SUPPLEMENT

Supplementary Materials

Deep Learning Algorithm

Considering the computationally demanding nature of the algorithm training using MR images with various spatial resolutions and diverse gray-scale intensity levels, these images were standardized through a preprocessing step. MR images were resized to a fixed resolution of 384 x 384 x 80 voxels using the Scipy three-dimensional (3D) resize tool [1] and normalized using a Gaussian normalization function to adjust the grayscale intensity level across MR images as follows:

$$X = \frac{\chi - \mu}{\sigma}$$

where X is the normalized pixel value, χ is the original pixel value, μ is the mean value for all pixels other than pixels in the background, and σ is the standard deviation of the values for all pixels other than the pixels in the background.

We designed the architecture of our deep learning algorithm (DLA) based on DeepLab V3+ [2], which is a convolutional neural network optimized for segmentation tasks on two-dimensional images. We implemented a 2.5D input set-up [3] to DeepLab V3+ in order to exploit the three-dimensional information of MR images in a time- and resource-efficient manner. The 2.5D input setup utilized the input axes for 2D red, green, and blue (RGB) images, which were composed of width, height, and RGB channel axes. The 2.5D input setup imports three consecutive sections of MR images (i.e., an MR image of interest, as well as MR image sections above and below the MR image of interest) in the axis of the RGB channel. This enables the DLA to perform segmentation tasks on a given MR image by utilizing the 3D spatial information of the three consecutive sections of the MR images.

The architecture of our DLA consists of an encoder and decoder. The encoder includes two blocks of convolution and a rectified linear unit, a modified Xception model [4], and an atrous separable spatial pyramid pooling (ASSPP) unit. The modified Xception model comprises entry, middle, and exit blocks of the down-sampling layers. The ASSPP unit is a pooling technique with an atrous depth-wise convolution that is known to be robust in capturing multi-scale information in each channel [2]. The ASSPP unit consists of a 1 x 1 convolution layer, depth convolution layers with different convolution rates, and a global average pooling layer. By combining the 2.5D input and ASSPP in our DLA, we created a synergy in capturing 3D spatial information from the three consecutive input images that were imported through RGB channels. The decoder part includes bilinear up-sampling layers followed by a 1 x 1 convolution. All the convolution layers were followed by a batch normalization layer. The final output of the algorithm is a three-channel (liver, spleen, and background) logit map with the same size as that of the input MR image.

For training the DLA, we used cross-entropy with regularization loss as the loss function. Because the cross-entropy loss may not be sufficiently robust for a segmentation task for images with a class imbalance (i.e., different numbers of pixels belonging to the liver, spleen, and background), we additionally employed a masked Dice-coefficient loss [5] as the regularization loss to improve the segmentation performance of the algorithm at the boundary of each class. This regularization loss was calculated for 10 pixels from the border of the segmented objects.

The final DLA was trained for 50 epochs. The algorithm training took approximately 60 GPU hours on a computer equipped with one Nvidia Geforce GTX 1080 Ti (Nvidia), two Intel Xeon E5-2695 (Intel), and 256 GB of RAM. Analysis of one MR image section took approximately 350 ms on the same computer, resulting in a computation time of approximately 25 seconds for an hepatobiliary phase (HBP) MR image set containing 70 image sections. The final DLA was implemented using web-based digital imaging and communications in medicine (DICOM) viewer system (GoCDSS; SmartCareworks Inc.).

Adjusted Liver-to-Spleen SI Ratio Corrected for MR Parameter Effects

The assessment of liver function using liver signal intensity (SI) on gadoxetic acid-enhanced HBP MRI is based on the assumption that liver SI reflects hepatocyte contrast uptake function. As an MRI SI is measured in arbitrary units, the liver-to-spleen SI ratio (LSSR), that is, liver SI normalized using the spleen SI as the internal reference, is commonly used as

a liver functional metric on HBP MRI [6-8]. However, as MRI parameters affect the image contrast, LSSRs measured using different MRI parameters may not be comparable to each other, which hampers the clinical application of LSSR to MRI examinations obtained using various parameters. To overcome this limitation, we attempted to correct for MR parameter effects on the LSSR through a computer simulation.

The SI on a spoiled gradient echo MRI is expressed by the following equation:

$$SI = M\frac{\sin\alpha(1-e^{-TR/T1})}{(1-(\cos\alpha)e^{-TR/T1})}e^{-TE/T2*} \quad (1)$$

where M is the net magnetization, α is the flip angle (FA), TR is the repetition time, T1 denotes the T1 relaxation time, TE is the echo time, and T2* is the T2* relaxation time. Hence, the LSSR is expressed as:

$$\frac{S_{Liver}}{S_{Spleen}} = (M_{Liver}\frac{\sin\alpha(1-e^{-TR/T1_{Liver}})}{(1-(\cos\alpha)e^{-TR/T1_{Liver}})}e^{-TE/T2^*_{Liver}}) / (M_{Spleen}\frac{\sin\alpha(1-e^{-TR/T1_{Spleen}})}{(1-(\cos\alpha)e^{-TR/T1_{Spleen}})}e^{-TE/T2^*_{Spleen}}) \quad (2)$$

where S_{Liver} and S_{Spleen} are signal intensities, $T1_{Liver}$ and $T1_{Spleen}$ denote the T1 relaxation times, and $T2^*_{Liver}$ and $T2^*_{Spleen}$ are the T2* relaxation times of the liver and spleen, respectively.

Since M is dependent on the proton density and magnetic field strength, given the nearly equal proton densities of the liver and spleen [9], M_{Liver} and M_{Spleen} can be canceled out from equation [2]. On T1-weighted spoiled gradient echo MRI using a short TE, the contribution of the T2* relaxation effect on the LSSR may be negligible. Thus, the signal equation for LSSR can be simplified as a function of $T1_{Liver}$, $T1_{Spleen}$, TR, and FA as follows:

$$\frac{S_{Liver}}{S_{Spleen}} = (\frac{\sin\alpha(1-e^{-TR/T1_{Liver}})}{(1-(\cos\alpha)e^{-TR/T1_{Liver}})}) / (\frac{\sin\alpha(1-e^{-TR/T1_{Spleen}})}{(1-(\cos\alpha)e^{-TR/T1_{Spleen}})}) \quad (3)$$

Using equation [3], a computer simulation was conducted to calculate LSSRs at all possible combinations of $T1_{Liver}$, $T1_{Spleen}$, TR, and FA over the reported ranges for T1Liver (range, 200–1000 ms; interval, 10 ms) and T1Spleen (range, 500–1200 ms; interval, 10 ms) on gadoxetic acid-enhanced HBP images [10-13], and for TR (2.0–5.0 ms) and FA (5°–30°) for clinical HBP MRI. This simulation yielded a large data library of various combinations of $T1_{Liver}$, $T1_{Spleen}$, TR, and FA and their corresponding LSSRs. Using this data library, we obtained spleen-to-liver T1 ratios (i.e., the reciprocals of liver-to-spleen T1 ratios) that matched with a given LSSR, TR, and FA. Unlike the LSSR, the spleen-to-liver T1 ratio is theoretically independent of the MRI parameters. Therefore, we calculated the adjusted LSSR (aLSSR) as the median spleen-to-liver T1 ratio (Supplementary Fig. 2). We used the spleen-to-liver T1 ratio instead of the liver-to-spleen T1 ratio for aLSSR calculation because the liver-to-spleen T1 ratio was inversely correlated with the LSSR. The algorithm to calculate the aLSSR is available as a web calculator (https://i-pacs.com/adjustedSIR).

Clinical Feasibility Study

During January 2021, liver MRI examinations using two 3T scanners in our institution (Magnetom Vida and Skyra; Siemens Healthiniers) were routinely performed, including two HBP-MRI sets at FAs of 10° and 19° as part of our quality assessment process for protocol optimization. Among the 230 consecutive patients in the present study who underwent liver MRI during this period, nine subjects were excluded from further analysis because of large or multiple liver tumors that precluded liver signal measurements (n = 8) or poor image quality (n = 2). The remaining 221 patients (164 male and 57 female; mean age, 62.5 (range, 36–93) years were included in the evaluation of the feasibility of aLSSR in correcting for the MRI parameter effects on the LSSR. The clinical indications for liver MRI included the evaluation of a focal hepatic lesion detected by other examinations (n = 92), follow-up after treatment for hepatocellular carcinoma (n = 84), evaluation of liver cirrhosis (n = 32), and follow-up for liver metastasis (n = 13). HBP-MRI was performed using a breath-hold T1-weighted fat-suppressed three-dimensional spoiled gradient-echo sequence acquired 20 minutes after a bolus injection of 0.025 mmoL/kg gadoxetic acid (Bayer Health Care). The scan parameters for the first HBP image set were a FA of 10°, echo

time of 1.33 ms, and repetition time of 3.39 ms, and for the second HBP image set were a FA of 19°, echo time of 1.39 ms, and repetition time of 3.83 ms. The other parameters were identical for both sets of HBP images as follows: a matrix of 312 x 384, a field of view of 308 x 379 mm, a slice thickness of 3 mm, and no gap. Fat suppression was performed using a frequency-selective fat-suppression technique.

An abdominal radiologist (8 years of experience) measured both the liver and spleen SI on the two sets of HBP images using 2 cm^2 circular regions of interest (i.e., six and three regions of interest for the liver and spleen, respectively) using a commercial DICOM viewer (RadiAnt DICOM viewer; Medixant). The LSSR was calculated for each of the two sets of HBP images, and the corresponding aLSSR values were calculated using the aforementioned algorithm. The correlations between the LSSR and aLSSR values obtained using the FA 10° image set and those obtained using the FA 19° image set were then evaluated.

REFERENCES

1. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:261-272
2. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European Conference on Computer Vision (ECCV); 2018 Sep 8-14; Munich, Germany: ECCV; 2018; p. 801-818
3. Roth HR, Lu L, Seff A, Cherry KM, Hoffman J, Wang S, et al. A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. *Med Image Comput Comput Assist Interv* 2014;17:520-527
4. Chollet F. Xception: deep learning with depthwise separable convolutions. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition; 2017 Jul 21-26; Honolulu, HI, USA; IEEE; 2017; p. 1251-1258
5. Milletari F, Navab N, Ahmadi SA. V-net: fully convolutional neural networks for volumetric medical image segmentation. Proceedings of 2016 Fourth International Conference on 3D Vision; 2016 Oct 25-28; Stanford, CA, USA; IEEE; 2016; p. 565-571
6. Yamada A, Hara T, Li F, Fujinaga Y, Ueda K, Kadoya M, et al. Quantitative evaluation of liver function with use of gadoxetate disodium-enhanced MR imaging. *Radiology* 2011;260:727-733
7. Araki K, Harimoto N, Kubo N, Watanabe A, Igarashi T, Tsukagoshi M, et al. Functional remnant liver volumetry using Gd-EOB-DTPA-enhanced magnetic resonance imaging (MRI) predicts post-hepatectomy liver failure in resection of more than one segment. *HPB (Oxford)* 2020;22:318-327
8. Kim DK, Choi JI, Choi MH, Park MY, Lee YJ, Rha SE, et al. Prediction of posthepatectomy liver failure: MRI with hepatocyte-specific contrast agent versus indocyanine green clearance test. *AJR Am J Roentgenol* 2018;211:580-587
9. Nyman R, Ericsson A, Hemmingsson A, Jung B, Sperber G, Thuomas KA. T1, T2, and relative proton density at 0.35 T for spleen, liver, adipose tissue, and vertebral body: normal values. *Magn Reson Med* 1986;3:901-910
10. Haimerl M, Fuhrmann I, Poelsterl S, Fellner C, Nickel MD, Weigand K, et al. Gd-EOB-DTPA-enhanced T1 relaxometry for assessment of liver function determined by real-time 13C-methacetin breath test. *Eur Radiol* 2018;28:3591-3600
11. Yoon JH, Lee JM, Kang HJ, Ahn SJ, Yang H, Kim E, et al. Quantitative assessment of liver function by using gadoxetic acid–enhanced MRI: hepatocyte uptake ratio. *Radiology* 2019;290:125-133
12. Yoon JH, Lee JM, Kim E, Okuaki T, Han JK. Quantitative liver function analysis: volumetric T1 mapping with fast multisection B1 inhomogeneity correction in hepatocyte-specific contrast-enhanced liver MR imaging. *Radiology* 2017;282:408-417
13. Kim JE, Kim HO, Bae K, Choi DS, Nickel D. T1 mapping for liver function evaluation in gadoxetic acid-enhanced MR imaging: comparison of look-locker inversion recovery and B1 inhomogeneity-corrected variable flip angle method. *Eur Radiol* 2019;29:3584-3594

Supplementary Table 1. MRI Techniques Used for the Development and Test Datasets

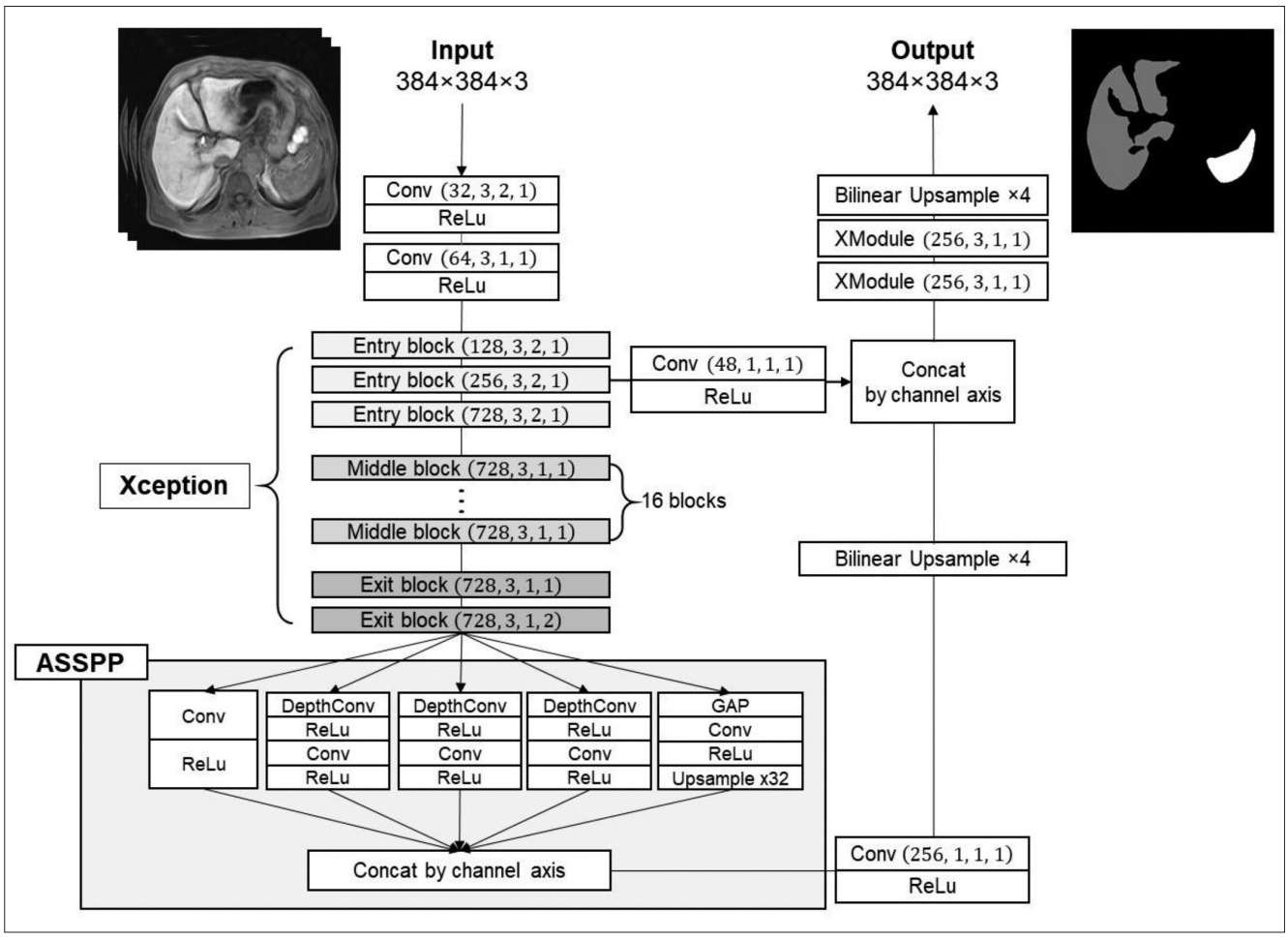
MRI Techniques	DLA Development and Test		Study Population for Clinical Utility Evaluation (n = 276)
	Development Dataset* (n = 1014)	Test Dataset (n = 200)	
Field strength			
1.5T	438 (43.2)	60 (30.0)	49 (17.8)
3T	575 (56.7)	140 (70.0)	227 (82.2)
Vendor			
Siemens	838 (82.6)	155 (77.5)	253 (91.7)
Philips	157 (15.5)	41 (20.5)	18 (6.5)
GE	18 (1.8)	4 (2.0)	5 (1.8)
Scanner name			
Avanto	423 (41.7)	56 (28.0)	46 (16.7)
Skyra	407 (40.1)	91 (45.5)	206 (74.6)
Achieva	40 (3.9)	15 (7.5)	3 (1.1)
Ingenia	115 (11.3)	26 (13.0)	15 (5.4)
Others	28 (2.8)	12 (6.0)	6 (2.1)
Slice thickness, mm†	4 (2.0–8.0)	3.5 (2.0–8.0)	3.0 (2.0–6.0)
Flip angle, °†	18.1 (9.0–30.0)	16 (9.0–35.0)	20.0 (9.0–25.0)
Echo time, ms†	1.5 (1.0–3.1)	1.5 (1.0–3.1)	2.0 (1.0–3.2)
Repetition time, ms†	4.1 (2.7–6.3)	4.1 (3.0–6.3)	4.5 (2.7–6.9)
Matrix†			
Phase encoding	238 (128–340)	230 (128–319)	240 (140–260)
Frequency encoding	320 (200–512)	320 (320–384)	384 (248–416)
Field of view, mm†			
Phase encoding	296 (218–379)	285 (262–308)	296 (206–357)
Frequency encoding	379 (330–440)	379 (214–384)	379 (320–440)
Pixel size, mm†			
Phase encoding	1.4 (1.0–2.1)	1.2 (1.0–2.1)	1.2 (1.0–2.0)
Frequency encoding	1 (0.7–1.5)	1.1 (0.9–1.6)	1.0 (0.8–1.7)
Bandwidth, Hz/pixel†	480 (244–1796)	480 (244–1800)	480 (244–1800)

Unless otherwise indicated, data are number of subjects, and data in parentheses are percentages. *MRI technique information was missing for one examination in the development dataset, †Data are median values (range). DLA = deep learning algorithm

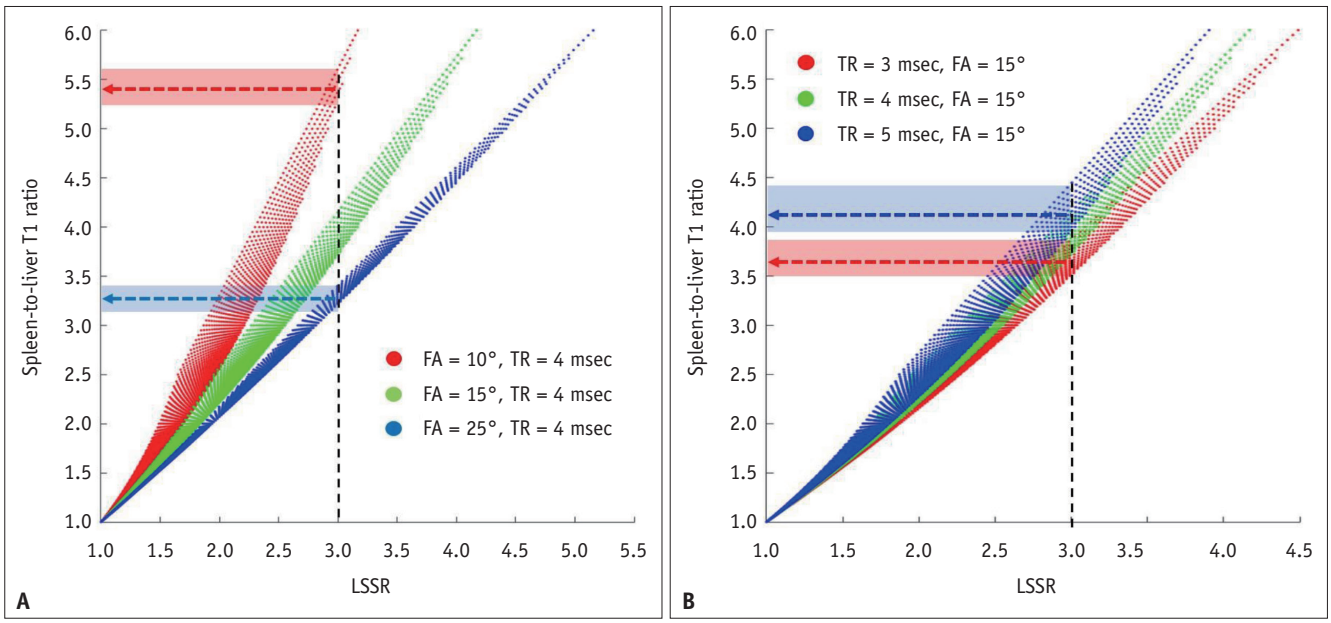
Supplementary Table 2. Characteristics of the Test Datasets according to MRI Data Source

	All	Internal Data	External Data
Number of patients	200	110	90
Male:female	153:47	80:30	73:17
Age, years*	55 ± 13	56 ± 13	54 ± 12
Underlying liver disease, %			
None	32 (16.0)	14 (12.7)	18 (20.0)
Hepatitis B	142 (71.0)	84 (76.4)	58 (64.4)
Hepatitis C	8 (4.0)	6 (5.5)	2 (2.2)
Alcohol-induced	4 (2.0)	2 (1.8)	2 (2.2)
NAFLD	4 (2.0)	1 (0.9)	3 (3.3)
Other†	10 (5.0)	3 (2.7)	7 (7.8)
Hepatic mass			
Absent, %	28 (14.0)	14 (12.7)	14 (15.6)
Present, %	172 (86.0)	96 (87.3)	76 (84.4)
Malignant	169 (84.5)	95 (86.4)	74 (82.2)
Benign‡	3 (1.5)	1 (0.9)	2 (2.2)
Largest mass size, cm*	3.1 ± 1.7	3.3 ± 1.8	2.9 ± 1.6
Pathologic liver fibrosis stage, %			
F0	60 (30.0)	28 (25.5)	32 (35.6)
F1	21 (10.5)	14 (12.7)	7 (7.8)
F2	29 (14.5)	19 (17.3)	10 (11.1)
F3	40 (20.0)	23 (20.9)	17 (18.9)
F4	50 (25.0)	26 (23.6)	24 (26.7)

Unless otherwise indicated, data are number of subjects, and data in parentheses are percentages. *Data are mean values ± standard deviation, †Including toxic hepatitis, hepatitis A, autoimmune hepatitis, Wilson's disease, and hepatolithiasis, ‡Including hemangioma and IgG4-related cholangitis. NAFLD = non-alcoholic fatty liver disease

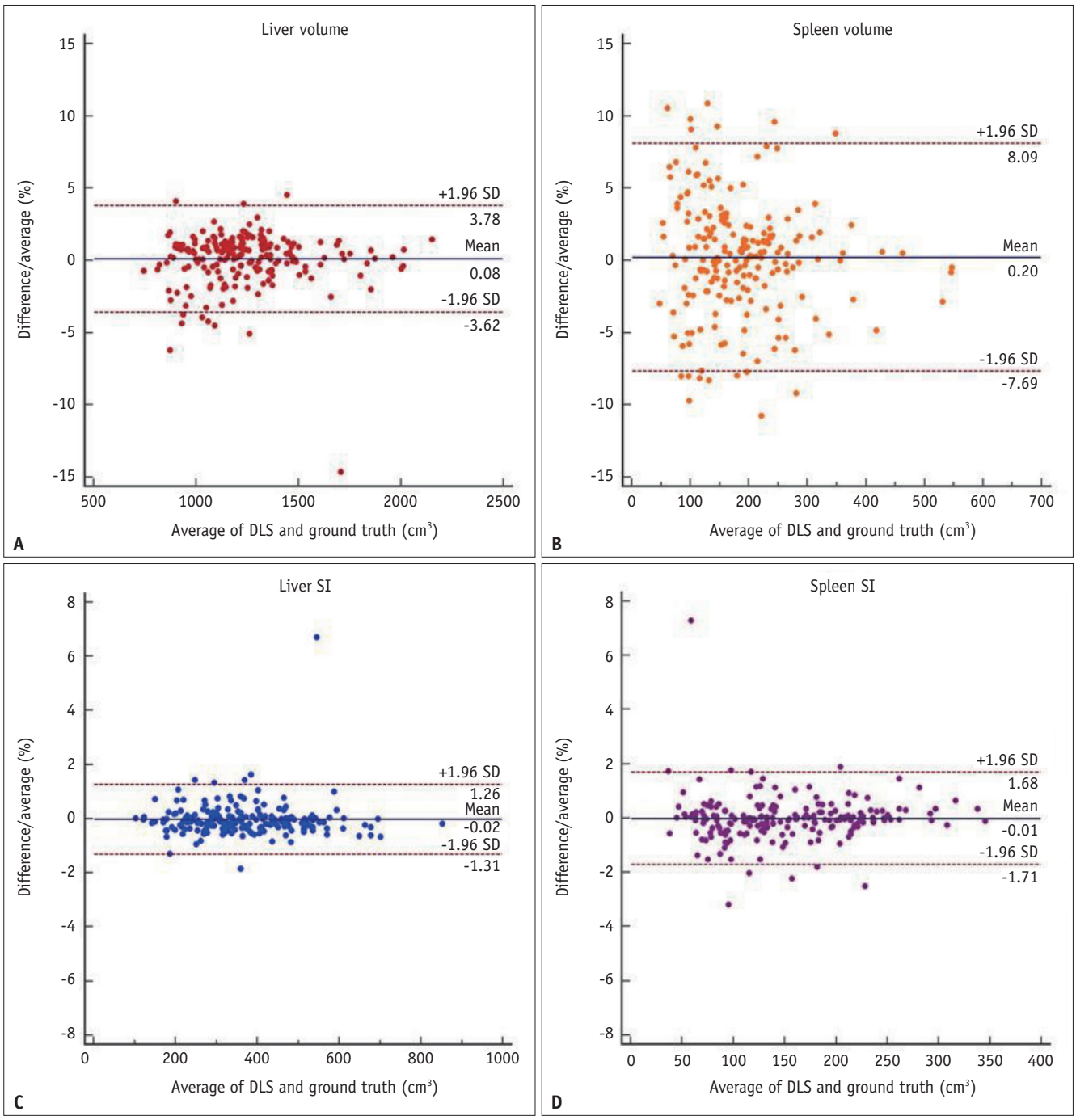


Supplementary Fig. 1. Schematic diagram of the architecture of the deep learning algorithm. The algorithm receives three consecutive MR images using a 2.5-dimensional input setup. The input images are processed by the encoder part of the algorithm, including two blocks of Conv and a ReLU, a modified Xception model, and an ASSPP unit (1 x 1 Conv layer, DepthConv layers, and a GAP layer). These layers are then Concat using the channel axis. The decoder part of the algorithm upsamples the output of the ASSPP unit. The final output of the algorithm is a three-class (liver, spleen, and background) logit map with the same size as that of the input MR image. All Conv layers are followed by a batch normalization layer. The numbers in parentheses indicate the hyperparameters (filter, kernel, stride, and rate). ASSPP = atrous separable spatial pyramid pooling, Concat = concatenated, Conv = convolution, DepthConv = depth convolution, GAP = global average pooling, ReLU = rectified linear unit

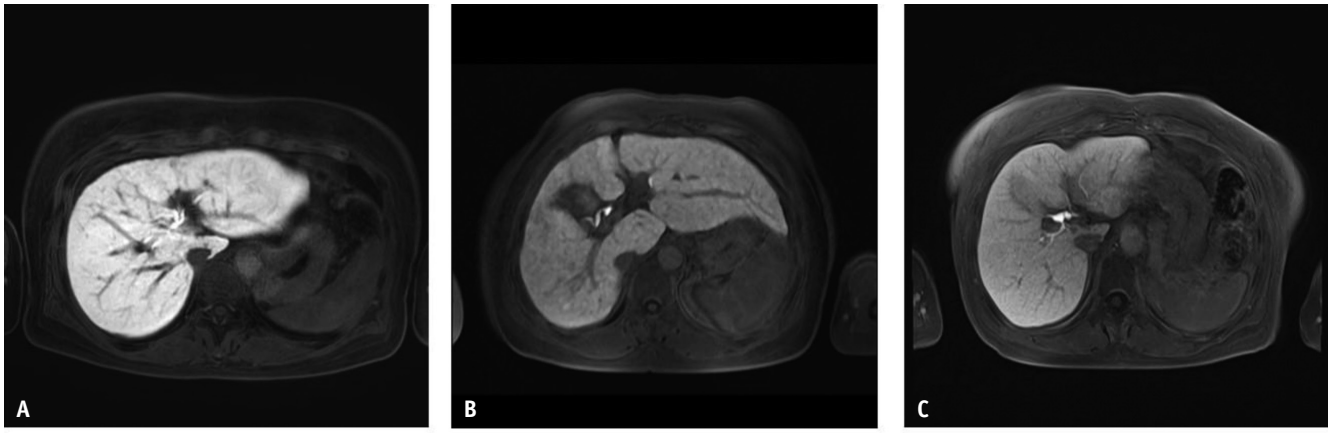


Supplementary Fig. 2. Graphs showing the association between the LSSR and spleen-to-liver T1 ratio in accordance with different MRI parameters.

A, B. The graphs depict the LSSR on the X-axis versus the spleen-to-liver T1 ratio on the Y-axis according to various FAs at a fixed TR of 4.0 ms **(A)** and various TRs at a fixed FA of 15° **(B)**. Note that the range of the spleen-to-liver T1 ratio corresponds to the LSSR. The aLSSR is calculated as the median spleen-to-liver T1 ratio at a given LSSR, FA, and TR. When a fixed TR of 4 ms is used **(A)**, for example, the LSSR of 3 (vertical dashed line) corresponds to the estimated spleen-to-liver T1 ratios ranging from 5.3 to 5.7 at an FA of 10° (red box) and those ranging from 3.2 to 3.4 at an FA of 25° (blue box). The aLSSRs (i.e., median spleen-to-liver T1 ratio values) are 5.45 (red horizontal dashed line) and 3.28 (blue horizontal dashed line) at the FAs of 10° and 25°, respectively. When a fixed FA of 15° is used **(B)**, the LSSR of 3 (vertical dashed line) corresponds to the estimated spleen-to-liver T1 ratios ranging from 3.5 to 3.8 at a TR of 3 ms (red box) and from 3.9 to 4.4 at a TR of 5 ms (blue box). The aLSSRs are 3.63 (red horizontal dashed line) and 4.12 (blue horizontal dashed line) at TRs of 3 ms and 5 ms, respectively. aLSSR = adjusted LSSR, FA = flip angle, LSSR = liver-to-spleen signal intensity ratio, TR = repetition time



Supplementary Fig. 3. Bland-Altman plots of the agreement between the deep learning algorithm and the ground truth data for liver volume (A), spleen volume (B), liver SI (C), and spleen SI (D). Solid lines indicate the mean differences, and dashed lines indicate the upper and lower limits of the 95% limits of agreement. DLS = deep learning system, SD = standard deviation, SI = signal intensity



DLA-estimated liver volume	1004.0 cm^3	DLA-estimated liver volume	1441.6 cm^3	DLA-estimated liver volume	730.8 cm^3
aLSSR	5.66	aLSSR	2.33	aLSSR	2.32
aLSSR x LV_{BSA}	3373.6 cm^3/m^2	aLSSR x LV_{BSA}	1957.0 cm^3/m^2	aLSSR x LV_{BSA}	1099.2 cm^3/m^2
ICG-R15	8.0%	ICG-R15	20.5%	ICG-R15	24.8%

Supplementary Fig. 4. Representative examples of DLA-assisted functional liver capacity assessments.
A. A gadoxetic acid-enhanced hepatobiliary phase image from a 62-year-old female with chronic hepatitis B. The DLA-estimated liver volume is 1004.0 cm^3. The aLSSR is 5.66. The aLSSR x LV_{BSA} is 3373.6 cm^3/m^2. The ICG-R15 is 8.0%, which is within the normal reference range (i.e., < 10%). **B.** A gadoxetic acid-enhanced hepatobiliary phase image from a 59-year-old female with chronic hepatitis B. The liver volume estimated by the DLA is 1441.6 cm^3, and the aLSSR is 2.33, resulting in an aLSSR x LV_{BSA} of 1957.0 cm^3/m^2. The ICG-R15 of this patient is 20.5%, which is considered a contraindication for major hepatectomy. **C.** A gadoxetic acid-enhanced hepatobiliary phase image from a 61-year-old female with chronic hepatitis B infection. The liver volume estimated by the DLA is 730.8 cm^3, and the aLSSR is 2.32. As a result, the aLSSR x LV_{BSA} is 1099.2 cm^3/m^2, which is far below the cut-off value (1957.0 cm^3/m^2) for detecting an ICG-R15 ≥ 20%. Her ICG-R15 is 24.8%, which is considered a contraindication for major hepatectomy. aLSSR = adjusted liver-to-spleen signal intensity ratio, DLA = deep learning algorithm, ICG-R15 = indocyanine green retention rate at 15 minutes, LV_{BSA} = liver volume normalized by body surface area