



Title	AIセグメンテーションによる肝臓の形態解析と体積の経時的变化に関する研究
Author(s)	木村, 理奈
Description	配架番号 : 2987
Degree Grantor	北海道大学
Degree Name	博士(医学)
Dissertation Number	甲第16894号
Issue Date	2026-03-25
DOI	https://doi.org/10.14943/doctoral.k16894
Doc URL	https://hdl.handle.net/2115/99739
Type	doctoral thesis
File Information	KIMURA_Rina.pdf, 全文



学位論文

AI セグメンテーションによる
肝臓の形態解析と体積の経時的変化に
関する研究

(A Study on Hepatic Morphology and Longitudinal
Volume Changes Using AI-Based Segmentation)

2026 年 3 月

北海道大学
木村 理奈

学位論文

AI セグメンテーションによる
肝臓の形態解析と体積の経時的変化に
関する研究

(A Study on Hepatic Morphology and Longitudinal
Volume Changes Using AI-Based Segmentation)

2026 年 3 月

北海道大学
木村 理奈

目次

発表論文目録および学会発表目録	1
要旨	2
略語集	5
全体の緒言	6
第一章	
緒言	11
方法	13
結果	20
考察	26
第二章	
緒言	28
方法	30
結果	32
考察	38
第三章	
緒言	40
方法	45
結果	47
考察	56
全体の結論	58
謝辞	61
利益相反	62
引用文献	63

発表論文目録及び学会発表目録

本研究の一部は、以下の論文に発表した

- ・ Kimura R, Hirata K, Tsuneta S, Takenaka J, Watanabe S, Abo D, Kudo K.
Evaluation of artificial-intelligence-based liver segmentation and its
application for longitudinal liver volume measurement.
Abdom Radiol (NY) . 2025 Jun 10. doi: 10.1007/s00261-025-05050-3.

本研究の一部は、以下の学会に発表した

1. 木村 理奈、平田 健司、常田 慧徳、竹中 淳規、渡邊 史郎、阿保
大介、工藤 與亮

AIによる肝臓セグメンテーションの性能評価と肝臓体積の経時的変化

第59回 日本医学放射線学会秋季臨床大会 徳島県徳島市 2023年9月
15日-17日

2. 木村 理奈、平田 健司、工藤 與亮

AIを利用した肝臓セグメンテーションの性能評価

第6回 日本メディカルAI学会学術集会 愛知県名古屋市 2024年6月
21日-22日

要旨

【背景と目的】CT 画像における臓器領域の自動抽出はセグメンテーションと呼ばれ、臓器の正確な三次元体積を測定することが可能となり、手術計画、移植ドナー評価、治療効果判定など臨床の様々な場面で重要な指標となる。臓器体積は臓器機能と密接に関連し、特に肝臓に関して多くの報告がなされている。一方で、従来の手動セグメンテーションは膨大な時間を要し、さらに観察者間・観察者内でのばらつきが大きい。これに対して近年、人工知能（artificial intelligence : AI）が急速に発展し、自動セグメンテーションの精度と効率は大幅に向上している。

本研究では肝体積の測定を通じた疾患理解を目指し、第一章では AI モデルの性能評価として、肝体積の測定精度を様々な CT 造影相で比較検討した。次に肝臓は加齢により体積が徐々に低下するとされているが、長期間の縦断的研究は少ないため、第二章では第一章で検証した AI モデルを用いて肝疾患を持たない肝体積を測定し、加齢による肝体積の経時的変化を調べた。これをベースラインとして第三章では種々の肝疾患を持つ症例において肝体積を縦断的に計測し、加齢に対する上乗せ影響を検討した。

○第一章

【対象と方法】2017 年 1 月から 2021 年 12 月に当院で術前 CT 検査を受けた生体肝移植ドナー 34 例を対象とした。CT は非造影、早期動脈相、後期動脈相、門脈相、平衡相の 5 相が撮像された。米国 NIH が開発し GitHub 上で公開されている AI モデルを使用し、肝臓のセグメンテーションを実施した。Part1 ではランダムに抽出した 10 例の 5 相を対象とした。AI モデルでセグメンテーションを行い、放射線科医 A、B の 2 名が AI 推定結果を修正したセグメントを基準に、セグメント一致度をダイス類似度係数（Dice similarity coefficient : DSC）で評価した。また、放射線科医 2 名による解析で高い一致度が得られたため、Part2 では後期動脈相 34 例を対象とし、AI モデルと放射線科医 A のみのセグメントの一致度を評価した。Part3 では AI モデルのみを用いて 34 例 5 相全ての体積を算出し、造影相間の一致度を評価した。

【結果】Part1 では全測定者と全時相における平均肝体積は 1373 ± 319 mL、DSC は 0.978 以上、ICC は 0.813 以上を示し、AI モデルと放射線科医の作

成したセグメントは高い相関が認められた。Part2 では後期動脈相 34 例における AI と放射線科医 A の DSC は 0.986 ± 0.017 と高い一致度が得られた。Part3 では、AI モデルが算出した各相の体積に有意差を認めず、ICC は 0.833 以上を示した。

【考察】AI モデルの性能は放射線科医のセグメンテーションと比較して DSC は 0.978 以上を示し、高い一致度を示した。また、CT の各相においてもセグメントの一致度は高く、非造影 CT、造影 CT のどの時相でも使用可能であることが確認された。肝臓は加齢により体積が徐々に低下するとされているが、縦断的研究が行われた報告は少ない。第一章の AI モデルを用いて、第二章では肝疾患のない患者群における肝体積の縦断的研究を行うこととした。

○第二章

【対象と方法】加齢による肝体積変化の調査のため、肝疾患が無く、薬剤性肝障害を引き起こしうる化学療法歴のない症例のうち、2000 年から 2022 年の間に CT で 5 年以上の経過観察が行われた 39 例（計 527 検査）を対象とした。第一章で検証した AI モデルを用いて CT から肝体積を自動算出し、加齢に伴う肝体積変化を縦断的に解析した。体積変化の解析には直線近似および指数関数近似を用い、それぞれ単位時間あたりの体積変化率を算出した。

【結果】観察期間は 5-20 年、中央値は 13.7 年であった。各症例の肝体積の変化率は、直線近似では $0.95 \pm 1.1\%$ ／年の減少、指数関数近似では $1.1 \pm 1.3\%$ ／年の減少であった。

【考察】縦断的解析の結果、肝疾患のない症例における肝体積は平均して年 1% 程度減少する傾向が示された。第二章で、肝疾患がない場合でも肝体積は加齢に伴って減少していくことが明らかになったため、これを踏まえ第三章では肝疾患と肝体積の関連について検討することとした。

○第三章

【対象と方法】2004 年から 2022 年の間に肝疾患の診断で CT 検査を受け、5 年以上の経過観察が可能であった 200 例（男性 93 例、平均年齢 64 歳、観察期間中央値 2,902.5 日 [IQR: 1,577.8-4,285.3]）を抽出し、合計 3,656 件の CT 画像を対象とした。第二章で用いた AI モデルでは正常肝では精度が高かったものの、肝疾患症例ではエラーが多く確認されたため、新たに開発されたバーゼル大学が開発したオープンソースソフトウェアである

TotalSegmentator を用いた自動セグメンテーションを実施した。対象疾患は、B 型肝炎、C 型肝炎、自己免疫性肝炎、代謝機能障害関連脂肪性肝炎／代謝機能障害関連脂肪肝炎、原発性胆汁性胆管炎、Wilson 病、アルコール性肝障害、うっ血肝、薬剤性肝炎であった。

【結果】疾患ごとの肝体積の年間変化率を算出したところ、多くの疾患で減少傾向が認められた一方、B 型肝炎／B 型肝硬変、MASH、Wilson 病では増加傾向が観察された。両側 t 検定の結果、PBC+AIH 群のみが有意な減少 ($p<0.05$) を示したが、その他の疾患群では統計学的に有意な変化率とはいえなかった。

【考察】B 型肝炎や MASH、Wilson 病の一部の症例では炎症や代謝異常に伴う体重増加がみられた一方、多くの慢性肝疾患では線維化や肝硬変進展に伴う減少傾向を示した。しかし、本研究は症例数の不足や治療介入の影響から統計的有意差を得るには至らなかった。今後はより大規模な研究を通じて、肝体積の変化を血液検査や肝予備能評価と統合し、臨床的意義を明確にすることが求められる。

【全体の結論】第一章では肝臓セグメンテーションを行う AI モデルの性能評価を実施し、いずれの造影時相においても高精度の肝臓セグメンテーションが可能であり、放射線科医の手動セグメンテーションと比較しても同等の精度を有することを示した。第二章では、この AI モデルを肝疾患のない集団に応用して縦断的解析を行い、加齢に伴い肝体積が年間 1.1%減少することを明らかにした。さらに第三章では、新たに開発された AI モデルである TotalSegmentator を用い、肝疾患症例における肝体積の推移を解析し、疾患特異的な体積変化の評価を試みた。多くの肝疾患で減少傾向が見られたが、統計的有意差は見られなかった。AI を活用することで、多数例に対して自動かつ効率的にセグメンテーションを行うことが可能となり、大規模データを基盤とした体積解析が現実的に実施できるようになった。今後は体積変化と肝機能や予後との関連を精緻に評価し、肝体積計測の臨床応用がさらに進展すると考えられる。

略語表

本文中および図中で使用した略語は以下の通りである。

AI	artificial intelligence
AIP	autoimmune hepatitis
CNN	convolutional neural network
CT	X-ray Computed Tomography
CycleGAN	Cycle-consistent Generative Adversarial Network
DICOM	Digital Imaging and Communication in Medicine
DL	deep learning
DSC	Dice similarity coefficient
ICC	intraclass correlation coefficient
IQR	interquartile range
JI	Jaccard Index
NIfTI	The Neuroimaging Informatics Technology Initiative
MASH	metabolic dysfunction associated steatohepatitis
MASLD	metabolic dysfunction associated steatotic liver disease
MRI	magnetic resonance imaging
PBC	primary biliary cholangitis
SD	standard deviation
SVR	sustained virological response

全体の緒言

1. 臓器セグメンテーションの概要と AI 利用

(1) 臓器のセグメンテーションとは

X-ray computed tomography (CT)、Magnetic Resonance Imaging (MRI) などの画像診断モダリティーが発達し、日常診療で広く用いられている。得られた画像は定性的あるいは定量的に評価され、治療方針決定に生かされている。定量的解析法の 1 つが臓器体積の測定である。断層画像から臓器体積を測定するには、セグメンテーション (Segmentation)、すなわち特定の臓器や病変の境界を正確に決定し、対象領域を他の構造から分離する処理が必要である。これにより、対象臓器の体積、形状、位置といった構造的特徴を客観的かつ定量的に把握することができ、術前計画や治療効果判定、疾患進行のモニタリングなど、臨床的意思決定を支援する基盤情報として活用される。肝臓においては、肝切除や生体肝移植の術前評価、放射線治療計画、疾患進行や治療効果のモニタリングなど、多岐にわたる臨床場面で精密なセグメンテーションが必要とされる (Ferrero et al., 2007; Kubota K et al., 1997; Kennedy PA et al., 1997)。肝切除術や生体肝移植においては、術後肝不全を回避するために残肝体積の評価が必須であり、術前計画の基盤となる (Cohnert TU et al., 1997; Urata K et al., 1995; Martel G et al., 2015; Um EH et al., 2015)。従来は体表面積や身長・体重を用いた推定式による肝体積評価が行われてきた (Yoshizumi T et al., 2003)。しかし、正確性に欠けるという限界がある (Linguraru MG et al., 2012; Poovathumkadavil A et al., 2010)。そのため、CT 画像を用いた三次元セグメンテーションによる体積測定が、より信頼性の高い方法として普及してきた。しかしながら、従来の手動または半自動セグメンテーションは時間と労力を要し、また観察者間のばらつきが問題となる (Cai W et al., 2016)。したがって、効率的かつ正確に肝体積を算出できる自動化技術の開発と臨床応用は極めて重要である (Molière S et al., 2024)。

(2) AI による医用画像解析

近年、医療分野において人工知能 (Artificial Intelligence : AI) の応用が急速に進展している。特に画像診断領域では、深層学習 (Deep Learning :

DL) 技術の発展により、コンピューターによる画像解析の精度の向上、業務効率化に貢献している。深層学習は、多層の人工ニューラルネットワークを用いてデータの特徴を自動的に抽出・学習する手法であり、その中でも畳み込みニューラルネットワーク (convolutional neural network : CNN) は画像認識やパターン解析に極めて有効である。CNN は、畳み込み層により局所的特徴を抽出し、プーリング層で位置ずれや縮尺変化に対する頑健性を獲得し、最終的に分類や領域分割などのタスクを行う構造を持つ。この技術は、放射線画像、病理画像、眼底写真、内視鏡画像、手術映像など、多様な医用画像解析に広く応用されている (Litjens et al., 2017)。さらに AI の導入により、セグメンテーション作業の自動化は精度の向上だけでなく、作業時間短縮や専門医の負担軽減にも寄与し、医療現場のマンパワー不足解消に資する技術として注目されている。

(3) セグメンテーション技術の発展

医用画像セグメンテーション技術は、1970～1980 年代に登場した古典的な画像処理アルゴリズムに端を発する。初期の方法は、画素値 (輝度値) の範囲を基準に抽出する閾値処理 (thresholding) が中心であり、対象と背景のコントラストが明瞭な場合に有効であった。しかし、造影効果や病変の影響で濃度に変化する臓器では、正確な境界抽出が困難であった (Gotra A et al., 2017)。

1990 年代には、ユーザーが指定した点から近傍画素を条件に応じて拡張する領域拡張法 (region growing) や、画像のエッジを連続的にたどる輪郭追跡法 (contour tracking) などの半自動化手法が普及した。これらは手動作業を部分的に軽減したものの、閾値設定や境界の曖昧さにより、依然として精度や作業効率に課題を残していた (Sharma N et al., 2010)。

2000 年代に入ると、統計的形状モデル (statistical shape model) やアクティブ輪郭モデル (active contour model) など、臓器の形態情報を組み込んだモデルベースの手法が登場した。これらはノイズや不連続境界の影響を軽減したが、肝硬変や腫瘍による形態変化が大きい症例では精度が低下する問題があった (Malmberg F et al., 2014)。

2010 年代前半には、機械学習によるセグメンテーションが実用化され、ランダムフォレストやサポートベクターマシンを用いた画素分類が行われるようになった。しかし、この時代の機械学習では、エッジ強度や濃度値、テクスチャなどの特徴量を人手で設計する必要があり、モデル性能は設計

者の知識や経験に依存していた (Jie L et al., 2017)。

2012 年に犬や猫などの一般画像分類の深層学習の国際大会 (ILSVRC) で後の 2024 年ノーベル物理学賞受賞者である Geoffrey E. Hinton 博士 (トロント大学) 率いるチーム” SuperVision” が発表した AlexNet が圧倒的なレベル差で優勝し、深層学習が注目されるきっかけとなった (Krizhevsky A et al., 2012)。DL の登場はセグメンテーション技術を大きく転換させた。特に U-Net は、エンコーダ部で抽出した多段階の特徴をデコーダ部で空間解像度を回復させながらスキップ接続で統合することで、高精度なピクセル単位分類を可能にした。さらに、3D U-Net により CT や MRI のボリュームデータ解析も精度が向上した (Huynh HT et al., 2014)。近年では generative adversarial networks (GAN) を用いたデータ拡張や、CycleGAN による造影条件変換が導入され、限られた学習データや相間の差異による性能低下を補う手法も開発されている (Sandfort V et al., 2019)。

2. 既知の部分と未解明の部分

肝体積の正確な評価は、肝切除術の適応判断や術後の残肝機能予測、生体肝移植ドナー評価など、肝疾患における臨床的意思決定において極めて重要である (Nakayama et al., 2006)。以前の肝体積測定は用手的セグメンテーションに依存しており、手作業には多大な時間と労力がかかり、加えて観察者間および観察者内でのばらつきが問題とされてきた (Hermoya L et al., 2005)。最近になって、AI による自動的な肝体積測定を可能にしたソフトウェアが使用できるようになってきた。それらを利用した新たな知見、例えば肝腫大の新たな定義に関することなどが報告され始めている (Tallam et al., 2022; Perez et al., 2022)。

しかし、CT 検査における各造影相 (非造影相、動脈相、門脈相、平衡相など) ごとに、AI モデルがどの程度の精度で肝臓のセグメンテーションを行うことができるかについては、十分な検討がなされていない。従来の研究では、血流コントラストが明瞭な門脈相で最も高い精度が得られるとされてきた報告が多いが、近年では非造影相や平衡相などでも同等の精度を維持できる可能性が指摘されており、造影相による性能差を系統的に評価することが求められている (Nayantara PV et al., 2024)。非造影相と造影相で肝体積は同一であるため、AI によるセグメンテーション精度も同等であることが期待されるが、造影剤の有無による組織コントラストの違いや、他臓器との境界の明瞭さ、血流動態に応じた実質内の濃度分布などが AI の特徴抽出に影響を与える可能性がある。また、多くの AI モデルが造影相の

データを中心に学習されている点も、非造影 CT での性能低下の可能性を示唆する。

他方、加齢や肝疾患に伴う肝体積の経時的変化を、AI セグメンテーションを利用して解析した研究（すなわち縦断的解析）は依然として少ないのが現状である。加齢による変化は横断的研究の報告はあるが、縦断的研究はほぼ見られない（Harada L et al., 2021）。肝疾患に伴う変化は縦断的研究の報告はあるが、対象数は 22 名と少なく、観察期間は 6 ヶ月と短かった（Fitzpatrick JA et al., 2016）。

3. 本研究で何を明らかにしようとしているか

本研究では、3D U-Net および CycleGAN を組み合わせたオープンソースの AI モデルを用いて、肝臓のセグメンテーションを行った。第一章では AI モデルの性能評価を行った。同一症例の非造影および各造影相の CT に対して、AI モデルが放射線科医とどの程度一致した精度で肝臓セグメンテーションを実行できるかを検証した。本研究では AI モデルが放射線科医と同等の精度を有し、造影の有無や各造影相（非造影相、各造影相）にかかわらず一貫した性能を示す、という仮説を立てて検証を行った。第二章ではこの AI モデルを縦断的データに応用し、肝疾患を有さない対象における加齢による肝体積の変化を定量的に評価した。肝体積が加齢に伴い減少するという既報の横断的研究の知見を踏まえ、同一被験者を経時的に追跡した縦断的研究においても同様の減少傾向がみられると仮説を立てた。加齢に伴う正常な肝体積変化を明らかにし、疾患による異常な変化を判断する基準とすることを目的とした。第三章では肝疾患を有する肝体積は減少が報告されているものの縦断的研究は少なく、様々な肝疾患で肝体積が減少するという仮説を立てた。しかし、第一章、第二章で使用した AI ソフトでは肝疾患を有する症例でのセグメンテーションエラーが多発した。そこで 2022 年にバーゼル大学のチームが発表して以来急速に広まっている新しい AI セグメンテーションモデルである TotalSegmentator を用いることとした。TotalSegmentator は既知の AI ソフトと同等あるいはそれ以上の精度を有していたため、病的肝を対象としたセグメンテーションにも適用可能であると判断した。特に、形態変化や腹水などの影響を受けにくく、従来モデルでは誤認されやすかった領域の抽出精度が向上していた。TotalSegmentator を用いて肝疾患のある肝体積の変化の推移を観察した。これらの研究により、AI モデルの臨床的有用性、特に加齢による肝体積の推移、肝疾患別の体積の変化を求め、肝機能変化の客観的指標として

の可能性を検証した。本研究の仮説は、AIモデルが放射線科医と同等の精度でセグメンテーションを行い、健常肝では加齢に伴って体積が減少し、肝疾患群では疾患特異的な体積変化がみられるというものである。

第一章

肝臓の AI セグメンテーションモデルの性能評価

緒言

肝臓は多様な血流動態と複雑な形態を有する臓器であり、CT 撮像においては非造影相から平衡相まで造影効果の変化が大きい。このため、画像上のコントラストや境界の明瞭さが撮像相ごとに異なり、自動セグメンテーションの精度にも影響を及ぼす可能性がある (Nayantara PV et al., 2024)。臨床現場では、撮像条件や造影相が統一されていないことが多く、AI を実用的に導入するためには、どの造影相でも安定した性能を示すことが求められる。

これまでの研究により、AI モデルによる肝臓セグメンテーションは、特定の条件下では放射線科医と同等の精度を達成できることが示されている。しかし、各 CT 撮像相（非造影、動脈相、門脈相、平衡相など）における性能差を、同一モデルで直接比較し、放射線科医による手動計測を基準として精度評価した報告は限られている。特に、臨床現場においては撮像条件が一定ではないため、モデルがどの相で最も安定して高精度を発揮するのかを明らかにすることは、実装・運用の上で重要な課題である (Heimann T et al., 2009)。

AI セグメンテーションを臨床的に応用するためには、放射線科医の手動セグメンテーションと同等の精度を有し、かつ造影条件によらず再現性の高い結果を提供できることが不可欠である。特に、造影相間での精度差が小さい場合、臨床業務において最も取得しやすい時相を柔軟に選択できる利点がある。

本章では、3D U-Net および CycleGAN を組み合わせたオープンソース AI モデルを用い、非造影および複数の造影相における肝臓セグメンテーション精度を、放射線科医による手動セグメンテーションと比較することを目的とした。評価にはダイス類似度係数 (Dice Similarity Coefficient : DSC) や級内相関係数 (Intraclass Correlation Coefficient : ICC) などの客

観的指標を用い、CT の各相の性能差を明らかにすることで、AI モデルの臨床応用における最適条件を提示する。

方法

1.対象

本研究は、北海道大学病院自主臨床研究審査委員会での承認を得て行われた（研究番号：023-0035）。研究参加に際しては患者からの個別同意取得の省略が同委員会によって認められた。

診療録に基づいて、2017年1月から2021年12月の間に当院規定のプロトタイプによる術前精査目的のCT検査を受けた生体肝移植ドナー90例のうち、以下の基準により除外された56例（異なるCT装置：39例、重複：18例、非造影CTのみ：9例、術後CT：6例、肝疾患あり：1例、一部症例は2つ以上の理由による）を除き、最終的に34例を対象とした。これらをランダムに10例のコホート1Aおよび24例のコホート1Bに分類した（図1-1）。放射線科医A、Bがコホート1Aとコホート1Bで行った解析を表1-1に示す。

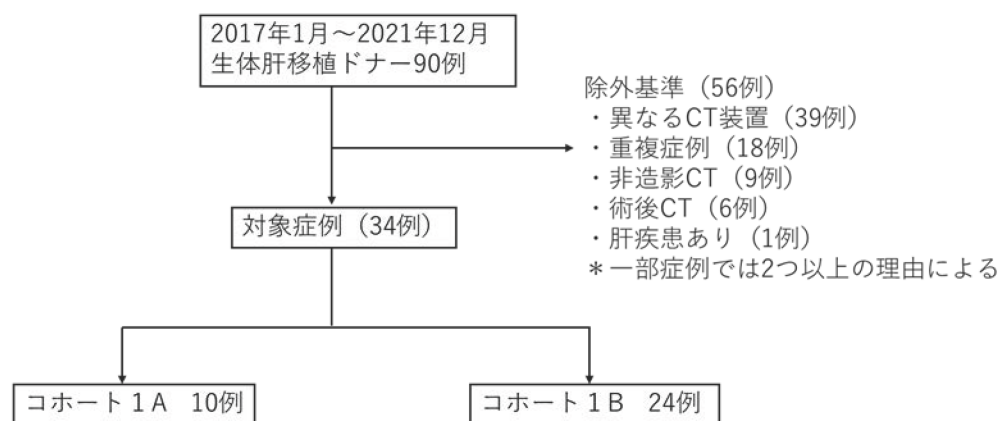


図 1-1：症例選択のフローチャート

表 1-1：コホート毎の放射線科医の解析

	コホート 1A	コホート 1B
症例数	10	24
放射線科医 A	5 相にセグメントを作成	後期動脈相のみにセグメントを作成
放射線科医 B	5 相にセグメントを作成	関与なし

2.CT 画像

当院での肝移植ドナー術前 CT プロトコルでは、肝臓を含む範囲について造影前、早期動脈相、後期動脈相、門脈相、平衡相の 5 相が撮像された。

CT 撮影には Aquilion ONE Vision（キヤノンメディカルシステムズ社製、日本・大田原）を用い、スライス厚 5 mm、管電圧 120 kVp で撮影が行われた。静脈内造影剤には Iopamiron®（イオパミドール、バイエル社製、日本・大阪）または Iomeron®（イオメプロール、エーザイ社製、日本・東京）が使用された。

撮像パラメータを表 1-2 に示す。

表 1-2：CT の撮像パラメータ

項目	値
管電圧 (kVp)	120
スキャンコリメーション (列×スライス厚)	320×0.5
回転時間 (秒)	0.5
マトリックス・サイズ (行×列)	512×512
ピッチ (mm)	0.813
スライス厚 (mm)	5
再構成スライス厚 (mm)	1

3.プロトコル

本研究は、AI を用いた肝臓セグメンテーションの精度およびその汎用性を体系的に検証するため、Part 1～Part 3 の三部構成で実施した（表 1-3 参照）。研究初期段階では、セグメンテーション結果の妥当性をどのように評価すべきかが明確でなく、まずは AI と放射線診断専門医（以下、単に放射線科医とする）2 名および放射線科医間の一致度の検証を行った。これを少数例（N=10）で実施したものが Part 1 である。

Part 1 の結果から、放射線科医間のセグメンテーション一致度が極めて高いことが確認され、以降の検証では一名の専門医による正解ラベル作成で十分であると判断した。また、各造影相の比較により、放射線科医同士の一一致度が最も高かった後期動脈相を至適造影相として特定した。

Part 2 では、Part 1 の知見をもとに症例数を拡大し（N=34）、後期動脈相に限定して AI と放射線科医によるセグメンテーション結果を比較した。こ

ここでは、DSCおよびICCを用いて、AIモデルのセグメンテーション精度と再現性を定量的に評価した。

さらに、Part 3 では、臨床応用を想定し、AI モデルが全ての造影相において一貫した精度を発揮できるかを検証した。Part 2 で放射線科医が後期動脈相で測定したセグメンテーションを真の値とし、他の造影相におけるAI 推定体積との間で ICC を算出した。なお、後期動脈相以外の造影相では真のセグメント（すなわち放射線科医が作成したセグメント）が存在しないため、DSC の算出は不可能であった。

また、Part 3 における ICC の評価は、セグメント境界が完全に一致していなくても体積が偶然一致してしまう可能性を内包する。このため、Part 1 および Part 2 で DSC による境界精度を十分に確認した上で、Part 3 では造影相の違いに対するセグメンテーション体積の一貫性を検証することを目的とした。

以上の考察から、Part 1～3 の実験計画を以下のようにデザインした。

- **Part 1** では、コホート 1A の 10 例について、2 名の放射線科医がそれぞれ独立に、5 相の CT 画像に対して肝臓全体のセグメンテーションを作成した。その後、以下の 3 つの組み合わせについて肝体積およびダイス類似度係数（DSC）を比較した：(1) AIモデル vs. 放射線科医 A（経験 15 年）、(2) AIモデル vs. 放射線科医 B（経験 9 年）、(3) 放射線科医 A vs. B。放射線科医 2 名は、AI によるセグメンテーション結果を手動で修正し、真の肝臓領域を作成した（図 1-2 参照）。
- **Part 2** では、コホート 1B の 24 例において、放射線科医 A が後期動脈相のみに肝臓全体の臓器境界を作成した。コホート 1A + 1B において、放射線科医 A と AI との比較を行った。
- **Part 3** では、コホート 1A および 1B の全 34 例を対象に、全 CT 5 相（非造影および造影）における AI モデルの肝体積測定結果を、Part 2 で得られた真の値（後期動脈相）と比較し、相間での精度を検証した。

表 1-3：研究のプロトコル

	Part 1	Part 2	Part 3
コホート（人数）	1A（10）	1A+1B（34）	1A+1B（34）
セグメントの作成者	AI と放射線科医 2 名	AI と放射線科医 1 名	AI のみ
測定基準	ダイス類似度係数、体積	ダイス類似度係数、体積	体積
造影相	5 相	後期動脈相	5 相
真の値	なし	放射線科医 A	放射線科医 A

4.AI システム

本研究では、米国 NIH が開発し GitHub 上で公開されている AI モデル (<https://github.com/rsummers11/CADLab/tree/master/CT%20Liver%20Segmentation%20Software>) を使用した。この AI は 3D U-Net アーキテクチャに基づいており、CycleGAN（サイクル整合型生成敵対ネットワーク）を組み合わせて、造影 CT 画像を非造影画像に変換することでデータ拡張を行っている。DICOM 形式の CT 画像を NIfTI 形式に変換した後、コマンドラインを用いて AI モデルを実行し、肝臓のセグメンテーションを行った。実行環境は Windows11、CPU Core i7-10700（2.90GHz）、メインメモリー 32GB、GPU GeForce GTX 1660 SUPER であった。

図 1-2 に非造影 CT 画像上での肝臓のスライスおよび AI が出力したセグメンテーション結果を示す。(a) は元の CT 画像であり、(b) は AI モデルにより抽出された肝臓領域を赤色で示したものである。(c) および (d) は、それぞれ放射線科医 A および B が、AI モデルで生成された肝臓領域を修正した結果を示す。放射線科医が AI セグメンテーションから除外した部位には、胆嚢（赤矢印）、消化管（黄矢印）、腹直筋（青矢印）が含まれる。

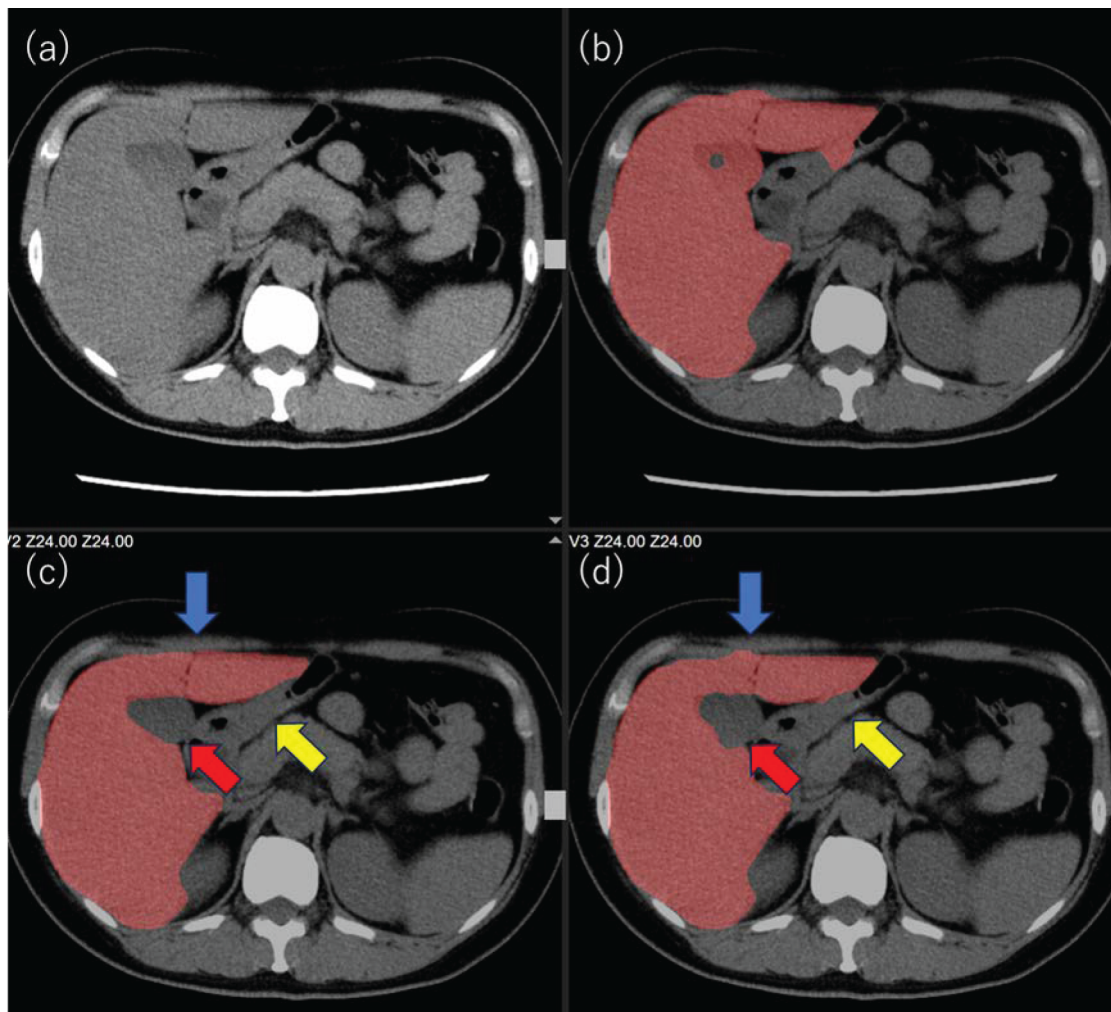


図 1-2：肝臓セグメンテーションを行った CT 画像

(a) セグメンテーションを行う前の非造影 CT、(b) AI が肝臓のセグメンテーションを行った部分を赤色で表示、下段；AI がセグメンテーションを行った画像を (c) 放射線科医 A、(d) 放射線科医 B がそれぞれ修正した画像。青矢印は腹直筋、赤矢印は胆嚢、黄色矢印は消化管を AI が誤って肝臓と認識した部位を放射線科医が修正した部分

また、AI モデルで誤って心臓が囲われており、放射線科医が修正を加えた症例を図 1-3 に示す。

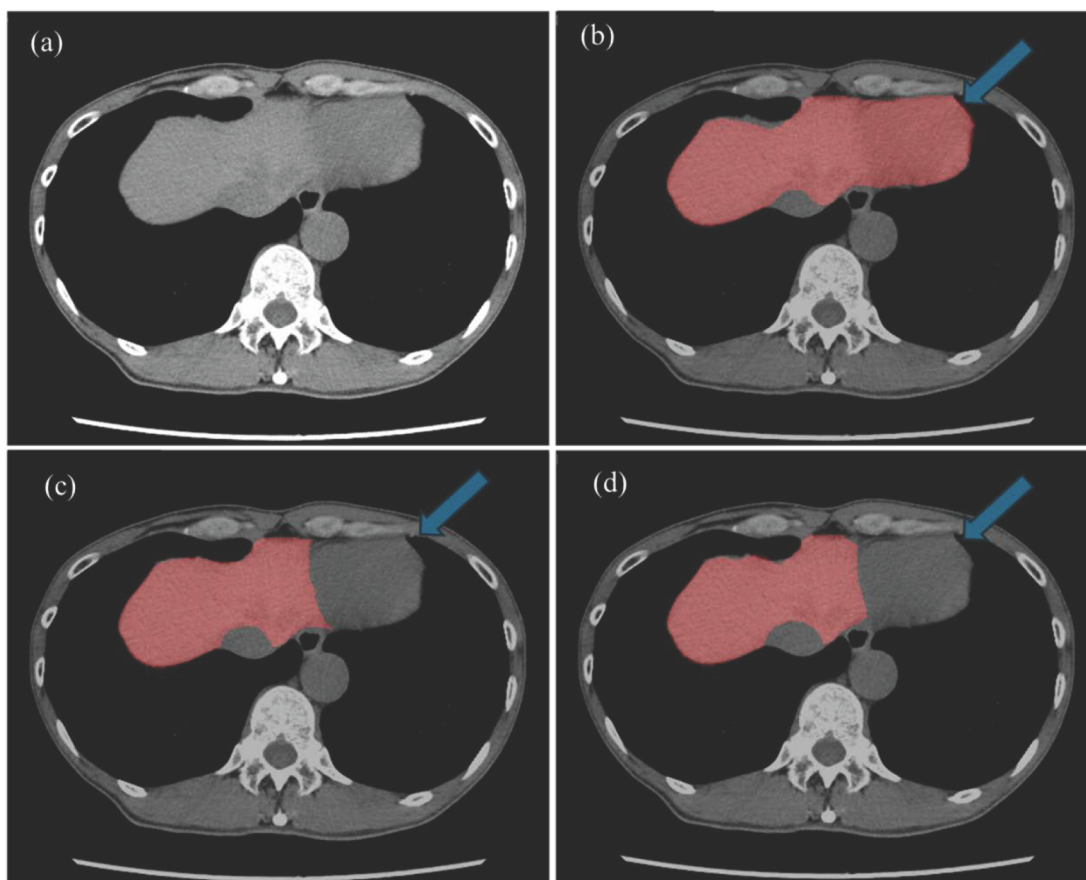


図 1-3：肝臓セグメンテーションを行った CT 画像（心臓）

(a) セグメンテーションを行う前の非造影 CT、(b) AI が肝臓のセグメンテーションを行った部分を赤色で表示、下段；AI がセグメンテーションを行った画像を (c) 放射線科医 A、(d) 放射線科医 B がそれぞれ修正した画像。青矢印は心臓を修正した部分

5. 性能評価

セグメンテーション結果の一致度は、ダイス類似度係数（DSC）を用いて定量的に評価した。DSC は、2 つのセグメント A および B の一致度を示す指標である。

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad \dots \dots \dots \text{式 1}$$

ここで A, B はセグメント A, B のボクセル集合を表し、|X| はセグメント X のボクセル個数を表す。DSC は $0 \leq DSC \leq 1$ の範囲をとり、DSC=1 はセグメント A, B の完全一致を、DSC=0 は一致領域を有さないことを表す。医用画像分野ではセグメンテーション精度の評価に広く用いられている。

セグメンテーション領域の一致度を評価するための別の指標として、

Jaccard Index (JI) も用いた (Müller et al., 2022)。JI は、予測領域と基準領域の積集合のボクセル数を、それらの和集合のボクセル数で割った比として定義され、DSC と同様に 0 (全く重ならない) から 1 (完全一致) の範囲をとる。

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \cdot \cdot \cdot \cdot \cdot \quad \text{式 2}$$

式 1 と式 2 から、DSC と Jaccard Index との間には以下の式が成り立つ。

$$DSC(A,B) = \frac{2J(A,B)}{1+J(A,B)} \cdot \cdot \cdot \cdot \cdot \quad \text{式 3}$$

定義域[0,1]では式 3 は単調増加するため、DSC と JI は互換性を有する。定義式が直感的である JI よりも、小さな差異に対して敏感に反応する DSC が、医用画像解析では使用される頻度が高いが、本論文においては両者を記載する。

6. 統計解析

統計解析は、JMP® Pro 17.0 ソフトウェア (SAS Institute Inc., Cary, NC, 1989–2024) を用いて実施した。カテゴリ変数については頻度および割合で、連続変数については、相関係数を計算した変数はすべて正規分布性がなく、正規分布性が無い場合には中央値と四分位範囲 (Interquartile Range : IQR) で記述し、中央値[IQR]と角括弧で示した。

対応のあるデータ間の比較には Kruskal-Wallis 検定を用い、複数比較の補正には Bonferroni 補正を適用した。肝体積測定の一貫度は ICC により評価した。ICC は複数の測定者間、あるいは同一測定者による繰り返し測定間の一貫度を定量化する指標であり、0 から 1 の値をとる。一般的に、ICC が 0.75 以上であれば「良好」、0.90 以上であれば「非常に高い一貫度」を示すと解釈される (Koo & Li, 2016)。2 群間の平均誤差は、AI あるいは放射線科医による測定値との差の平均値として算出した。値は平均 ± 標準偏差 (Standard Deviation : SD) で表記し、括弧内には相対誤差 (%) を示した。本研究では、AI モデルによる自動セグメンテーションと放射線科医による手動セグメンテーションの体積測定値の一貫度を ICC で評価した。0.05 未満の p 値を統計学的に有意とした。

結果

1. 症例の特徴

コホート 1A (n=10) の年齢の中央値は 27 歳 [四分位範囲：24.5–37.75]、コホート 1B (n=24) の年齢の中央値は 38.5 歳 [28.3–55.3] であった。両群を合わせたコホート 1A+1B (n=34) の年齢の中央値は 36.5 歳 [26.0–50.8] であった (表 1-4)。

表 1-4：症例の特徴

	コホート 1A	コホート 1B	コホート (1A+1B)
人数	10	24	34
年齢 (歳)	27 [24.5-37.75]	38.5 [28.3-55.3]	36.5 [26.0-50.8]
男性	4 (40%)	14 (58%)	18 (53%)
身長 (cm)	164.3 [160.9-175.9]	168.1 [162.7-171.3]	168.05 [162.15-174.5]
体重 (kg)	59.5 [54.1-66.9]	61.9 [58.2-67.6]	61.7 [55.8-67.3]

2. Part1：コホート 1A における AI と放射線科医 2 名による体積と DSC の検討

セグメンテーションを行う際の実行時間は症例により異なるが、概ね 1～2 分であった。10 名の健常者において、放射線科医 A および B による肝体積の平均は 1373 ± 319 mL であった。AI モデルおよび 2 名の放射線科医による各 CT 相での肝体積の測定値の散布図を図 1-4 に示す。上段は AI と放射線科医 A、中段は AI と放射線科医 B、下段は放射線科医 A と放射線科医 B を比較した。列は左から順に造影前、早期動脈相、後期動脈相、門脈相、平衡相を示す。グラフ内の赤線は $y=x$ の直線 (line of identity) である。相関係数および平均誤差も併記した。

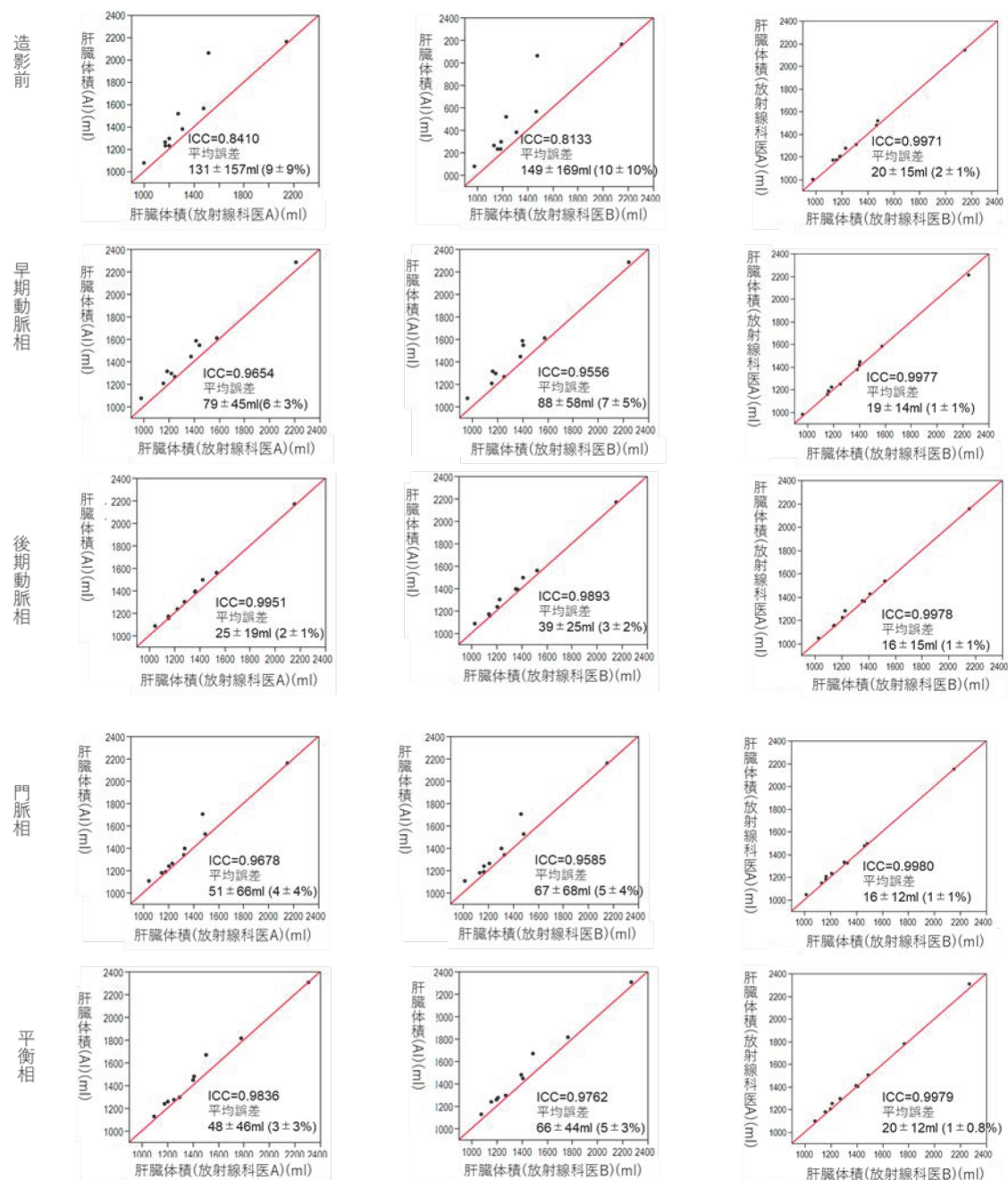


図 1-4： AI モデルと放射線科医による肝体積測定値の相関散布図と ICC

多くの症例において、AI モデルは放射線科医よりも大きな体積を出力する傾向にあった。ICC は造影前で 0.8133、造影後は 0.9556 以上であり、高い一致度と考えられた。

次に DSC の結果は表 1-5 および図 1-5 に示す。

表 1-5 : AI と放射線科医 2 名の各相のダイス類似度係数

造影相	比較対象			中央値
	AI vs. R _A	AI vs. R _B	R _A vs. R _B	
造影前	0.966 (0.842-0.995)	0.961 (0.831-0.997)	0.983 (0.971-0.996)	0.978 (0.831-0.997)
早期動脈相	0.967 (0.934-0.993)	0.966 (0.931-0.990)	0.986 (0.931-0.990)	0.979 (0.931-0.996)
後期動脈相	0.990 (0.970-0.999)	0.986 (0.961-0.996)	0.992 (0.972-0.997)	0.990 (0.961-0.999)
門脈相	0.986 (0.927-0.998)	0.979 (0.920-0.998)	0.991 (0.980-0.998)	0.987 (0.920-0.998)
平衡相	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_A : 放射線科医 A、R_B : 放射線科医 B

統計学的有意差 : $p < 0.01$ ((AI vs. R_A) vs. (R_A - R_B))、 $p < 0.001$ ((AI vs. R_B) vs. (R_A vs. R_B))、 $p < 0.0002$ (非造影相 vs. 遅延動脈相)、 $p < 0.0096$ (非造影相 vs. 門脈相)、 $p < 0.0433$ (非造影相 vs. 平衡相)

表中の丸括弧内の数値は範囲を示す。

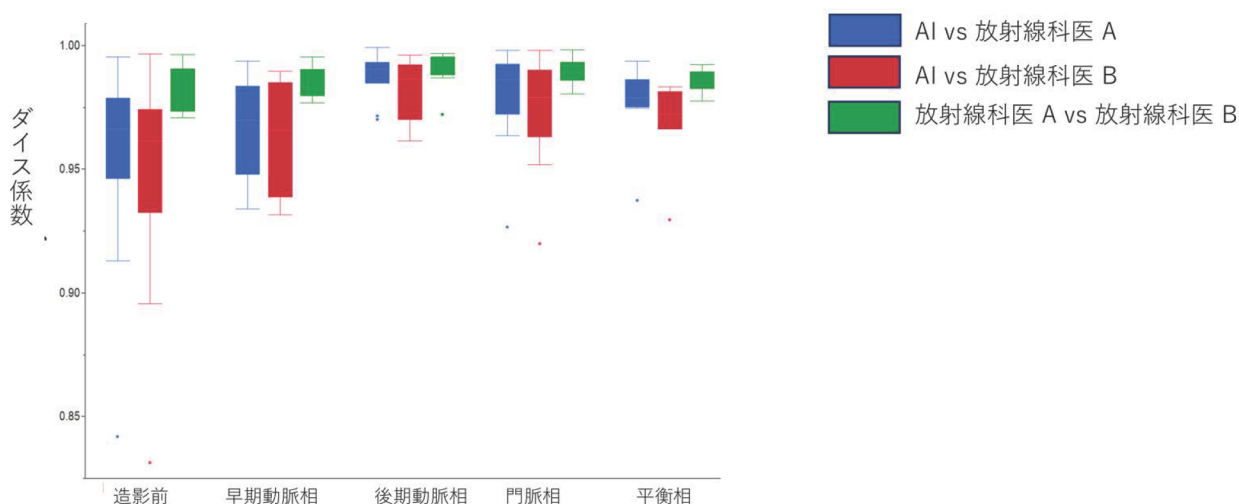


図 1-5 : AI と放射線科医 2 名の造影相ごとのダイス類似度係数

DSC は AI と放射線科医 2 名、放射線科医同士でいずれの造影相でも中央値で 0.978 以上を示し、いずれも高い一致度が得られた。

Jaccard Index (JI) を表 1-6 に示す。

表 1-6 : Jaccard Index

	AI vs. R _A	AI vs. R _B	R _A vs. R _B
造影前	0.935 (0.899-0.958)	0.926 (0.874-0.950)	0.967 (0.949-0.981)
早期動脈相	0.942 (0.901-0.968)	0.934 (0.885-0.971)	0.973 (0.961-0.981)
後期動脈相	0.981 (0.970-0.987)	0.973 (0.942-0.984)	0.985 (0.977-0.991)
門脈相	0.973 (0.946-0.985)	0.959 (0.929-0.980)	0.983 (0.972-0.987)
平衡相	0.959 (0.952-0.973)	0.946 (0.935-0.964)	0.971 (0.966-0.979)

R_A : 放射線科医 A、R_B : 放射線科医 B

丸括弧内は範囲を示す。

Jaccard Index はいずれの時相、どの組み合わせでも 0.9 以上を示しており、高い一致度と考えられた。

放射線科医 A と放射線科医 B の DSC や ICC が十分高いことから、複数の放射線科医の結果は高い精度で一致すると考えられ、以後の解析では放射線科医 A が 1 名がセグメンテーションを行うこととした。また、最も高い DSC が得られたのは後期動脈相であったため、後期動脈相での放射線科医 A のセグメンテーションを真の値とした。

3. Part2 : コホート 1A+1B における AI と放射線科医のセグメンテーションを後期動脈相で比較

症例数を増やすために、Part1 で使用したコホート 1A にさらにコホート 1B の 24 例を追加し、合計 34 例の後期動脈相における放射線科医 A と AI モデルの肝体積を比較した。AI モデルは一貫して放射線科医 A よりも平均 4%体積を大きく出力する傾向があった (図 1-6)。

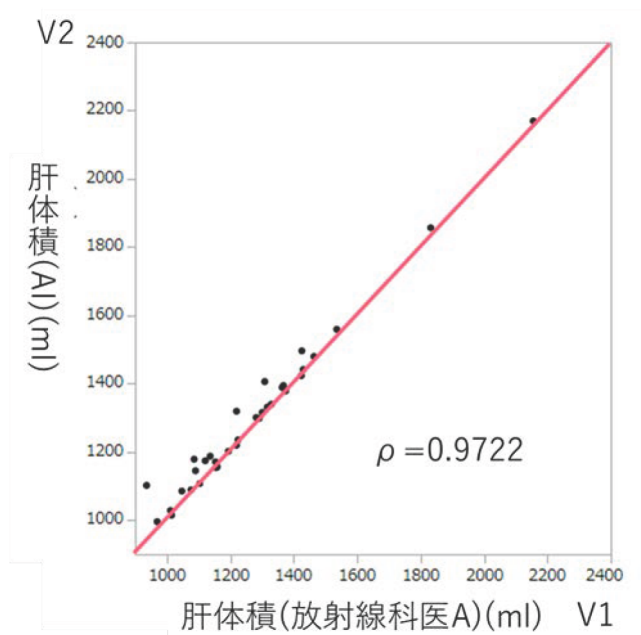


図 1-6 : AI と放射線科医 A の肝体積の比較（後期動脈相）
赤線は $V1=V2$ （line of identity）を示す。

DSC は 0.986 ± 0.017 、ICC は 0.980 と、いずれも非常に高い一致度を示した。後期動脈相において、AI モデルは放射線科医が肝臓のセグメンテーションを行うものと極めて高いと考えられた。次に各相での肝体積セグメンテーションの一致度に関して、解析を行った。

4. Part3 : コホート 1A+1B で AI モデルのみで CT 5 相での比較

コホート 1A+1B の全 34 例について、AI モデルによる各 CT 相の肝体積を測定した。造影相ごとの自動セグメンテーション結果を図 1-7 に示す。

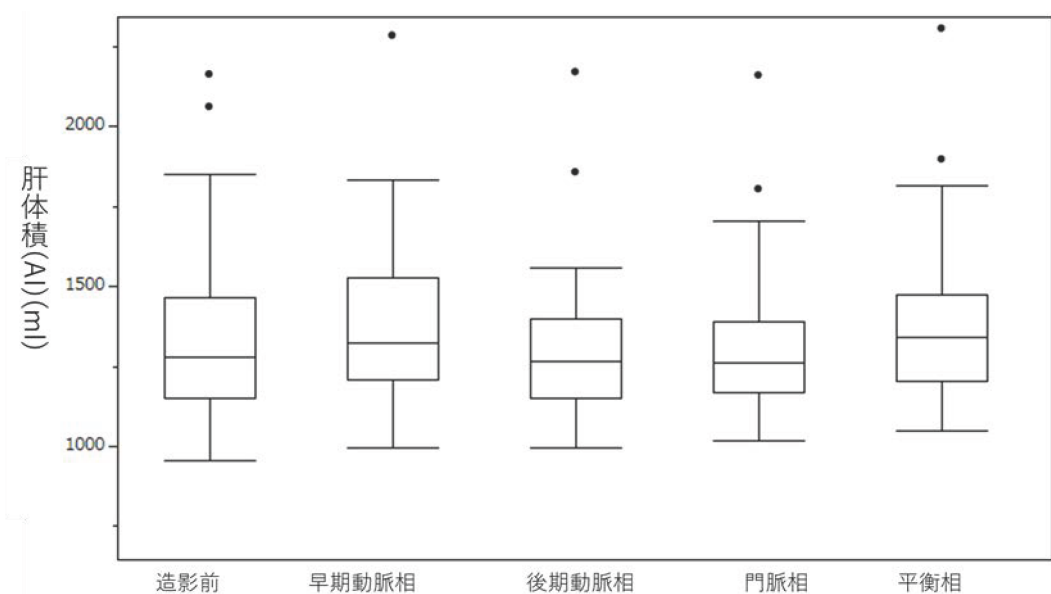


図 1-7：AI モデルのみの各造影相の肝体積

Kruskal–Wallis 検定の結果、 $p = 0.4359$ であり、各造影相の間で有意な差は認められなかった。

Part2 で放射線科医 A が後期動脈相で測定を行った肝体積を真の値として、各造影相での AI が測定した肝体積との間で ICC を算出した。表 1-7 に各造影相の ICC を示す。

表 1-7：放射線科医 A（後期動脈相）と AI の ICC

造影相	ICC
造影前	0.833
早期動脈相	0.870
後期動脈相	0.980
門脈相	0.949
平衡相	0.882

すべての ICC が 0.8 を超え、高い一致度が確認されたことから、AI モデルはどの造影相において応用可能であると考えられた。

考察

本研究では、全自動の定量的肝セグメンテーション AI モデルの性能を評価した。その結果、AI モデルは放射線科医によるセグメンテーションと比較して高い一致度を示した。DSC は遅延動脈相で 0.988 ± 0.010 、門脈相で 0.979 ± 0.021 、平衡相で 0.977 ± 0.015 、早期動脈相で 0.968 ± 0.019 、非造影相で 0.954 ± 0.045 であった。これらの値は、先行研究における動脈相での報告値 (DSC = 0.948 ± 0.003) を上回っており、十分に高い精度であると考えられた (Perez AA et al., 2022)。前述の先行研究では肝疾患を有する被験者が含まれていたのに対し、本研究では健常者のみを対象としたことが、精度向上の一因と考えられる。

AI によるセグメンテーションは、放射線科医と比較してやや大きな体積を出力する傾向があった。これは、AI が肝臓外の構造物を取り込んでしまう場合があるためであり、放射線科医がそうした部分をトリミングすることによって差が生じたと考えられる。逆に、肝臓実質を過小評価する例は少なく、真の肝実質がセグメントに含まれないケースは稀であった。ICC は非造影 CT で 0.8 以上、造影 CT ではすべての相で 0.9 以上となり、すべての相で高い精度が示された。

一方で、AI によるセグメンテーションにはいくつかの課題も認められた。特に、肝門部や腸管（横行結腸等、特に肝弯曲部）付近では、構造が複雑であるため誤認識が生じやすかった。また、内臓脂肪が乏しい症例では、肝臓と腹壁筋の境界が不明瞭となり、非造影 CT 画像で筋肉を肝臓と誤認する例も見られた。極めて稀ではあるが、大腰筋のように肝臓から離れた筋肉が肝臓と誤認された症例もあった。過剰セグメンテーションは肝周囲臓器で起こりやすく、逆に低濃度やコントラストの低い領域では過小評価が生じやすいとされている (Fan M. et al., 2025)。本研究において後期動脈相で最も高い DSC が得られたのは、腎臓や消化管との境界が明瞭に描出されるためと考えられる。

本研究にはいくつかの限界がある。第一に、単施設の後向き観察研究であり、対象集団の規模が比較的小さく、複数の施設から多様な年齢層のサンプルを含む大規模データベースを用いて検証することが今後の課題である。第二に、本研究で使用した AI モデルは肝区域ごとの体積測定には対

応しておらず、肝硬変によって右葉の萎縮や左葉の肥大が生じることを考慮すると、区域ごとの解析も今後の検討課題である。第三に、肝疾患を有する症例での AI の精度評価を行っていない点である。複数の肝疾患に対して、本研究と同様の AI によるセグメントと放射線科医によるセグメントの比較を行い、精度を保証する必要がある。第 4 に真の値を放射線科医が測定した後期動脈相の肝体積としていたが、造影剤による影響が考慮されていないため、造影前の肝体積は厳密には異なる可能性がある。

第一章においては、3D U-Net および CycleGAN を組み合わせたオープンソース AI モデルを用い、非造影および各造影相における肝臓セグメンテーション性能を放射線科医による手動セグメンテーションと比較した。その結果、全ての相で高い一致度 ($ICC > 0.8$) が確認され、特に後期動脈相で最も優れた精度を示した。さらに、非造影相を含め、臨床で一般的に取得される全ての CT 相で ICC および DSC が高値を示し、本モデルが多様な撮像条件下で安定した性能を発揮できることが明らかとなった。本研究からは、AI モデルを用いた自動肝体積測定が、異なる造影相や長期にわたる経時的データにも適用可能であることを示している。

そこで第二章では、この AI モデルを臨床例の縦断的解析に適用し、肝疾患を有さない対象者における加齢に伴う肝体積の経時的変化を定量的に明らかにすることを目的とする。

第二章

AI モデルを用いた肝臓セグメンテーションによる 肝疾患を有さない集団における 肝臓体積の加齢性変化の検討

緒言

肝臓は、代謝、解毒、胆汁産生など多岐にわたる機能を担う重要臓器であり、その体積は肝機能や全身代謝状態と密接に関連する。加齢や疾患により肝重量・肝細胞数が減少することは古くから知られており、病理解剖や断面解剖学的研究からも高齢者では肝体積の縮小が認められている (Kwo PY et al., 1998)。放射線画像を用いた報告では、肝体積はおおむね成人期をピークに緩やかに減少し、その減少速度は加齢、栄養状態、アルコール摂取、脂肪肝の有無などの影響を受けることが示唆されている (Tang A et al., 2015)。

これまでの報告では、加齢に伴い肝重量や肝細胞数が減少し、肝体積も縮小することが知られているが、その変化率や経時的パターンを長期間にわたり定量的に評価した研究は限られている (Mtomomi K et al., 1990)。特に、肝疾患を有さない集団における正常加齢変化の基準値は十分に確立されておらず、加齢に伴う肝機能変化や疾患発症リスク評価の指標として活用するには、より詳細な縦断的データが必要である。

一方、従来の肝体積測定は、手動または半自動によるセグメンテーションが必要であり、数百例規模の長期追跡データに適用するには多大な時間と労力を要した。第一章で性能を検証した AI モデルは、異なる CT 相においても高精度かつ安定した肝臓セグメンテーションが可能であり、大規模・長期の縦断的解析への応用が現実的となった。

そこで本章では、肝疾患や治療歴を有さない症例を対象に、最大 20 年に及ぶ CT データを、第一章で検証した AI モデルで解析し、肝体積の経時的

変化を定量的に評価した。これにより、正常加齢に伴う肝体積変化の実態を明らかにし、今後の肝疾患予防や予後予測に資する基礎データを提供することを目的とする。

方法

1. 対象

本研究は、北海道大学病院自主臨床研究審査委員会での承認を得て行われ（研究番号：023-0035）、研究参加に際しては患者から個別同意取得の省略が同委員会によって認められた。

診療録に基づいて、2021 年時点で肝疾患および化学療法歴のない症例を抽出した。抽出作業は 2021 年に行い、画像の取り出しは 2022 年に行った。2000 年から 2022 年の間に 1 回以上 CT が撮像された症例のうち、5 年以上の期間で CT の撮像が行われた 39 症例を選定した。観察期間は 5～20 年で、観察期間中に肝生検、肝臓切除術や化学療法などの肝臓に対して侵襲的処置は実施されていないものを対象とした。最終的に 39 症例の計 527 回の CT 撮像を解析対象とした。

2. CT 画像

非造影 CT および様々な造影相の CT 画像を使用した。撮影に使用された CT 装置には、Aquilion ONE Vision、Aquilion PRIME（いずれもキヤノンメディカルシステムズ社製、日本・大田原）、および Brilliance iCT Elite（フィリップス社製、オランダ・アムステルダム）が含まれた。スライス厚は 5 mm、管電圧は 120 kVp で統一された。造影剤は Iopamiron®および Iomeron®を使用した。撮影パラメータの詳細は表 2-1 に示す。

表 2-1：CT の撮像パラメータ

	Aquilion ONE	Aquilion PRIME	Brilliance iCT Elite
管電圧 (kVp)	120	120	120
スキャンコリメーション (列×スライス厚)	320×0.5	80×0.5	256×0.625
回転時間 (秒)	0.5	0.35	0.5
マトリックス (行×列)	512×512	512×512	512×512
ピッチ (mm)	0.813	1.388	0.900
スライス厚 (mm)	5	5	5
再構成スライス厚 (mm)	1	1	1

3. AI システム

第一章と同一の AI モデルを使用した。

4. 統計解析

統計解析は、JMP® Pro 17.0 ソフトウェア (SAS Institute Inc., Cary, NC, 1989–2024) および Python 3.12、NumPy 1.26、Pandas 2.2、Matplotlib 3.8 を用いて実施した。2 変数の相関解析には Spearman の順位相関係数を用いた。P<0.05 を有意水準とした。

結果

1. 症例の特徴

対象症例（n=39）の 2021 年時点での平均年齢は 74 歳 [68.8–81.3] であった。患者背景の詳細は表 2-1 に示す。

表 2-1：症例の特徴

項目	値
人数	39
年齢（歳）	74 [68.8-81.3]
男性（人、%）	32（80%）
身長（cm）	161.3 [149.6-164.3]
体重（kg）	57.4 [52.1-64.5]

年齢、身長、体重は中央値、角括弧は IQR を示す

39 例において合計 527 件の肝体積測定を実施し、1 症例あたり平均 13.2 回の測定が得られた。観察期間は 5～20 年で、中央値は 13.7 年であった。

観察開始時の肝体積は $1252 \pm 431.7 \text{ mL}$ であった。年齢と肝体積から作成した散布図を図 2-1 に示す。同一症例の同日の症例で複数の造影相がある場合には肝体積の平均値をプロットした。Spearman の相関係数は $\rho = -0.451$ であり、年齢が高いほど肝体積が小さい傾向が認められた。

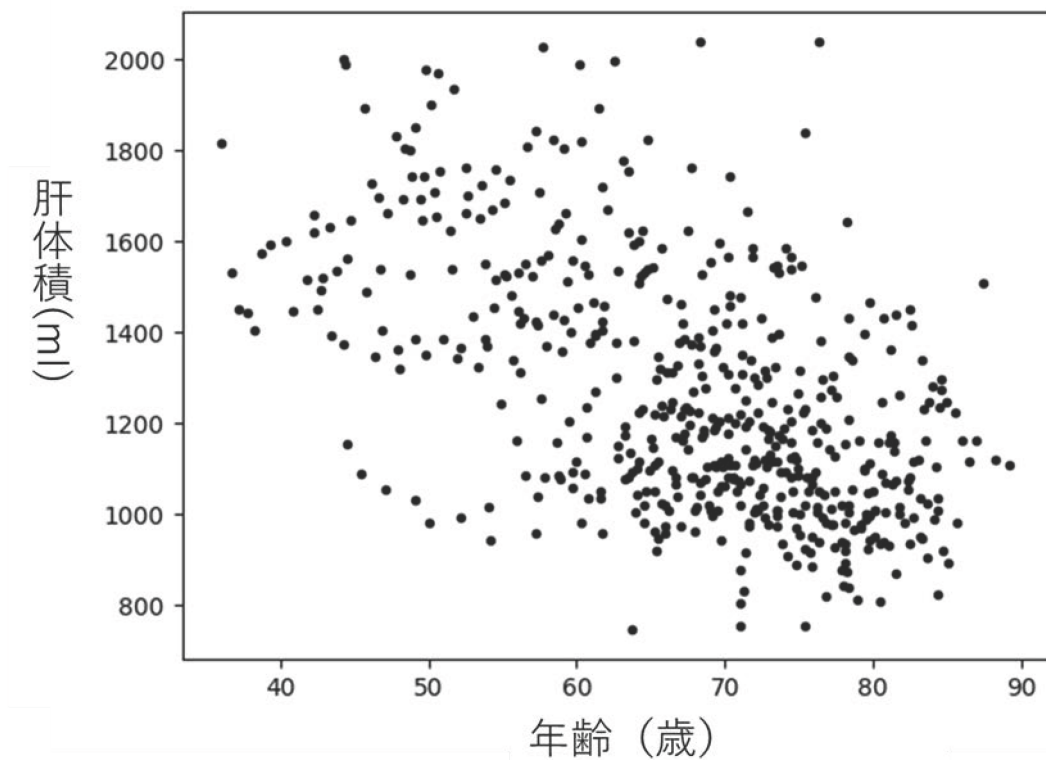


図 2-1：肝体積と年齢の散布図

患者の年齢ごとに肝体積をプロットした散布図。

ついで、同一患者の肝体積と年齢の関係を明らかにするため、同一患者のサンプルを線で結んだ図を 2-2 に示す。

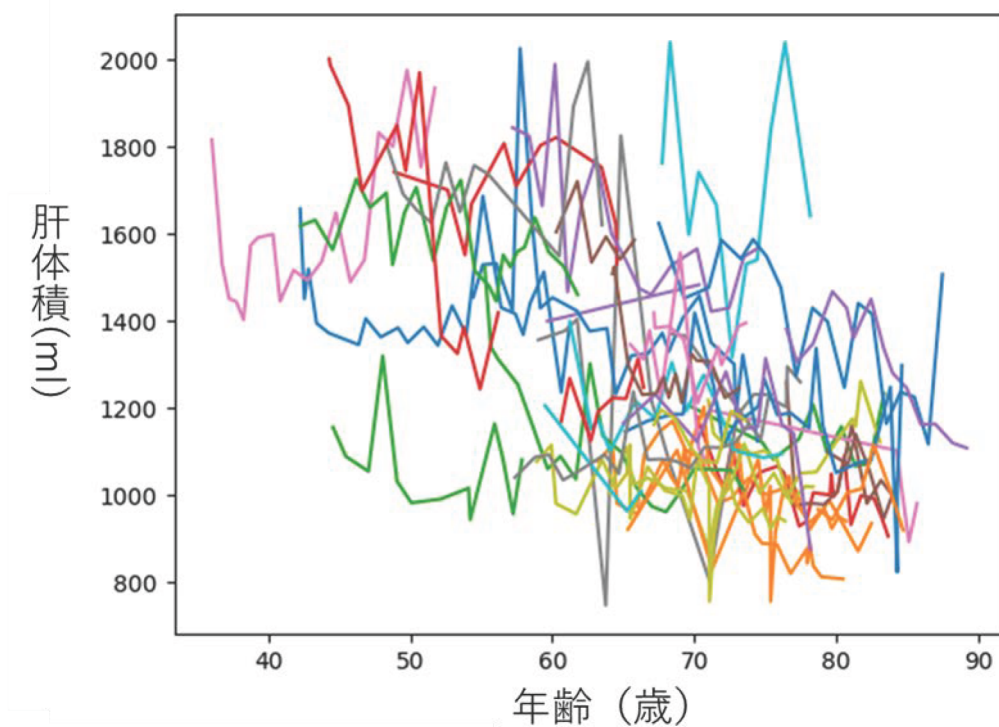


図 2-2：症例ごとの肝体積と年齢のグラフ

図 2-1 での散布図で同一患者を線で結んだ。

図 2-2 のグラフでは肝体積と年齢の推移が不明瞭であった。そこで、最初に CT が撮像された体積を 1.0 として、その体積に対して増減を比率で参照できる相対的肝体積を算出した。そして、相対的肝体積を観察期間に対してプロットしたグラフを図 2-3 に示す。

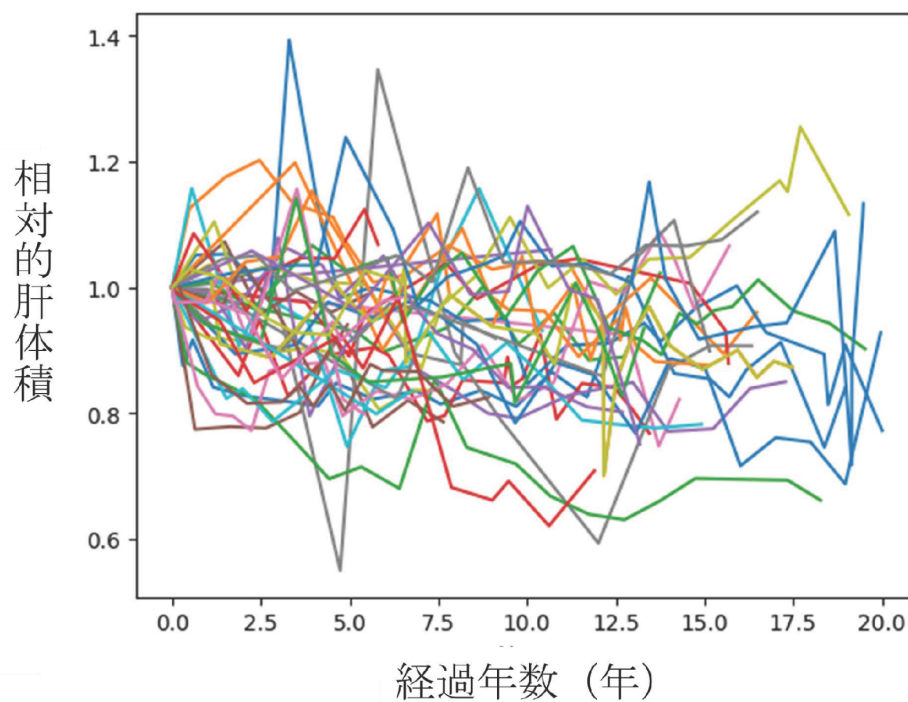


図 2-3：相対的肝体積と経過年数のグラフ

初めて CT を撮像した時の体積を 1.0 として、その後の変化率をグラフとした。0 年は最初に CT を撮像した時であり、経過を年数で示している。同一患者は直線で結ばれている。

図 2-3 ではある程度の傾向が掴めたが、グラフに凹凸が見られて、関係が未だ不明瞭であった。この問題の解決のため、同一患者のグラフに対して傾きを算出し、それをプロットした後に全体の傾きを算出した。それを図 2-4 に示す。

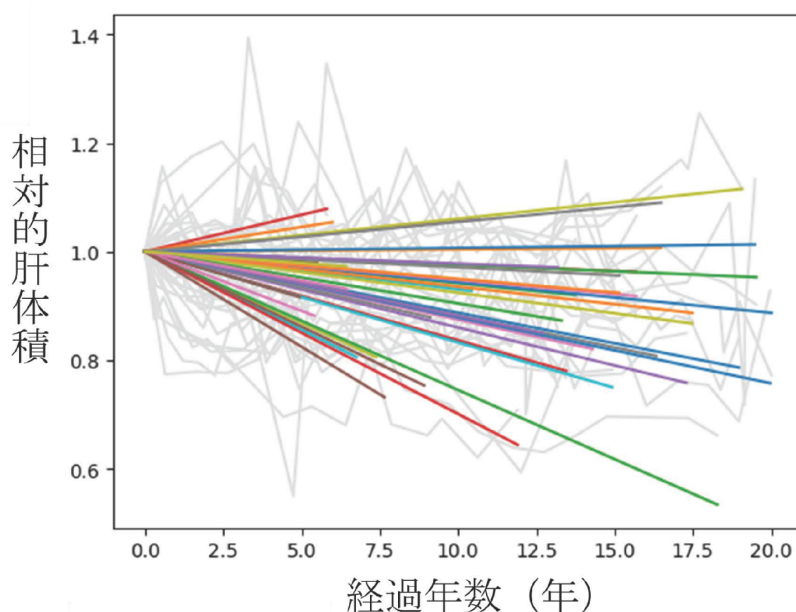


図 2-4：相対的肝体積の傾きと経過年数のグラフ

図 2-3 で示したグラフは灰色で表示している。直線は同一症例の経時的な肝体積の傾きを示している。

症例ごとに傾きを算出した結果、平均の傾きは $-0.95 \pm 1.1\%$ /年であり、肝体積は年間平均 0.95% 減少していたことが示された。ただし、この直線モデルは観測開始時の体積に対する変化率が一定であることを仮定している。別のモデルとして、常にその時点での体積に対する変化率が一定であることを仮定する指数関数モデル ($y=a \exp (-bx)$) も考慮される。そこで、肝体積の対数を取り、観測開始時点からの差を縦軸にプロットしたものが図 2-5 である。この直線の傾きを常数に戻したところ、年間 $1.1 \pm 1.3\%$ の減少率を得た。直線モデルと比べて大きな差がなく、直線モデルよりもやや大きな減少率が得られている点は、指数関数が下に凸となることから考えて妥当である。

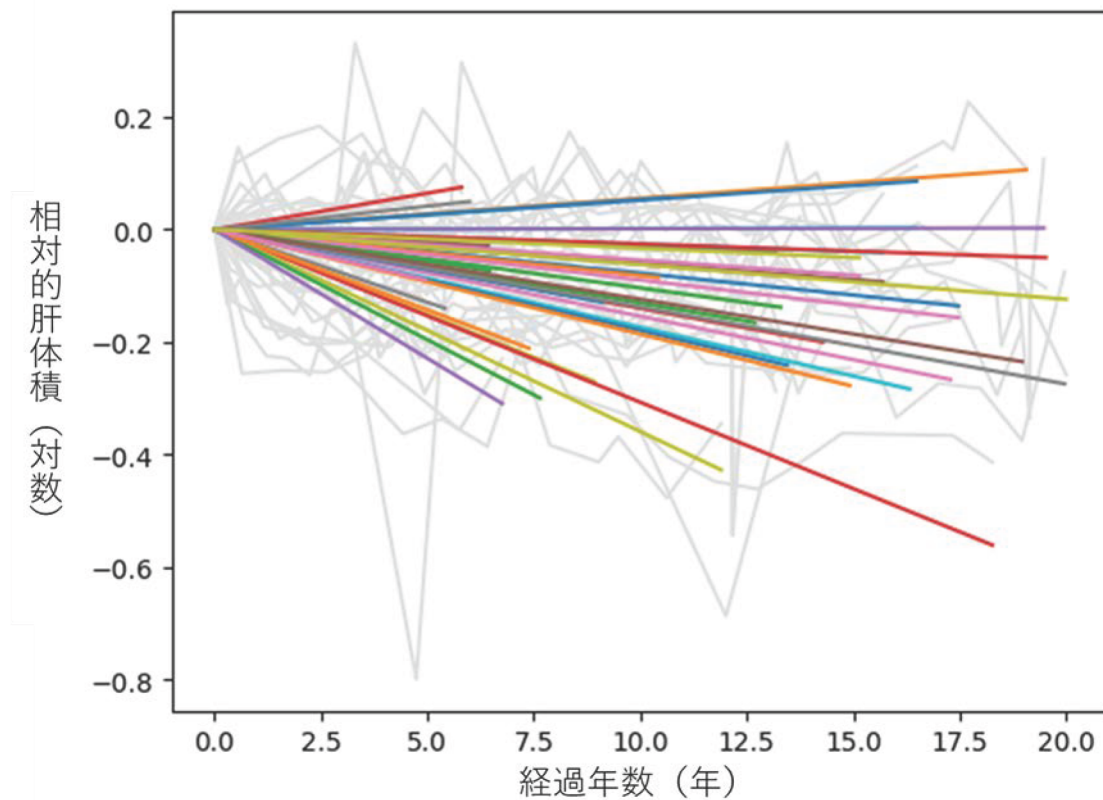


図 2-5：相対的肝体積（対数）の経年変化

縦軸は肝体積の対数（底 e）の経過観察開始時からの差分。症例ごとに直線へのあてはめを行った。

考察

今回の研究では加齢による肝体積の減少傾向を縦断的に示すことができた。先行研究では、肝体積は身長、体重、年齢と独立に関連していると報告されている。今回の結果においても、肝体積は1年ごとに約1%の割合で減少し、加齢に伴う肝の萎縮がCT画像からも示唆された。

これまでの報告では、肝重量は10年間で約100g減少し、80歳代では60歳代に比べて肝重量が10～30%低下し、肝臓と体重の比率も減少することが知られている (Mtomomi K et al., 1990)。これは肝細胞数の減少や血流低下に起因すると考えられており、肝細胞の肥大や多倍体化も再生能の低下を反映している (Childs, B. G. et al., 2015)。また、ASTやALTの血中値は加齢によって著明に悪化しないが、薬物代謝やアルブミン合成といった肝機能は40歳以降、年間0.8%ずつ低下するという報告もあり、肝体積の変化はこれらの機能的変化を反映している可能性がある (Soejima K et al., 2002)。

本研究のAIモデルは、症例数の多い解析を可能にするという点で大きな利点を持つ。たとえば、先行研究ではAIを用いて3,065例の肝体積を測定し、体重に依存した肝腫大のカットオフ値 ($14.0 \times \text{体重}[\text{kg}] + 979\text{mL}$) が導かれている (Perez AA et al., 2022)。市販のワークステーションでも半自動計測が可能であるが、本研究で使用したAIモデルは無料であり、バッチ処理によって大量のデータ解析が可能であるため、大規模コホート研究において特に有用である。

本章の研究の限界は、①単施設後ろ向き観察研究であること、②AIによる測定誤差に対して放射線科医による修正を行っていないため、一定のばらつきが含まれている可能性があることである。③本研究では、肝臓体積と身長、体重、年齢、性別などの因子との関連について多変量解析を試みたが、特に体重データの欠損が多く、十分な統計的解析を行うことは困難であった。そのため、本章では加齢に伴う体積変化の検討を中心に報告した。それでもなお、AIモデルは高精度であり、多数例に統計解析を行って全体の動向を把握する目的には十分に意味を持つと考えられる。

本モデルは無料で使用可能であり、大量症例を効率的に解析できる点で大きな利点を有する。今後は、肝疾患を有する症例における肝体積変化と

肝機能との関係性を明らかにすることで、肝疾患の予後予測や早期診断への応用が期待される。特に、肝硬変などの疾患に対する新たなバイオマーカーとして、肝体積変化率の活用が検討されるべきである。

肝疾患における体積変化と予後の関連については十分に解明されていない部分が多い。肝体積の減少が肝機能や予後に関連するのではないかと考えた。そこで、引き続き第三章では、肝疾患患者における肝体積の臨床的有用性を検討するため、AI を用いて肝体積の縦断的にセグメンテーションを行うこととした。

第三章

AI モデルを用いた肝臓セグメンテーションによる 肝疾患のある肝体積の縦断的研究

緒言

第一章および第二章では、既存の 3D U-Net と CycleGAN を組み合わせた AI モデルを用いて、健常肝の CT 画像に対するセグメンテーション性能を評価し、さらに長期縦断解析を通じて加齢に伴う肝体積の変化を明らかにした。これにより、本モデルが正常肝において高い精度と再現性を有することが示された。

慢性肝疾患においては、線維化や肝硬変の進行に伴い肝体積は変化し、特に右葉の萎縮や左葉・尾状葉の代償性肥大が生じることが知られている。これらの体積変化は単なる形態的变化にとどまらず、肝機能予備能や予後に直結する重要な指標となる。

実臨床においても、肝切除や生体肝移植の術前評価では、残肝体積の計測が術後肝不全回避に不可欠であることが広く認識されている。さらに、近年の研究では、肝体積と Child-Pugh 分類、MELD スコアといった従来の肝機能評価指標との関連が報告されており、肝体積そのものが肝不全リスクや長期予後予測に有用である可能性が示唆されている。特に、急性肝不全や進行肝硬変症例において、体積の減少や区域性の不均衡は臨床転帰に大きく影響する (Zaborn et al., 2018)。

従来、これらの体積測定は手動または半自動によるセグメンテーションに依存しており、解析可能な症例数は限られていた。そのため、肝疾患における体積変化と予後の関連については十分に解明されていない部分が多い。しかし、近年の人工知能 (AI) の進歩により、高精度かつ自動的に肝体積を測定できる環境が整いつつあり、従来困難であった大規模かつ長期

の縦断解析が可能となっている。

しかし、臨床現場においては、肝硬変、肝腫瘍、脂肪肝、術後変化など、形態や内部構造が大きく変化した肝臓を解析する機会が多い。これらの病的肝では、境界の不明瞭化や構造の変形により、健常肝を対象とした AI モデルの精度が低下する可能性がある。実際に本モデルを肝疾患症例に適用したところ、腹水や胸水を肝実質として誤認し、肝体積を過大評価をしてしまう事例が認められた（図 3-1）。

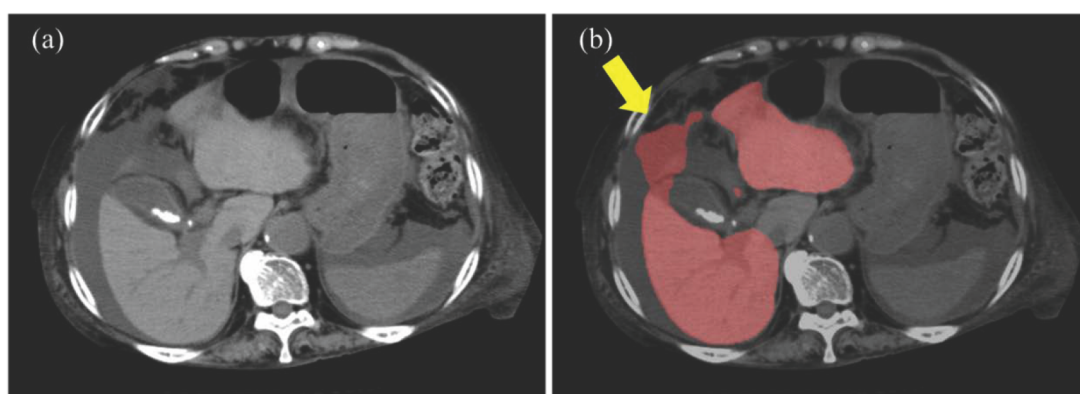


図 3-1：AI モデルが腹水を肝臓の一部と誤認識した症例

腹水が見られる症例。(a) は CT のみの画像。腹水の貯留を認める。(b) はセグメンテーションを行った画像で赤色の領域が肝臓としてセグメンテーション処理された箇所。黄色矢印の領域が腹水だが、肝臓として誤ってセグメンテーション処理をされてしまった箇所。

セグメンテーションで抽出された領域の CT 値を解析し、明らかに肝実質の値を示さない部分については除外を行った（図 3-2）。具体的には、腹水や胸水を肝実質として誤認するセグメンテーションエラーが多発したため、腹水は肝臓より CT 値が低いという特性を利用し、0 HU 付近の画素が多数を占める領域を除外対象とした（図 3-3）。しかし、正常肝の造影前の CT 値はおおよそ 50～60 HU 程度とされるものの、脂肪肝やヘモジデロシスなど異常 CT 値を呈する肝臓においては、この方法は必ずしも適切では無かった。また、実際には肝実質と腹水の CT 値が重なり合う領域が多く、エラーの十分な低減には至らなかった。さらに、肝臓と CT 値が近似している他の実質臓器（例：筋肉や脾臓など）を誤って抽出するケースもあり、閾値処理のみで完全に除外することは困難であった。このことは、従来モデルが病的肝や随伴所見に対して必ずしも適応的ではないことを示

していた。

ボクセル数

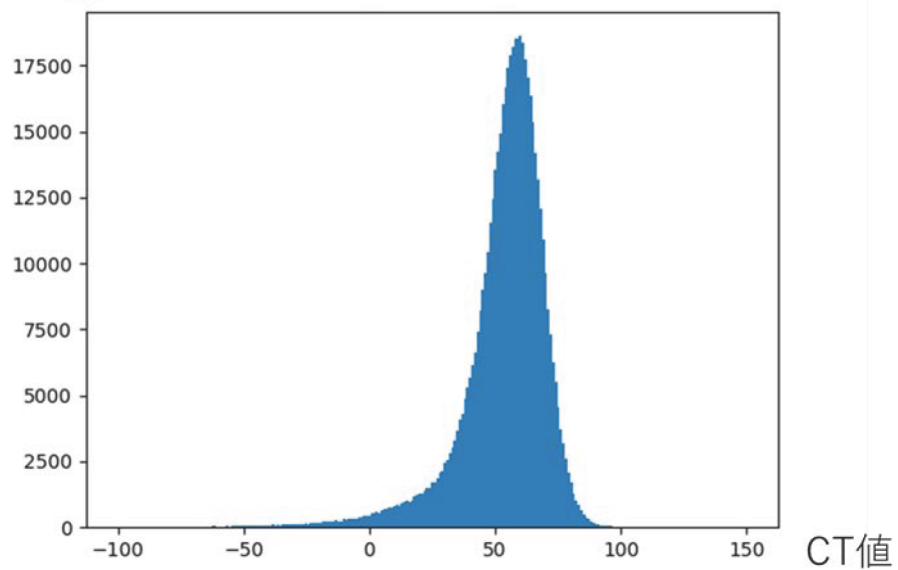


図 3-2： 正常な肝臓セグメンテーションが行われた時の CT 値の分布

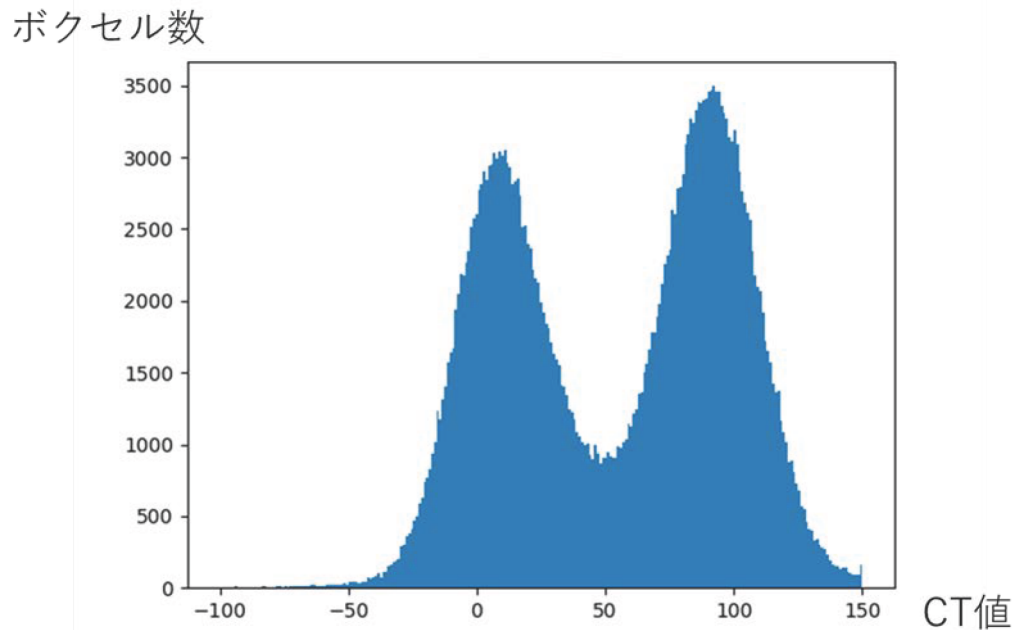


図 3-3：肝臓と腹水のセグメンテーションが行われてしまった症例の造影 CT 値の分布

80HU にピークのある部分が肝臓のセグメンテーション、0HU にピークのある部分が腹水のセグメンテーションを示す。

近年では、セグメンテーションモデルは日進月歩であり、新しいアーキテクチャや大規模データ学習の導入により、精度と汎用性が向上している。その代表例が TotalSegmentator である (Wasserthal, J. et al., 2023)。TotalSegmentator は汎用的な 3D 医学画像セグメンテーションモデルであり、臓器単位だけでなくサブセグメントや病変構造も自動抽出可能とされる (図 3-4)。多様な疾患症例やマルチセンターデータで学習されており、既存モデルと比較して病的形態変化への適応力が高いことが期待される。DSC は様々な造影相や造影前を含む CT のセグメンテーションにおいて、全身で 0.943 と高い値を示している (Wasserthal, J. et al., 2023)。TotalSegmentator はデータセットでは非造影を含む様々な造影相が含まれており、臨床的応用に十分な精度を有すると報告されている。本研究でも同様に、TotalSegmentator による肝臓セグメンテーションは既存の AI モデルと比較して境界の整合性が高く、過剰・過小抽出の抑制が確認された。これらの結果から、TotalSegmentator は肝疾患を有する症例においても信

頼性の高いセグメンテーションが可能であり、既存モデルに代わる妥当な解析手法であると判断し、TotalSegmentator の精度検証は割愛した。

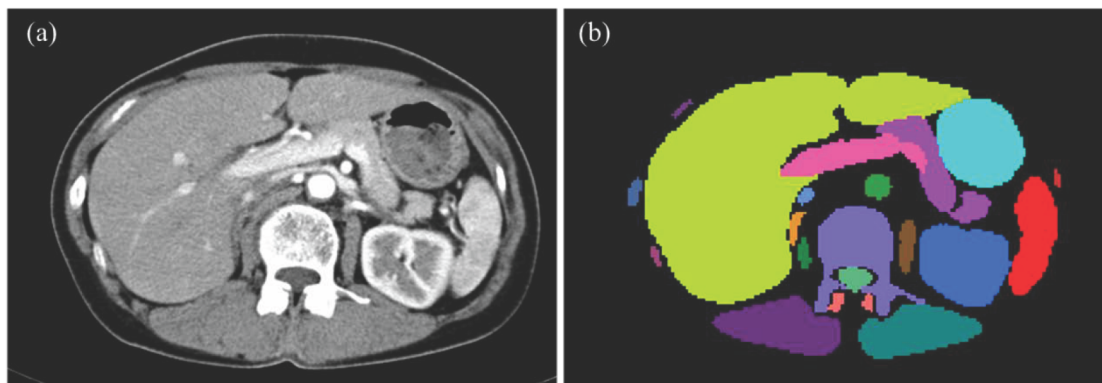


図 3-4 : TotalSegmentator によるセグメンテーション

(a) は CT のみの画像。(b) は TotalSegmentator によるセグメンテーションが行われた図。肝臓は黄緑色で表示されている。その他の臓器（脾臓、脾臓、腎臓、胃、血管、筋肉、骨）もそれぞれの色でセグメンテーション処理が行われている。

本章では、新規 AI モデル（TotalSegmentator）を用いて肝疾患患者の CT 画像から肝体積を自動計測し、その縦断的推移と予後予測、肝障害の程度評価、さらに肝不全に至る可能性との関連を明らかにすることを目的とした。

方法

1. 対象

本研究は、北海道大学病院自主臨床研究審査委員会での承認を得て行われ（研究番号：023-0035）、研究参加に際しては個別の同意を取らなかった。

診療録に基づいて 2004 年から 2022 年で依頼書に肝疾患を有するものを対象とし、肝腫瘍がある症例を除外した。これは、肝腫瘍の存在が全肝体積に対する影響を及ぼす可能性があり、また腫瘍部分を肝機能に含めるか否かについては一定の見解が得られていないためである（Bilic P et al. 2023）。CT 検査で 5 年以上の経過がある CT を 200 例抽出した。合計 3656 件の CT を AI を用いたセグメンテーションを行い、縦断的な観察を行った。

対象とした肝疾患は、B 型肝炎、C 型肝炎、自己免疫性肝炎（Autoimmune hepatitis:AIH）、代謝機能障害関連脂肪肝炎（Metabolic Dysfunction Associated Steatohepatitis : MASH）／代謝機能障害関連脂肪性肝疾患（Metabolic Dysfunction Associated Steatotic Liver Disease : MASLD）、原発性胆汁性胆管炎（primary biliary cholangitis : PBC）、Wilson 病、アルコール性肝障害、うっ血性肝障害、薬剤性肝炎である。

2. CT 撮像

非造影 CT および様々な造影相の CT 画像を使用した。撮影に使用された CT 装置には、Aquilion ONE Vision、Aquilion PRIME（いずれもキヤノンメディカルシステムズ社製、日本・大田原）、および Brilliance iCT Elite（フィリップス社製、オランダ・アムステルダム）が含まれた。スライス厚は 5 mm、管電圧は 120 kVp で統一された。造影剤は Iopamiron®および Iomeron®を使用した。撮影パラメータの詳細は表 3-1 に示す。

表 3-1 : CT の撮像パラメータ

	Aquilion ONE	Aquilion PRIME	Brilliance iCT Elite
管電圧 (kVp)	120	120	120
スキャンコリメーション (列×スライス厚)	320×0.5	80×0.5	256×0.625
回転時間 (秒)	0.5	0.35	0.5
マトリックス (行×列)	512×512	512×512	512×512
ピッチ (mm)	0.813	1.388	0.900
スライス厚 (mm)	5	5	5
再構成スライス厚 (mm)	1	1	1

3. AI システム

TotalSegmentator は、全身 CT 画像から 100 以上の解剖学的構造（臓器・骨・血管など）を自動抽出できるオープンソースのセグメンテーションツールである。nnU-Net フレームワークに基づいて構築されており、多様な撮像条件に対して高い精度と汎用性を示す (Isensee et al., 2022)。GitHub 上で公開されている (<https://github.com/wasserth/TotalSegmentator>)。TotalSegmentator を用いて 3656 件の CT 画像で肝臓のセグメンテーションを行った。解析環境は次の通りである。OS: Ubuntu 22.4, CPU: Xeon Silver 4214 (12Core x 2CPU)、メインメモリー: 96GB、GPU: 3x NVIDIA TITAN RTX 24GB である。

4. 統計解析

統計解析は Microsoft Excel (Microsoft Corp., Redmond, WA, USA) を用いて実施した。連続変数は正規分布を示す場合には、値は平均 ± 標準偏差 (SD)、正規分布性が無い場合には中央値と四分位範囲 (IQR) で記述し、中央値[IQR]と角括弧で示した。有意水準は両側 t 検定を用いて、 $p < 0.05$ を統計学的有意と判断した。

結果

1. 症例特性

症例特性を表 3-1 に示す。全体の年齢の中央値は 64[56-71]、観察期間は 2002 年 1 月 1 日から 2023 年 1 月 31 日、観察期間の中央値は 2902.5[1577.75-4285.25]日であった。

表 3-1：症例特性

疾患	項目	値
全体	症例数（人）	200
	年齢（歳）	64[56-71]
	男性（人、割合）	93（46.5%）
	観察期間（日）	2902.5
		[1577.75-4285.25]
AIH	症例数（人）	8
	年齢（歳）	68.5 [64-71.25]
	男性（人、割合）	1（12.5%）
	観察期間（日）	3037.5 [1566-4655.25]
B 型肝炎／B 型肝硬変	症例数（人）	47
	年齢（歳）	59 [52.5-65.5]
	男性（人、割合）	25（53.19%）
	観察期間（日）	3201 [1920.5-5082]
C 型肝炎／C 型肝硬変	症例数（人）	73
	年齢（歳）	64 [56-72]
	男性（人、割合）	34（46.58%）
	観察期間（日）	2408 [1379.5-4066.75]
MASLD	症例数（人）	2
	年齢（歳）	60.5 [51.75-69.25]
	男性（人、割合）	0（0%）
	観察期間（日）	3174 [2656.5-3691.5]
MASLD＋うっ血	症例数（人）	1

	年齢（歳）	79
	男性（人、割合）	1（100%）
	観察期間（日）	5081
MASH	症例数（人）	15
	年齢（歳）	68 [59-71.5]
	男性（人、割合）	5（33.3%）
	観察期間（日）	2965 [2140-5784]
MASH+B型肝炎／B型肝炎硬変	症例数（人）	1
	年齢（歳）	60
	男性（人、割合）	0（0%）
	観察期間（日）	5662
PBC	症例数（人）	15
	年齢（歳）	68 [60.5-72.5]
	男性（人、割合）	0（0%）
	観察期間（日）	3304 [2178-5825.5]
PBC+AIH	症例数（人）	3
	年齢（歳）	67 [52-68.5]
	男性（人、割合）	0（0%）
	観察期間（日）	1363 [1340.5-4178.5]
Wilson 病	症例数（人）	1
	年齢（歳）	28
	男性（人、割合）	1（100%）
	観察期間（日）	7034
アルコール性	症例数（人）	24
	年齢（歳）	66 [61.75-72]
	男性（人、割合）	20（83.3%）
	観察期間（日）	2435 [1812.75-3890.25]
うっ血肝	症例数（人）	9
	年齢（歳）	67 [38-77]
	男性（人、割合）	6（66.7%）
	観察期間（日）	2261 [967-3125]
薬剤性	症例数（人）	1
	年齢（歳）	59
	男性（人、割合）	0（0%）

年齢、観察期間は中央値、角括弧内は IQR を示す。

2. TotalSegmentator と第一章・第二章で使った AI モデルの比較

同一の肝疾患症例に対して、第一章および第二章で使った既存の AI モデルと TotalSegmentator を用いて肝臓セグメンテーションを実施した。目視による比較評価の結果、TotalSegmentator は既存モデルと比較してより正確に肝臓領域を抽出できていた。特に、腹水や胸水を伴う症例において、既存モデルではこれら低吸収領域を肝実質として誤って包含する過剰セグメンテーションが散見された。一方、TotalSegmentator では腹水を除外し、肝実質の輪郭をより正確に抽出していた。

図 3-5 に、両モデルのセグメンテーション結果が異なった代表症例を示す。既存 AI モデルでは腹水や脾臓を肝臓領域として誤認していたが、TotalSegmentator では肝実質のみを正確に囲っており、病的肝を含む症例においても高い適応性を示した。

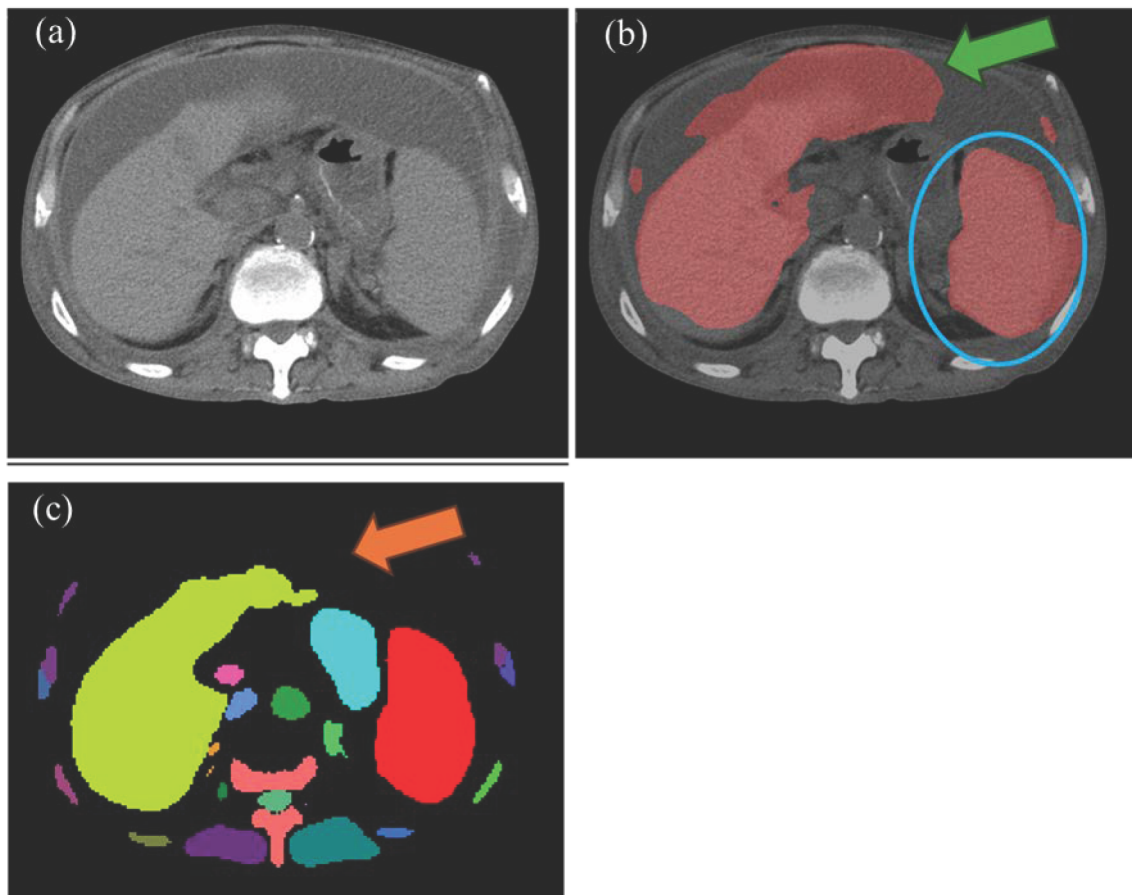


図 3-5：既存の AI モデルと TotalSegmentator のセグメンテーション比較

(a) は CT のみの画像。(b) は既存の AI モデルで肝臓のセグメンテーションを行った図。赤色の領域が肝臓としてセグメンテーション処理された領域を示す。緑矢印が腹水の部分が肝臓として誤認されている領域。水色の丸が脾臓の部分が肝臓として誤って囲まれている領域。(c) は TotalSegmentator でセグメンテーション処理された図。黄緑色の肝臓は CT と比較して、正しく領域設定されている。橙色矢印は既存の AI モデルで腹水として囲まれていた部分は、エラーが見られない。また、脾臓は赤い領域でセグメンテーション処理がされている。

3. 肝疾患毎の肝体積

TotalSegmentator を用いてセグメンテーション解析を行い、肝体積を検出した後、疾患毎に年間の変化率を線形モデルで算出した結果を表 3-2 に示す。

表 3-2：肝疾患毎の肝体積の年間変化率

肝疾患	人数	肝体積の年間変化率 (%)	p 値
AIH	8	-1.26±0.08	0.07
B 型肝炎／B 型肝硬変	46	0.55±0.47	0.68
C 型肝炎／C 型肝硬変	72	-0.73±0.20	0.10
MASLD	2	-0.83±0.07	0.65
MASLD＋うっ血肝	1	-0.65±0	-
MASH	15	0.44±0.13	0.52
MASH＋B 型肝炎／B 型肝 硬変	1	-1.225±0	-
PBC	15	-0.52±0.11	0.37
PBC＋AIH	3	-1.55±0.10	0.04*
Wilson 病	1	0.29±0	-
アルコール性	23	-1.353±0.24	0.18
うっ血肝	19	-0.44±0.11	0.58
薬剤性	1	-2.61±0	-

両側 t 検定を行い、 $p < 0.05$ を統計学的有意とした

多く疾患で肝体積は減少を示したが、B 型肝炎／B 型肝硬変、MASH、Wilson 病では増加が見られた。また、両側 t 検定を行ったところ、PBC+AIH のみ $p < 0.05$ で有意な変化と考えられたが、その他では有意な変化とは言えなかった。

4. 実際の肝体積のセグメンテーションを行った症例の代表例

症例 1：60 歳代、女性。X 年に検診にて C 型肝炎ウイルス陽性を指摘された。X+1 年にインターフェロン α 900 万単位を 4 週連投、12 週間欠で治療を行ったが、ウイルス学的著効（sustained virological response：SVR）は得られなかった。X+4 年、インターフェロン α 2a とリバビリン併用療法、X+5 年までペグインターフェロン少量投与が行われた。X+20 年に肝機能の増悪により永眠。X+8 年-X+20 年の肝体積の推移のグラフを図 3-6、実際のセグメンテーションの画像を図 3-7 に示す。

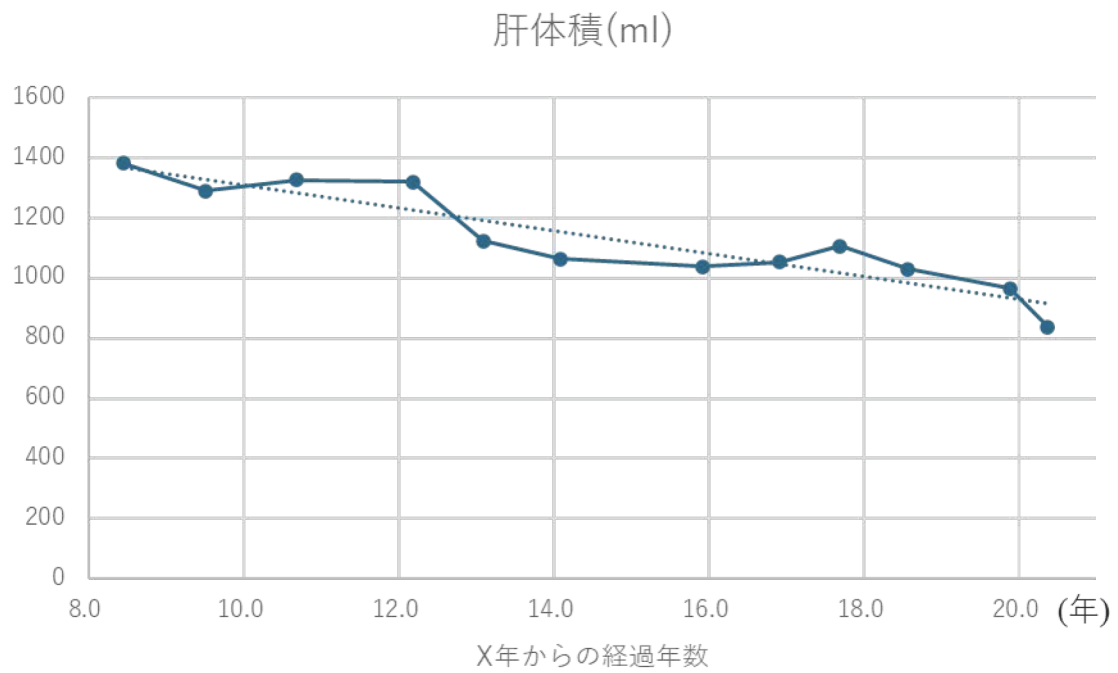


図 3-6: 症例 1 の肝体積のグラフ

縦軸は肝体積、横軸は X 年からの経過年数、折れ線は実際の体積値、点直線は近似直線を示す。年間の減少率は 3.3%であった。

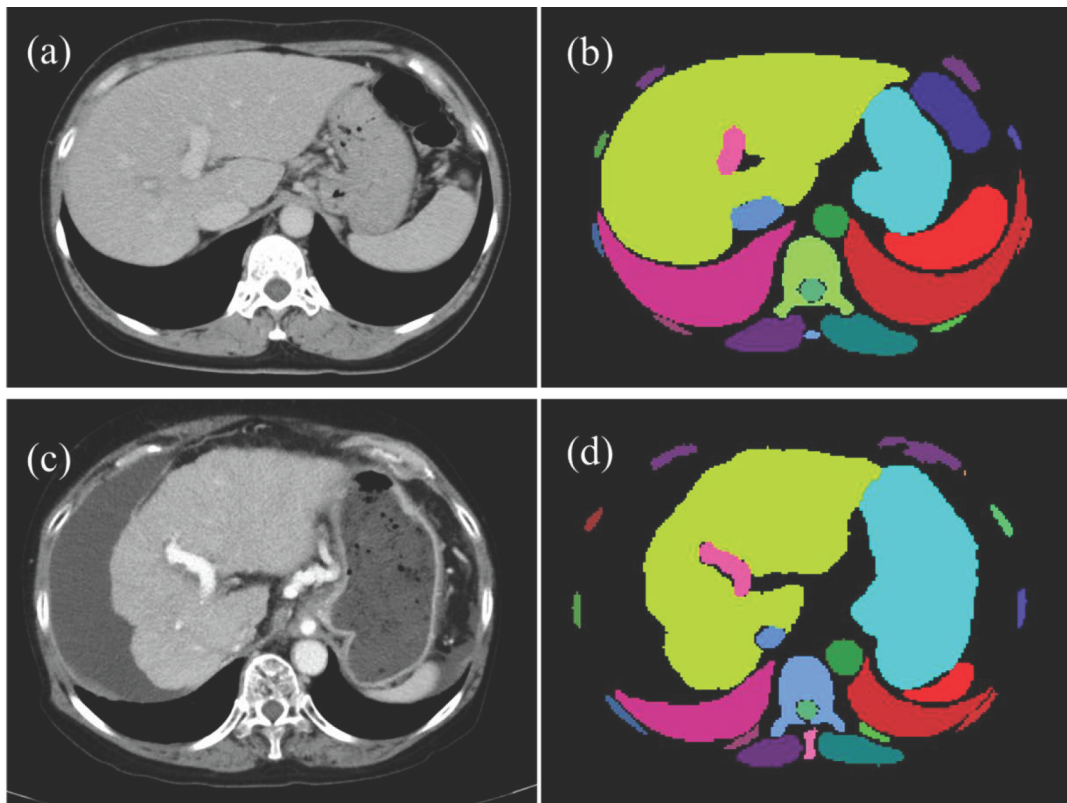


図 3-7：症例 1 のセグメンテーションが施行された図

上段は X+8.5 年。(a) は実際の CT 画像、(b) はセグメンテーション処理された図。黄緑色が肝臓を示す。下段は X+20.4 年。(c) は実際の CT 画像、(d) はセグメンテーション処理された図。黄緑色が肝臓を示す。

CT では 12 年の間で肝臓の体積が減少しており、肝形態も変化し、肝硬変の進行が確認された。右葉の萎縮の進行が見られた。

症例 2：60 歳代、男性。

X 年に検診の肝機能障害にて、B 型慢性肝炎の診断をされた。X 年にインターフェロン α を 1 ヶ月間施行、X+14 年インテカビル+インターフェロン β 併用療法開始した。SVR は得られなかった。肝体積の推移のグラフを図 3-8、実際のセグメンテーションの画像を図 3-9 に示す。

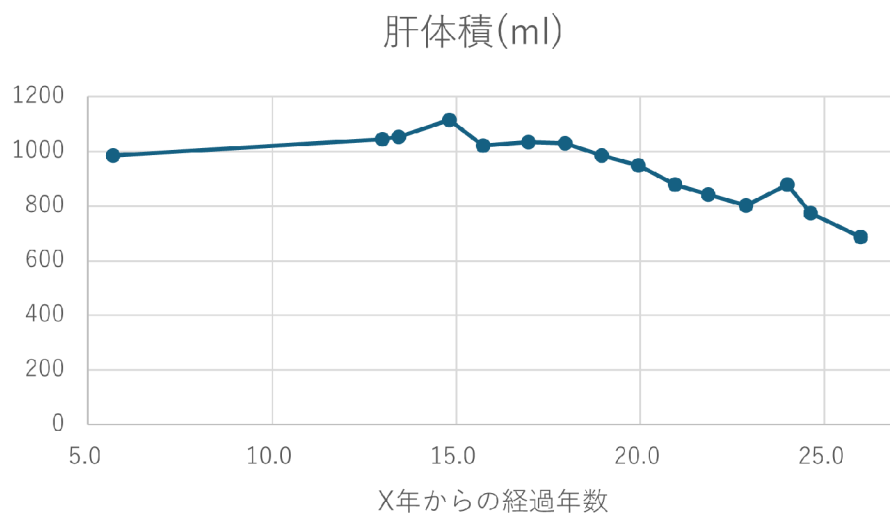


図 3-8：症例 2 の肝体積のグラフ

縦軸は肝体積、横軸は X 年からの経過年数、折れ線は実際の体積値を示す。減少に転じた X+15 年からの年間の減少率は 3.8%であった。

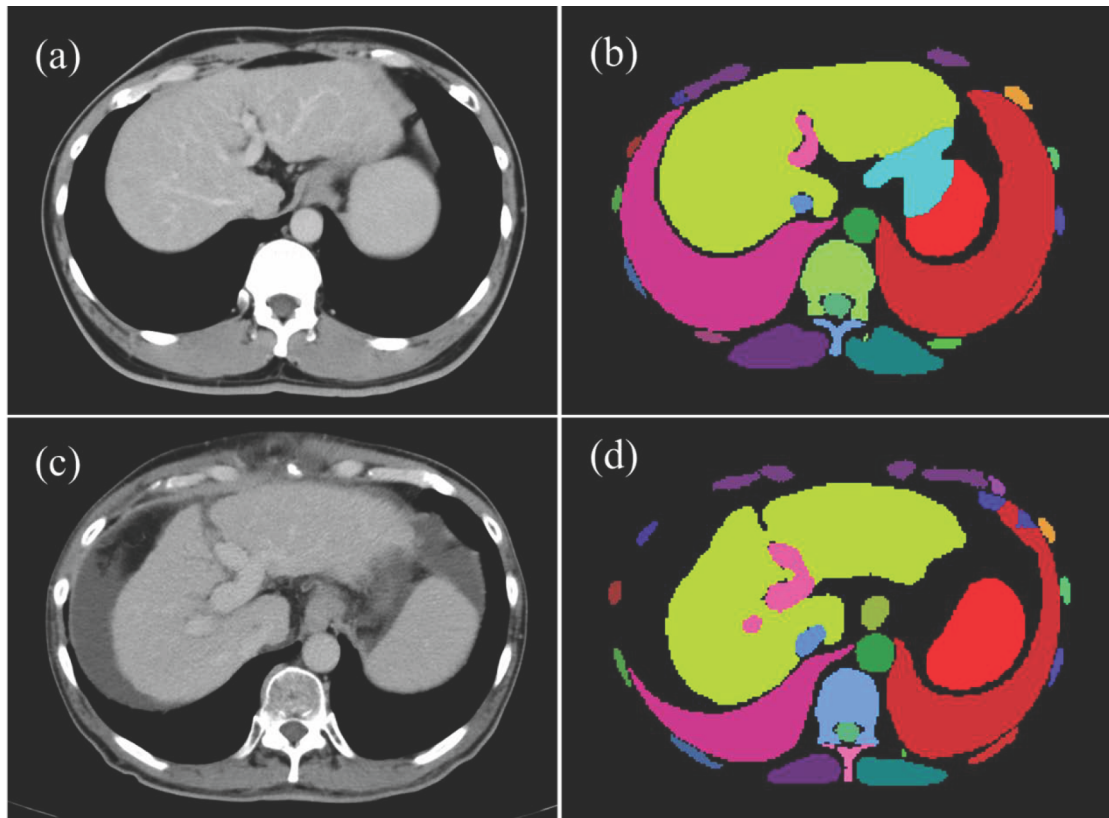


図 3-9：症例 2 のセグメンテーションが施行された図

上段は X+5.7 年。(a) は実際の CT 画像、(b) はセグメンテーション処理された図。黄緑色が肝臓を示す。下段は X+26.0 年。(c) は実際の CT 画像、(d) はセグメンテーション処理された図。黄緑色が肝臓を示す。

CT では X+5-10 年の期間では肝体積の変動はほぼ見られなかったが、X+15 年以降は肝体積が減少しており、肝形態も変化し、肝硬変の進行が確認された。

考察

本研究において、肝疾患患者の肝体積の経時的変化を TotalSegmentator を用いて解析したところ、B 型肝炎、MASH、Wilson 病では肝体積がやや増加する傾向を示した一方で、それ以外の多くの疾患では減少傾向がみられた。しかし、両側 t 検定による解析の結果、ほぼ全ての群において $p > 0.05$ であり、統計学的に有意な変化率とはいえなかった。また、PBC+AIH 群で有意差が見られたものの 3 例と症例数が少なく、統計的ばらつきの影響を受けやすいため、この結果の解釈には慎重を要すると考えられた。抽出した 2 例の縦断的検討では、年間 3.3~3.8% の肝体積減少が認められ、第二章で得られた加齢による生理的減少率（約 1.1%）を上回っていた。このことは、肝疾患に伴う線維化進行や慢性炎症が、加齢変化に加えて体積減少を促進している可能性を示唆する。しかしながら、症例数が少なく、統計学的に有意な変化とみなすには至らなかったため、本結果は傾向としての解釈に留める必要がある。今後は症例数を増やし、疾患別に体積変化率を比較することで、肝体積変化と病勢進行の関連をより明確にすることが必要である。

肝体積の測定は、肝移植、術前評価、放射線治療計画、術後肝不全の予防において重要な役割を担ってきた (Urata K et al., 1995)。これまでは身長・体重・体表面積に基づく推定式が用いられていた。近年では、SYNAPSE SAI Viewer（富士フイルムメディカル社製）のように DICOM ビューア上でセグメンテーション機能を備えるものも存在するが、臨床的有用性が確立されていないことから、読影レポートに反映されることは稀である。日常の画像検査で肝臓のみならず脾臓、膵臓、腎臓の体積も測定可能であり、特に脾臓・膵臓体積の研究も一部報告されているが、依然として限られている (Perez AA et al., 2023; Tallam H et al., 2022)。肝臓疾患と肝臓の体積の関係については多くの研究がある。肝臓の容積と肝線維化の評価として、右葉左葉は減少するものの尾状葉のみが増大する (Li et al., 2010)、線維化のステージとともにクイノー区域 1-3 対クイノー区域 4-8 が上昇するという報告がある (Pickhardt et al., 2017)。急性肝炎やアルコール性肝炎などの予後予測にも肝体積が有用である可能性がある (Zabron A et al., 2018; Zhou X ping et al., 2007)。脂肪肝では肝体積が増加することが報

告されており (Choi JY et al., 2023)、正常範囲の解釈や病的意義の解明は今後の課題である。さらに、放射線治療によって肝体積が変化することも知られており、AI による自動測定はこの変化の評価にも有用と考えられる (Cazoulat G et al., 2023)。また、腫瘍を除外することで、肝実質量だけを対象とした肝体積評価が可能となる。肝機能は主に肝実質細胞に依存するため、腫瘍を含めると体積と機能との乖離が生じやすい。縦断解析においても、腫瘍の成長や縮小が実質量の変化を覆い隠し得る。腫瘍を除外すれば、肝実質の萎縮や加齢性減少率をより精度良く評価でき、残肝評価・術前予後予測や治療効果モニタリング精度を高めることが期待される (Bilic P et al. 2023)。

統計的には有意な差を得られなかったものの、本研究で示された疾患ごとの傾向は臨床的に注目に値する。肝障害の初期では一過性の肝腫大が生じることが既報でも報告されている (Sherigar JM et al. 2018; Pickhardt PJ et al. 2016)。進行性の線維化を伴う疾患群では体積減少が主体であり、これは肝予備能の低下や予後不良と関連する可能性がある。今後、より大規模かつ縦断的なデータを用いることで、疾患ごとに「体積増加から減少へ移行する転換点」を同定できれば、治療方針決定や予後予測に有用な指標となり得る。

本研究にはいくつかの限界が存在する。後ろ向き研究で、単一施設であり、症例数が十分ではなく、疾患ごとの詳細な比較解析を行うには検出力が不足していた。当院では、SVR が達成された症例では、以降の経過観察に CT ではなく超音波検査が主に用いられていたため、本研究における CT 検査症例には SVR 未達成例が多く含まれていた。TotalSegmentator の精度を検証していない。また、TotalSegmentator は肝臓全体の体積しか測定できず、右葉や左葉や区域ごとのセグメンテーションを行えない。また、観察期間中に行われた治療介入や生活習慣の変化が結果に影響した可能性を完全には排除できない。さらに、セグメンテーションに基づく体積測定には依然として測定誤差やばらつきが存在し、個々の症例における変化が過小評価または過大評価される可能性がある。今後は、より多数例を対象とした前向き研究に加え、肝体積の推移と血液検査結果や肝予備能評価との統合解析を行うことで、肝疾患における体積変化の臨床的意義をより明確にする必要がある。

全体の結論

1. 本研究から得られた新知見

○第一章

- ・ 従来の手動や半自動による肝臓セグメンテーションの課題に対し、AIモデルを用いた自動セグメンテーションの性能を評価し、非造影および造影各相において、放射線診断専門医によるセグメンテーションと高い一致度を示し、DSC や ICC がいずれも高値を示した。
- ・ AI モデルが放射線診断専門医と同等の精度での肝体積測定を実現できることを実証した。

○第二章

- ・ 第一章で性能が確認された AI モデルを応用し、肝疾患を有さない症例を対象に最大 20 年間の縦断的解析を行い、加齢に伴い肝体積は年間約 1%減少することが明らかとなった。
- ・ 従来断面的にしか把握されていなかった加齢による肝体積変化を、縦断的かつ定量的に裏付けた初めての成果であり、今後の正常基準値の構築や加齢に伴う肝機能変化の解釈に寄与するものと考えられる。

○第三章

- ・ 第三章では、近年発表された新たな AI モデルである **TotalSegmentator** を導入し、肝疾患患者の CT 画像を対象に肝体積の縦断的变化を解析した。
- ・ 従来モデルでは腹水や胸水を肝実質と誤認するセグメンテーションエラーが多発したが、**TotalSegmentator** を用いることで精度がある程度向上し、病的肝に対しても安定した計測が可能となった。
- ・ 疾患ごとの肝体積の推移を把握することで、肝障害の進行度や予後との関連を検討する基盤を得ることができ、肝疾患症例における AI 活用の臨床的可能性が示された。

2. 新知見の意義

肝臓の体積の研究は今まで数多くなされていたが、手動や半手動の時代では症例数が限られていた。また、年齢と肝臓の体積の関係の研究では横断的研究が多く見られていた。本研究では第一章、第二章では公開されて

いる AI モデル、第三章では公開されている **TotalSegmentator** を用いて、大規模な肝体積の縦断的研究を行い、同一患者の肝体積の推移を検討することができた。加齢による影響は一年あたり約 1%減少することが示された。また、肝疾患を有する肝体積の同一患者での推移を検討することができた。

3. 本研究で得られた新知見から今後展開されうる研究

AI を用いることでかつて小規模で行われていた研究を大規模に検討することが可能であることが考えられる。また、日常の読影でも肝臓を含め、様々な臓器の体積を測定することができるようになったものの、その病的意義に関しては不明な部分が多い。臓器毎に病的意義のある臓器体積、臓器体積の加齢性変化を検討することで、体積測定の意義を検証する必要がある。

4. 今後の研究課題

今後の課題として、まず肝体積の縦断的推移と肝機能指標との関連をさらに深く検討する必要がある。具体的には、採血データに基づく生化学的指標や肝予備能評価で広く用いられている **Child-Pugh** 分類、アルブミン・ビリルビンスコアとの相関を解析することで、体積変化がどの程度機能低下を反映するかを明らかにすることが期待される。このような検討により、単なる形態学的指標である体積を、機能的アウトカムと結びつける新たなバイオマーカーとして活用できる可能性がある。

さらに、読影の現場において肝体積を測定した際、その増減がどの程度で病的意義を持つのか、すなわち臨床的に意味のある閾値（カットオフライン）を設定することも重要である。

加えて、**TotalSegmentator** は肝臓に限らず、脾臓、骨格筋、骨など多様な臓器の自動セグメンテーションに対応しているため、肝臓以外の臓器体積の縦断的変化を同時に評価することも可能である。例えば、筋肉量の低下は肝疾患患者だけでなく心疾患、神経疾患、悪性腫瘍など多くの疾患の予後に関わる重要な因子である。骨密度や骨体積の評価もまた、長期的な代謝異常や栄養状態の影響を反映する可能性がある。

以上より、今後は AI を活用した多臓器同時解析により、肝疾患に限らず包括的な臓器ネットワークの変化を捉える研究が求められる。これによ

り、肝体積の推移を単独で評価するのではなく、他臓器の変化と組み合わせることで、疾患の進行や治療効果をより精緻に予測する新たな臨床応用の可能性が拓かれると考える。

謝辞

本稿を結ぶに当たり、終始懇切なる指導を賜りました指導教官の北海道大学大学院医学研究院放射線科学分野画像診断学教室 工藤與亮 教授に深い感謝の意を表します。

併せて、本研究全般に渡って直接的なご指導を賜りました北海道大学大学院医学研究院放射線科学分野画像診断学教室平田健司 准教授に深い感謝の意を表します。

加えて、画像診断及び解析にご協力を頂いた、北海道大学病院 核医学診療科 渡邊 史郎 助教、北海道大学大学院歯学研究院口腔病態学分野放射線学教室 常田慧徳 助教、北海道大学大学院医学研究院 画像診断学教室 竹中 淳規 客員研究員に深く感謝申し上げます。

また、臨床的な観点から様々な助言をいただきました、北海道大学放射線部 阿保大介准教授に深く感謝いたします。

最後に、本論文は臨床データを使用させていただいた患者の皆様、当該分野内外の先生方、その他多くの方々のご指導とご協力に支えられて完成したものであり、研究に携わったすべての方々への心からの感謝の意を表します。

利益相反

本研究において開示すべき利益相反はありません。

引用文献

Bilic P, Christ P, Li HB, Vorontsov E, Ben-Cohen A, Kaissis G, Szeskin A, Jacobs C, Humpire Mamani GE and Chartrand G, et al. (2023) The Liver Tumor Segmentation Benchmark (LiTS) . *Med Image Anal*, 84, 102680.

Cai W, He B, Fan Y, Fang C and Jia F (2016) Comparison of liver volumetry on contrast-enhanced CT images: one semiautomatic and two automatic approaches. *J Appl Clin Med Phys*, 17, 118–127.

Cazoulat G, Gupta AC, Al Taie MM, Koay EJ and Brock KK (2023) Analysis and prediction of liver volume change maps derived from computational tomography scans acquired pre- and post-radiation therapy. *Phys Med Biol*, 68, 205009.

Childs BG, Durik M, Baker DJ and van Deursen JM (2015) Cellular senescence in aging and age-related disease: from mechanisms to therapy. *Nat Med*, 21, 1424–1435.

Choi JY, Lee SS, Kim NY, Park HJ, Sung YS, Lee Y, Yoon JS and Suk H-I (2023) The effect of hepatic steatosis on liver volume determined by proton density fat fraction and deep learning-measured liver volume. *Eur Radiol*, 33, 5924–5932.

Cohnert TU, Rau HG, Buttler E, Hernandez-Richter T, Sauter G, Reuter C and Schildberg FW (1997) Preoperative risk assessment of hepatic resection for malignant disease. *World J Surg*, 21, 396–400; discussion 401.

Fan M, Wang T, Lei Y, Patel PR, Dresser S, Ghavidel BB, Qiu RLJ, Zhou J, Luca K and Kayode O, et al. (2025) Evaluation and failure analysis of four commercial deep learning-based autosegmentation software for abdominal organs at risk. *J Appl Clin Med Phys*, 26, e70010.

Ferrero A, Viganò L, Polastri R, Muratore A, Eminefendic H, Regge D and Capussotti L (2007) Postoperative liver dysfunction and future remnant liver: where is the limit? results of a prospective study. *World J Surg*, 31, 1643–1651.

Fitzpatrick JA, Kim JU, Cobbold JFL, McPhail MJW, Crossey MME, Bak-Bol AA, Zaky A and Taylor-Robinson SD (2016) Changes in liver volume in patients with chronic hepatitis C undergoing antiviral therapy. *J Clin Exp Hepatol*, 6, 15–20.

Gotra A, Sivakumaran L, Chartrand G, Vu K-N, Vandenbroucke-Menu F, Kauffmann C, Kadoury S, Gallix B, de Guise JA and Tang A (2017) Liver segmentation: indications, techniques and future directions. *Insights Imaging*, 8, 377–392.

Harada K, Ishinuki T, Ohashi Y, Tanaka T, Chiba A, Numasawa K, Imai T, Hayasaka S, Tsugiki T and Miyanishi K, et al. (2021) Nature of the liver volume depending on the gender and age assessing volumetry from a reconstruction of the computed tomography. *PLoS One*, 16, e0261094.

Heimann T, van Ginneken B, Styner MA, Arzhaeva Y, Aurich V, Bauer C, Beck A, Becker C, Beichel R and Bekes G, et al. (2009) Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imaging*, 28, 1251–1265.

Hermoye L, Laamari-Azjal I, Cao Z, Annet L, Lerut J, Dawant BM and Van Beers BE (2005) Liver segmentation in living liver transplant donors: comparison of semiautomatic and manual methods. *Radiology*, 234, 171–178.

Jie L, Defeng W, Lin S and Pheng AH (2012) Automatic liver segmentation in CT images based on support vector machine. *Proceedings of 2012 IEEE-EMBS, Hong Kong, China, 5–7 January 2012*.
<https://ieeexplore.ieee.org/document/6211581>

Kennedy PA and Madding GF. *Surgical Anatomy of the Liver*. (1997)

Surgical Clinics of North America. 57, 233-244.

Koo, T. K. and Li, M. Y. (2016) . A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine, 15, 155–163.

Krizhevsky A, Sutskever I and Hinton GE. (2012) ImageNet Classification with Deep Convolutional Neural Networks. Adv Neural Inf Process Syst. 25. Accessed April 20, 2024. <http://code.google.com/p/cuda-convnet/>

Kubota K, Makuuchi M, Kusaka K, Kobayashi T, Miki K, Hasegawa K, Harihara Y and Takayama T (1997) Measurement of liver volume and hepatic functional reserve as a guide to decision-making in resectional surgery for hepatic tumors. Hepatology, 26, 1176–1181.

Kwo PY, Ramchandani VA, O'Connor S, Amann D, Carr LG, Sandrasegaran K, Kopecky KK and Li TK (1998) Gender differences in alcohol metabolism: relationship to liver volume and effect of adjusting for body mass. Gastroenterology, 115, 1552–1557.

Lee S, Elton DC, Yang AH, Koh C, Kleiner DE, Lubner MG, Pickhardt PJ and Summers RM (2022) Fully automated and explainable liver segmental volume ratio and spleen segmentation at CT for diagnosing cirrhosis. Radiol Artif Intell, 4, e210268.

Li, W. X., Zhao, X. T., Chai, W. M., Zhu, N. Y., Du, L. J., Huang, W., Ling, H. W., Chen, K. M. and Xie, Q. (2010) Hepatitis B virus-induced liver fibrosis and cirrhosis: the value of liver and spleen volumetry with multi-detector spiral computed tomography. Journal of Digestive Diseases, 11, 215–223.

Linguraru MG, Sandberg JK, Jones EC, Petrick N and Summers RM. (2012) Assessing Hepatomegaly. Automated Volumetric Analysis of the Liver. Acad Radiol. 19, 588-598.

Litjens G, Kooi T, Ehteshami Bejnordi B, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B and Sánchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Anal*, 42, 60–88.

Malmberg F, Nordenskjöld R, Strand R and Kullberg J. (2014) SmartPaint: a tool for interactive segmentation of medical volume images. *Comput Methods Biomech Biomed Eng Imaging Vis* 5, 36–44.

Martel G, Cieslak KP, Huang R, van Lienden KP, Wiggers JK, Belblidia A, Dagenais M, Lapointe R, van Gulik TM and Vandenbroucke-Menu F. (2015) Comparison of techniques for volumetric analysis of the future liver remnant: implications for major hepatic resections. *HPB (Oxford)* , 17, 1051–1057.

Molière S, Hamzaoui D, Granger B, Montagne S, Allera A, Ezziane M, Luzurier A, Quint R, Kalai M, Ayache N, et al. (2024) Reference standard for the evaluation of automatic segmentation algorithms: quantification of inter observer variability of manual delineation of prostate contour on MRI. *Diagn Interv Imaging*, 105, 65–73.

Motomi Katayama, Kiyotaka Yamazumi, Kenji Kino, Masanobu Tsuru, Toshio Fukuzawa and Hiroyuki Shimada. (1990) Change of Liver Weight in the Elderly. *Japanese Journal of Geriatrics*. 25, 584-588.

Müller, D., Soto-Rey, I. and Kramer, F. (2022) Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15, 1–8.

Nakayama Y, Li Q, Katsuragawa S, Ikeda R, Hiai Y, Awai K, Kusunoki S, Yamashita Y, Okajima H, Inomata Y, et al. (2006) . Automated hepatic volumetry for living related liver transplantation at multisection CT. *Radiology*, 240, 743–748.

Nayantara PV, Kamath S, Kadavigere R and Manjunath KN. (2024) Automatic Liver Segmentation from Multiphase CT Using Modified SegNet

and ASPP Module. SN Comput Sci. 5 (4) .

Perez AA, Noe-Kim V, Lubner MG, Graffy PM, Garrett JW, Elton DC, Summers RM and Pickhardt PJ. (2022) . Deep learning CT-based quantitative visualization tool for liver volume estimation: defining normal and hepatomegaly. Radiology, 302, 336–342.

Perez AA, Noe-Kim V, Lubner MG, Somsen D, Garrett JW, Summers RM and Pickhardt PJ. (2023) . Automated deep learning artificial intelligence tool for spleen segmentation on CT: defining volume-based thresholds for splenomegaly. AJR Am J Roentgenol, 221, 611–619.

Pickhardt PJ, Malecki K, Hunt OF, Beaumont C, Klope J, Ziemlewicz TJ and Lubner MG. (2017) . Hepatosplenic volumetric assessment at MDCT for staging liver fibrosis. Eur Radiol, 27, 3060–3068.

Poovathumkadavil A, Leung KF, Al Ghamdi HM, El Hassan Othman I and Meshikhes AW. (2010) Standard formula for liver volume in Middle Eastern Arabic adults. Transplant Proc. 42, 3600-3605.

Sandfort V, Yan K, Pickhardt PJ and Summers RM. (2019) Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Scientific Reports 9, 1-9.

Sharma N and Aggarwal LM. (2010) Automated medical image segmentation techniques. J Med Phys 35, 3–14.

Sherigar JM, de Castro J, Yin YM, Guss D and Mohanty SR. (2018) Glycogenic hepatopathy: A narrative review. World J Hepatol. 10, 172-185.

Soejima K and Sato H, Hisaka A. (2002) Age-Related Change in Hepatic Clearance Inferred from Multiple Population Pharmacokinetic Studies: Comparison with Renal Clearance and Their Associations with Organ Weight and Blood Flow. Clin Pharmacokinet. 61, 295-305.

- Tallam H, Elton DC, Lee S, Wakim P, Pickhardt PJ and Summers RM. (2022) . Fully automated abdominal CT biomarkers for type 2 diabetes using deep learning. *Radiology*, 304, 85–95.
- Tang A, Chen J, Le TA, Changchien C, Hamilton G, Middleton MS, Loomba R and Sirlin CB. (2015) . Cross-sectional and longitudinal evaluation of liver volume and total liver fat burden in adults with nonalcoholic steatohepatitis. *Abdom Imaging*, 40, 26–37.
- Urata K, Kawasaki S, Matsunami H, Hashikura Y, Ikegami T, Ishizone S, Momose Y, Komiyama A and Makuuchi M. (1995) . Calculation of child and adult standard liver volume for liver transplantation. *Hepatology*, 21, 1317–1321.
- Um EH, Hwang S, Song GW, Jung DH, Ahn CS, Kim KH, Moon DB, Park GC and Lee SG. (2015) . Calculation of standard liver volume in Korean adults with analysis of confounding variables. *Korean J Hepatobiliary Pancreat Surg*, 19, 133–138.
- Vauthey, J.N., Abdalla, E.K., Doherty, D.A., Gertsch, P., Fenstermacher, M.J., Loyer, E.M., Lerut, J., Materne, R., Wang, X., Encarnacion, A. et al. (2002) Body surface area and body weight predict total liver volume in Western adults, *Liver Transplantation*, 8,233–240.
- Wasserthal, J., Breit, H.-C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S. et al. (2023) TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 5, e230024.
- Yoshizumi, T., Gondolesi, G.E., Bodian, C.A., Jeon, H., Schwartz, M.E., Fishbein, T.M., Miller, C.M. and Emre, S. (2003) A simple new formula to assess liver weight. *Transplant Proc.* 3:1415-1420.
- Zabron, A., Quaglia, A., Fatourou, E., Peddu, P., Lewis, D., Heneghan, M., Willars, C., Auzinger, G., Heaton, N., Wendon, J. et al. (2018) Clinical

and prognostic associations of liver volume determined by computed tomography in acute liver failure. *Liver International*. 38, 1592-1601.

Zhou XP, Lu T, Wei YG and Chen XZ. (2007) . Liver volume variation in patients with virus-induced cirrhosis: findings on MDCT. *AJR Am J Roentgenol*, 189, W153–159.