



# HOKKAIDO UNIVERSITY

|                  |                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title            | Evaluation of artificial-intelligence-based liver segmentation and its application for longitudinal liver volume measurement                                                                                                                                                                                                                                                                                       |
| Author(s)        | Kimura, Rina; Hirata, Kenji; Tsuneta, Satonori et al.                                                                                                                                                                                                                                                                                                                                                              |
| Citation         | Abdominal Radiology, 50(12), 6192-6200<br><a href="https://doi.org/10.1007/s00261-025-05050-3">https://doi.org/10.1007/s00261-025-05050-3</a>                                                                                                                                                                                                                                                                      |
| Issue Date       | 2025-06-10                                                                                                                                                                                                                                                                                                                                                                                                         |
| Doc URL          | <a href="https://hdl.handle.net/2115/97904">https://hdl.handle.net/2115/97904</a>                                                                                                                                                                                                                                                                                                                                  |
| Rights           | This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature' s AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <a href="https://doi.org/10.1007/s00261-025-05050-3">https://doi.org/10.1007/s00261-025-05050-3</a> |
| Type             | journal article                                                                                                                                                                                                                                                                                                                                                                                                    |
| File Information | manuscript_114392.pdf                                                                                                                                                                                                                                                                                                                                                                                              |



**Manuscript title:**

**Evaluation of artificial-intelligence-based liver segmentation and its application for longitudinal liver volume measurement**

**Authors**

Rina Kimura <sup>1,2</sup>, Kenji Hirata <sup>1,2,3,4,5</sup>, Satonori Tsuneta <sup>2,6</sup>, Junki Takenaka <sup>3</sup>, Shiro Watanabe <sup>1,3</sup>,  
Daisuke Abo <sup>1,2</sup>, Kohsuke Kudo <sup>1,2,4,5</sup>

- 1 Department of Diagnostic Imaging, Faculty of Medicine, Hokkaido University, Sapporo, Japan
- 2 Department of Diagnostic and Interventional Radiology, Hokkaido University Hospital, Sapporo, Japan
- 3 Department of Nuclear Medicine, Hokkaido University Hospital, Sapporo, Japan
- 4 Division of Medical AI Education and Research, Faculty of Medicine, Hokkaido University, Sapporo, Japan
- 5 Healthcare AIX Innovation Center (HAIXIC), Hokkaido University, Sapporo, Japan
- 6 Department of Radiology, Graduate School of Dental Medicine, Hokkaido University, Sapporo, Japan

**Address for correspondence author:**

Kenji HIRATA, MD, PhD, FANMB, FJCR

Department of Diagnostic Imaging, Graduate School of Medicine, Hokkaido University

Kita 15, Nishi 7, Kita-Ku, Sapporo, Hokkaido, Japan 060-8638

Phone: 81-11-706-7779

Fax: 81-11-706-7408

email: [khirata@med.hokudai.ac.jp](mailto:khirata@med.hokudai.ac.jp)

## **Evaluation of artificial-intelligence-based liver segmentation and its application for longitudinal liver volume measurement**

### **Abstract**

**Background:** Accurate liver-volume measurements from CT scans are essential for treatment planning, particularly in liver resection cases, to avoid postoperative liver failure. However, manual segmentation is time-consuming and prone to variability. Advancements in artificial intelligence (AI), specifically convolutional neural networks, have enhanced liver segmentation accuracy. We aimed to identify optimal CT phases for AI-based liver volume estimation and apply the model to track liver volume changes over time. We also evaluated temporal changes in liver volume in participants without liver disease.

**Methods:** In this retrospective, single-center study, we assessed the performance of an open-source AI-based liver segmentation model previously reported, using non-contrast and dynamic CT phases. The accuracy of the model was compared with that of expert radiologists. The Dice similarity coefficient (DSC) was calculated across various CT phases, including arterial, portal venous, and non-contrast, to validate the model. The model was then applied to a longitudinal study involving 39 patients without liver disease (527 CT scans) to examine age-related liver volume changes over 5 to 20 years.

**Results:** The model demonstrated high accuracy across all phases compared to manual segmentation. Among the CT phases, the highest DSC of  $0.988 \pm 0.010$  was in the arterial phase. The intraclass correlation coefficients for liver volume were also high, exceeding 0.9 for contrast-enhanced phases and 0.8 for non-contrast CT. In the longitudinal study, the model indicated an annual decrease of 0.95%.

**Conclusion:** This model provides high accuracy in liver segmentation across various CT phases and offers insights into age-related liver volume reduction. Measuring changes in liver volume may help with the early detection of diseases and the understanding of pathophysiology.

### **Abbreviations**

artificial intelligence (AI)

Cycle-consistent Generative Adversarial Network (CycleGAN)

Dice similarity coefficient (DSC)

Digital Imaging and Communications in Medicine (DICOM)

interquartile range (IQR)

intraclass correlation coefficient (ICC)

**Word count: 3241words**

**Keywords:** Liver, Volumetric, Computed Tomography (CT), Artificial Intelligence, Neural Networks

## Introduction

Segmentation, the process of identifying and delineating the boundaries of organs or lesions from medical images such as CT scans, is a critical task in medical imaging. Liver segmentation, in particular, is of great importance due to the liver's vital role in sustaining life and its function being highly dependent on its volume<sup>1</sup>. Clinical applications include liver volume measurement, blood vessel extraction, and tumor detection.

In cases requiring living liver transplantation donors or liver resection, preoperative liver volume measurement via CT is essential for determining the treatment plan and predicting the remaining liver volume after resection, helping to prevent postoperative liver failure<sup>2-4</sup>.

However, in routine clinical practice, liver volume measurement is not performed for every CT examination. The standard evaluation of liver size in radiology reports is often subjective and qualitative and does not accurately reflect three-dimensional liver volume<sup>5</sup>. Although studies have been conducted on liver volume<sup>6</sup>, manual segmentation remains time-consuming and subjective<sup>7-11</sup>, leading to intra- and inter-observer variability and limiting study populations<sup>12</sup>. Few studies have addressed temporal changes in liver volume in the same patient<sup>13,14</sup>; longitudinal studies are needed to further examine age-related liver volume changes.

This AI model, based on 3D U-Net architecture, achieves high accuracy by augmenting training images with Cycle-consistent Generative Adversarial Networks (CycleGANs) to enhance image quality. Additionally, the model uses machine learning techniques trained on a large, diverse dataset, making it well-suited for clinical applications.

The model was trained using contrast-enhanced CT data, which is expected to provide high segmentation accuracy across all CT phases. Recently, there have been some reports of segmentation using CT multiphase images, but few studies have directly compared the segmentation accuracy of each CT phase with that of expert radiologists as the ground truth<sup>15</sup>. Furthermore, the natural course of liver volume changes with aging has not been fully elucidated. Therefore, we aimed to identify which CT phases, non-contrast and dynamic, are most effective for use with the AI model. We also sought to evaluate the accuracy of the model and, based on these findings, apply it in real-world cases to investigate temporal changes in liver volume over time.

## Methods

This retrospective, single-center study was approved by the institutional review board of Hokkaido University Hospital (#023-0035); the requirement for written informed consent was waived.

The study consisted of two parts: Analysis 1 and Analysis 2. In Analysis 1, we compared the AI model with radiologist-measured segmentation using non-contrast and dynamic CT scans of healthy liver transplant donors. In Analysis 2, the AI model validated in Analysis 1 was used for liver segmentation on

CT scans of patients without liver disease. We calculated liver volumes and investigated temporal changes in liver volume.

### *Study participants*

Analysis 1 included 34 healthy liver transplant donors who underwent preoperative CT examinations at our institution between January 2017 and December 2021. Of the 90 eligible participants, 56 were excluded: different CT protocol (n=39), Duplication (n=18): Multiple CT scans from the same patient were excluded, retaining only the initial scan to avoid redundancy, cases in which only non-contrast CT scans were acquired (n=9), postoperative CT scans (n=6), and liver disease (n=1). The remaining 34 participants were randomly assigned to Cohort 1A (n=10) and Cohort 1B (n=24) to minimize selection bias and ensure representative sampling (Fig. 1).

In Analysis 2, we selected 39 patients with no liver disease or chemotherapy history from CT scans in 2021 who had been followed for 5 to 20 years between 2000 and 2022; 527 scans were used. Radiological assessments by a board-certified radiologist confirmed the absence of liver pathology, such as tumors, fatty liver, or cirrhosis, although benign hepatic cysts were present in some cases. Additionally, medical records were reviewed to exclude patients with prior liver disease diagnoses or abnormal liver function tests. No invasive procedures, such as liver biopsy or chemotherapy, were performed during the observation period.

The protocol for Analysis 1 is outlined in Table 1 and consisted of three parts. In Part 1, two board-certified radiologists independently delineated the whole liver regions on 5-phase CT scans of 10 patients from Cohort 1A. Liver volumes and Dice similarity coefficient (DSCs) were then compared across three combinations: (1) between the model and Radiologist A (15 years of experience), (2) between the model and Radiologist B (9 years of experience), and (3) between Radiologists A and B. The radiologists refined liver segmentations by correcting the AI-generated liver regions (Fig. 2). The DSC was calculated by comparing the AI-generated segmentations before any manual **correction** with the final segmentations corrected by Radiologist A and Radiologist B. This allowed us to quantitatively evaluate the agreement between the uncorrected AI output and the radiologist-defined ground truth.

Given the substantial effort and time required for two radiologists to manually generate ground truth data across all five phases, we aimed to streamline the process. In Part 2, a single radiologist delineated the liver regions on the late arterial phase for 24 cases from Cohort 1B.

In Part 3, we compared liver segmentations generated by the AI model across all CT phases, including contrast-enhanced and non-contrast scans, for 34 cases from Cohorts 1A and 1B against the ground truth volumes established by Radiologist A.

After demonstrating sufficiently high performance of the AI model in Analysis 1, we proceeded to Analysis 2. In this phase, the AI model was used to measure liver volumes in patients without liver disease, and temporal changes were analyzed. Cohort 2 comprised the target group for Analysis 2. No

manual corrections by radiologists were performed in this analysis; therefore, no ground truth volume values were established, and all evaluations relied solely on the unaltered AI-generated segmentations. For cases in which CT scans from different phases were acquired on the same day, the volumes were averaged. The liver volume at the start of observation was standardized to 100%, and this measure was termed the liver volume index.

#### *Image acquisition (CT Protocol)*

Details are provided in the Supplementary Documents.

#### *AI system*

We used the AI model with freely available source code on GitHub (<https://github.com/rsummers11/CADLab/tree/master/CT%20Liver%20Segmentation%20Software>)<sup>16</sup>. This model was based on a 3D U-Net architecture. It used a variant of the GAN known as CycleGAN to convert contrast-enhanced CT scans into non-contrast CT scans, thereby augmenting the dataset. We converted the DICOM files into NIfTI format then ran the model using command lines to segment the liver. Fig. 2 shows a non-contrast CT slice of the liver and the corresponding AI-generated segmentation. The AI model used in this study was a pre-trained 3D U-Net model, and no further training was performed. This model achieved a mean DSC of  $0.887 \pm 0.006$  (standard deviation)<sup>9</sup>. The median processing time for liver segmentation per study was reported to be 55 seconds<sup>9</sup>.

#### *Performance evaluation*

The agreement of the segmented regions was quantitatively evaluated using the DSC<sup>17</sup>, a metric commonly used in medical image segmentation to assess similarity between two segments. Specifically, it is calculated as the ratio of twice the volume of the intersection of segments A and B to the sum of the volumes of segments A and B. The symbol  $|X|$  denotes the cardinality, i.e., the total number of voxels within the region X. The DSC value ranges from 0 to 1, with a value of 1 indicating a perfect match between the two segments.

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

Another metric used to evaluate the agreement of the segmented regions was the Jaccard Index(JI). Details of the calculation and interpretation are provided in the Supplementary Materials.

#### *Statistical analysis*

Statistical analyses were performed using JMP Pro 17.0 (SAS Institute Inc., Cary, NC, 1989–2024). Demographic, clinical, and radiologic characteristics of participants are presented as frequencies

and percentages for categorical variables. Continuous variables are expressed as means  $\pm$  standard deviations or medians with interquartile ranges (IQRs). For comparisons of paired data, the Kruskal–Wallis test was used. To account for multiple comparisons, we applied the Bonferroni correction. Spearman’s rank correlation coefficient was used to assess the correlation between two sets of data. The intraclass correlation coefficient (ICC) was used to evaluate the level of agreement between liver volume measurements. In Analysis 2, we plotted age against liver volume to determine the slope and assess the amount of change in liver volume with each additional year of age. A p-value of less than 0.05 was considered statistically significant.

## Results

### *Patient characteristics*

In Analysis 1, Cohort 1A (n=10) had a mean patient age of 27 years (IQR 24.5–37.75), and Cohort 1B (n=24) had a mean age of 38.5 years (IQR 28.3–55.3). The combined Cohort 1A+1B (n=34) had a mean age of 36.5 years (IQR 26.0–50.8). In Analysis 2, Cohort 2 (n=39) had a mean patient age of 74 years (IQR 68.8–81.3). The characteristics of the patients are summarized in Table 2.

### *Analysis 1 on Cohort 1A (Part 1)*

The average liver volume of 10 healthy individuals, as measured by Radiologists A and B, was  $1373 \pm 319$  mL. Scatter plots showing the volumes measured using each CT phase by the AI model and the two radiologists are presented in Fig. S1. The correlation coefficients and mean errors are also provided. In most cases and across most phases, the model yielded larger volume measurements than the radiologists. The correlation coefficients were mostly around 0.9, indicating a strong correlation. The graphs and tables for the DSC are shown in Fig. 3(a) and Table 3. The correlation coefficients were above 0.96 for contrast-enhanced CT and above 0.88 for non-contrast CT, indicating a strong correlation (Fig. 3(b)). The JI results are provided in Supplementary Table 2.

When comparing the agreement between the two radiologists’ segmentations across all five phases, the DSC was greater than 0.97 in all phases, indicating very high agreement. The highest DSC was observed in the late arterial phase; for subsequent analyses, the volumes segmented by Radiologist A in this phase were used as the ground truth.

### *Analysis 1 on Cohort 1A+1B (Part 2)*

After adding 24 cases (Cohort 1B), we compared the segmentation between the model and Radiologist A using only the late arterial phase. The AI model tended to produce slightly larger volumes than those of Radiologist A (Fig. 4(a)). The DSC was  $0.986 \pm 0.017$ , and the ICC was 0.9802, both indicating very high agreement.

### *Analysis 1 on Cohort 1A+1B (Part 3)*

Liver volumes were measured using the AI model for each CT phase, targeting Cohort 1A+1B. The measurements from AI auto-segmentation for each phase are shown in Fig. 4(b). A Kruskal–Wallis test was conducted, yielding a  $p$  value of 0.4359, indicating no significant difference across the phases. ICCs were calculated against the ground truth (Table 4). The ICC values were 0.833 for non-contrast CT, 0.870 for the early arterial phase, 0.980 for the arterial phase, 0.949 for the portal venous phase, and 0.882 for the equilibrium phase. All ICCs were greater than 0.8 (high agreement), suggesting the AI model is applicable across all phases.

### *Analysis 2*

A total of 527 measurements were taken from 39 patients, with an average of 13.175 measurements per patient. The observation period ranged from 5 to 20 years (median follow-up time 13.73 years). The liver volume at the first scan was  $1252 \pm 431.7$  mL. A scatter plot of age versus liver volume was created using the raw data, with a correlation coefficient ( $\rho$ ) of -0.451 (Fig. 5(a)). The liver volume index was plotted against the number of years since the start of observation (Fig. 5 (b-d)). The slope for each case was calculated, indicating liver volume decreased by an average of 0.95% per year (95% CI = 0.70 - 1.21% per year).

## **Discussion**

In this study, we evaluated a fully automated AI model for liver segmentation. The model demonstrated high accuracy, achieving DSC values of  $0.988 \pm 0.010$  (late arterial),  $0.979 \pm 0.021$  (portal),  $0.977 \pm 0.015$  (equilibrium),  $0.968 \pm 0.019$  (early arterial), and  $0.954 \pm 0.045$  (non-contrast). Notably, the arterial phase DSC ( $0.948 \pm 0.003$ ) exceeded prior reports, likely due to our cohort of healthy individuals<sup>16</sup>. The AI-generated segments tended to be slightly larger than radiologist-drawn ones, occasionally including non-liver tissues. However, the AI rarely missed true liver tissue. The ICC exceeded 0.8 for non-contrast CT and surpassed 0.9 for all contrast-enhanced phases, confirming high accuracy. A longitudinal analysis revealed a liver volume reduction of 0.95% per year over 20 years.

Segmentation errors were more frequent near the hepatic hilum and intestine, particularly in low-fat individuals where the liver-muscle boundary was unclear. In rare cases, the AI misclassified distant structures, such as the psoas major muscle. These findings align with previous reports of peri-hepatic over-segmentation and under-segmentation in areas with inhomogeneous density<sup>18</sup>. The high DSC in the arterial phase likely reflects improved delineation of the liver from surrounding structures, including the kidneys, digestive tract, muscles, and gallbladder.

In Analysis 2, we confirmed an annual liver volume decline of 0.95%, consistent with known age-related reductions in liver size and function<sup>19,20</sup>. Although manual corrections were not applied in Cohort 2, segmentation errors such as slight over- or under-segmentation occurred in both directions with similar

frequency. Therefore, while the absolute liver volumes may include minor inaccuracies, we believe the analysis of temporal trends—such as the observed annual decrease of approximately 0.95%—remains valid and robust. Previous research indicates liver weight decreases by 100 g per decade, with hepatocyte size increasing and regeneration capacity declining<sup>21</sup>. AI facilitates large-scale analyses, with prior studies using AI models in over 3,000 individuals to establish liver volume thresholds based on body weight. Unlike commercial semi-automatic tools, our AI model is free and enables batch processing, making it valuable for large-scale research.

Liver segmentation is critical for transplantation, surgical planning, and oncology<sup>22</sup>. Historically, liver volume estimation relied on height and weight-based formulas, but modern DICOM viewers offer segmentation tools. However, volumetry is not routinely included in radiology reports due to limited clinical applications. Liver volume assessment may aid prognosis in conditions like acute and alcoholic hepatitis. Further research is needed to clarify its role in disease diagnosis and progression<sup>23</sup>. Automatic liver volume measurement is also relevant for monitoring radiation therapy effects and fatty liver disease, though interpreting normal and pathological ranges remains challenging. If information on temporal changes in liver volume is available during image interpretation, it may help assess the progression of chronic diseases, identify pre-disease conditions, and predict prognosis. Additionally, longitudinal assessment of liver volume may contribute to the early detection of liver pathology, identification of age-related functional decline, and prediction of prognosis in liver disease.

The study has some limitations. First, the study population was relatively small. A larger, more diverse cohort stratified by age and disease is needed to build a comprehensive database of temporal changes in liver volume. Second, the model used cannot measure the volume of intrahepatic segments. It is well-established that cirrhosis leads to right lobe atrophy and left lobe enlargement, making it crucial to study regional volume changes in addition to total liver volume. Advanced software capable of measuring individual liver segments has been developed, and future studies should incorporate such tools<sup>24</sup>. Third, segmentation errors were not corrected by a radiologist in Analysis 2, which may have introduced variability into the measured data. Despite this, the model demonstrated high accuracy, and statistical measures for large groups are considered meaningful.

## **Conclusion**

This model demonstrated high accuracy in liver segmentation across various CT phases and provided valuable insights into liver volume reduction with aging. We observed that liver volume decreased by 0.95% per year with age. The relationship between liver volume and liver function should be further investigated to predict the prognosis of liver diseases.

## **Relevance Statement**

AI-driven liver volume quantification shows strong agreement with radiologists and reveals a

0.9% annual decline with aging, emphasizing its potential in liver function assessment and disease prognosis.

## References

1. Ferrero A, Viganò L, Polastri R, et al. Postoperative Liver Dysfunction and Future Remnant Liver: Where Is the Limit? *World J Surg.* 2007;31(8):1643-1651. doi:10.1007/S00268-007-9123-2
2. Kubota K, Makuuchi M, Kusaka K, et al. Measurement of liver volume and hepatic functional reserve as a guide to decision-making in resectional surgery for hepatic tumors. *Hepatology.* 1997;26(5):1176-1181. doi:10.1002/HEP.510260514
3. Cohnert TU, Rau HG, Buttler E, et al. Preoperative Risk Assessment of Hepatic Resection for Malignant Disease. *World J Surg.* 1997;21(4):396-401. doi:10.1007/PL00012260
4. Urata K, Kawasaki S, Matsunami H, et al. Calculation of child and adult standard liver volume for liver transplantation. *Hepatology.* 1995;21(5):1317-1321. doi:10.1016/0270-9139(95)90053-5
5. Martel G, Cieslak KP, Huang R, et al. Comparison of techniques for volumetric analysis of the future liver remnant: implications for major hepatic resections. *HPB.* 2015;17(12):1051-1057. doi:10.1111/HPB.12480
6. Kennedy PA, Madding GF. Surgical Anatomy of the Liver. *Surgical Clinics of North America.* 1977;57(2):233-244. doi:10.1016/S0039-6109(16)41180-1
7. Um EH, Hwang S, Song GW, et al. Calculation of standard liver volume in Korean adults with analysis of confounding variables. *Korean J Hepatobiliary Pancreat Surg.* 2015;19(4):133. doi:10.14701/KJHBPS.2015.19.4.133
8. Linguraru MG, Sandberg JK, Jones EC, Petrick N, Summers RM. Assessing Hepatomegaly. Automated Volumetric Analysis of the Liver. *Acad Radiol.* 2012;19(5):588-598. doi:10.1016/J.ACRA.2012.01.015
9. Perez AA, Noe-Kim V, Lubner MG, et al. Deep Learning CT-based Quantitative Visualization Tool for Liver Volume Estimation: Defining Normal and Hepatomegaly. *Radiology.* 2022;302(2):336-342. doi:10.1148/RADIOL.2021210531/ASSET/IMAGES/LARGE/RADIOL.2021210531FIG5.JPEG
10. Poovathumkadavil A, Leung KF, Al Ghamdi HM, El Hassan Othman I, Meshikhes AW. Standard formula for liver volume in Middle Eastern Arabic adults. *Transplant Proc.* 2010;42(9):3600-3605. doi:10.1016/J.TRANSPROCEED.2010.07.098
11. Kwo PY, Ramchandani VA, O'Connor S, et al. Gender differences in alcohol metabolism: Relationship to liver volume and effect of adjusting for body mass. *Gastroenterology.* 1998;115(6):1552-1557. doi:10.1016/S0016-5085(98)70035-6

12. Cai W, He B, Fan Y, Fang C, Jia F. Comparison of liver volumetry on contrast-enhanced CT images: One semiautomatic and two automatic approaches. *J Appl Clin Med Phys*. 2016;17(6):118-127. doi:10.1120/jacmp.v17i6.6485
13. Tang A, Chen J, Le TA, et al. Cross-sectional and longitudinal evaluation of liver volume and total liver fat burden in adults with nonalcoholic steatohepatitis. *Abdom Imaging*. 2015;40(1):26-37. doi:10.1007/S00261-014-0175-0/TABLES/3
14. Tang A, Chen J, Le TA, et al. Cross-sectional and longitudinal evaluation of liver volume and total liver fat burden in adults with nonalcoholic steatohepatitis. *Abdom Imaging*. 2015;40(1):26-37. doi:10.1007/S00261-014-0175-0
15. Nayantara PV, Kamath S, Kadavigere R, Manjunath KN. Automatic Liver Segmentation from Multiphase CT Using Modified SegNet and ASPP Module. *SN Comput Sci*. 2024;5(4). doi:10.1007/S42979-024-02719-2
16. Sandfort V, Yan K, Pickhardt PJ, Summers RM. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports 2019 9:1*. 2019;9(1):1-9. doi:10.1038/s41598-019-52737-x
17. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945;26(3):297-302. doi:10.2307/1932409
18. Heimann T, Van Ginneken B, Styner MA, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imaging*. 2009;28(8):1251-1265. doi:10.1109/TMI.2009.2013851
19. Perez AA, Noe-Kim V, Lubner MG, et al. Deep Learning CT-based Quantitative Visualization Tool for Liver Volume Estimation: Defining Normal and Hepatomegaly. *Radiology*. 2022;302(2):336-342. doi:10.1148/RADIOL.2021210531/ASSET/IMAGES/LARGE/RADIOL.2021210531FIG5.JPEG
20. Yoshizumi T, Gondolesi GE, Bodian CA, et al. A simple new formula to assess liver weight. *Transplant Proc*. 2003;35(4):1415-1420. doi:10.1016/S0041-1345(03)00482-2
21. Motomi Katayama, Kiyotaka Yamazumi, Kenji Kino, Masanobu Tsuru, Toshio Fukuzawa, Hiroyuki Shimada. Change of Liver Weight in the Elderly. *Japanese Journal of Geriatrics*. 1990;27(5):584-588.
22. Urata K, Kawasaki S, Matsunami H, et al. Calculation of child and adult standard liver volume for liver transplantation. *Hepatology*. 1995;21(5):1317-1321. doi:10.1016/0270-9139(95)90053-5
23. Zabron A, Quaglia A, Fatourou E, et al. Clinical and prognostic associations of liver volume determined by computed tomography in acute liver failure. *Liver International*.

- 2018;38(9):1592-1601. doi:10.1111/LIV.13725
24. Lee S, Elton DC, Yang AH, et al. Fully Automated and Explainable Liver Segmental Volume Ratio and Spleen Segmentation at CT for Diagnosing Cirrhosis. *Radiol Artif Intell.* 2022;4(5). doi:10.1148/ryai.210268

## Figure legends

Fig. 1: Flow diagram illustrating the inclusion and exclusion criteria for the cohort in Analysis 1

Fig. 2: Example of liver auto-segmentation performed using the artificial intelligence (AI) model. (a) Non-contrast CT scan. (b) Liver region segmented by the AI model, highlighted in red. Parts (c) and (d) show corrections made by Radiologist A and Radiologist B, respectively, by modifying the AI-generated liver region. Areas where AI segments were removed by radiologists include the gallbladder (*blue arrow*), digestive tract (*red arrow*), and rectus abdominis (*yellow arrow*).

Fig. 3: (a) Comparison of the artificial intelligence (AI) model with two radiologists using Dice similarity coefficients (DSC). The vertical axis represents the DSC, while the horizontal axis shows comparisons between the AI model and Radiologist A (*blue*), the AI model and Radiologist B (*red*), and between Radiologist A and Radiologist B (*green*). The DSC values are shown across different CT phases. (b) Comparison of correlation coefficients between the artificial intelligence (AI) model and radiologists across different CT phases. The correlation coefficients between the AI model and Radiologist A and Radiologist B were above 0.96 for contrast-enhanced CT and above 0.88 for non-contrast CT, indicating a strong correlation and good agreement between the AI model and both radiologists.

Fig. 4: (a) Scatter plot comparing liver volumes measured by the artificial intelligence (AI) model and Radiologist A in the late arterial phase. AI tended to produce slightly larger volumes compared to Radiologist A. The correlation coefficient ( $\rho$ ) is indicated. The red line represents the identity line ( $y=x$ ), while the blue line represents the regression line. (b) Results of liver volume auto-segmentation by the artificial intelligence (AI) model for each phase, showing almost identical volumes across all phases. A Kruskal–Wallis test was performed, yielding  $p = 0.4359$ , indicating no significant difference in liver volumes across different phases.

Fig. 5: (a) Scatter plot showing liver volume in relation to age in Analysis 2. A gradual decrease in liver volume with age is observed.

(b-d) Temporal changes in liver volume for normal subjects in Analysis 2. (b) The vertical axis represents liver volume (mL), and the horizontal axis represents age. (c) The vertical axis represents the liver volume index, normalized to 100% of the patient's liver volume at the start of the observation period. The horizontal axis shows the number of years of follow-up. (d) The regression line is shown in graph b. By calculating the slope for each individual case and averaging the results, we observed an annual decrease in liver volume of 0.95%.

Fig. 1

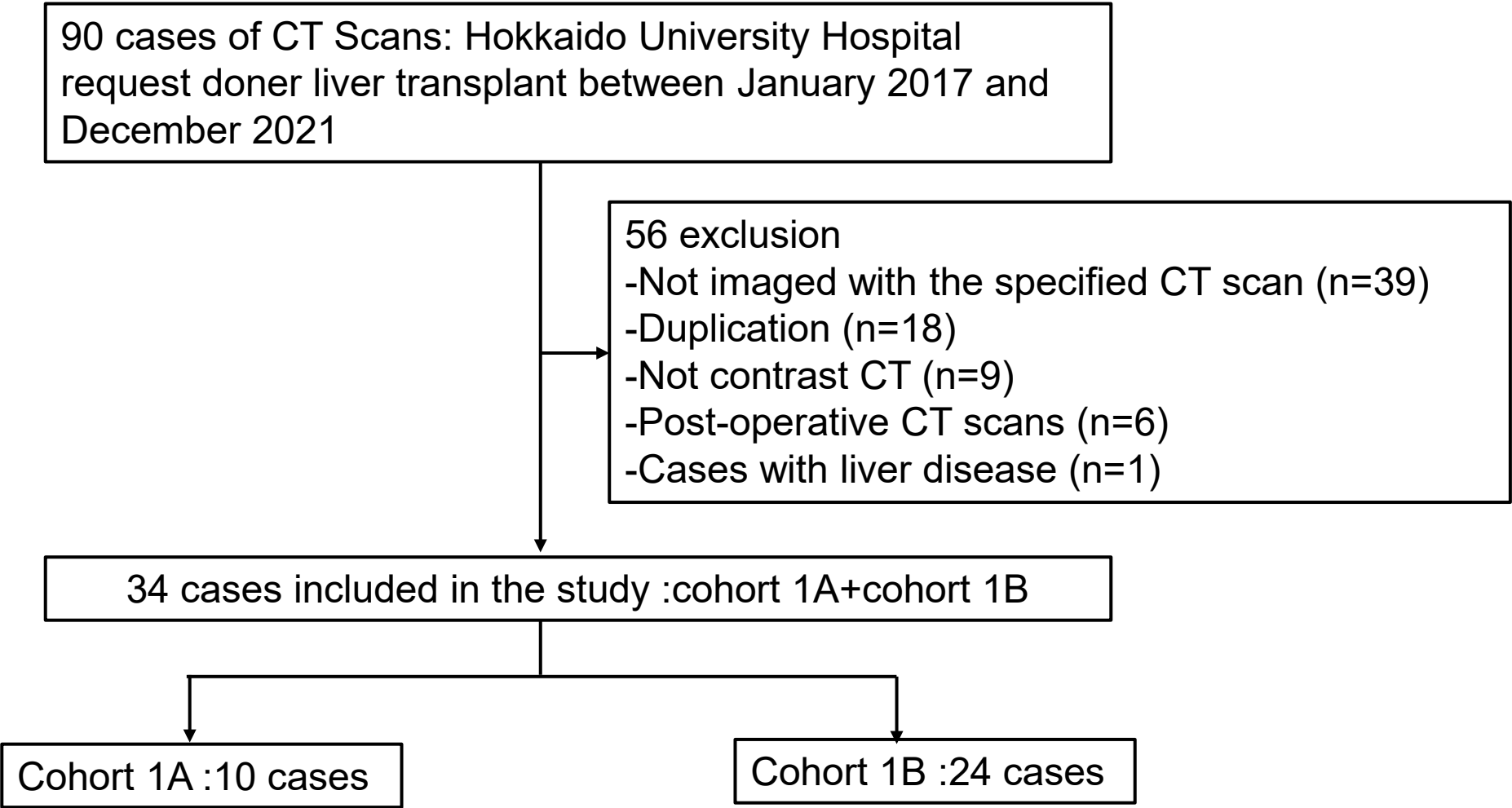


Fig. 2

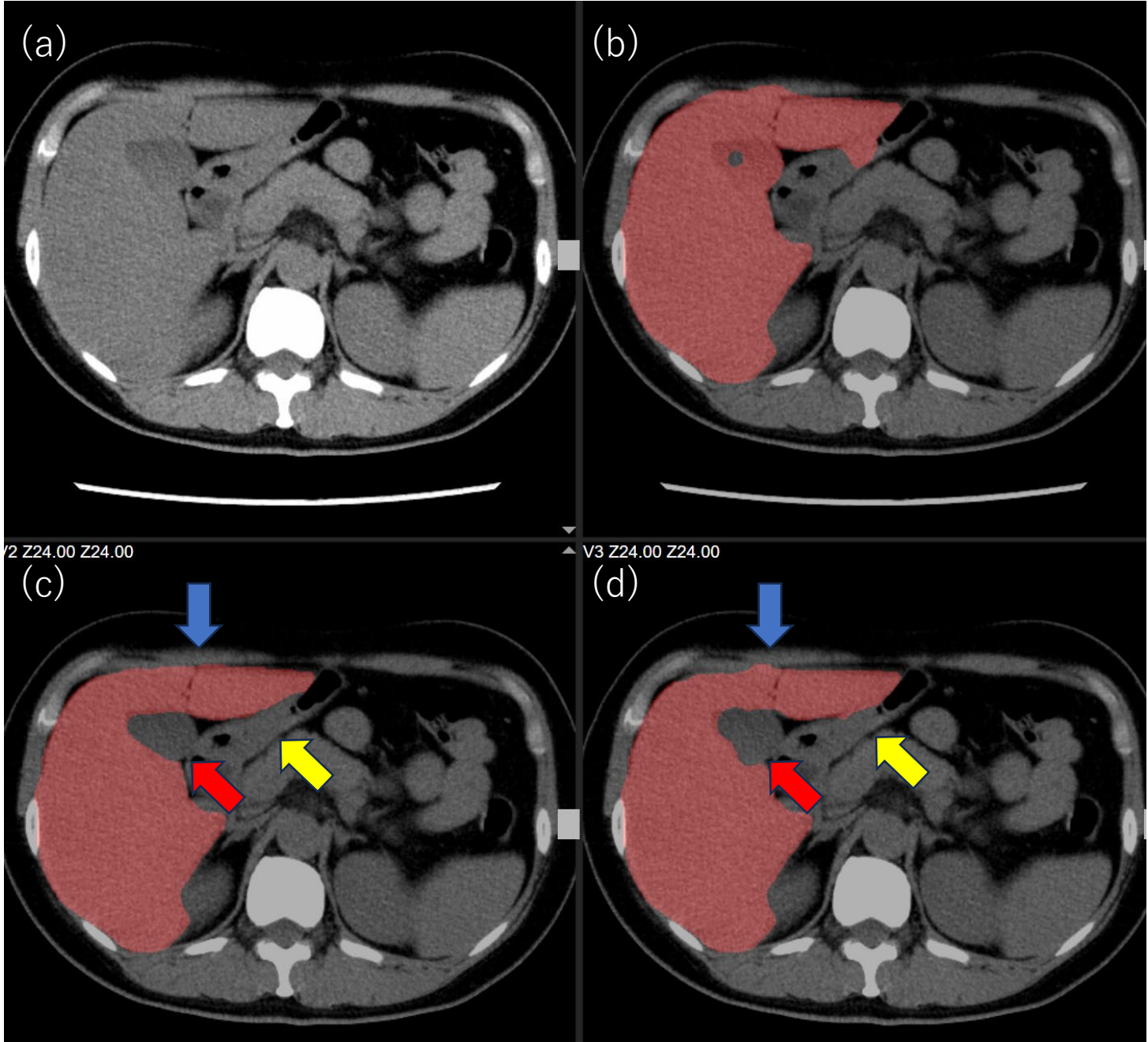
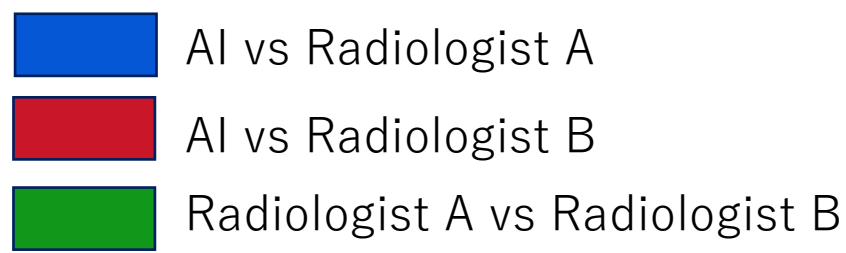
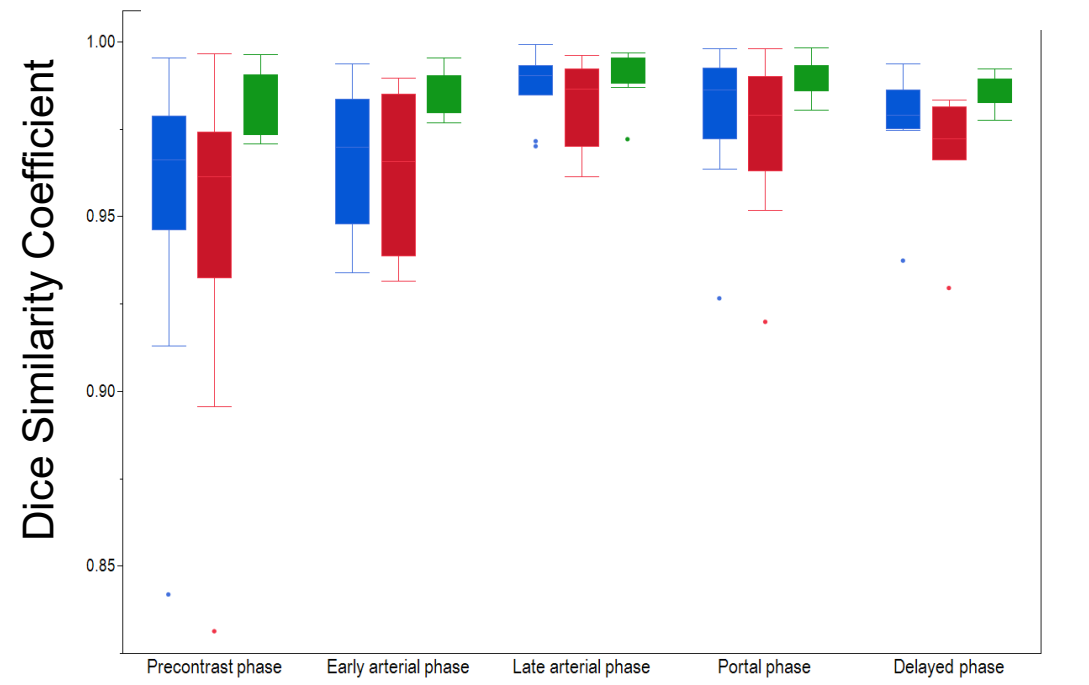


Fig. 3

(a)



(b)

Correlation coefficient of liver volume

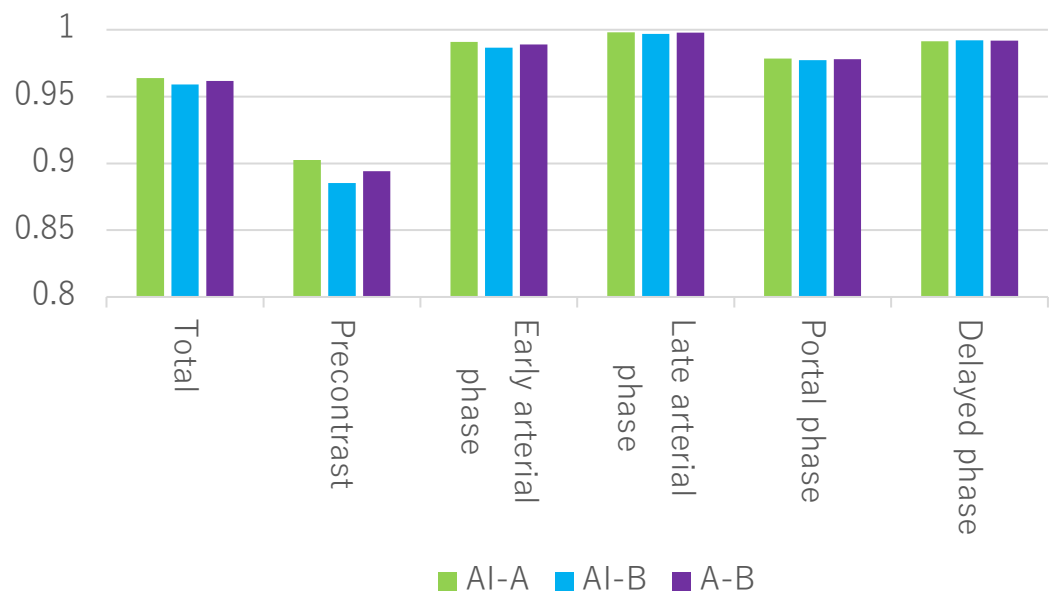
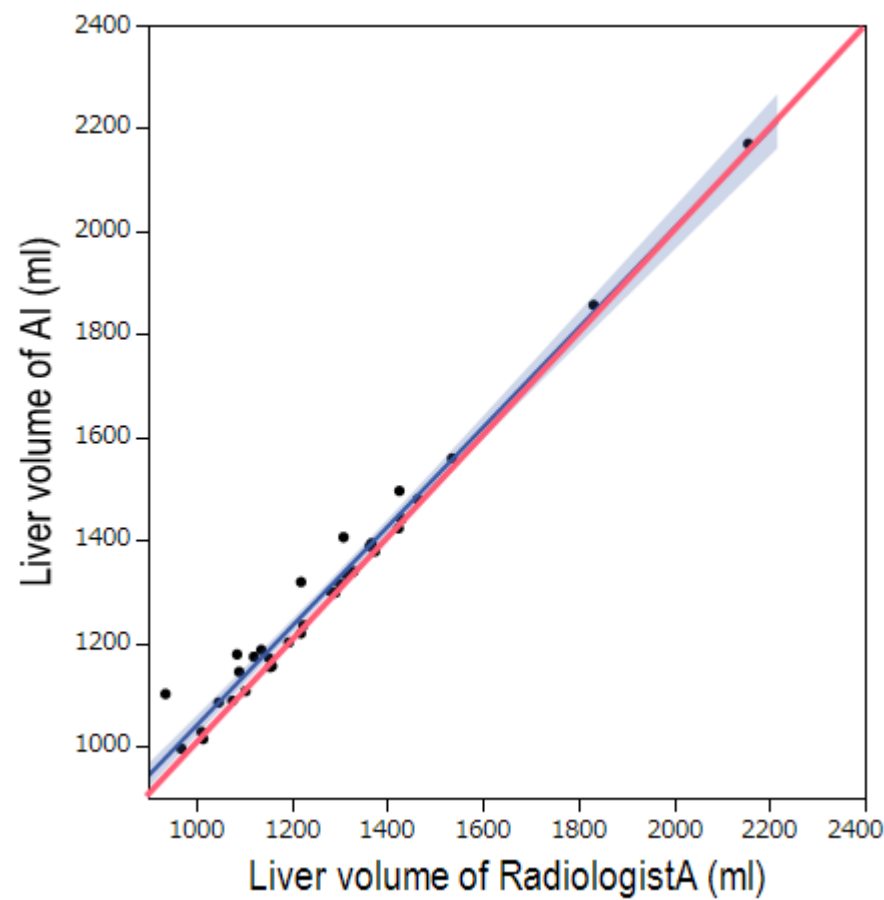


Fig. 4

(a)



(b)

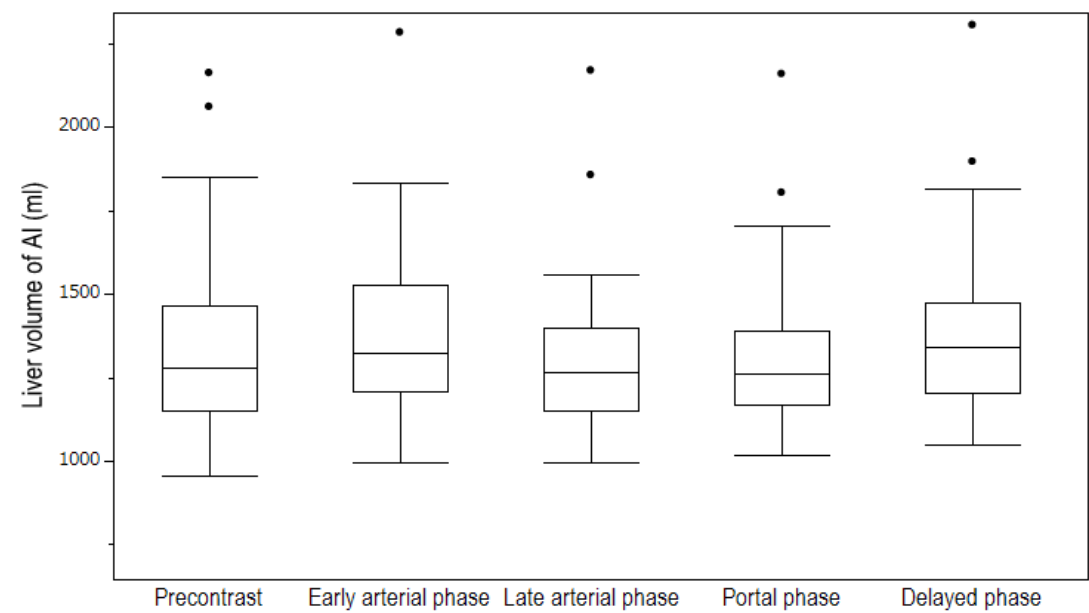
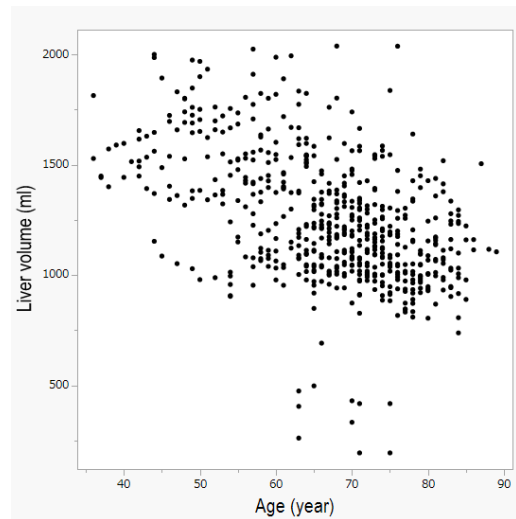
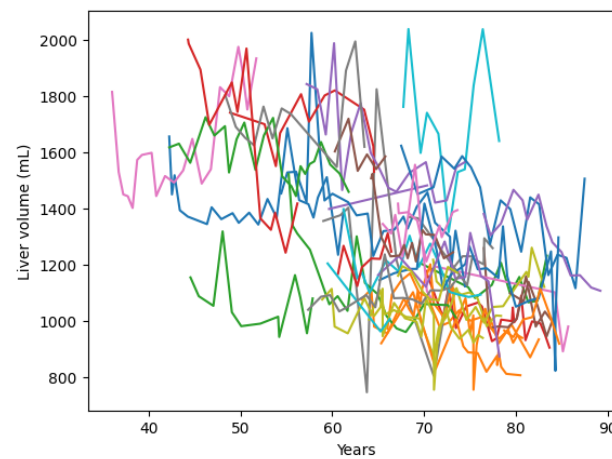


Fig. 5

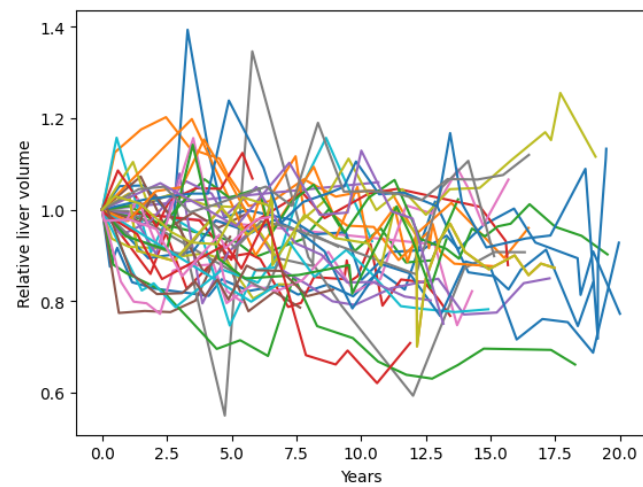
(a)



(b)



(c)



(d)

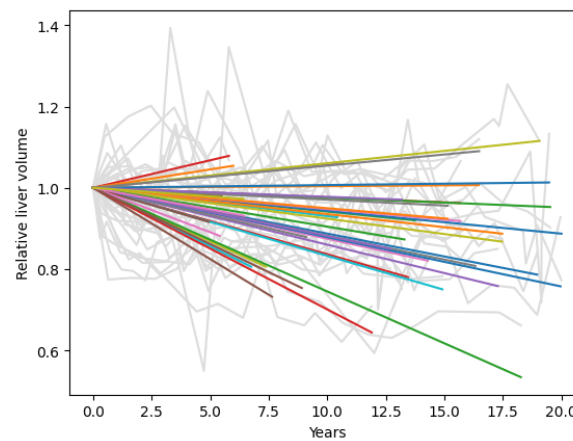


Table 1: This is the research plan for Analysis 1

|                         | <b>Part 1</b>        | <b>Part 2</b>       | <b>Part 3</b> |
|-------------------------|----------------------|---------------------|---------------|
| <b>Cohort (n)</b>       | 1A (10)              | 1A+1B (34)          | 1A+1B (34)    |
| <b>Who made segment</b> | AI, two radiologists | AI, one radiologist | AI only       |
| <b>Metrics</b>          | DSC, volume          | Volume              | Volume        |
| <b>Phase</b>            | Five phases          | Late arterial phase | Five phases   |
| <b>Ground truth</b>     | Not defined          | Radiologist A       | Radiologist A |

Table 2: Patient Demographics

|            | Cohort 1A           | Cohort 1B           | Cohort (1A+1B)        | Cohort 2*           |
|------------|---------------------|---------------------|-----------------------|---------------------|
| N          | 10                  | 24                  | 34                    | 39                  |
| Age(years) | 27 [24.5-37.75]     | 38.5 [28.3-55.3]    | 36.5 [26.0-50.8]      | 74 [68.8-81.3]      |
| Male       | 4 (40%)             | 14 (58%)            | 18 (53%)              | 32 (80%)            |
| Height(cm) | 164.3 [160.9-175.9] | 168.1 [162.7-171.3] | 168.05 [162.15-174.5] | 161.3 [149.6-164.3] |
| Weight(kg) | 59.5 [54.1-66.9]    | 61.9 [58.2-67.6]    | 61.7 [55.8-67.3]      | 57.4 [52.1-64.5]    |

\*The data for Cohort 2 were obtained from the start date of the observation period.

Table 3: The median and range of the DSC between the AI model and the radiologists

| <b>Phase</b>                | <b>AI-R<sub>A</sub></b> | <b>AI-R<sub>B</sub></b> | <b>R<sub>A</sub> - R<sub>B</sub></b> | <b>Median (Range)</b> |
|-----------------------------|-------------------------|-------------------------|--------------------------------------|-----------------------|
| <b>Precontrast phase</b>    | 0.966 (0.842-0.995)     | 0.961 (0.831-0.997)     | 0.983 (0.971-0.996)                  | 0.978 (0.831-0.997)   |
| <b>Early arterial phase</b> | 0.967 (0.934-0.993)     | 0.966 (0.931-0.990)     | 0.986 (0.931-0.990)                  | 0.979 (0.931-0.996)   |
| <b>Late arterial phase</b>  | 0.990 (0.970-0.999)     | 0.986 (0.961-0.996)     | 0.992 (0.972-0.997)                  | 0.990 (0.961-0.999)   |
| <b>Portal phase</b>         | 0.986 (0.927-0.998)     | 0.979 (0.920-0.998)     | 0.991 (0.980-0.998)                  | 0.987 (0.920-0.998)   |
| <b>Delayed phase</b>        | 0.979 (0.937-0.994)     | 0.972 (0.930-0.983)     | 0.985 (0.977-0.992)                  | 0.981 (0.930-0.994)   |

R<sub>A</sub>: Radiologist A, R<sub>B</sub>: Radiologist B

Statistical significance:  $p < 0.0101$  (AI-R<sub>A</sub> vs. R<sub>A</sub> - R<sub>B</sub>),  $p < 0.001$  (AI-R<sub>B</sub> vs. R<sub>A</sub> - R<sub>B</sub>),  $p < 0.0002$  (Precontrast phase vs. Late arterial phase),  $p < 0.0096$  (Precontrast phase vs. Portal phase),  $p < 0.0433$  (Precontrast phase vs. Delayed phase)

Table 4: Interclass correlation coefficient of the arterial phase by the radiologist vs. ICC of each phase by AI

| <b>Phase</b>                | <b>ICC</b> |
|-----------------------------|------------|
| <b>Precontrast phase</b>    | 0.833      |
| <b>Early arterial phase</b> | 0.870      |

|                            |       |
|----------------------------|-------|
| <b>Late arterial phase</b> | 0.980 |
| <b>Portal phase</b>        | 0.949 |
| <b>Delayed phase</b>       | 0.882 |