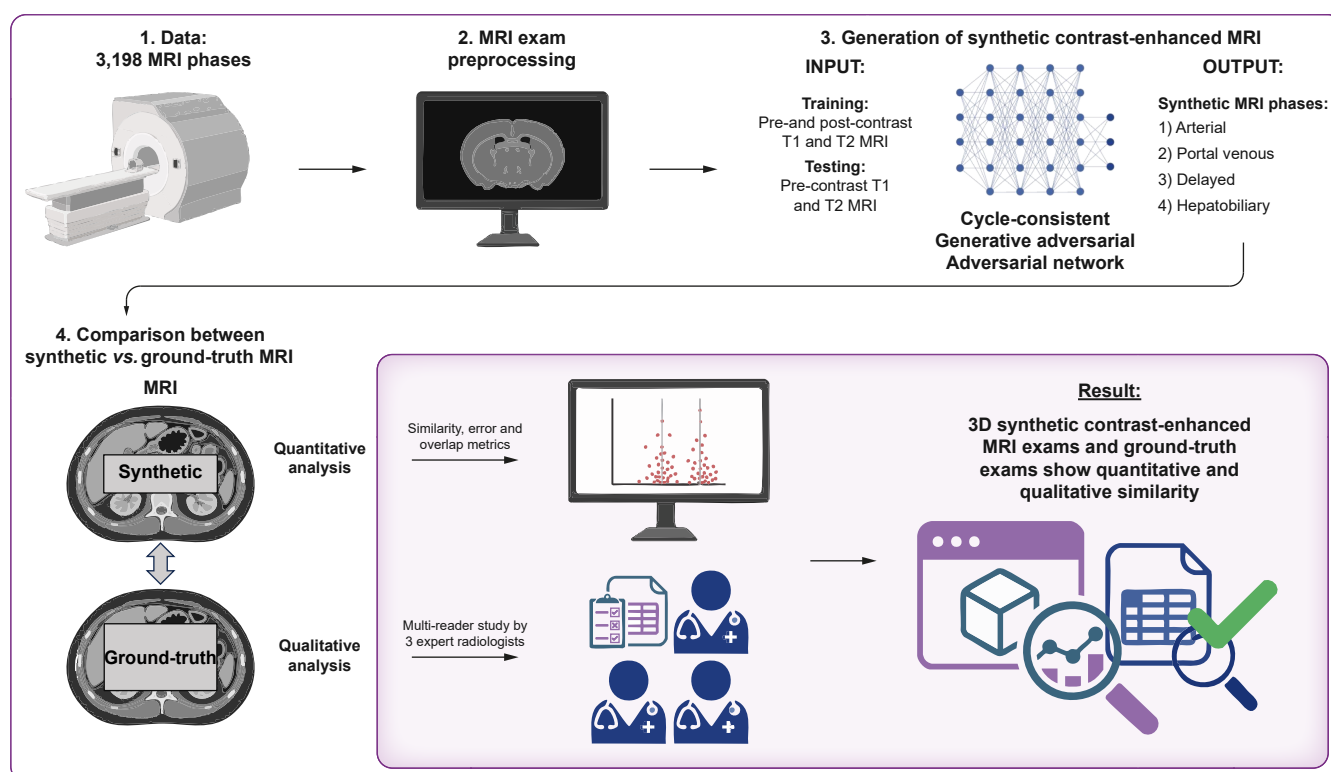
Authors

Sara A. Abosabie, Salma A.S. Abosabie, Weicheng Dai, ..., David C. Madoff, James S. Duncan, Julius Chapiro

Correspondence

julius.chapiro@yale.edu (J. Chapiro).

Graphical abstract



Highlights:

- Deep learning generated synthetic 3D contrast-enhanced liver MRI using only pre-contrast T1- and T2-weighted input images.
- Synthetic post-contrast liver MRI showed high similarity to ground-truth (structural similarity index 0.86 ± 0.03 , Dice 0.97 ± 0.05).
- Radiologists rated synthetic MRI similar to ground-truth MRI for diagnosability, quality, artifacts, anatomy, and visual Turing test.
- LI-RADS hepatocellular carcinoma tumor features were well preserved in synthetic MRI (accuracy 0.76–0.86).
- Deep learning-based approaches may reduce contrast medium use, scanning time, and costs in contrast-enhanced liver MRI workflows.

Impact and implications:

This work demonstrates the early feasibility of generating high-quality, 3D contrast-enhanced multiphasic liver MRI exams from pre-contrast sequences, with synthetic exams showing strong agreement with ground-truth across quantitative metrics and key qualitative criteria, including the visual Turing test, image quality, disease diagnosability, anatomic accuracy, artifact severity, and HCC Liver Imaging Reporting and Data System features. Despite the model currently representing a proof of concept based on a moderate single-center dataset, with a need for larger multi-center studies and external validation, the results highlight the potential to transform liver MRI workflows by reducing contrast media costs and potential side effects, significantly shortening acquisition time, especially the prolonged 20-min hepatobiliary phase, and improving accessibility for patients unable to tolerate contrast-enhanced MRI because of renal impairment, contrast agent allergy, or claustrophobia.

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

Sara A. Abosabie^{1,2}, Salma A.S. Abosabie^{1,2,3,4}, Weicheng Dai¹, Junlin Yang¹, Moritz Gross^{1,2}, Jeffrey Weinreb¹, Margarita V. Revzin¹, Gaurav Parmar¹, Chenyu You^{1,5}, MingDe Lin^{1,6}, Olivia Gaddum^{1,2}, Bernhard Gebauer², Lynn Jeanette Savic^{2,7,8}, David C. Madoff^{1,9,10}, James S. Duncan^{1,11,12}, Julius Chapiro^{1,11,12,*}

JHEP Reports 2026. vol. 8 | 1–11



Background & Aims: There is growing interest in reducing contrast medium use and the lengthy scan duration in liver imaging. This proof of concept study evaluated the feasibility of deep learning-based generation of synthetic 3D liver contrast-enhanced multiphasic magnetic resonance imaging (MRI) exams, which are similar to ground-truth exams in hepatocellular carcinoma and cirrhosis.

Methods: MRI exams from patients with hepatocellular carcinoma (HCC) or cirrhosis at a single academic center were retrospectively collected. A 3D cycle-consistent generative adversarial network was trained to generate synthetic 3D T1-weighted contrast-enhanced multiphasic liver MRI exams, including arterial, portal venous, delayed, and hepatobiliary phases, using two pre-contrast T1-weighted and T2-weighted input phases. Quantitative performance evaluated similarity, error, and overlap metrics between synthetic and ground-truth exams. For the qualitative multireader study, three blinded radiologists assessed the ground-truth and synthetic MRI exams using a comprehensive questionnaire. Questionnaire tasks 1–5 comprised: visual Turing test (ground-truth vs. synthetic nature), image quality, anatomic accuracy, disease diagnosability, and artifacts. Task 6 comprised Liver Imaging Reporting and Data System features.

Results: The study included 3,198 MRI phases from 533 MRI exams from 185 patients with HCC (mean age, 62.1 years $\pm$ 9.7 [SD]; 141 men) and 182 patients with cirrhosis (54.4 years $\pm$ 10.0; 111 men). Synthetic MRI exams achieved high quantitative and qualitative similarity to ground-truth exams. Quantitative analysis demonstrated high structural similarity index (0.86 ± 0.03), overlap (0.97 ± 0.05), and low symmetric mean absolute percent error ($0.63 \pm 0.23\%$). The qualitative multireader study showed no significant difference in tasks 1–5 ($p = 0.06$ – 0.50) and high performance metrics in task 6 (accuracy: 0.76–0.86; precision: 0.96–1.00) with moderate to perfect Fleiss's Kappa inter-rater agreement (0.58 – 1.00 , $p < 0.001$).

Conclusions: Deep learning enabled the generation of synthetic 3D liver contrast-enhanced multiphasic MRI exams from pre-contrast sequences, achieving high quantitative and qualitative similarity to ground-truth images.

© 2026 The Authors. Published by Elsevier B.V. on behalf of European Association for the Study of the Liver (EASL). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Three-dimensional liver contrast-enhanced multiphasic magnetic resonance imaging (CE-mpMRI) has a key role in the diagnosis and evaluation of a variety of pathological liver conditions, including hepatocellular carcinoma (HCC) and chronic liver disease, particularly cirrhosis.^{1–3} Patients with cirrhosis have a high risk of developing HCC, and complications of portal hypertension are common causes of death, necessitating regular surveillance for early HCC detection and timely management.⁴ Together, these needs create substantial demand for frequent liver CE-mpMRI in clinical practice for patients with cirrhosis and HCC.

While gadolinium-based contrast media enhance diagnostic quality in many liver imaging applications, they also pose significant drawbacks, including longer scan duration,

especially when acquiring the hepatobiliary phase, nephrotoxicity, potentially severe allergic-like reactions, nausea, headache, risk of anaphylaxis, and environmental concerns.^{5,6} Contrast media use is contraindicated in patients with advanced renal disease or during pregnancy, and there is uncertainty about the long-term effects of gadolinium intracranial deposition.^{7,8} In addition, abdominal CE-mpMRI has limited accessibility and high costs.^{9,10} The required long scan duration with multiple-phase acquisitions and breath holding^{9,10} could even cause motion artifacts and failure to capture the optimal phase during acquisition.¹¹ Thus, shorter scan durations, less contrast media injection, and reduced need for repeated breath-holds, particularly difficult for older, pediatric, or critically ill patients, can improve patient comfort and compliance, reduce the need for sedation, and help avoid costly rescans.¹²

* Corresponding author. Address: Department of Radiology and Biomedical Imaging, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA. Tel.: +1 203 785 2428.

E-mail address: julius.chapiro@yale.edu (J. Chapiro).

<https://doi.org/10.1016/j.jhepr.2026.101813>



In response to the growing interest in reducing reliance on lengthy imaging protocols and contrast media use, without compromising diagnostic accuracy, recent studies have explored deep learning-based generation of synthetic contrast-enhanced MRI in regions such as the brain, breast, and nasopharynx.^{13–18} Similarly, there is a growing need for liver CE-mpMRI synthesis from pre-contrast MRI phases, which could potentially shorten scan duration, especially for hepatobiliary phase acquisition, address contrast media shortages scenarios,¹⁹ increase patient comfort, lower contrast media exposure and associated costs and risks, and improve daily liver MRI throughput. However, deep learning-based generation of synthetic liver CE-mpMRI remains largely unexplored.

This proof of concept study aimed to fill this unmet need by evaluating the feasibility of using 3D cycle-consistent generative adversarial network (CycleGAN) deep learning models,²⁰ which use pre-contrast input phases to generate synthetic 3D T1-weighted liver CE-mpMRI phases (arterial, portal venous, delayed, and hepatobiliary), with comparable image quality to ground-truth ones, in patients with HCC and cirrhosis.

Patients and methods

Study group

The institutional review board approved this retrospective Health Insurance Portability and Accountability Act-compliant study and provided a waiver for the requirement for written informed consent. Inclusion criteria were patients aged 18 years or older; with at least one treatment-naïve HCC lesion (HCC cohort) OR only cirrhosis without HCC (cirrhosis cohort) confirmed by clinical and radiological evaluation; who had liver CE-mpMRI, including T1-weighted and T2-weighted pre-contrast phases and T1-weighted post-contrast phases (arterial, portal venous, delayed, and hepatobiliary phase acquired using intravenous gadoxetate disodium [Eovist; Bayer, New Jersey, US]). There was no overlap of the MRI exams used between the HCC and cirrhosis cohorts. Exclusion criteria included severe MRI artifacts obscuring liver structures, infiltrative HCC, incomplete imaging phases, or liver metastases from another primary malignancy. Eligible patients were identified and included consecutively using Epic (Epic Systems Corporation, USA) and Montage (Montage Healthcare Solutions, USA) search tools for liver CE-mpMRI between 2009 and 2023.

Model training and inputs

For all CycleGAN models explained in this section and the upcoming methodology sections, 80% of the MRI dataset was used for the training set, and 20% was used for the completely independent test set. Within the 80% training set, a 10-fold cross-validation was used. Pre-contrast T1-weighted and T2-weighted phases were used as inputs for the models.

The ‘target model’, named as such for its maximum efficiency, was a ‘multiphase combined cohort’ model trained using 10-fold cross-validation to simultaneously generate the four T1-weighted post-contrast phases (arterial, portal venous, delayed, and hepatobiliary) for both HCC and cirrhosis MRI exams (‘combined cohort’). Details, including MRI pre-processing, data augmentation, liver and tumor

segmentations, and CycleGAN architecture as adapted from Zhu, Park *et al.*²⁰ are included in [Appendix S1](#).

Quantitative analysis for target model: Similarity and error metrics

Similarity and error metrics were calculated on the 20% test set across the 10 cross-validation folds between the ground-truth and corresponding synthetic CE-mpMRI exams from the target model in the liver and tumor regions using a published method.¹⁴ Similarity metrics included structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR). Error metrics included normalized root mean square error (NRMSE), symmetric mean absolute percentage error (SMAPE), log accuracy ratio (LogAC), median symmetric accuracy (MedSymAC), and dissimilarity index (DSSIM). Higher values in similarity metrics (SSIM and PSNR) and lower values in error metrics (NRMSE, SMAPE, LogAC, MedSymAC, and DSSIM) indicated better performance.

Quantitative analysis for target model: Dice overlap

The Dice Similarity Coefficient for overlap was used to quantify the spatial overlap between ground-truth and corresponding synthetic liver and tumor segmentation masks.

Single and multiphase generation for separate and combined cohorts

To assess whether the ‘multiphase combined cohort’ target model offered optimal performance and to ensure that it did not underperform compared with alternative CycleGAN models that generate a single post-contrast phase or multiple post-contrast phases for either separate (HCC only or cirrhosis only) or combined cohorts, 14 additional models were trained, each for the first cross-validation fold. Their similarity and error metrics were compared with those of the first cross-validation fold of the target model, as detailed in [Appendix S1](#).

Relative contribution of input phases

To evaluate the relative contribution of the pre-contrast T1-weighted and T2-weighted input phases to the target model, two additional CycleGAN models were trained for the first cross-validation fold using either only pre-contrast T1-weighted or only pre-contrast T2-weighted phases as input to generate all four postcontrast phases for the combined cohort. Their similarity and error metrics were compared with those of the first cross-validation fold of the target model.

Qualitative analysis for target model: Multireader study

A multireader study was conducted by three blinded expert radiologist readers (JW, MVR, and GP) with 11, 22, and 40 years of experience in abdominal imaging, respectively, including liver MRI interpretation. 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. A comprehensive six-task questionnaire was developed based on a detailed review of established literature^{13–16,21} and modified as necessary to fit the specific aims of this study. The

final questionnaire was reviewed and refined by a board-certified radiologist (JC), with over 15 years of experience, who was not involved as a study reader. The six questionnaire tasks aimed to assess visual Turing test (ground-truth vs. synthetic nature), image quality, anatomic accuracy, disease diagnosability, artifacts, and the presence of major HCC Liver Imaging Reporting and Data System (LI-RADS) features.²² Details are provided in Table 1 and Appendix S1.

Statistical analysis

Quantitative metrics were compared using two-tailed Wilcoxon signed-rank or Mann–Whitney *U* test for paired and unpaired data, respectively. Multireader study questionnaire evaluations were evaluated using a two-tailed McNemar test. Fleiss's Kappa was used to assess interrater agreement and reliability. A *p* value <0.05 was considered statistically significant, with Bonferroni correction applied to control for multiple comparisons, when applicable. Statistical analyses were performed using SciPy (version 1.13.1, www.scipy.org) and statsmodels (version 0.14.2, www.statsmodels.org).

Results

Study group

The study included 3,198 MRI phases (1,560 HCC MRI phases and 1,638 cirrhosis MRI phases) from a total of 533 liver CE-mpMRI exams, comprising 260 HCC and 273 cirrhosis CE-mpMRI exams (six phases: pre-contrast T1-weighted, pre-contrast T2-weighted, arterial, portal venous, delayed, and hepatobiliary in each MRI exam). The MRI exams were from 185 patients with HCC (mean age, 62.1 years ± 9.7 [SD]; 141 men) and 182 patients with cirrhosis (mean age, 54.4 years ± 10.0 [SD]; 111 men). Fig. 1 shows the selection flowchart of the MRI exams. Patient demographics information and HCC tumor characteristics are provided in Tables 2 and 3.

Model training and inputs

For all models, 2,556 MRI phases (1,248 HCC MRI phases and 1,308 cirrhosis MRI phases) from a total of 426 CE-

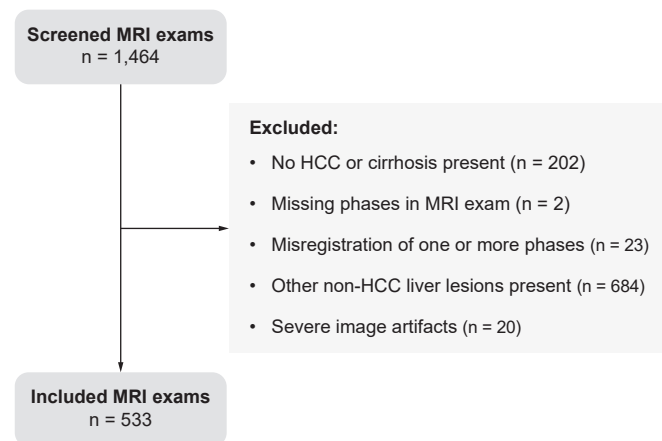


Fig. 1. Selection flowchart of MRI exams included in the study. HCC, hepatocellular carcinoma; MRI, magnetic resonance imaging.

Table 1. Questionnaire of the multireader study.

Task	Rationale	Image dataset presented to each reader	Questionnaire	Possible answers in questionnaire
Task 1	Visual Turing test (ground-truth vs. synthetic nature): to determine whether readers identify MRI exam as ground-truth or synthetic	Tasks 1–5: 107 ground-truth and synthetic liver T1-weighted CE-mpMRI exams of 20% test set*	Is this MRI exam ground-truth or synthetic?	Ground-truth/synthetic
Task 2	Image quality assessment: to assess qualitative image quality of MRI exam	• 52 HCC MRI exams • 55 cirrhosis MRI exams	What is the image quality of this exam?	Likert scale: 1–5†
Task 3	Radiological diagnosability of disease: to determine whether HCC or cirrhosis can be diagnosed on MRI exam		Are you able to diagnose HCC (in HCC cohort) or cirrhosis (in cirrhosis cohort) on the MRI exam?	Yes/no
Task 4	Liver anatomic accuracy: to determine whether liver anatomy is accurate on MRI exam		Is the liver anatomy accurate on this exam?	Yes/no
Task 5	Artifact assessment: to evaluate severity of artifacts on MRI exam		How severe are the artifacts on this exam?	Likert scale: 0–3‡
Task 6	HCC characterization and LI-RADS assessment: to evaluate LI-RADS features within target HCC lesion in MRI exam	52 HCC ground-truth and synthetic liver T1-weighted CE-mpMRI exams of 20% test set*	Does the target HCC lesion demonstrate: A. APHE on arterial phase? B. non-peripheral washout on portal venous/delayed phases? C. enhancing capsule surrounding the lesion on portal venous/delayed phases? D. hypointensity on hepatobiliary phase?	A–D: yes/no

APHE, non-rim arterial hyperenhancement; CE-mpMRI, contrast-enhanced multiphasic magnetic resonance imaging; HCC, hepatocellular carcinoma; LI-RADS, Liver Imaging Reporting and Data System.

*Each CE-mpMRI exam comprised either four ground-truth or four synthetic post-contrast phases (arterial, portal venous, delayed, and hepatobiliary).

†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.

‡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’.

Table 2. Patient clinical characteristics.

Characteristic	Value for HCC cohort	Value for cirrhosis cohort
No. of patients	185	182
No. of MRI phases	1,560	1,638
No. of CE-mpMRI exams	260	273
Age (year)*	62.1 ± 9.7	54.4 ± 10.0
Sex		
Female	44 (23.8)	71 (39.0)
Male	141 (76.2)	111 (61.0)
Cirrhosis present	152	182
Cirrhosis etiology		
HCV	51 (33.6)	55 (30.2)
HBV	2 (1.3)	4 (2.2)
Alcoholic	11 (7.2)	46 (25.3)
Alcoholic/HCV	6 (3.9)	4 (2.2)
MASH	9 (5.9)	15 (8.2)
Autoimmune	0 (0.0)	5 (2.7)
Cryptogenic	73 (48.0)	53 (29.1)

Unless otherwise noted, data are the number of patients, with percentages in parentheses.

CE-mpMRI: contrast-enhanced multiphase magnetic resonance imaging; HCC, hepatocellular carcinoma; MASH, metabolic dysfunction-associated steatohepatitis.

*Age is presented as mean ± SD.

mpMRI exams (208 HCC and 218 cirrhosis exams) were used for the 80% training set. In addition, 642 MRI phases (312 HCC MRI phases and 330 cirrhosis MRI phases) from a total of 107 CE-mpMRI exams (52 HCC and 55 cirrhosis exams) were used for the 20% test set.

Quantitative analysis for target model: similarity and error metrics

Quantitative analysis in the form of similarity and error metrics for the target model showed high SSIM, moderate PSNR, and low error metrics between ground-truth and synthetic post-contrast images, both in the liver (combined cohort) and within the tumor in the HCC cohort (Fig. 2 and Table S1). For the liver region, the mean ± SD for SSIM was 0.862 ± 0.030 , for PSNR was 22.394 ± 2.988 dB, for NRMSE was 0.081 ± 0.030 , for SMAPE was $0.625\% \pm 0.229\%$, for LogAC was 0.013 ± 0.005 , for MedSymAC was 0.001 ± 0.001 , and for DSSIM was 0.069 ± 0.015 . For the tumor region, the mean value ± SD for SSIM was 0.758 ± 0.107 , for PSNR was 19.436 ± 4.196 dB, for NRMSE was 0.120 ± 0.063 , for SMAPE was $1.043 \pm 0.733\%$, for LogAC was 0.021 ± 0.015 , for MedSymAC was 0.004 ± 0.008 , and for DSSIM was 0.121 ± 0.054 . In all metrics, there was no significant difference in the liver quantitative metrics between the HCC and cirrhosis cohorts (two-tailed Mann-Whitney *U* test: $p = 0.192$ – 0.992). In the HCC cohort, all liver metrics significantly outperformed tumor metrics (Fig. 3; two-tailed Wilcoxon signed-rank test: all $p < 0.001$).

Table 3. HCC lesion characteristics.

Characteristic	Value
No. of patients	185
No. of target HCC lesions	260
No. of MRI phases	1,560
No. of CE-mpMRI exams	260
Target HCC lesion location	
Right lobe	164 (63.1)
Left lobe	92 (35.4)
Bilobar	4 (1.5)

CE-mpMRI, contrast-enhanced multiphase magnetic resonance imaging; HCC, hepatocellular carcinoma.

Quantitative analysis for target model: Dice overlap

The Dice Similarity Coefficient for overlap between ground-truth and corresponding synthetic liver segmentation masks demonstrated a mean overlap of 0.97 ± 0.05 for the liver in the combined cohort. In the HCC cohort, the mean Dice Similarity Coefficient for overlap for tumor segmentation was 0.85 ± 0.14 .

Relative contribution of input phases

The target model with pre-contrast T1-weighted and T2-weighted phases as inputs significantly outperformed the pre-contrast T2-weighted-only model in both liver and tumor regions in all similarity and error metrics (two-tailed Wilcoxon signed-rank test: all $p < 0.001$) except for MedSymAC in the tumor, where the models performed similarly. In addition, the target model with pre-contrast T1-weighted and T2-weighted phases as inputs significantly outperformed the pre-contrast T1-weighted-only model in the DSSIM, SSIM, and PSNR of the liver region, but the pre-contrast T1-weighted-only model outperformed the target model in the liver MedSymAC metric (two-tailed Wilcoxon signed-rank test: all $p < 0.001$). The target model performed similarly to the pre-contrast T1-weighted-only model in the tumor region except for PSNR, where the target model significantly outperformed the latter (two-tailed Wilcoxon signed-rank test: $p = 0.002$). The results are shown in Fig. S1.

Single and multiphase generation for separate and combined cohorts

The target model and all 14 additional models yielded synthetic post-contrast T1-weighted images that were quantitatively similar to ground-truth post-contrast images (high SSIM and low NRMSE in the liver). The target model performed similarly to, or significantly outperformed, other models (eight of 14 models) more often than being outperformed by other models (two-tailed Wilcoxon signed-rank test: all $p < 0.001$), as shown in Fig. S2. This supports its selection as the optimal model, because it uses the largest input data volume (both pre-contrast T1-weighted and T2-weighted phases as inputs) for the combined HCC and cirrhosis cohorts, enabling greater

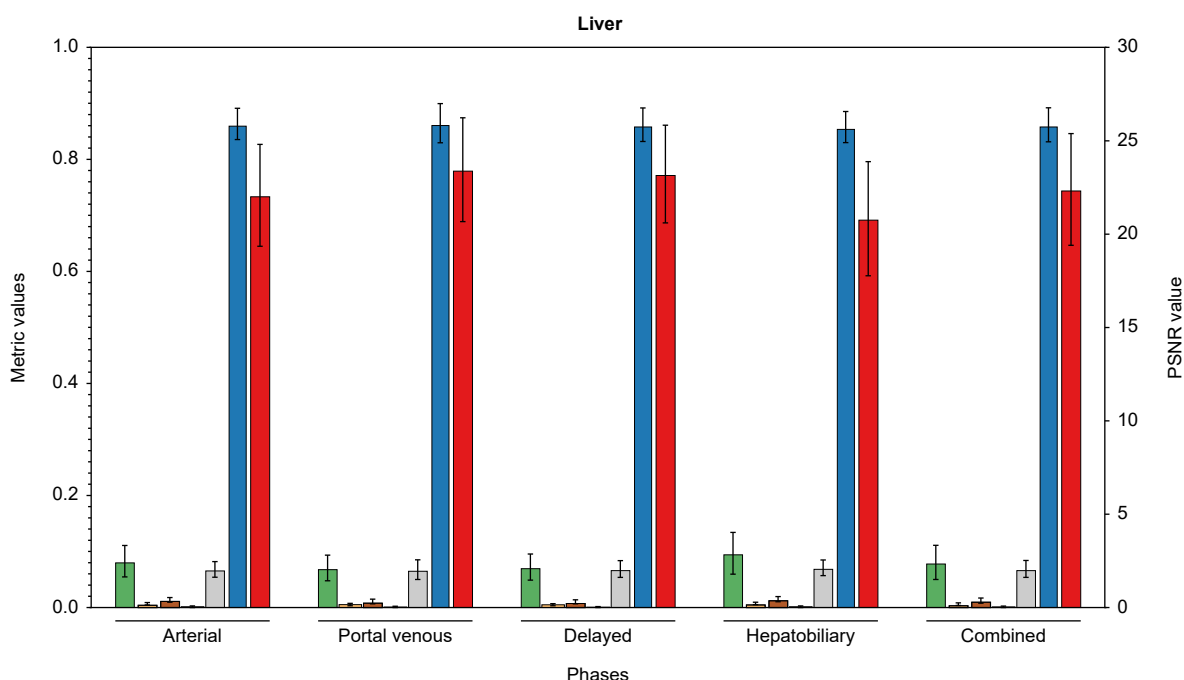
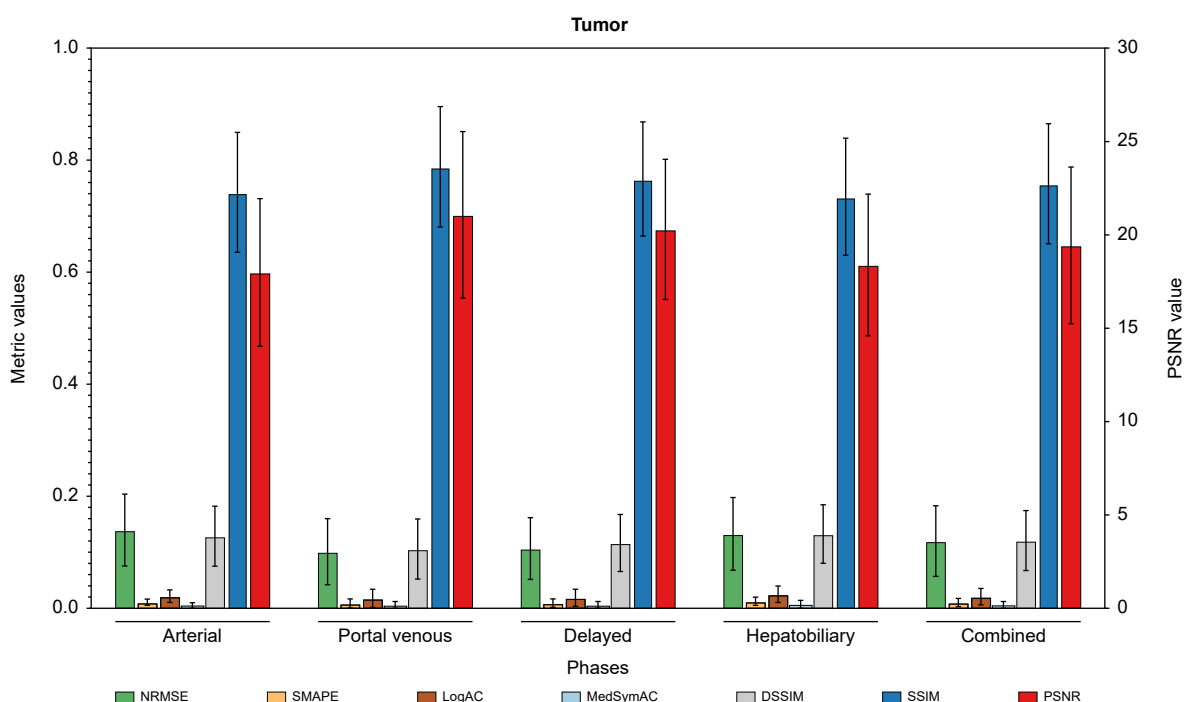
A**B**

Fig. 2. Quantitative results of the target model. Quantitative metric results of the target model between the four post-contrast phases of ground-truth and corresponding synthetic CE-mpMRI exams of the 20% test set are shown in (A) for the liver of the combined HCC and cirrhosis cohort and (B) for the tumor in the HCC cohort. The metrics are shown for each individual phase and on average across the four phases combined (average value per full CE-mpMRI exam). 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. CE-mpMRI, contrast-enhanced multiphasic Magnetic Resonance Imaging; DSSIM, structural dissimilarity index; HCC, hepatocellular carcinoma; LogAC, log accuracy ratio; MedSymAC, median symmetric accuracy; NRMSE, normalized root mean square error; PSNR, peak signal-to-noise ratio; SMAPE, symmetric mean absolute percent error; SSIM, structural similarity index.

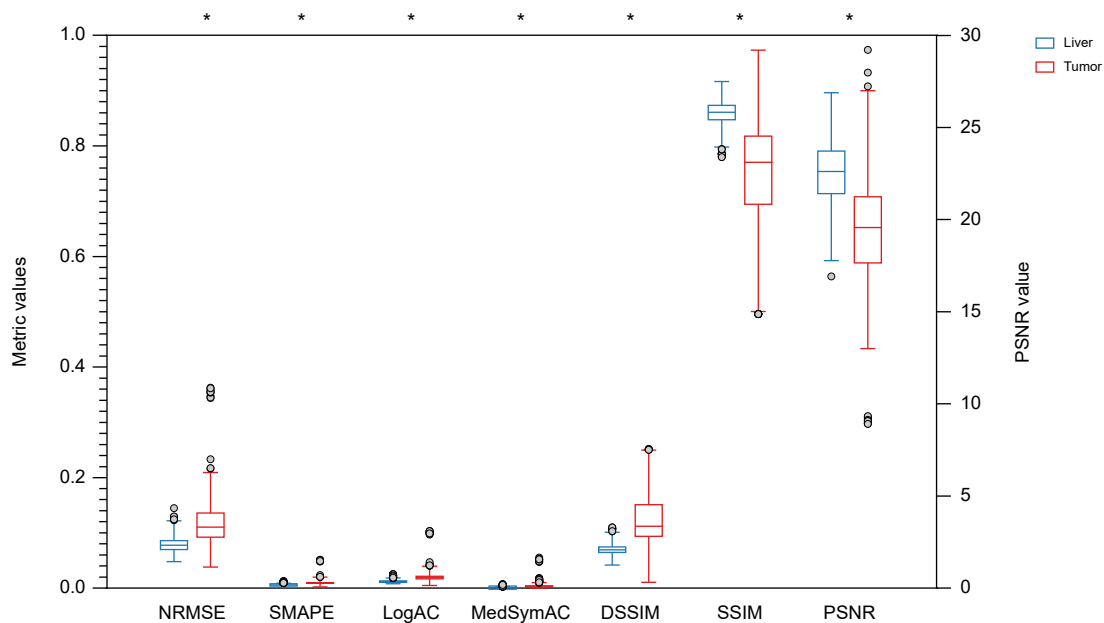


Fig. 3. Quantitative results of the liver and tumor in the HCC cohort. Similarity and error metrics between the HCC ground-truth and corresponding synthetic CE-mpMRI exams for the target model are shown. Statistical comparisons were performed between HCC liver and tumor metrics using the two-tailed Wilcoxon signed-rank test. The tests were run based on average results of the combined four post-contrast phases (arterial, portal venous, delayed, and hepatobiliary), that is, average per full CE-mpMRI exam, of the target model. *Indicates that all metrics were significantly better in the liver than in the tumor (two-tailed Wilcoxon signed-rank test: all $p < 0.001$). 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, and DSSIM) indicate better performance. CE-mpMRI, contrast-enhanced multiphase magnetic resonance imaging; DSSIM, structural dissimilarity index; HCC, hepatocellular carcinoma; LogAC, log accuracy ratio; MedSymAC, median symmetric accuracy; NRMSE, normalized root mean square error; PSNR, peak signal-to-noise ratio; SMAPE, symmetric mean absolute percent error; SSIM, structural similarity index.

generalizability and maximum efficiency in synthesizing all four post-contrast MRI phases for the combined cohort simultaneously. This eliminates the need to train separate models for individual HCC or cirrhosis cohorts, or to synthesize only a single post-contrast phase per model, thereby saving significant time, GPU memory, and computational resources.

Qualitative analysis for target model: multireader study

In the multireader study, all 107 ground-truth and synthetic CE-mpMRI exams of the 20% test set of the target model were evaluated by the three readers. On average across the three readers, ground-truth MRI exams in the combined cohort were more often evaluated favorably than were synthetic ones in tasks 1–5, with higher rates of being perceived as ground-truth (97.20% vs. 91.90%), showing ‘diagnostic quality’ (quality scores 3–5) (100% vs. 95.64%), ‘high diagnostic quality’ (quality scores 4–5) (93.46% vs. 85.98%), radiological diagnosability of disease (97.20% vs. 91.90%), liver anatomic accuracy (100.00% vs. 96.26%), and ‘minimal artifacts’ (artifact scores 0–1) (96.26% vs. 92.52%) for ground-truth and synthetic MRI exams, respectively. However, statistically, there was no significant difference between ground-truth and synthetic MRI exams across tasks 1–5 ($p = 0.06$ – 0.50) except that significantly more ground-truth MRI exams were rated as ‘high diagnostic quality’ compared with synthetic ones by reader 1 ($p = 0.01$) (Table 4). In task 6, the model successfully reproduced major LI-RADS features²² (non-rim arterial phase

hyperenhancement, portal venous/delayed washout, and enhancing capsule) and the ancillary hepatobiliary phase hypointensity feature (accuracy 0.76–0.86, precision 0.96–1.00, F1 score 0.85–0.92, sensitivity 0.75–0.87, and specificity 0.73–1.00) (Table 4). Inter-rater agreement was moderate to perfect across all tasks, as indicated by Fleiss’s Kappa values ranging from 0.58 to 1.00 ($p < 0.001$) (Table 4).

Successful visual examples of ground-truth and synthetic CE-mpMRI exams for HCC and cirrhosis are presented in Figs 4 and 5. Examples of specific model failures in generating synthetic versus ground-truth MRI exams are shown in Fig. 6.

Discussion

This study presents a deep learning approach for the generation of synthetic liver MRI exams in patients with HCC or cirrhosis. Our primary results show the feasibility of using 3D deep learning CycleGAN models to directly generate fully volumetric 3D liver synthetic post-contrast CE-mpMRI exams that are quantitatively and qualitatively similar to ground-truth exams using pre-contrast input phases. First, in terms of quantitative metrics, the synthetic postcontrast CE-mpMRI exams of the target model demonstrated high SSIM (mean 0.86 ± 0.03 vs. 0.76 ± 0.11), high overlap (0.97 ± 0.05 vs. 0.85 ± 0.14), and low error (NRMSE 0.08 ± 0.03 vs. 0.12 ± 0.06 ; SMAPE $0.63 \pm 0.23\%$ vs. $1.04 \pm 0.73\%$) compared with ground-truth exams in liver and tumor regions, respectively. Second, the multireader qualitative study (visual Turing test,

Table 4. Results of the multi-reader study of McNemar test and interrater agreement (Fleiss's Kappa) results of the questionnaire tasks 1–5 (for the entire cohort) and average performance metrics based on the three readers' evaluations and interrater agreement (Fleiss's Kappa) results of the LI-RADS features of the HCC cohort in questionnaire task 6.

McNemar test and interrater agreement (Fleiss's Kappa) results of questionnaire tasks 1–5 (for entire cohort)										
Task	Rationale	McNemar test <i>p</i> values								
		Reader 1	Reader 2	Reader 3	Interrater agreement (Fleiss's Kappa)*					
Task 1	Visual Turing Test (ground-truth vs. synthetic)	0.13	0.09	0.18	0.58 (0.34–0.76)					
Task 2	Diagnostic quality (quality scores 3–5)	0.06	0.06	0.13	0.93 (0.71–1.00)					
Task 2	High diagnostic quality (quality scores 4–5)	0.01	0.11	0.07	0.88 (0.78–0.96)					
Task 3	Radiological diagnosability of disease	0.06	0.07	0.07	0.70 (0.48–0.84)					
Task 4	Liver anatomic accuracy	0.06	0.06	0.50	0.75 (0.49–1.00)					
Task 5	Minimal artifacts (artifact scores 0–1)	0.22	0.22	0.22	1.00 (1.00–1.00)					
Average performance metrics based on the three readers' evaluations and interrater agreement (Fleiss's Kappa) results of the LI-RADS features of the HCC cohort in questionnaire task 6										
Task 6: HCC LI-RADS Feature		MRI phase		Accuracy	Precision	F1 score	Sensitivity	Specificity	Interrater agreement (Fleiss's Kappa)*	
Non-rim APHE		Arterial		0.86	0.97	0.92	0.87	0.73	0.92 (0.80–1.00)	
Non-peripheral washout		Portal venous/delayed		0.76	0.97	0.85	0.75	0.80	0.94 (0.86–1.00)	
Enhancing capsule		Portal venous/delayed		0.84	0.96	0.86	0.77	0.95	0.96 (0.91–1.00)	
Hypointensity		Hepatobiliary		0.83	1.00	0.91	0.83	1.00	0.88 (0.75–0.98)	

APHE, non-rim arterial hyperenhancement; HCC, hepatocellular carcinoma; LI-RADS, Liver Imaging Reporting and Data System; MRI, magnetic resonance imaging.

*All Fleiss's Kappa $p < 0.001$. Significant McNemar p values are bolded.

image quality, disease diagnosability, anatomic accuracy, and artifact severity) showed no significant difference between ground-truth and synthetic MRI exams ($p = 0.06$ – 0.50) and high performance metrics (accuracy: 72–86%; precision: 96–100%) for HCC LI-RADS features.

Interestingly, the target model demonstrated high SSIM but moderate PSNR (mean 22.39 ± 2.99 dB, 19.44 ± 4.20 dB in liver and tumor, respectively). This is expected because SSIM measures image structural properties, provides an accurate assessment of MR image quality, and correlates most with qualitative evaluation by expert readers.²³ In addition, although the quantitative metrics of HCC tumors also showed good performance, all quantitative metrics in the liver region were significantly better ($p < 0.001$), consistent with the greater heterogeneity and complexity of the tumor.²⁴

In contrast to our full 3D MRI generation approach, previous deep learning studies^{16–18,25–28} mainly focused on 2D-based image synthesis across various modalities (X-ray, computed tomography, and MRI) because of the complexity, high computational, and GPU memory demands of 3D synthesis; however, 2D-based image synthesis approaches can negatively affect performance because they are prone to slice-to-slice inconsistencies.²⁹ Previous deep learning studies with 3D post-contrast MRI synthesis were mainly in brain or breast imaging, focused on a single disease, synthesis of only one post-contrast phase, and required five to seven pre-contrast inputs for the model.^{13,14} By contrast, this study was considerably more challenging: using only two pre-contrast input phases, we synthesized all four post-contrast phases of liver MRI exams, across two diseases (HCC and cirrhosis), and evaluated the synthetic images using a more comprehensive questionnaire covering a wider range of image characteristics. To the best of our knowledge, no previous studies have successfully generated 3D liver CE-mpMRI exams using pre-contrast phases.

As used in previous studies,^{16,30} we adopted a GAN architecture, known for high fidelity and speed,²⁹ and specifically used a more advanced variant, CycleGAN, with two generators and two discriminators for better structural and contextual fidelity of the synthetic images.²⁰ Prior studies, including a multicenter effort generating synthetic brain MRI, demonstrated that a GAN-based deep learning architecture significantly outperformed a U-Net-based architecture in producing realistic synthetic MRI exams (SSIM 0.818, $p < 0.0001$).³¹ Similarly, a study using a convolutional neural network for brain MRI synthesis encouraged future studies to use GANs to enhance image quality; notably, the synthetic images from that study exhibited apparent blurring, an issue not observed in our GAN-generated images.¹⁴

This study has several limitations. First, this was a single-center study, limiting the generalizability to broader patient populations or datasets acquired using different imaging protocols and scanner types. Second, we focused on HCC and cirrhosis only, while the applicability in normal liver tissue and other common malignant and benign liver lesions remains unexplored. Finally, while our models demonstrate promising performance metrics, the variability in liver pathology, including differences in lesion size, vascularity, and liver function, presents challenges for consistent model performance, particularly in cases with complex or atypical presentations.

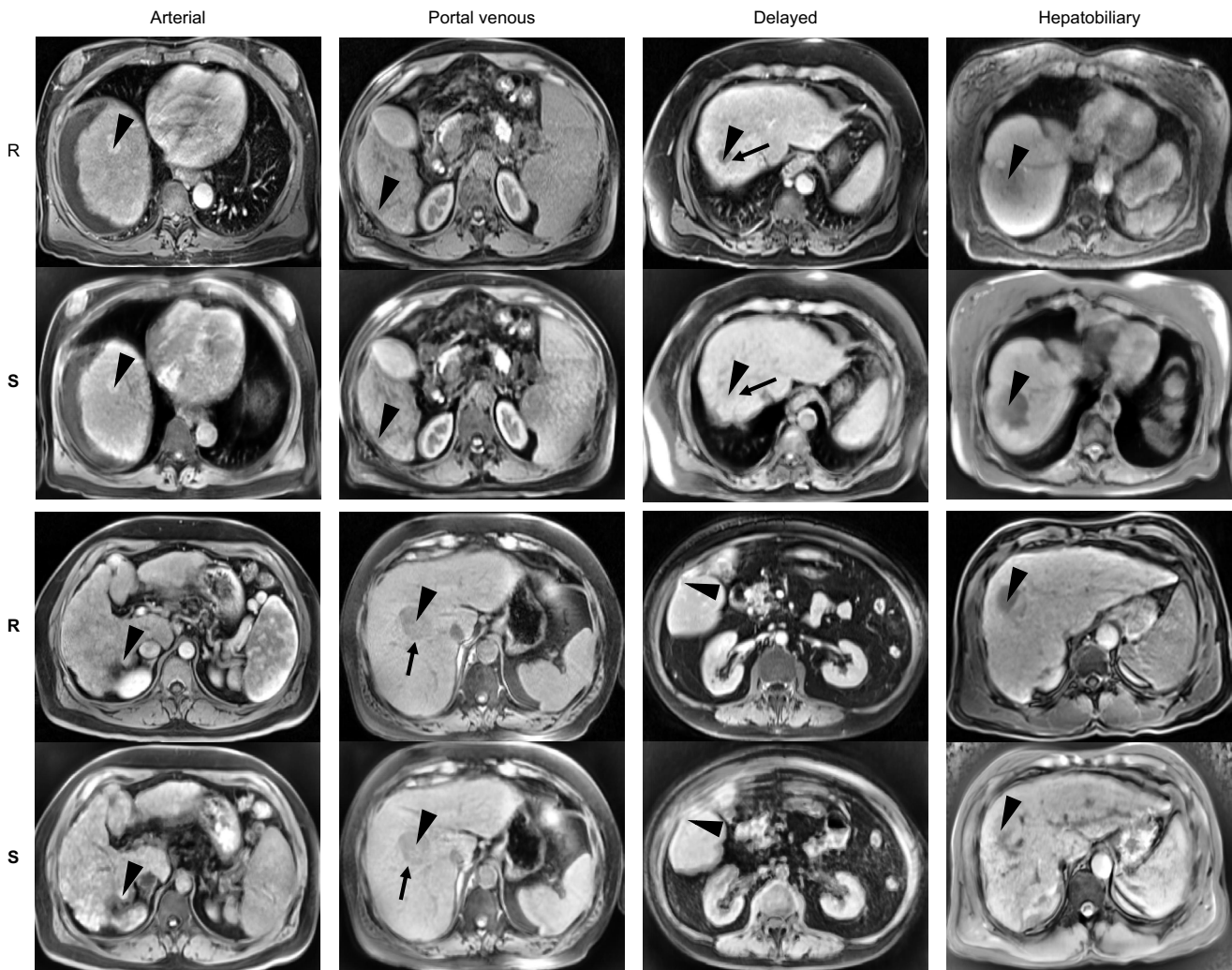


Fig. 4. HCC ground-truth and synthetic MRI. Examples are shown of ground-truth (real) (R) and synthetic (S) liver post-contrast MRI phases of patients with HCC, demonstrating LI-RADS features. Both ground-truth and synthetic phases show major LI-RADS features of non-rim arterial hyperenhancement (arrowhead) on the arterial phase, as well as non-peripheral washout (arrowhead) and an enhancing capsule (arrow) surrounding the HCC lesion on the portal venous and/or delayed phases. The ancillary LI-RADS feature of hypointensity on the hepatobiliary phase (arrowhead) is also shown. HCC, hepatocellular carcinoma; LI-RADS: Liver Imaging Reporting and Data System; MRI, magnetic resonance imaging.

These highly promising findings highlight the potential to transform liver MRI workflows by reducing contrast media cost and potential side effects, significantly shortening acquisition time, especially the prolonged hepatobiliary phase acquisition of 20 min, and improving accessibility for patients unable to tolerate it because of renal impairment, contrast agent allergy, or claustrophobia. Despite these promising results, it is important to emphasize that our deep learning model currently represents a proof of concept rather than a clinically deployable tool. The dataset size remains limited, and no external validation was performed, which restricts the generalizability. Nevertheless, it establishes a strong proof of concept framework and paves the way for future studies. Looking ahead, further research can build on these findings by model adaptation for other liver diseases, such as metastases or benign liver lesions, validation in large-scale, multicenter prospective studies, and additional pre-contrast input phases to potentially enhance model performance and enable the generation of the same four post-contrast phases

used in this study (or more, if clinically relevant). Furthermore, a dedicated hallucination analysis, for example by systematically cross-checking synthetic and ground-truth image pairs, will be important to explicitly assess and mitigate the hallucination risk of the model. A cost-effectiveness analysis is also required in future studies to evaluate potential quantitative cost reductions. Finally, this deep learning model could be adapted for related tasks, such as postcontrast CT synthesis or even intermodality MRI generation from computed tomography and vice versa.

In conclusion, the proposed deep learning model demonstrates early feasibility of using pre-contrast MRI sequences to generate high-quality 3D synthetic postcontrast CE-mpMRI, potentially offering a pipeline toward a viable alternative for conventional CE-mpMRI. The synthetic MRI exams were similar to ground-truth ones across quantitative metrics and several key qualitative aspects, including visual Turing test, image quality, disease diagnosability, anatomic accuracy, artifact severity, and LI-RADS features.

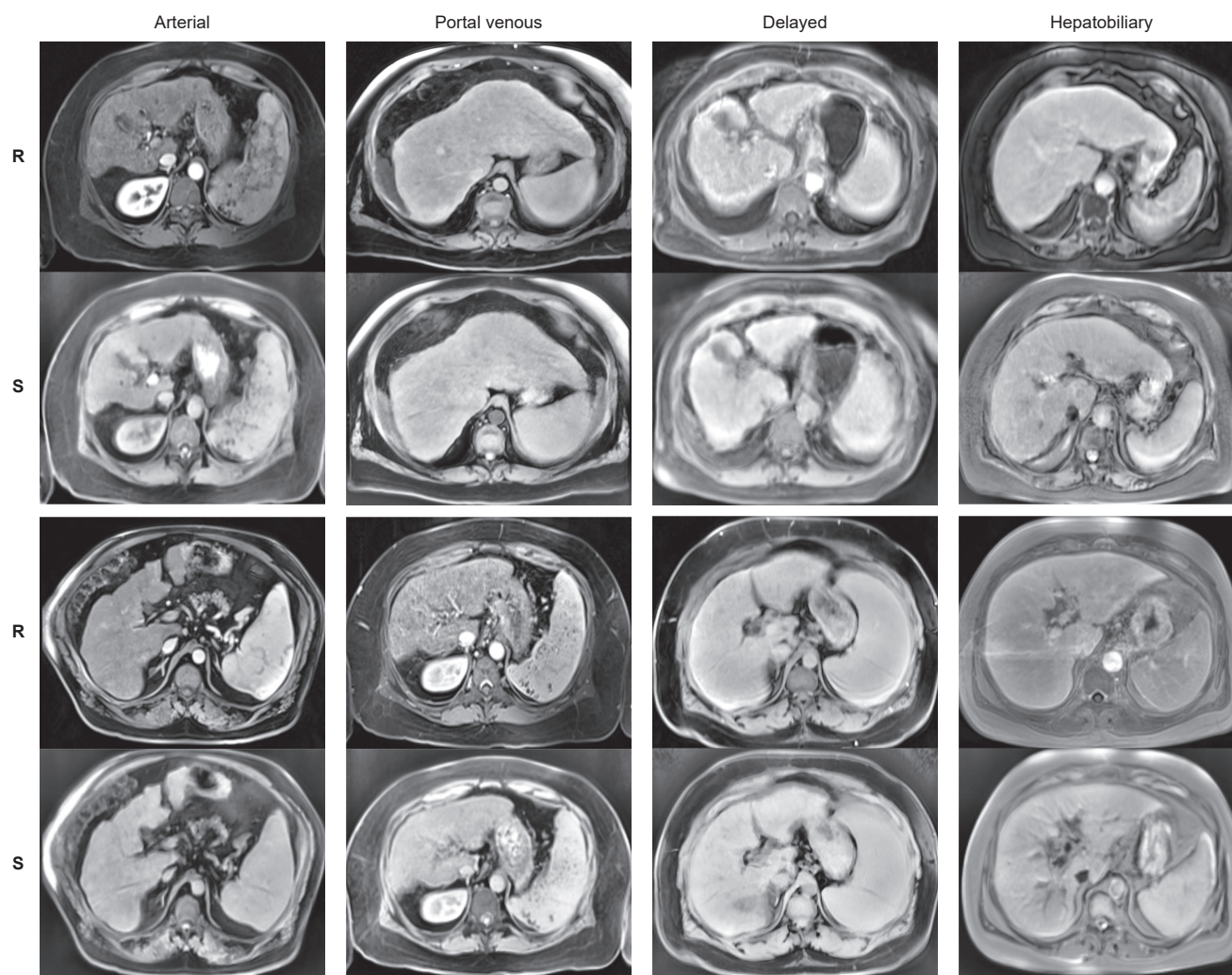


Fig. 5. Cirrhosis ground-truth and synthetic MRI. Examples are shown of ground-truth (real) (R) and synthetic (S) liver post-contrast MRI phases of cirrhosis patients. MRI, magnetic resonance imaging.

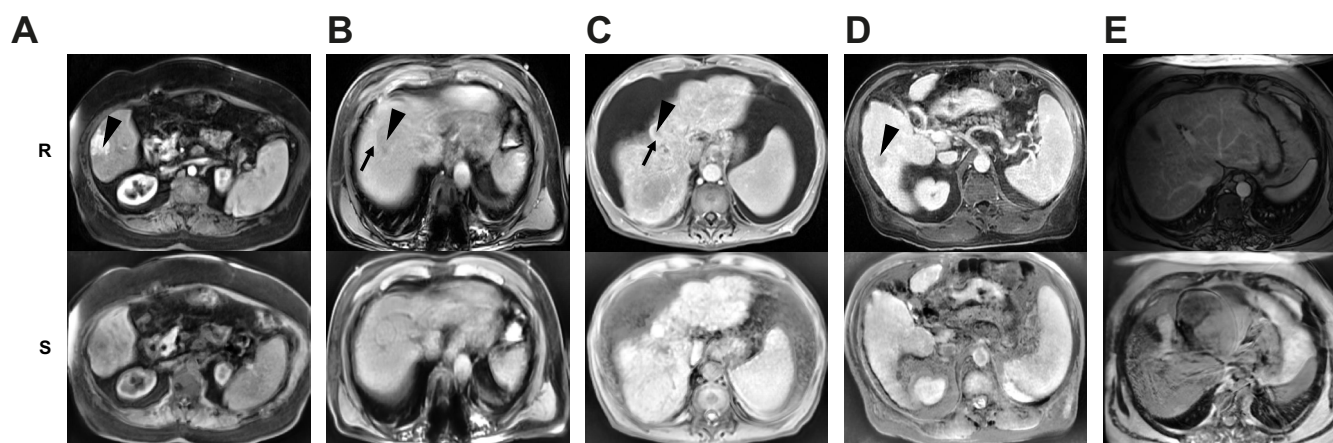


Fig. 6. Cases of model failure in synthetic MRI. Examples are shown of model failure in synthetic MRI phases. (A–D) Instances where the model did not correctly reproduce HCC LI-RADS features present in ground-truth (real) (R) MRI phases in their corresponding synthetic (S) phases. LI-RADS features shown are (A) non-rim arterial hyperenhancement (arrowhead) on the arterial phase, (B,C) non-peripheral washout (arrowhead) and an enhancing capsule (arrow) surrounding the HCC lesion on the portal venous (B) and delayed (C) phases, and (D) hypointensity on the hepatobiliary phase. (E) Synthetic delayed MRI phase of a patient with cirrhosis with incorrect liver anatomy, undiagnosable cirrhosis, and more artifacts compared with its corresponding ground-truth phase. HCC, hepatocellular carcinoma; LI-RADS, Liver Imaging Reporting and Data System; MRI, magnetic resonance imaging.

Affiliations

¹Department of Radiology and Biomedical Imaging, Yale University School of Medicine, New Haven, CT 06520, USA; ²Department of Radiology, Charité-Universitätsmedizin Berlin, Campus Virchow Klinikum, 13353 Berlin, Germany; ³Institute of Experimental Biomedicine, University Hospital Wuerzburg, 97080 Wuerzburg, Germany; ⁴Rudolf Virchow Center for Integrative and Translational Bioimaging, Julius Maximilians University of Wuerzburg, 97080 Wuerzburg, Germany; ⁵Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA; ⁶Visage Imaging, Inc., San Diego, CA 91230, USA; ⁷Berlin Institute of Health at Charité-Universitätsmedizin Berlin, 10115 Berlin, Germany; ⁸Experimental Clinical Research Center (ECRC) at Charité-Universitätsmedizin Berlin and Max-Delbrück-Centrum für Molekulare Medizin (MDC), 13125 Berlin, Germany; ⁹Department of Medicine, Section of Medical Oncology, Yale University School of Medicine, New Haven, CT 06520, USA; ¹⁰Department of Surgery, Section of Surgical Oncology, Yale University School of Medicine, New Haven, CT 06520, USA; ¹¹Department of Medicine, Liver Center, Section of Digestive Diseases, Yale University School of Medicine, New Haven, CT 06520, USA; ¹²Department of Biomedical Engineering, School of Engineering and Applied Sciences, New Haven, CT 06520, USA

Abbreviations

CE-mpMRI, contrast-enhanced multiphase magnetic resonance imaging; CycleGAN, cycle-consistent generative adversarial network; DSSIM, dissimilarity index; HCC, hepatocellular carcinoma; LI-RADS, Liver Imaging Reporting and Data System; LogAC, log accuracy ratio; MedSymAC, median symmetric accuracy; PSNR, peak signal-to-noise ratio; NRMSE, normalized root mean square error; SMAPE, symmetric mean absolute percentage error; SSIM, structural similarity index.

Financial support

This study was funded by the National Institute of Health (NIH/NCI R01 CA206180). This study was supported, in part, by the Yale Liver Center award NIH P30 DK034989 Clinical and Translational core.

Conflicts of interest

SAA and OG received a travel stipend from Rolf W. Günther Stiftung. ML is a Visage Imaging, Inc. employee and stockholder and reports institutional grants from the National Institute of Health (NIH/NCI R01 CA206180 grant). BG received honoraria and travel support over the past 10 years from Parexel/CALYX, C.R. BARD/BD, SIRTex Medical, St. Jude Medical, COOK, AngioDynamics, Pharmcept, Guerbet, Ewimed, Terumo, Roche, Merck, 3M, Beacon Bioscience/ICON, IPSEN, Bayer, Pfizer, Eisai, MSD, and INARI. LJS is a fellow of the BIH Clinician Scientist Program funded by the Charité-Universitätsmedizin Berlin and the Berlin Institute of Health and reports research grants from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG): CRC 1340 'Matrix in Vision' (Project ID 372486779) and DFG research unit FOR5628 'Onco-MRE' (Project ID 513752256), research grants from Berliner Krebsgesellschaft e.V., Charité 3R Replace Reduce Refine, and the Cardiovascular and Interventional Radiological Society of Europe (CIRSE), and research grants and honoraria from Guerbet. JSD and DCM received a NIH/NCI R01 CA206180 grant from the National Institute of Health. JC reports institutional grants from the National Institute of Health (NIH/NCI R01 CA206180 grant), SIO, Guerbet, and Boston Scientific and consulting fees from, and participates in, data safety or advisory boards of AstraZeneca, Eisai, Bayer, Guerbet, and Genetech. The remaining authors declare no conflicts of interest.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Investigation, conceptualization, validation, visualization, resources, software, project administration, data curation, supervision, methodology, formal analysis, manuscript writing: SAA, SASA. Methodology, software, data curation, validation, visualization: WD. Methodology, software, validation: JY. Data curation, methodology, formal analysis, validation: JW, MRV, GP. Software, resources: CY. Methodology, project administration, validation, funding acquisition: ML. Data curation: OG. Software, supervision, project administration, investigation, validation: BG, LJS. Supervision, project administration, resources, validation, funding acquisition: JSD. Conceptualization, supervision, project administration, investigation, software, resources, funding acquisition, formal analysis, validation: DCM, JC. Reviewed the manuscript: All authors. Approved the final manuscript submitted: All authors.

Data availability

Individual deidentified participant data cannot be shared to protect patient privacy and comply with institutional and ethical regulations.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhepr.2026.101813>.

References

Author names in bold designate shared co-first authorship

- [1] Ramalho M, Matos AP, AlObaidy M, et al. Magnetic resonance imaging of the cirrhotic liver: diagnosis of hepatocellular carcinoma and evaluation of response to treatment - Part 1. *Radiol Bras* 2017;50:38–47.
- [2] Taouli B, Ehman RL, Reeder SB. Advanced MRI methods for assessment of chronic liver disease. *AJR Am J Roentgenol* 2009;193:14–27.
- [3] Sangro B, Argemi J, Ronot M, et al., on behalf of the European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* 2025;82(2):315–374.
- [4] Gines P, Krag A, Abralides JG, et al. Liver cirrhosis. *Lancet* 2021;398:1359–1376.
- [5] Beckett KR, Moriarty AK, Langer JM. Safe use of contrast media: what the radiologist needs to know. *Radiographics* 2015;35:1738–1750.
- [6] Dekker HM, Stroomberg GJ, Prokop M. Tackling the increasing contamination of the water supply by iodinated contrast media. *Insights Imaging* 2022;13:30.
- [7] Gulani V, Calamante F, Shellock FG, et al. Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol* 2017;16:564–570.
- [8] Ramalho J, Semelka RC, Ramalho M, et al. Gadolinium-based contrast agent accumulation and toxicity: an update. *Am J Neuroradiol* 2016;37:1192–1198.
- [9] Canellas R, Rosenkrantz AB, Taouli B, et al. Abbreviated MRI protocols for the abdomen. *Radiographics* 2019;39:744–758.
- [10] Maung ST, Tanpowpong N, Satja M, et al. MRI for hepatocellular carcinoma and the role of abbreviated MRI for surveillance of hepatocellular carcinoma. *J Gastroenterol Hepatol* 2024;39:1969–1981.
- [11] Yoon JH, Nickel MD, Peeters JM, et al. Rapid imaging: recent advances in abdominal MRI for reducing acquisition time and its clinical applications. *Korean J Radiol* 2019;20:1597–1615.
- [12] Feng L, Chandarana H. Accelerated abdominal MRI: a review of current methods and applications. *J Magn Reson Imaging* 2025;62:654–672.
- [13] Chung M, Calabrese E, Mongan J, et al. Deep learning to simulate contrast-enhanced breast MRI of invasive breast cancer. *Radiology* 2023;306:e213199.
- [14] 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:e200276.
- [15] 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. *Am J Neuroradiol* 2017;38:1103–1110.
- [16] 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:e222211.
- [17] Li W, Xiao H, Li T, et al. Virtual contrast-enhanced magnetic resonance images synthesis for patients with nasopharyngeal carcinoma using multi-modality-guided synergistic neural network. *Int J Radiat Oncol Biol Phys* 2022;112:1033–1044.

- [18] Murugesan G, Yu FF, Achilleos M, et al. Synthesizing contrast-enhanced MR images from noncontrast MR images using deep learning. *Am J Neuroradiol* 2024;45:312–319.
- [19] Ananthakrishnan L, Kay FU, Zeikus EA, et al. What the baby formula and medical contrast material shortages have in common: insights and recommendations for managing the iodinated contrast media shortage. *Radiol Cardiothorac Imaging* 2022;4:e220101.
- [20] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv*. Published online March 30, 2017. <https://doi.org/10.48550/arXiv.1703.10593>.
- [21] 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;290:649–656.
- [22] 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;289:816–830.
- [23] Sagawa H, Itagaki K, Matsushita T, et al. Evaluation of motion artifacts in brain magnetic resonance images using convolutional neural network-based prediction of full-reference image quality assessment metrics. *J Med Imaging* 2022;9:015502.
- [24] Amin J, Anjum MA, Sharif M, et al. Liver tumor localization based on YOLOv3 and 3D-semantic segmentation using deep neural networks. *Diagnosics* 2022;12:823.
- [25] 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* 2025;9:494–506.
- [26] Sandfort V, Yan K, Pickhardt PJ, et al. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep* 2019;9:16884.
- [27] Pan S, Wang T, Qiu RLJ, et al. 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. *Phys Med Biol* 2023;68:105004.
- [28] Liu Y, Lei Y, Wang T, et al. MRI-based treatment planning for liver stereotactic body radiotherapy: validation of a deep learning-based synthetic CT generation method. *Br J Radiol* 2019;92:20190067.
- [29] Koetzier LR, Wu J, Mastrodicasa D, et al. Generating synthetic data for medical imaging. *Radiology* 2024;312:e232471.
- [30] Dar SU, Yurt M, Karacan L, et al. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans Med Imaging* 2019;38:2375–2388.
- [31] Jayachandran Preetha C, Meredig H, Brugnara G, et al. Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. *Lancet Digit Health* 2021;3:e784–e794.

Keywords: Deep learning; Generative adversarial networks; Synthetic 3D contrast-enhanced liver MRI; Contrast-enhanced multiphasic liver MRI; Hepatocellular carcinoma; Cirrhosis.

Received 10 December 2025; received in revised form 23 February 2026; accepted 26 February 2026; Available online 4 March 2026

Supplemental information

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 C. Madoff, James S. Duncan, and Julius Chapiro

**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

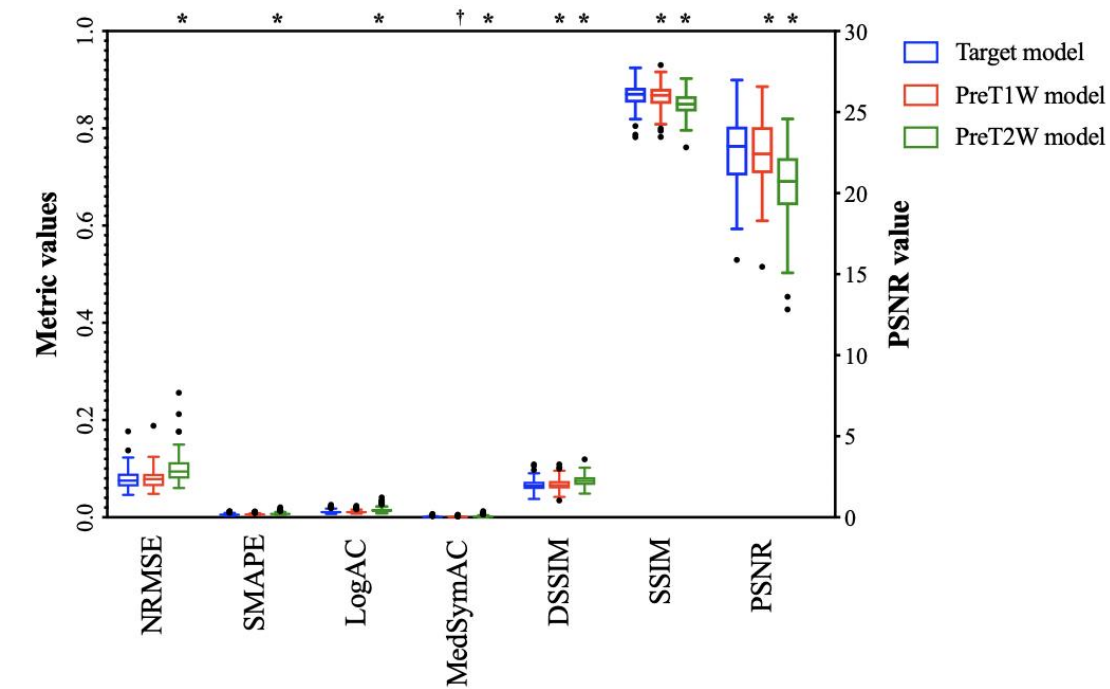
A multi-reader study was conducted by three blinded board-certified radiologist readers (J.W.), (M.V.R.), and (G.P.) with 11, 22, and 40 years of experience in abdominal imaging, respectively, including liver MRI interpretation. Readers were intentionally selected to represent a range of experience levels 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.

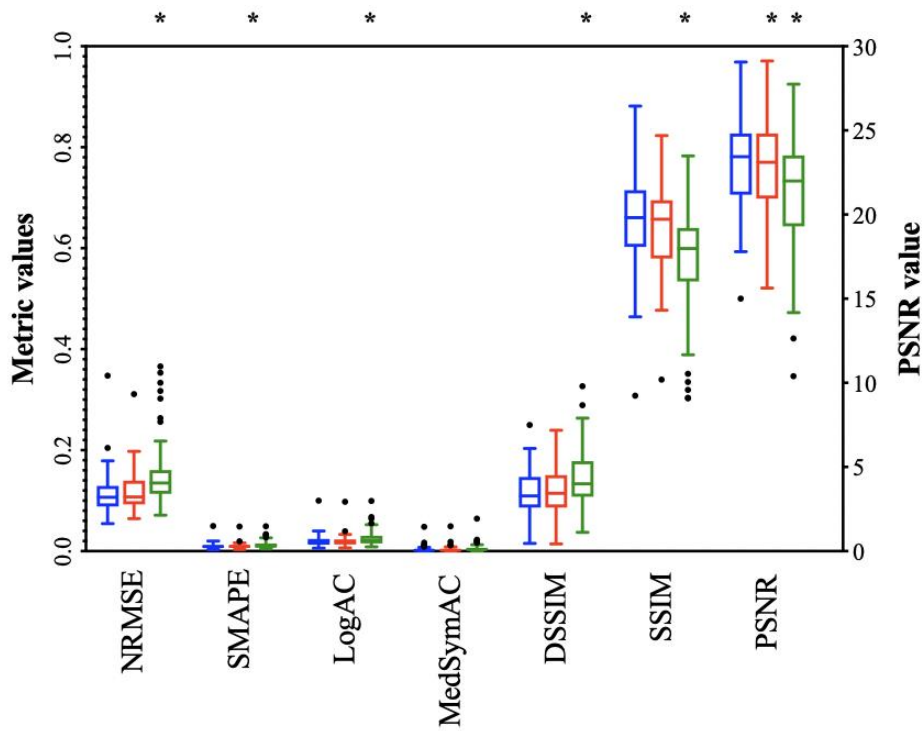
A comprehensive six-task questionnaire was developed to thoroughly assess the key aspects of the MRI exams based on a detailed review of established literature (8-13) and modified as necessary to fit the specific aims of this study. The final questionnaire was reviewed and refined by a board-certified 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.