



Title	作業分解手法を用いた腹腔鏡下S状結腸切除術における自動手術技術評価に関する研究
Author(s)	大淵, 佳祐
Description	配架番号 : 2979
Degree Grantor	北海道大学
Degree Name	博士(医学)
Dissertation Number	甲第16886号
Issue Date	2026-03-25
DOI	https://doi.org/10.14943/doctoral.k16886
Doc URL	https://hdl.handle.net/2115/99731
Type	doctoral thesis
File Information	OBUCHI_Keisuke.pdf, 全文



学 位 論 文

作業分解手法を用いた腹腔鏡下 S 状結腸切除
術における自動手術技術評価に関する研究

(Studies on automated surgical skill assessment
in laparoscopic sigmoidectomy using a task
decomposition approach)

2026 年 3 月

北 海 道 大 学

大 渕 佳 祐

学 位 論 文

作業分類手法を用いた腹腔鏡下 S 状結腸切除
術における自動手術技術評価に関する研究

(Studies on automated surgical skill assessment in
laparoscopic sigmoidectomy using a task
decomposition approach)

2026 年 3 月

北 海 道 大 学

大 渕 佳 祐

目 次

発表論文目録および学会発表目録	1 頁
要旨	2 頁
略語表	5 頁
緒言	6 頁
対象と方法	14 頁
結果	27 頁
考察	46 頁
結論	51 頁
謝辞	53 頁
利益相反	54 頁
引用文献	55 頁
付録 1 共通基準採点表【大腸】	61 頁
付録 2 臓器別評価項目【大腸】	66 頁

発表論文目録および学会発表目録

本研究の一部は以下の論文に発表した。

1. Keisuke Obuchi, Shin Takenaka, Daichi Kitaguchi, Kei Nakajima, Yuto Ishikawa, Hiroki Mitarai, Kyoko Ryu, Nobuyoshi Takeshita, Akinobu Taketomi, Masaaki Ito
Automatic surgical skill assessment using a task classification model in laparoscopic sigmoidectomy
Surgical Endoscopy. 2025 Oct 39(10).

本研究の一部は以下の学会に発表した。

1. 大渕 佳祐、竹中 慎、中島 啓、龍 喬子、北口 大地、竹下 修由、伊藤 雅昭
腹腔鏡下 S 状結腸切除術における作業分解的手法を用いた DEI 時間の可視化と AI 自動評価に関する研究
第 35 回日本内視鏡外科学会総会、2022 年 12 月・愛知県名古屋市
2. 大渕 佳祐、北口 大地、竹中 慎、竹下 修由、伊藤 雅昭
手術器具認識に基づく腹腔鏡下 S 状結腸切除術の自動手術技能評価, 第 124 回日本外科学会定期学術集会、2024 年 4 月・愛知県常滑市

要 旨

【背景と目的】

本研究は、腹腔鏡下 S 状結腸切除術（laparoscopic sigmoidectomy; Lap-S）の手術動画を対象に、作業分解手法（task decomposition approach）に基づく人工知能（artificial intelligence; AI）による自動手術技能評価モデルを開発し、その有用性を検証したものである。

従来、手術技能評価には OSATS（Objective Structured Assessment of Technical Skill）や GOALS（Global Operative Assessment of Laparoscopic Skills）などが広く用いられ、本邦では日本内視鏡外科学会が実施する技術認定制度（Endoscopic Surgical Skill Qualification System; ESSQS）が外科教育および技能認定の中核的役割を担ってきた。しかし、これらの評価法は熟練した評価者の主観に依存する部分が大きく、評価者間のばらつきや、評価の客観性・再現性・効率性に限界があると指摘されている。

一方、AI 技術の急速な進歩により、画像や動画から手術操作を自動認識・定量化する研究が国際的に進展している。特にコンピュータビジョンを応用した AI による手術解析は、術者の動作や鉗子操作、術野変化を高精度に識別することが可能となり、主観的評価を補完する新しい客観的技術評価法として注目されている。

本研究では、Lap-S 動画を用いて手術操作を剥離（Dissection: D）、展開（Exposure: E）、手術無効（Invalid: I）の 3 要素に分類し、手術習熟度を反映する AI 自動評価モデルを構築するとともに、外科教育に有用な定量的指標（パラメータ）を確立することを目的とした。

【対象と方法】

2016～2019 年度の ESSQS 審査用 Lap-S 動画 87 症例（+2SD 群 26 例、-2SD 群 31 例）および国内 8 施設から収集した初学者（Novice）群 30 例を解析対象とした。動画は 1 秒あたり 30 フレームの映像を 30 フレームごとに静止画として抽出し、AI によるフレーム単位の分類を実施した。

全手術動画を作業分解し、各フレームを剥離（D）、展開（E）、手術無効（I）の 3 要素に分類した。D 時間は電気メスや超音波凝固切開装置、自動縫合器、剥離鉗子などが出気にしている時間、E 時間は把持鉗子のみが出現している時間と定義した。I 時間は、カメラ体外時間（Outside）を除外し、鉗子が描出されず手術が進行していない（No grasper）時間のみを抽出した。これらの各時間（DEI 時間）および D 時間と E 時間の切替回数（DE transition）

を定量的指標として算出し、術者の熟練度との関連を解析した。

AIモデルは2種類構築した。① 手術鉗子8種類を自動認識し、各鉗子の出現状況から DEI を推定する Mask R-CNN (Region-based Convolutional Neural Network) ベースの鉗子認識モデル、② 手術画面全体を入力として鉗子情報に依存せず DEI を直接分類する Xception ベースの DEI 分類モデルである。

学習データはアノテーターおよび外科医 2 名によるダブルチェックのもと作成した教師データを用い、Precision、Recall、F1-score、および Pearson 相関係数で評価した。統計解析には EZR (Jichi Medical University Saitama Medical Center) を使用し、 $p < 0.05$ を統計学的有意とした。

【結果】

教師データ解析の結果、D 時間、E 時間、および DE transition は術者の習熟度を反映する指標であり、上級者になるにつれて D・E 時間の短縮および DE 切替回数の減少を示した。I 時間は中級者から上級者になる過程で初めて有意に短縮した。

鉗子認識モデルの全体 F1-score は 0.70 であり、DE 時間および DE transition は教師データと同様の傾向を示したが、Novice 群でばらつきが大きく、I 時間を過大推定する傾向を認めた。工程別解析では、Lap-S を 9 工程に分類したところ、内側授動 (Phase2)、外側授動 (Phase6)、直腸切離～吻合 (Phase9) において手術習熟度に有意差を認めた。

一方、画面全体を学習する DEI 分類モデルでは、F1-score 0.88、Pearson 相関係数 0.9 以上と高い精度を示した。Action recognition performance による視覚比較でも誤分類が少なく、教師データとの一致度が高かった。

【考察】

DEI 時間および DE transition は、術者の熟練度を反映しうる新たな客観的指標と考えられた。新規評価項目として DE transition を本研究で初めて提唱し、D と E の変遷の程度が手術評価へ有用であることが示唆された。上級者では 1 回の展開における剥離時間が長く、結果として DE の切替回数が減少することを示し、手術習熟に伴う作業構造の特徴を定量的に捉え得る指標であると推察された。

鉗子認識モデルに比して、DEI 分類モデルは高精度かつ汎用性に優れ、手術動画全体を解析対象とすることで安定した性能を示した。鉗子認識モデルの精度が低い要因として、未学習鉗子の存在や映像のブレ、鉗子のフレーム

アウトが主な要因と推定された。工程別解析の結果、内側・外側授動および直腸切離～吻合の工程が習熟度評価上で統計学的有意差がつきやすく、外科教育における重点学習領域を定量的に可視化できることが示唆された。

AI による動画解析は、主観的評価法を補完する客観的評価基盤として、外科技術標準化および教育効率化に寄与する可能性がある。今後、リアルタイム動画解析技術を応用することで、術中フィードバック型教育支援や、個々の術者の技能推移（learning curve）の自動可視化が可能になると考えられる。

【結論】

作業分解手法を用いた Lap-S における自動手術技術評価において、D 時間、E 時間、I 時間、そして本研究で新たに導入した DE transition が手術習熟度評価に有用なパラメータであった。また自動評価に関して、画面全体を教師データに用いた DEI 分類モデルは高精度に DEI を判別し、自動手術評価が可能であった。

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
D	dissection
E	exposure
I	Invalid
IMA	inferior mesenteric vein
IMV	inferior mesenteric artery
JSES	The Japan Society for Endoscopic Surgery
LapS	Laparoscopic sigmoidectomy
LCA	left colic artery
mAP	mean average precision
Mask R CNN	Mask Region-based Convolutional Neural Network

緒 言

1. 大腸癌の疫学、外科医を取り巻く環境の変化

日本内視鏡外科学会（The Japan Society for Endoscopic Surgery; JSES）の全国調査（Update on endoscopic surgery in Japan: Results of the 16th National Survey of endoscopic surgery）（Shiroshita et al, 2024）によれば、腹部領域では 2021 年の 1 年間に約 124,000 件の腹腔鏡手術が実施されている。腹腔鏡手術は、低侵襲性、整容性、術後回復の早さといった利点から急速に普及した。近年ではロボット支援下手術が保険収載され、前立腺・胃・直腸など多領域で導入が進んでいるものの、腹部領域では依然として腹腔鏡手術が主流である。とくに大腸癌は本邦で最も罹患数の多い悪性新生物であり、2020 年の全国がん登録では新規罹患患者数が約 16 万人、死亡者数が約 5 万人と報告されている（がん情報サービス（ganjoho.jp））。高齢化の進行および食生活の欧米化を背景に、今後も罹患率は上昇することが予測されている。

大腸癌に対する標準治療の中核をなすのが手術療法であり、腹腔鏡手術の普及によって低侵襲かつ安全な治療が広く行われるようになった。一方で、高齢者の増加により合併症を有する症例が増加し、術前評価・リスク管理・手技選択の複雑化が進んでいる。

しかしながら外科医の減少は深刻な社会的問題である。厚生労働省の「医師・歯科医師・薬剤師調査」によると、2002 年に比べ 2022 年の外科医数は約 20% 減少しており（医師・歯科医師・薬剤師統計の概況）、特に若手外科医の減少が顕著である（図 1）。日本消化器外科学会の会員数も 2011 年をピークに減少傾向を示し、日本消化器外科学会の推計では 2043 年には外科医数が現在の約半数である約 8,000 人にとどまると予測されている。外科は労働負担の大きさや家庭生活との両立困難を理由に敬遠され、さらに医師の都市集中により特に地方医療圏での人材不足が顕在化している。今後の外科医療を維持するには、限られた人材の中で効率的かつ効果的に外科技術を継承・発展させていく必要がある。

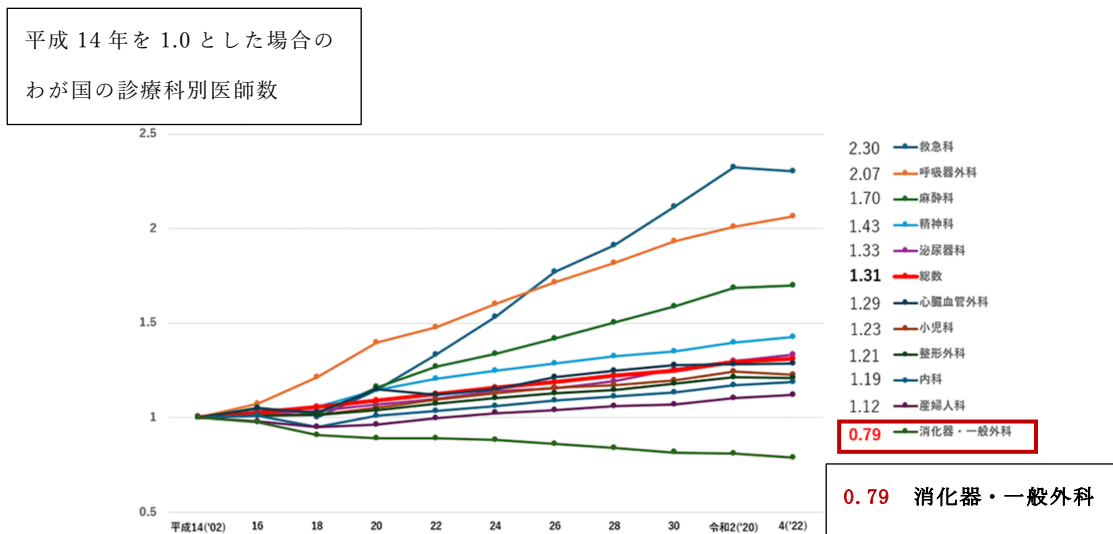


図 1 診療科別医師数推移グラフ 日本消化器外科学会
(<https://www.jsogs.or.jp/news/citizen/1914/>)

2. 外科教育と手術評価、人工知能を用いた自動手術評価

現在主流である腹腔鏡手術やロボット手術は、従来の開腹手術に比べて視野制限や鉗子操作の制約が大きく、空間認識能力・精緻な両手操作・チーム連携といった複合的スキルが求められる。これらの手技の習得には多大な訓練時間を要するため、教育効率が課題となっている。さらに、近年の働き方改革の進展により、医師の時間外労働に上限規制が導入されるなど、教育の現場にも大きな影響を及ぼし、限られた時間の中で効率的な教育体制を構築することが求められている。

手術習熟度に関しては術後合併症発生率 (Park et al, 2009) や術後入院期間 (Schlachta et al, 2001)、医療費 (Park et al, 2007)、さらには予後 (Birkmeyer et al, 2013; Stulberg et al, 2020) といった臨床成績に関係することが報告されてきた。すなわち、外科技術の向上は単に教育上の課題にとどまらず、患者アウトカムに直結する重要な医療課題である。教育学の観点からも、学習効果を最大化するために、適切なフィードバックが有効であると考えられている (木村ほか, 2023)。

近年腹腔鏡手術の普及により、手術動画を定量的データとして取り扱うことが可能となり、手術技術評価の標準化が進展した。従来から、objective

structured assessment of technical skill; OSATS (Martin et al, 1997) や Global Operative Assessment of Laparoscopic Skills; GOALS (Vassiliou et al, 2005) といった評価ツールが考案され、外科医の技能を構造化して評価する試みが行われてきた。OSATS は外科技術をステップごとに細分化して客観的に評価する方法であり、GOALS は「奥行き認識」「両手協調性」「効率性」「組織の取り扱い」「主体性」の 5 項目を 5 段階で採点するスコアリングシステムである。これらの評価法は教育現場で広く用いられ、手技の標準化や教育効果の検証に寄与しており、多くの術式でその有用性が報告されている (Humm et al, 2022; Schmidt et al, 2022; Vaillancourt et al, 2011; van Zwieten et al, 2022)。

JSES では、腹腔鏡手術の質を担保するための指導医資格として技術認定制度 (Endoscopic Surgical Skill Qualification System; ESSQS) が設けられている (Kimura et al, 2010; Mori et al, 2010)。2004 年に本格運用が開始され、大腸領域では主に S 状結腸切除または高位前方切除術を対象に評価が行われる。審査は提出動画をもとに 2 名の審査員が盲検下で 100 点満点評価を行い、一定基準を満たす術者に資格を付与する方式である。2024 年度の大腸領域合格率は 31%であり、肝臓・食道分野では 10~20%台とさらに難易度が高く (日本内視鏡外科学会)、本制度は外科医にとって最難関資格の一つである。この有資格者は非資格者と比較して術後短期成績が良好であることが報告されており (Ichikawa et al, 2020; Iwamoto et al, 2023; Kiyozumi et al, 2023)、本制度は腹腔鏡手術の質保証および教育評価の標準として重要な役割を果たしている。

しかしながら、ESSQS を含む手術評価にはいくつかの課題が挙げられる。まず評価が評価者の主観に依存しており、評価者間のばらつきが生じやすい点である。次に動画評価には多大な時間と労力を要し、評価者の負担が大きいことから、働き方改革の観点からも持続可能性に課題がある。さらに評価頻度を高めることが難しいため、教育的フィードバックの即時性にも制約がある。これらの理由から、主観的評価に依存しない客観的・定量的な技能評価法の確立が求められている。

このような背景のもと、人工知能 (artificial intelligence; AI) を用いた自動手術評価システムの応用が注目されている。AI はコンピューターシステムが人間の知的活動を模倣する能力を示す概念として認識され、情報処理技術を基盤として、知覚や推論、学習、問題解決など、人間が通常行う知的プロセスを実現することを目的としている。AI の定義は、機械学習を行い、データからパターンを学習し、それを未知のデータに当てはめることで予測

や意思決定を行うアルゴリズムの集合を指す。さらに機械学習の一分野として深層学習があり、脳構造のように多層のニューラルネットワークを形成して、高次元かつ複雑な特徴を自動解析して抽出する能力を示す。主に深層学習を用いた AI は産業・研究領域を超えて社会全体に急速に浸透し、医療に限らず交通・通信など多方面で実用化が進んでいる。

たとえば、Google の子会社である Waymo 社は 2018 年に完全自動運転タクシー「Waymo One」を商用運用し、都市交通における AI 活用の象徴的事例となった (Hawkins, 2023)。またスマートフォンにおいても AI 搭載機種が一般化し、自然言語処理技術を応用した ChatGPT (OpenAI) や Gemini (Google DeepMind) などの生成系 AI が、ビジネスから日常生活に至るまで様々な分野に広く利用されている (Daniel Rock, 2023; McAfee, 2017)。AI は今や社会の基盤的技術として位置付けられ (Ajay Agrawal, 2022)、その応用範囲は拡大の一途をたどっている。医療分野においても AI の導入は急速に進展しており、1970 年代に INTERNIST-1 と呼ばれる内科診断システムに初めて AI が導入されて以降、内科診断画像診断支援や病理画像解析、さらには予後予測モデルなどで有用性が示されている (Hosny et al, 2018; Komura & Ishikawa, 2018)。手術分野でも AI 研究が進められており (図 2)、外科手術視覚支援システムである EURIKATM α は、術中リアルタイムに粗性結合組織の位置や領域を協調表示することで重要な解剖学的構造の把握に有用である。その結果外科医の剥離層の認識を補助することで、手術精度の向上および若手外科医に対する教育を目的として、手術支援システムとして薬事承認された。同様に SurVis-Hys は、Jmees 社と国立がん研究センター東病院が共同開発した内視鏡手術支援プログラムで、内視鏡下子宮全摘術において、尿管や膀胱を AI でリアルタイムに協調表示を行う。これにより解剖学的構造の認識率が向上して、臓器損傷リスクの軽減につながる事が期待されている。

さらに近年では解剖学的認識に留まらず、手術動画や術者動作を解析して技能評価や自動スコアリングを行う試みが進んでいる (Hashimoto et al, 2018)。これにより評価の客観性や効率性の向上、教育現場での即時フィードバック実現など、新たな教育体系の構築が期待されている。

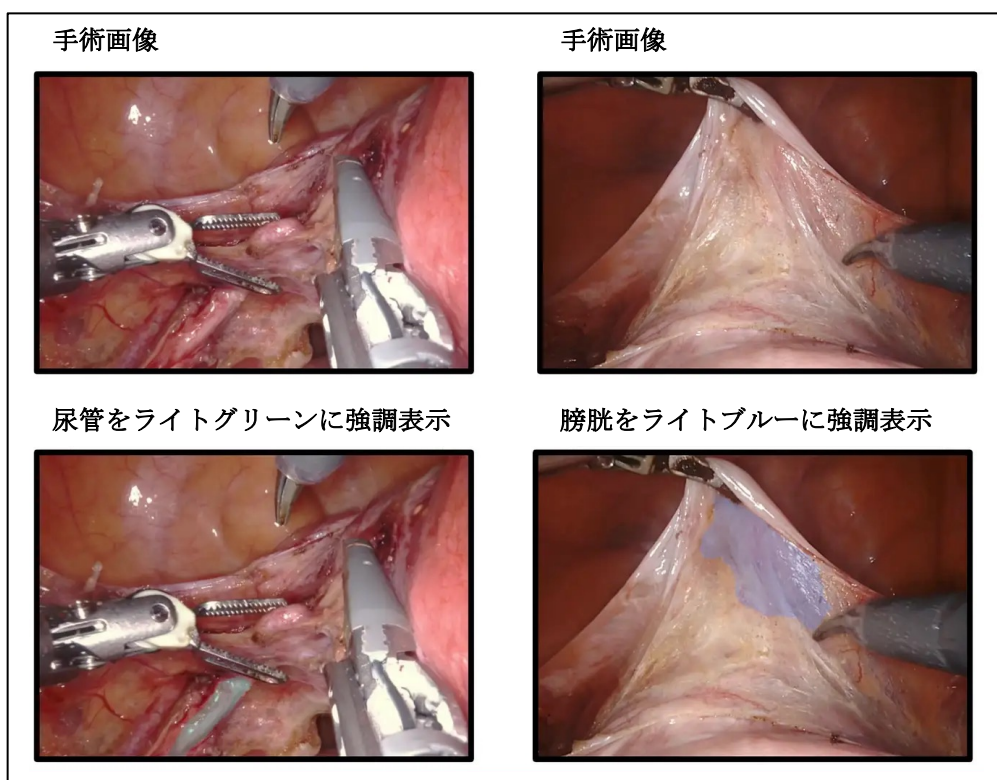
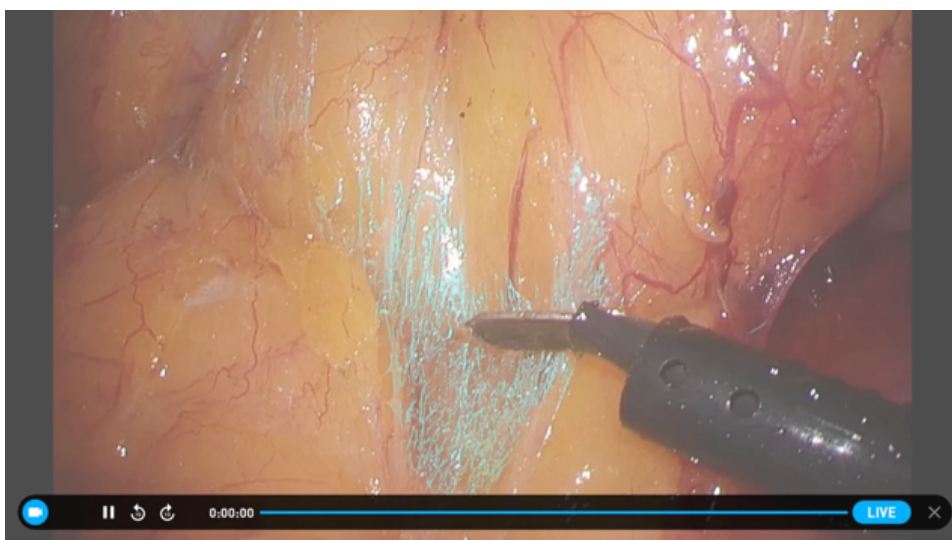


図 2 手術支援システム

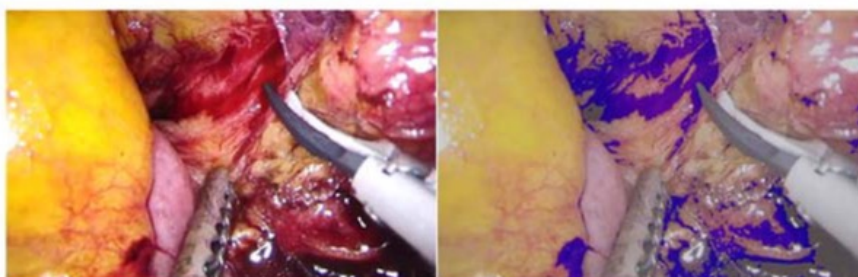
上 Anaut 社 EURIKA α (https://anaut-surg.com/eureka_alpha)
結合組織をライトグリーンで表示する。

下 Jmees 社 SurVis-Hys (<https://www.jmees-inc.com>)
尿管をライトグリーンに、膀胱をライトブルーに表示する。

国立がん研究センター東病院 NEXT 医療機器開発推進室では、JSES と共同で、腹腔鏡下 S 状結腸切除術（Laparoscopic sigmoidectomy; Lap-S）における ESSQS スコアと手術動画を用いた自動手術評価の研究開発を行い（Igaki et al, 2023a; Nakajima et al, 2024; Nakajima et al, 2025; Sasaki et al, 2023）、自動評価可能な複数のパラメータを抽出する取り組みを行っている。これまでに井垣らは手術全体を 9 工程に分類し、AI による工程認識の確信度（AI が各工程をどの程度の自信度で分類できたかの指標）をもとに習熟度の判別を行った（Igaki et al, 2023a）。また図 3 に示すように出血ピクセルを自動検知する AI では、ピクセル量と鉗子操作性の点数が逆相関することが明らかになった（Sasaki et al, 2023）。これらの試みは、AI による手術手技の客観的・定量的評価の可能性を示しており、今後の外科教育の在り方を大きく変革する可能性がある。



6,935 blood pixels



101,311 blood pixels

図 3 AI で検出された出血ピクセルの実際（Sasaki et al, 2023）

左は実際の手術場面で、右は AI が検出した出血ピクセルを青色で示す。上段は出血が少なく綺麗な視野で解剖認識が容易であるが、下段の症例では出血箇所が目立つ。

3. 作業分解手法について

本研究ではその手術自動評価のパラメータ探索の一環として、Miskovic らが提唱した作業分解手法に着目した (Miskovic et al, 2012)。手術は手術工程を電気メスや超音波凝固装置などが出現する剥離 (Dissection; D) 時間、組織を剥離するための前段階である展開 (Exposure; E) 時間から構成される。Lap-S においては S 状結腸および下行結腸を外側の腹膜から剥離する外側授動と呼ばれる工程において、D/E 比 (D 時間と E 時間の比) が手術効率を反映し、手術習熟度と関連することが示唆された (図 4)。手術上級者 (Expert) において、DE 比が他の 2 群と比較して高値であることが示された。

一般的に手術時間は手術習熟によって短縮していくことから、DE 比に加えて、これまで外科の暗黙知とされてきた習熟度に応じた D 時間・E 時間短縮の特徴を定量的に明らかにすることで、新たな手術評価の手法を確立できると考えられた。また自動手術評価を目的にこれらのパラメータを自動認識可能な AI モデルの開発を試みた。

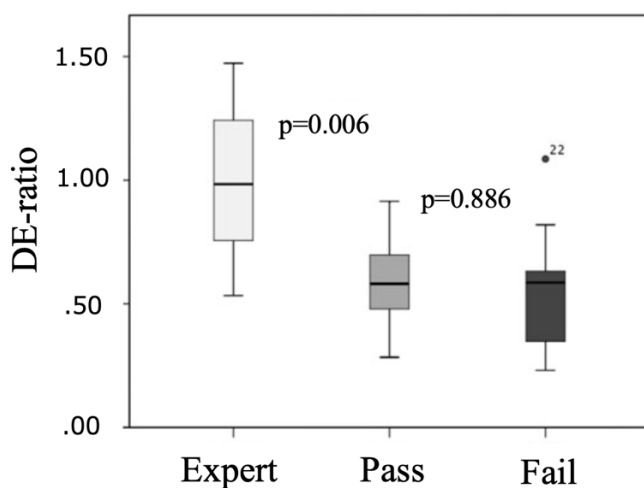


図 4 Lap-S における DE 比の比較
(Miskovic et al, 2012) より引用、一部改変

4. 本研究の目的

本研究の目的は、腹腔鏡下 S 状結腸切除術 (Lap-S) の手術動画に対して

作業分解手法に基づく DEI パラメータを手術習熟度別に解析して、手術評価可能なパラメータの検索を行うことである。また D 時間と E 時間を自動認識する AI モデルを構築し、自動手術評価の実現可能性を検討した。

対 象 と 方 法

1. 研究デザイン

本研究は、JSES が実施する内視鏡技術認定制度（ESSQS）に提出された Lap-S の手術動画を対象とした後ろ向き研究である。解析に用いたビデオデータは、診療の過程で記録された既存の手術動画であり、本研究の実施にあたり新たな侵襲や介入を加えるものではない。データの収集および研究利用に関しては、研究対象者全員から事前に文書による十分な説明を行い、同意を取得している。手術動画には、患者および術者個人を直接特定できる情報（氏名、生年月日、患者 ID 等）は含まれておらず、研究用データとして管理する際には匿名化处理が施されている。

すなわち、本研究は適切な説明同意のもとで収集された既存のビデオデータを用いた侵襲を伴わない研究であり、研究対象者に新たな不利益やリスクを生じるものではない。研究の実施にあたっては、国立がん研究センター東病院倫理委員会の審査を受け、承認を得たうえで実施した（承認番号：2020-315）。

2. データセット

本研究の対象は、2016 年から 2019 年に日本内視鏡外科学会（JSES）の技術認定制度（Endoscopic Surgical Skill Qualification System: ESSQS）に提出された腹腔鏡下 S 状結腸切除術（Lap-S）の審査結果および手術動画である。得点分布に基づき、全受験者を平均値に対して +2 標準偏差（SD）以上の高得点（+2SD）群および -2 標準偏差以下の低得点群（-2SD 群）の 2 群に分類した（図 5）。

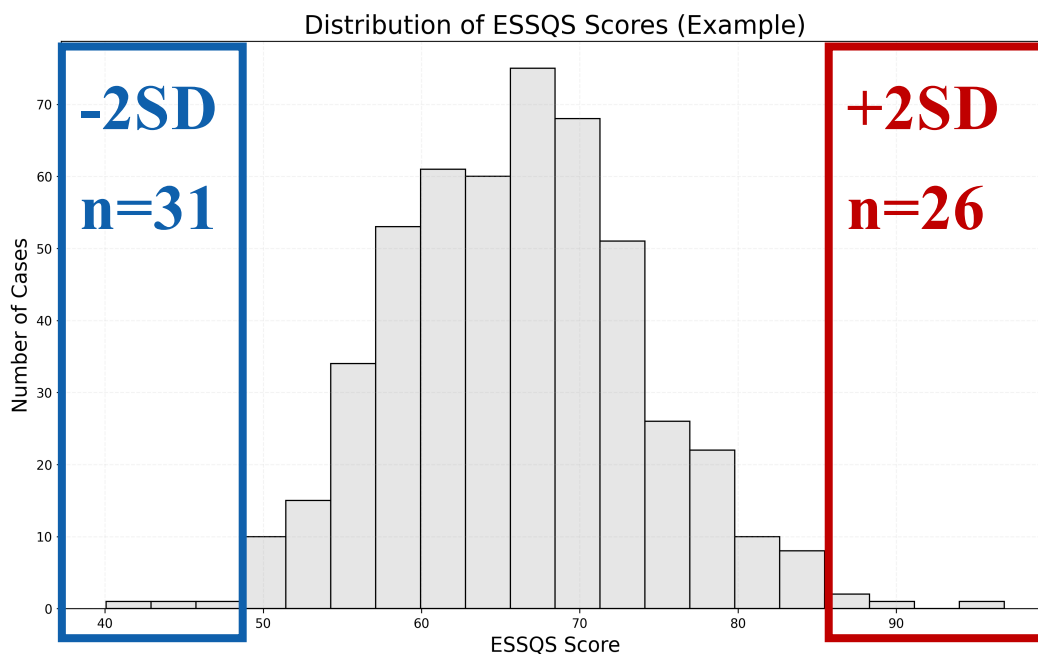


図 5 ESSQS 得点分布図と+2SD 群および-2SD 群のイメージ

技術認定資格は、日本外科学会専門医資格を有し、過去 3 年間に食道切除、胃切除、大腸切除などの高難度手術を 20 件以上施行している外科医を対象とする。審査は、2 名の認定審査員が匿名化された手術動画を評価し、日本内視鏡外科学会が定める評価基準に基づいて採点する。両審査員の評価がいずれも 100 点満点中 70 点以上の場合に認定される。審査項目は、①共通評価項目（基本的な内視鏡操作技能を評価）と、②臓器・術式別評価項目（対象臓器に特化した操作技能を評価）の 2 カテゴリーで構成される。大腸領域の採点表をこの論文の最後に付録として掲載した（日本内視鏡外科学会）。応募資格には一定の手術経験が明示されており、受験者の技量は少なくとも中級レベル以上に達していると判断される。さらに、比較対象として初学者（Novice）群を設定した。Lap-S の執刀経験が 5 例未満の外科修練医を対象とし、日本国内 8 施設から 30 症例の Lap-S 手術動画を収集した。

解析に際しては、腹腔鏡下 S 状結腸切除術の手術動画を機械学習モデルに入力できる形式に変換するためのアルゴリズムを作成した。具体的には、1 秒あたり 30 フレームの動画から 1 秒ごとに 1 フレームを抽出し、手術時間（秒）と同数の静止画を解析用データとして使用した。

3. 腹腔鏡下 S 状結腸切除術について

腹腔鏡下 S 状結腸切除術の工程を表 1 に示す。主に 9 工程 (phase) に分類することができる (Kitaguchi et al, 2020)。臍部に小切開をおき、カメラポートを留置し、腹腔内を観察後、臍の上下左右に 4 本のポートを挿入して 5 ポートで手術を行う (図 6)。砕石位で頭低位、右下ローテーションとして、小腸を右上腹部に受動する。その後、図 7 に提示するように直腸間膜付着部右側を衝立状に展開して、腹膜切開を行い、直腸固有筋膜を同定する TME (total mesorectal excision) と呼ばれる操作を行う (Phase1)。剥離層を保ちながら、岬角上に剥離ラインをつなげ、下腸間膜動脈 IMA (Inferior mesenteric artery; IMA) を同定する (Phase2)。IMA 周囲の腰内臓神経腸管枝を切離して IMA を露出する。進行癌症例であれば根部で切離して、周囲脂肪組織を #253 として切除側につける (Phase3)。その後外側および尾側に向かって、尿管下腹神経前 (Gerota) 筋膜と呼ばれる後腹膜成分を背側に授動し、尿管および性腺血管を剥離する (Phase4)。続いて下腸間膜静脈 IMV (Inferior mesenteric vein; IMV) ならびに左結腸動脈 LCA (Left colic artery; LCA) を切離する (Phase5)。展開を変えて外側授動を行う。ここでは下行結腸と S 状結腸移行部の腹壁との生理的癒着を外したのち、腹膜を切開して、内側授動との剥離層を交通させることで結腸授動が終了する (Phase6)。続いて直腸間膜左側の剥離を行い、Phase1 の剥離と交通させる (Phase7)。腫瘍占拠部位を確認し、十分な肛門側マージンを確保し、直腸間膜切離ラインを定める。直腸を損傷しないように注意して直腸間膜を切離する (Phase8)。その後自動縫合器を用いて直腸を切離し、検体を臍部から体腔外操作で摘出する。再気腹して経肛門的に自動吻合器を挿入して、結腸と直腸を吻合するまでの操作を Phase9 と定義する。

表 1 腹腔鏡下 S 状結腸切除術の工程

Phase	定義
Phase	TME（右側/後側）
Phase2	結腸の内側授動（IMA 切離前まで）
Phase3	IMA 切離
Phase4	IMA 切離後の内側授動
Phase5	IMV/LCA 切離
Phase6	外側授動
Phase7	TME（左側）
Phase8	直腸間膜切離
Phase9	直腸切離～吻合

TME; total mesorectal excision

IMA; inferior mesenteric artery

IMV; inferior mesenteric vein

LCA; left colic artery

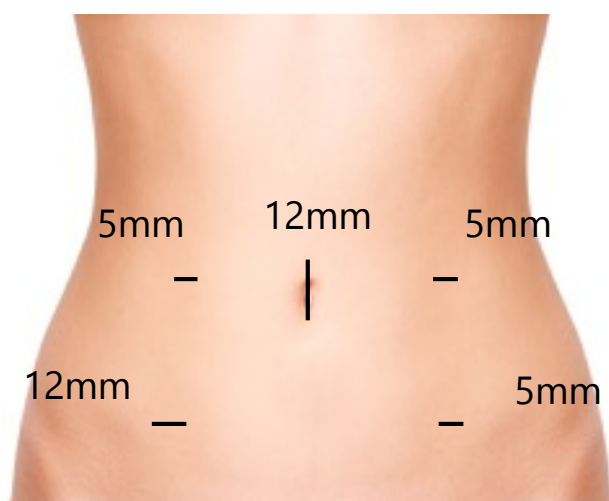


図 6 腹腔鏡下 S 状結腸切除術におけるポート配置

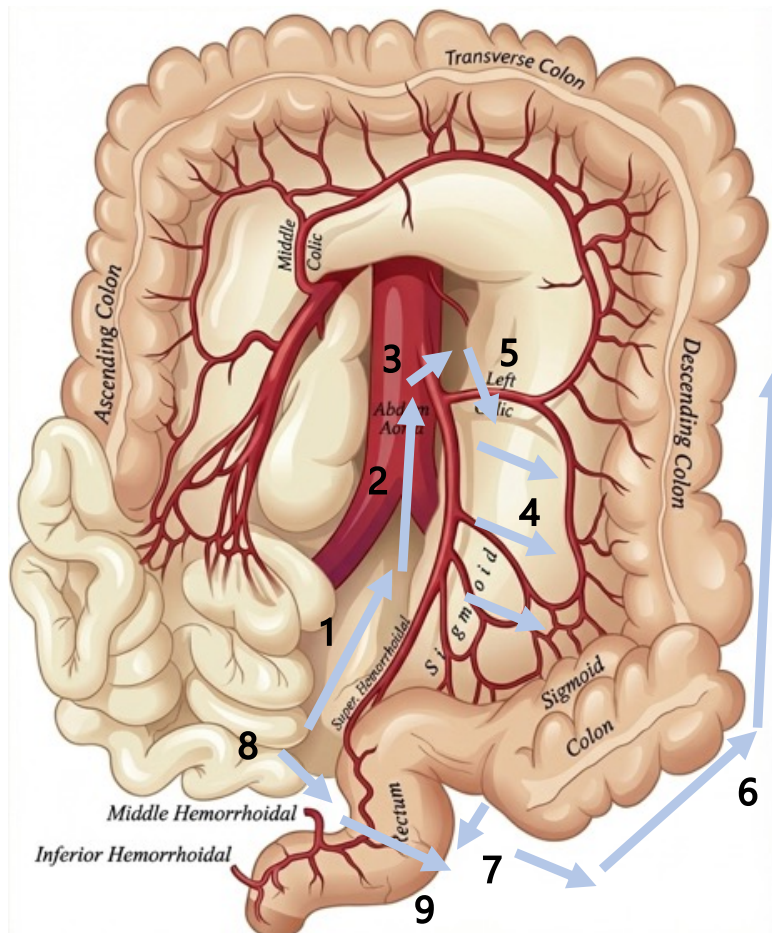


図 7 腹腔鏡下 S 状結腸切除術の解剖（図は Chat GPT 作成）
数字は表 1 の Phase1-9 に対応する。

4. パラメータの設定

本研究では、手術全体を剥離（Dissection: D）時間、展開（Exposure: E）時間、および手術無効（Invalid: I）時間の 3 つに分類した。Miskovic らの報告（Miskovic et al, 2012）では、手術を D 時間と E 時間の 2 つに区分している。しかし、実際の手術では、鉗子が描出されておらず手術操作が進行していない（No grasper）時間や、カメラのクリーニングなどでカメラが体外にある（Outside）時間が一定割合含まれる。そこで本研究では、これらの非操作時間を考慮し、I 時間を No grasper と Outside に細分化した（図 8）。これらの定義に基づき、解析に用いる各パラメータを設定した（表 2）。

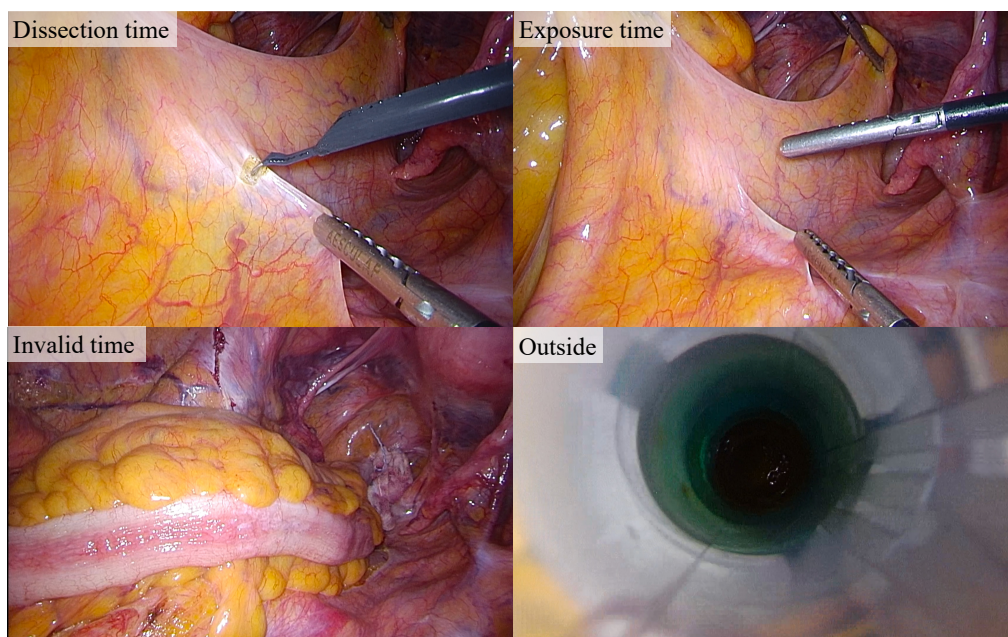


図 8 腹腔鏡下 S 状結腸切除（Lap-S）における DEI 時間及び Outside 時間の例

表 2 パラメータの定義

パラメータの名称	定義
Dissection (D) 時間	電気メスや超音波凝固切開装置、自動縫合器などで組織を剥離している時間
Exposure (E) 時間	組織把持鉗子で組織を展開している時間、止血操作を含む
Invalid (I) 時間	カメラが体腔外にある時間（Outside）を除いた、手術操作が行われていない時間
DE transition	D 時間と E 時間の切替回数
D/E 比	D 時間と E 時間の比
D/（E+I）比	D 時間と E+I 時間の比
E20s/30s	20 秒以上および 30 秒以上を要した E 時間の出現時間の総和

D 時間の定義は、電気メス、超音波凝固切開装置、剥離鉗子、自動縫合器などを用いて組織を剥離している時間とした。E 時間の定義は、術者または

助手が組織把持鉗子を用いて組織を展開している時間とした。止血操作は、手術進行に直接寄与する操作ではあるが、剥離操作には該当しないため E 時間に含めた。I 時間は、何も操作が行われていない (No grasper) 時間とカメラが体腔外にある (Outside) 時間を合わせたものである。ただし、ESSQS から収集された動画の中には、編集段階で体腔外操作時間が削除されている症例が含まれていた。そのため、本研究では公平性を担保する目的で、Outside 時間を工程分類モデル (Kitaguchi et al, 2020) を用いて除外し、No grasper 時間のみを I 時間として定義した。さらにこれらの分類に基づき、以下のパラメータを解析に用いた。

手術習熟度が高いほど、1 回あたりの E 時間の後に続く D 時間が長くなる傾向があり、したがって D 時間と E 時間を行き来する回数 (DE transition) は減少すると考えられる。この仮説に基づき、DE transition を技能評価の指標として検討した。次に先行研究で解析されていた D/E 比および、D/E 比から派生させた D 時間を E 時間と I 時間の和で割った $D/(E + I)$ 比を算出した。また 20 秒以上および 30 秒以上を要した E 時間の出現時間の総和を解析に含め、展開操作の効率性を評価した。

5. 教師データ (Ground truth)

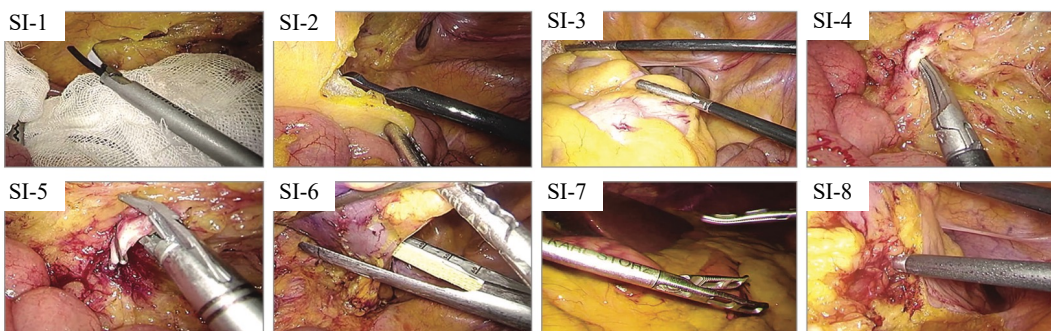
データセットで作成した静止画について、国立がん研究センター東病院医療機器開発センターに従事する専従のアノテーター複数名および同研究室で AI 研究に従事する消化器外科専門医を有する消化器外科医師 2 名が D、E、I 時間に分類し、ダブルチェックを行って教師データを作成した。そのラベル付けされたフレーム数を用いて算出された DEI 時間から各パラメータの値を求めた。これらのデータを真の正解 (Ground truth) として定義した。

6. 鉗子認識モデル

本研究では、手術鉗子を自動認識する AI モデルと、画面全体を操作段階 (DEI) に分類する AI モデルの 2 種類の深層学習モデルを解析した。鉗子認識モデルには、以前に報告した AI モデル (Kitaguchi et al, 2022) を用いた。このモデルは、超音波凝固切開装置、電気メス、把持鉗子、剥離鉗子な

ど代表的な 8 種類の内視鏡手術鉗子を自動認識するものである（図 9）。

Surgical instruments



Annotation

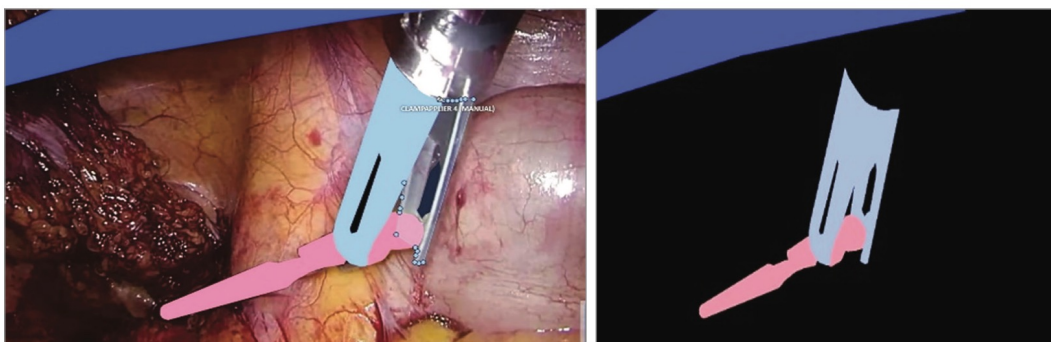


図 9 鉗子認識モデルの実際（Kitaguchi et al, 2022）

上段：AI で認識させて術具を示す。SI-1：超音波凝固切開装置、SI-2：モノポーラ型電気メス、SI-3：波型無傷把持鉗子、SI-4：剥離鉗子、SI-5：クリップ鉗子、SI-6：自動縫合器、SI-7：有窓無傷鉗子（クローチェ）、SI-8：ソフト凝固（送水吸引システム）

下段左：アノテーション作業中の画像 下段右：アノテーション完成画像

この学習には、337 本の手術動画から抽出した 38,628 枚の手術静止画像を用いた。これらの画像には、計 81,760 箇所の鉗子領域がピクセル単位でアノテーションされている。このデータを基に、Mask Region-based Convolutional Neural Network（Mask R-CNN）（He et al, 2020）を用いて物体検出モデルを構築した。畳み込みニューラルネットワーク（Convolutional Neural Network: CNN）とは、画像や映像などの二次元データから自動的に特徴を抽出し、分類を行う深層学習モデルである（Li et al, 2022）。CNN は入力画像全体を対象として特徴を学習し、画像全体のカテゴリを識別する。

一方、Region-based CNN (R-CNN) (Girshick et al, 2014) は、画像内の関心領域 (region of interest: ROI) を抽出し、各領域を個別に分類することで、対象物の位置を特定できるよう拡張された手法である。さらに Mask R-CNN は、R-CNN にセグメンテーションブランチ (画像内の各ピクセルが何に属するかを識別する経路) を追加し、検出対象の輪郭をピクセル単位で抽出できるよう改良したモデルである。これにより、物体の位置に加えて形状情報を高精度に取得することができる。具体例として、CNN は画像全体を「猫」や「犬」といったカテゴリに分類するのに対し、R-CNN は対象物を矩形 (バウンディングボックス) で囲んで位置を特定し、Mask R-CNN はその輪郭をピクセルレベルで抽出できる。

学習の評価には 5-fold クロスバリデーション (表 3) を用い、データを 4 つの学習用モデルと 1 つの検証用モデルに分割して 5 回繰り返し、精度の平均値を算出した。

表 3 5-fold クロスバリデーション

	All data				
First validation	Train	Train	Train	Train	Validation
Second validation	Train	Train	Train	Validation	Train
Third validation	Train	Train	Validation	Train	Train
Fourth validation	Train	Validation	Train	Train	Train
Fifth validation	Validation	Train	Train	Train	Train

評価指標には物体検出で一般的に用いられる mean Average Precision (mAP) を使用した。mAP は、分類や検出タスクにおけるモデルの総合的な性能を評価する指標である。各クラスについて Average Precision (AP) を算出し、これは Precision-Recall 曲線 (再現率と適合率の関係) から得られる面積 (Area Under Curve) として定義される。mAP は全クラスの AP の平均値として次式で表される。

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

mAP は単一の閾値に依存せず、再現率全体にわたる性能を統合的に評価できる点で、物体検出タスクにおいて最も広く用いられている指標の一つである。本モデルの mAP は 0.919 であった（図 10）。

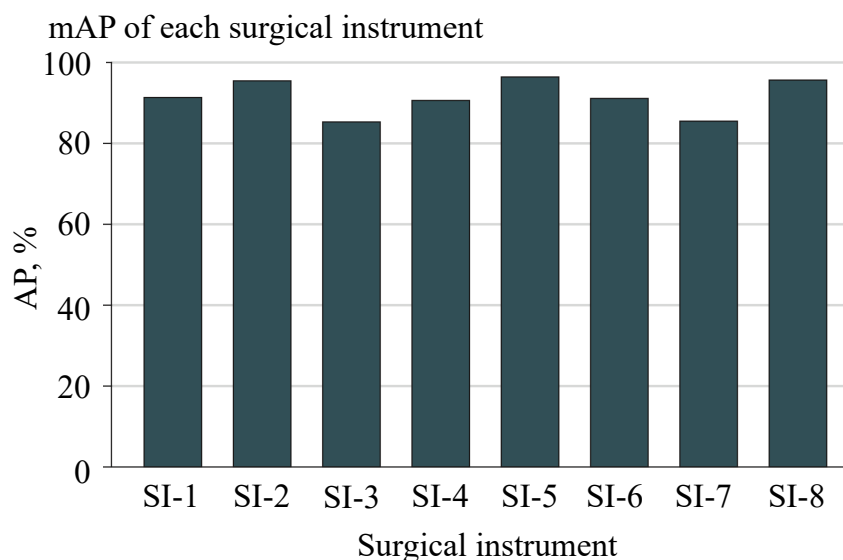


図 10 鉗子認識モデルの精度

横軸 SI 1-8 は、図 9 で示した鉗子の種類に対応する。SI 1-8 の AP の平均値が 0.919 と算出された（mAP）。

モデルの実装には Python 3.6 (Python Software Foundation) を使用した。AI 学習環境は、16 GB VRAM を搭載した Tesla T4 GPU (NVIDIA) および、48 GB RAM を有する Xeon Gold 6230 CPU (2.10 GHz, 20 cores; Intel Corp) を搭載したワークステーションを用いた。

本モデルで認識された鉗子の種類に基づき、剥離操作に関連する鉗子が出現している場合を D 時間、把持鉗子のみが出現している場合を E 時間、いずれの鉗子も検出されない場合を I 時間と自動分類し、この DEI 時間をもとに各種パラメータを自動算出した。

7. DEI 分類モデル

鉗子認識モデルでは、アノテーションされていない鉗子の存在により認識精度が低下する可能性が懸念された。そこで本研究では、新たなアプローチ

として、手術動画全体を対象に DEI 分類を行う AI モデルを開発した。本モデルは、個々の鉗子認識に依存せず、DEI にラベル付けされた教師データを用いて画面全体を学習データとして解析するものである。このモデルには、Xception (Chollet, 2017) と呼ばれる CNN を用いた。従来の CNN は、複雑な特徴を抽出する際に逐次的な畳み込み処理を行うため、層構造が深くなりやすく、計算量やパラメータ数の増大による過学習のリスクが指摘されている。この課題を解決するために、Inception (Christian et al, 2015) モデルが開発された。Inception は、同一層内で異なるサイズの畳み込みを並列に行う構造を採用することで、特徴抽出の多様性を高めつつ計算効率を改善した。さらに、Inception の構造を発展させ、畳み込みを空間方向とチャンネル方向に完全に分離して計算効率を極限まで高めたモデルが Xception である。Xception は、その高い表現力と計算効率により、近年の医用画像解析においても広く応用されている。本研究では、この Xception を基盤とし、データ拡張 (data augmentation) を行って DEI 分類モデルを構築した (表 4)。ただし、本研究ではアノテーション済みデータの制約により、新たな学習データの追加が困難であった。そのため、Novice 群の教師データを学習に用いたモデルとして構築したため、Novice 群を除く 2 群間での解析を行った。

コードの実装には Python 3.9 (Python Software Foundation) を使用した。AI 学習環境には、16 GB の VRAM を搭載した Tesla T4 GPU (NVIDIA) および、48 GB の RAM を有する Xeon Gold 6230 CPU (2.10 GHz、20 cores; Intel Corp) を搭載したワークステーションを使用した。

表 4 DEI 分類モデルの概要

Network	Xception
Optimizer	Radam (Learning rate 0.01)
Loss	Cross Entropy Loss
Minibatch Size	10
epoch	5
data augmentation	10% 平行移動 0.5-1.5 倍 拡大縮小 ±15 度回転

8. 統計解析

2 群間の比較には Welch 検定および Student の t 検定を用いた。3 群間の解析には、一元配置分散分析 (one-way ANOVA) および Kruskal–Wallis 検定を用いた。有意水準は両側検定で $P < 0.05$ を統計学的に有意と判定した。すべての統計解析は、EZR (自治医科大学附属さいたま医療センター) (Kanda, 2013) を用いて実施した。EZR は、R (The R Foundation for Statistical Computing, Vienna, Austria) のグラフィカルユーザーインターフェースである。本研究では、分類モデルの性能評価指標として Precision (適合率)、Recall (再現率)、および F1-score を用いた。これらの指標は、予測結果と真のラベルとの対応関係を示す混同行列 (confusion matrix) に基づき算出した。混同行列では、予測結果を真陽性 (True Positive: TP)、偽陽性 (False Positive: FP)、真陰性 (True Negative: TN)、および偽陰性 (False Negative: FN) の 4 区分に分類する。各指標は以下の式で定義される。

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Precision は、モデルが陽性と判定した事例のうち、実際に正しいものの割合を示す指標であり、「予測の正確さ」を表す。一方、Recall は、実際に陽性である事例のうち、モデルが正しく陽性と認識できた割合を示し、「見逃しの少なさ」を評価する指標である。F1-score は、Precision と Recall の調和平均であり、両者のバランスを総合的に評価する尺度として広く用いられている。これらの指標を併用することで、分類モデルの性能特性を多面的に定量化でき、モデルの臨床的有用性をより客観的に評価することが可能となる。また、他の AI モデル性能評価手法として、モデルの出力パラメータと教師データとの相関関係を Pearson 相関係数を用いて解析した。さらにモデルの精度を視覚的に評価する手法として、手術全体を作業時間別に色分けして表示した Action Recognition Performance と呼ばれる棒グラフを、教師データおよび 2 つの AI モデルについて作成して比較した。

これらの統計解析および性能評価指標を用いて、教師データに対する各

AI モデル（鉗子認識モデルおよび DEI 分類モデル）の分類精度と臨床的妥当性を検証した。本研究のアルゴリズムを図 11 に示す。

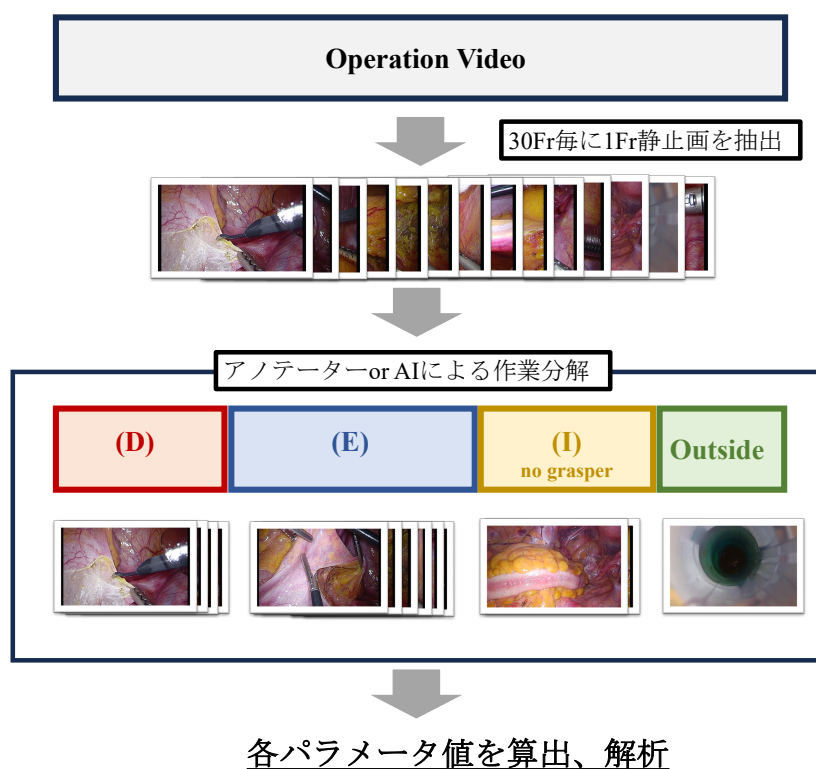


図 11 本研究のアルゴリズム

9. 本研究における筆者のタスク

本研究において筆者は、教師データ作成に際して手術動画のアノテーションおよびデータ品質の確認を担当した。AI モデルの実装（プログラミング）は研究室専任エンジニアが担当し、筆者はモデル構築に必要な要件定義・検討に参与した。解析で得られた各パラメータについて、筆者が統計学的解析および結果の解釈を行った。

結 果

1. データセット

本研究で解析対象とした腹腔鏡下 S 状結腸切除術 (Lap-S) 動画は計 87 症例であった。このうち、ESSQS スコアに基づく +2SD 群は 26 症例、-2SD 群は 31 症例であり、さらに Novice 群として 8 施設から収集された 30 症例を加えた。各年度 (2016~2019 年) ごとに ESSQS スコアの平均値および標準偏差 (SD) を算出し、 $\pm 2SD$ をカットオフ値として +2SD 群および -2SD 群を設定した。全体の平均得点は 66.6 点 (SD 8.4) であり、+2SD 群の平均得点は 85.5 点 (SD 2.7)、-2SD 群の平均得点は 43.2 点 (SD 5.4) であった。+2SD 群および -2SD 群の動画データは合計 386,721 フレーム (=秒数換算) で構成され、各時間区分の総時間は、D 時間 : 223,954 秒、E 時間 : 108,801 秒、I 時間 : 35,304 秒、Outside 時間 : 18,212 秒であった。一方、Novice 群の総解析時間は 358,759 秒であり、各区分の内訳は、D 時間 : 200,235 秒、E 時間 : 89,030 秒、I 時間 : 10,708 秒、Outside 時間 : 58,786 秒であった。

2. 教師データ (Ground Truth)

2-1. 教師データ結果

アノテーターおよび外科医による教師データの解析結果を図 12 に示す。D 時間の平均値は、それぞれ +2SD 群 : 3440 秒、-2SD 群 : 4339 秒、Novice 群 : 6674 秒であり、3 群間すべてにおいて統計学的に有意差を認めた ($p < 0.01$)。E 時間および I 時間の平均値は、それぞれ 1474 / 2273 / 3967 秒および 188 / 429 / 356 秒であった (いずれも $p < 0.01$)。E 時間についても D 時間と同様に 3 群間で有意差を認めたが、I 時間に関しては +2SD 群でのみ有意差を認め、-2SD 群と Novice 群との間では有意差を認めなかった。D/E 比では、3 群間に統計学的有意差を認めなかった ($p = 0.15$)。また、分母に I 時間を加えた D/(E+I) 比についても、D/E 比と類似した結果であり、有意差は認められなかった。DE transition の平均値は、+2SD 群 : 48 ± 19 回、-2SD 群 : 73 ± 29 回、Novice 群 : 116 ± 56 回 ($\pm SD$) であり、いずれの群間でも有意差を認めた ($p < 0.01$)。さらに、20 秒以上および 30 秒以上継続した E 時間を抽出して解析したところ、20 秒以上の条件では 3 群間に有

意差を認めなかったが、30 秒以上の条件では Novice 群と+2SD 群の間に有意差を認めた ($p < 0.05$)。

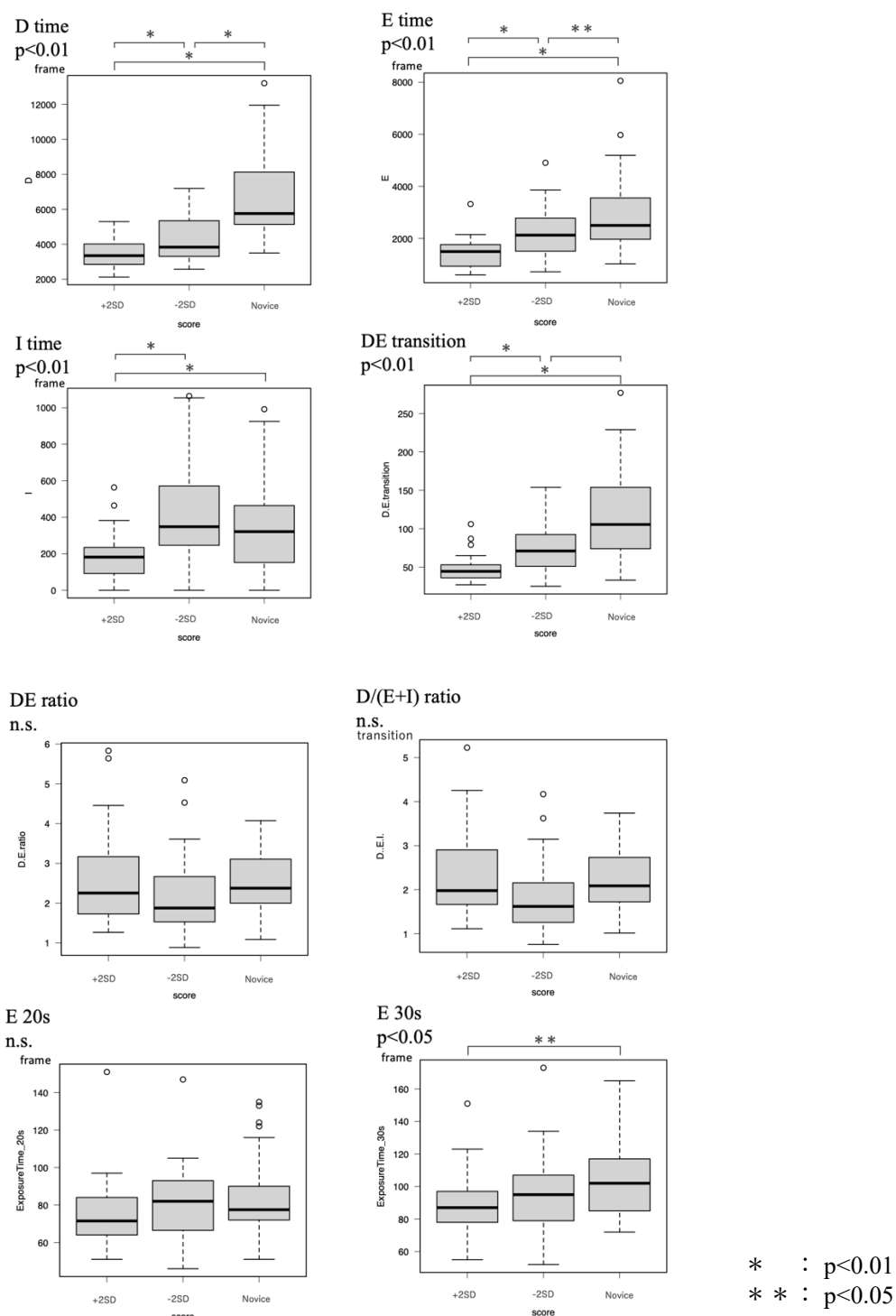


図 12 教師データにおける 3 群間での各パラメータの比較
 DEI 時間および 20s、30s の縦軸はフレーム数 (=秒数) を表す。2 群間の比較で有意差を認めたものについて、グラフ上部に提示した。

2-2. 教師データ小括

教師データの解析結果では、D 時間、E 時間、および DE transition のいずれにおいても、3 群間で統計学的に有意差が認められた。一方 I 時間に関しては、+2SD 群においてのみ有意差が示され、-2SD 群と Novice 群との間では有意差を認めなかった。Novice 群では、ハイボリュームセンターにおける症例が多く、指導的助手の支援下で手術が実施されていることから、特に I 時間においては-2SD 群との間で差が生じにくいと考えられた。また、D/E 比および D/ (E+I) 比については、先行研究と同様の傾向は得られず、3 群間に統計学的有意差を認めなかった。さらに、30 秒以上継続した E 時間 (E_30s) では、Novice 群と-2SD 群との間で有意差を認めたものの、E 時間そのものと比較して解析能は低かった。教師データの結果から DEI 時間および DE transition が手術習熟度の解析に有用であり、以降の検討ではこれら 4 パラメータを AI モデルの解析に用いた。

3. 鉗子認識モデル結果

3-1. 鉗子認識モデルの精度

表 5 に鉗子認識モデルの精度を示す。全体の精度は F1 スコアで 0.70 であった。Outside 時間は高い精度で予測可能であったが、I 時間については他の手技時間と比較してとくに適合率 (precision) の精度が低く、誤認識 (false positive) が多い結果となった。

表 5 鉗子認識モデルの精度 (教師データとの比較)

Parameter	Precision	Recall	F1-score	Support
D	0.96	0.63	0.76	434910
E	0.57	0.70	0.63	211162
I	0.23	0.87	0.36	38439
Outside	0.99	0.94	0.97	95532
Accuracy			0.70	780043
macro avg	0.69	0.78	0.68	780043
Weight avg	0.82	0.7	0.73	780043

また、精度を表す別の手法として散布図 (Scatter plot) を用いた (図 13)。横軸を教師データ、縦軸を鉗子認識モデルの数値 (いずれも単位はフレーム数) として、DEI 時間における Pearson 相関係数を算出した。とくに E や I 時間において教師データと AI モデルの数値に強い正の相関を認めた。

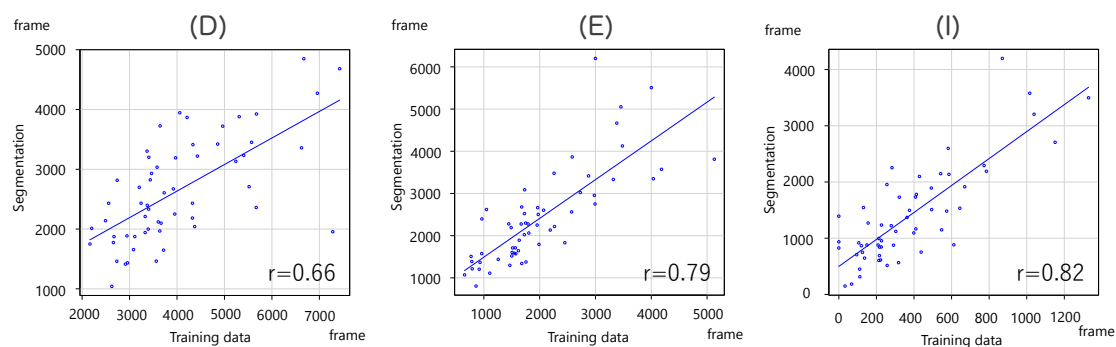


図 13 Pearson 相関係数を用いた鉗子認識モデルの精度
横軸：実際の DEI 時間、縦軸：AI が解析した DEI 時間

3-2. 鉗子認識モデルと教師モデルの各パラメータの比較

鉗子認識モデルを用いて算出した DEI 時間をもとに各パラメータを算出し、3 群間で比較を行った (図 14)。教師データの結果と比較すると、D 時間および DE transition では、-2SD 群と Novice 群との間に統計学的有意差は認められなかったものの、全体として類似した傾向を示した。E 時間についても、教師データと同様の傾向を示した。一方、I 時間はすべての群でフレーム数が増加し、特に Novice 群において顕著であった。

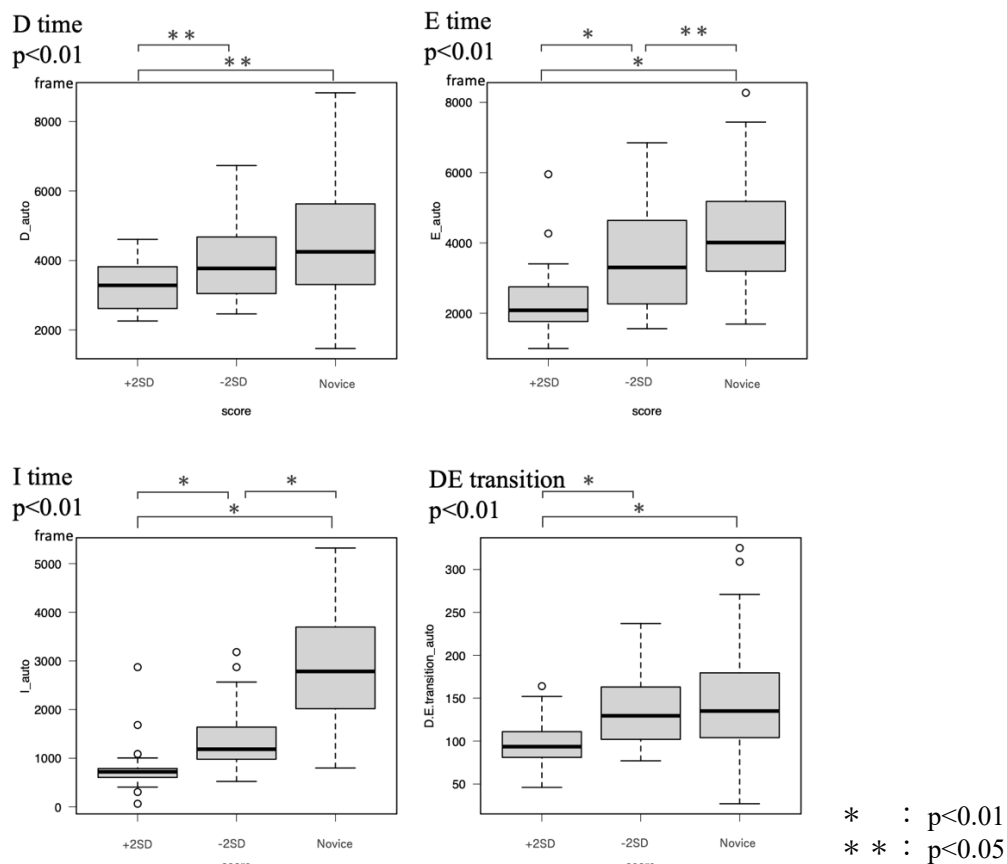


図 14 鉗子認識モデルにおける 3 群間での各パラメータの比較

3-3. 9 工程別の DEI 時間の比較検討

サブ解析として手術全体を以前に報告した工程分類モデル (Kitaguchi et al, 2020) を用いて 9 工程にわけて解析し、とくに DEI の有意差が出やすい工程を調べた。表 6 に 9 工程の分類及びモデルの精度を示す。解析対象症例は、エンジニアのエフォートに限りがあり 2016-2017 年の ESSQS に提出された 2 年分の症例で、+2SD 17 症例、-2SD 20 症例、Novice 30 症例を扱った。限られた条件下であるが、工程毎に 3 群で比較して、とくに習熟度の差が出やすい工程を検討した。

表 6 9 工程と定義と各工程の認識精度 ((Kitaguchi et al, 2020)より改変)

Phase	定義	Precision	Recall	F1 score
1	TME (右側/後側)	0.77	0.77	0.77
2	内側授動 (IMA 切離前)	0.79	0.82	0.81
3	IMA 切離	0.78	0.60	0.68
4	IMA 切離後の内側授動	0.80	0.71	0.75
5	IMV/LCA 切離	0.80	0.75	0.77
6	外側授動	0.84	0.86	0.85
7	TME (左側)	0.78	0.65	0.71
8	直腸間膜切離	0.75	0.81	0.78
9	直腸切離～吻合	0.87	0.90	0.88

TME; total mesorectal excision

IMA; inferior mesenteric artery

IMV; inferior mesenteric vein

LCA; left colic artery

9つの手術工程 (Phase 1～9) それぞれについて、DEI 時間別に 3 群間の比較解析を行い (図 15-1, 2, 3)、結果を表 7 にまとめた。まず、Phase 3 (IMA 切離) および Phase 7 (左側 TME) における剥離操作では、3 群間で有意差を認めなかった。また、Phase 5 (IMV および LCA 切離) においては、-2SD 群で体外操作により処理されている症例が含まれており、その結果、-2SD 群で D 時間が短縮していた。これらの結果から、Phase 3、Phase 5、Phase 7 は Lap-S の評価において手術習熟度の差が現れにくい工程である可能性が示唆された。一方で Phase 2 (岬角頭側から IMA 切離前までの内側授動)、Phase 6 (外側授動)、および Phase 9 (直腸切離から吻合まで) では、3 群間で統計学的に有意差を認めた。これらの工程は、手術習熟度の差が顕著に反映されやすい工程であることが示唆された。

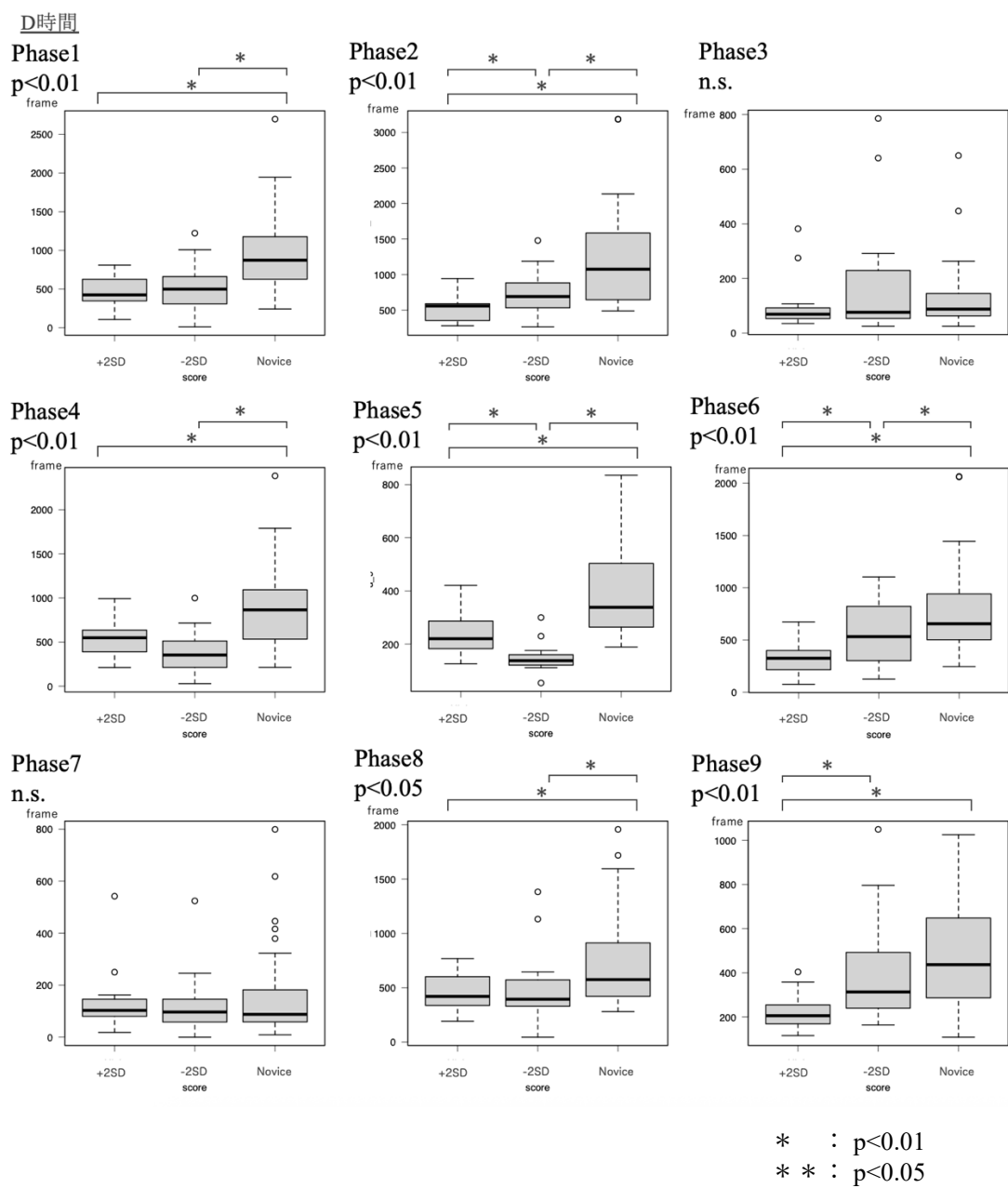
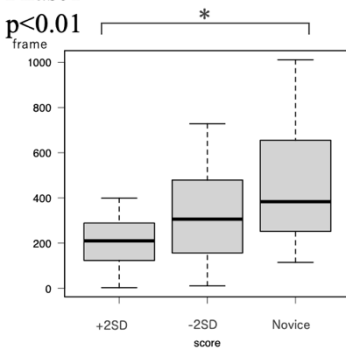


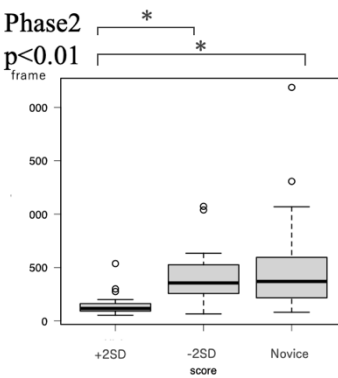
図 15-1 鉗子認識モデルにおける工程別の D 時間の比較

E時間

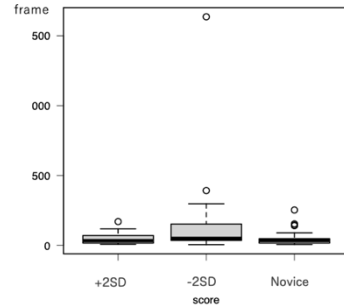
Phase1
p<0.01



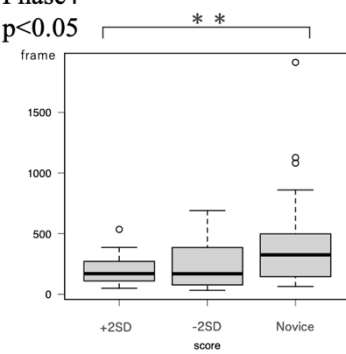
Phase2
p<0.01



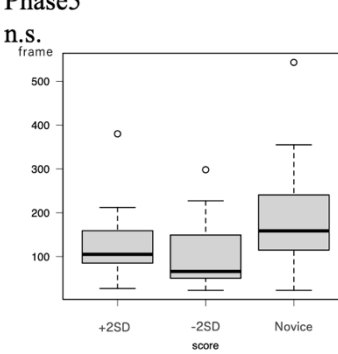
Phase3
n.s.



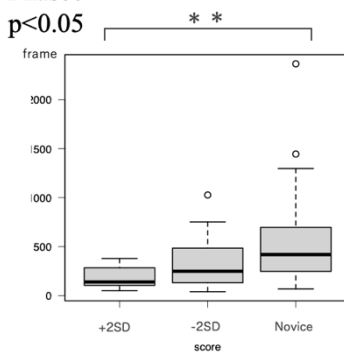
Phase4
p<0.05



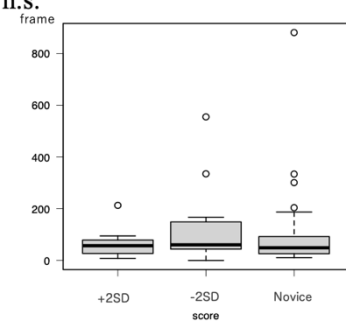
Phase5
n.s.



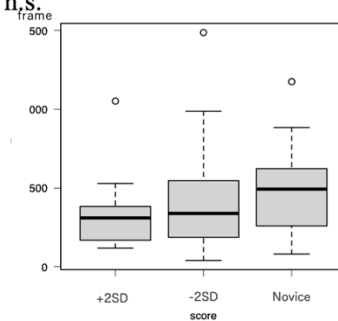
Phase6
p<0.05



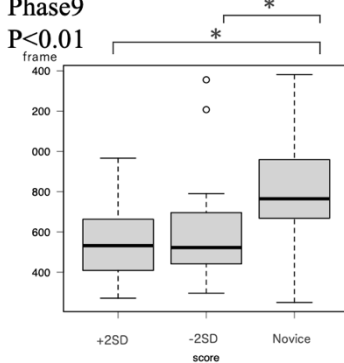
Phase7
n.s.



Phase8
n.s.



Phase9
P<0.01



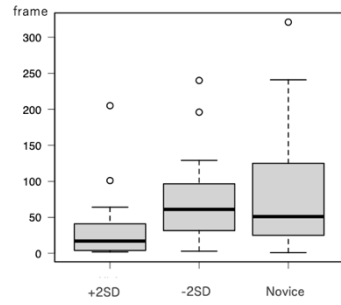
* : p<0.01
* * : p<0.05

図 15-2 鉗子認識モデルにおける工程別の E 時間の比較

I時間

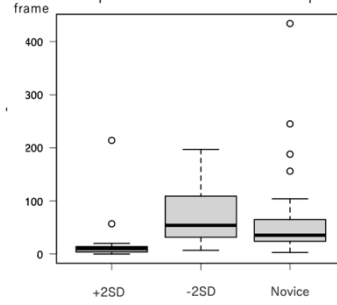
Phase1

n.s.



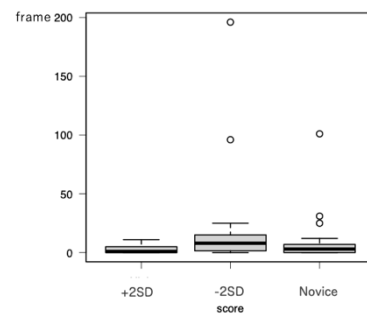
Phase2

p<0.01



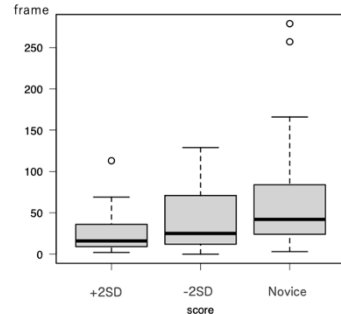
Phase3

n.s.



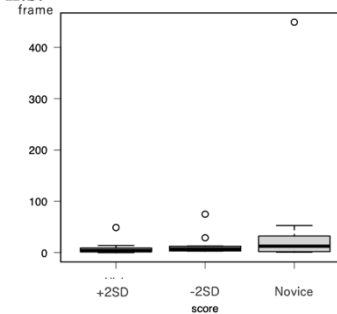
Phase4

n.s.



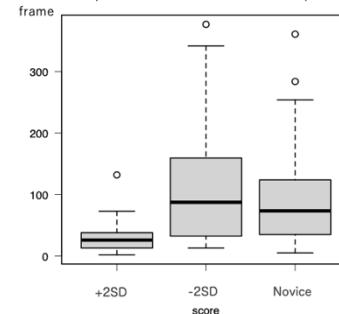
Phase5

n.s.



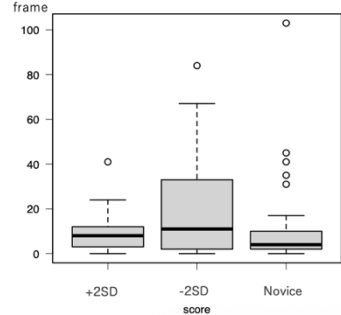
Phase6

n.s.



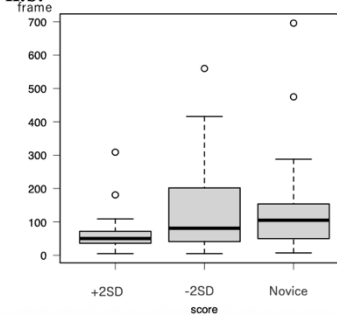
Phase7

n.s.



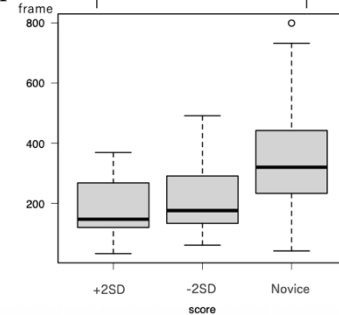
Phase8

n.s.



Phase9

p<0.01



* : p<0.01
* * : p<0.05

図 15-3 鉗子認識モデルにおける工程別の I 時間の比較

表 7 各工程における DEI 時間と 3 群の比較まとめ

Phase	D	E	I
1	+2SD = -2SD < Novice	+2SD < Novice	
2	+2SD < -2SD < Novice	+2SD < -2SD = Novice	+2SD < -2SD = Novice
3	-		
4	+2SD = -2SD < Novice	+2SD < Novice	
5	2SD < +2SD < Novice		
6	+2SD < -2SD < Novice	+2SD < Novice	+2SD < -2SD = Novice
7			
8	+2SD = -2SD < Novice		
9	+2SD < -2SD = Novice	+2SD = -2SD < Novice	+2SD = -2SD < Novice

3-4. 各パラメータと ESSQS スコアの相関に関する検討

本研究は、ESSQS における腹腔鏡下 S 状結腸切除術（Lap-S）の自動技術評価に有用なパラメータ探索が目的であり、鉗子認識モデルを用いて ESSQS スコアとの関連性を検討した。

自動評価システムの確立にあたり、井垣ら（Igaki et al, 2023b）がこれまでに報告されている代表的な外科技術評価法の中から、表 8 に示す 9 種類の技術評価ツールを抽出した（Aghazadeh et al, 2015; Champagne et al, 2017; Curtis et al, 2020; Feldman et al, 2020; Goh et al, 2012; Larson et al, 2005; Martin et al, 1997; Miskovic et al, 2013; Miskovic et al, 2012）。これらの評価ツールに含まれる評価項目を質的研究手法を用いて分析し、計 398 項目の最小単位動作（micro-tasks）に分解した。分解後の要素を再分類して、最終的に 5 つのカテゴリーに大別した。

例えば GOALS の Tissue Handling に関する記載に関して、左に示す文章を、手術手技を表現する最小単位の動作（Skill performance）で分解した（図 16）。表 8 に示す手術評価ツールすべての項目を同様に分解し、表 9 に示すように類似する 5 項目に再分類した。

表 8 代表的な 9 つの手術評価ツール

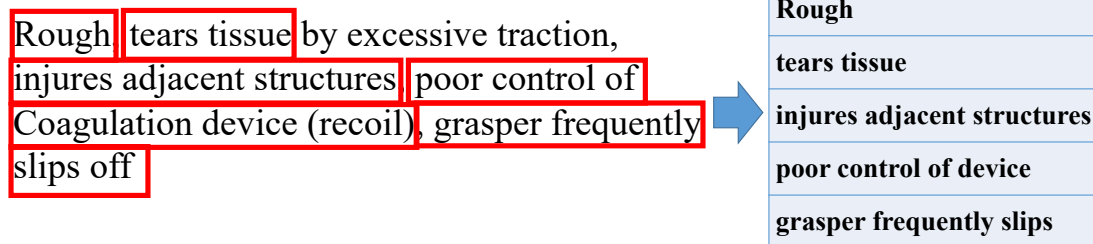
OSATS	Objective Structured Assessment Technical Skills
GOALS	Global Operative Assessment of Laparoscopic Skills
OPRS	The operative performance rating system
OCHRA	The observation clinical human reliability assessment
CAT	NTP Competency Assessment Tool
VBA	SAGES Video-Based Assessment
GEARS	Global Evaluative Assessment of Robotic Skills
LapTMEpt	The laparoscopic Total Mesorectal Excision Performance Tool
ASCRS	The Americal Society of Colon and Rectal Surgeons Assessment Tool

SAGES; Society of American Gastroenterological

NTP; National Training Programme (United Kingdom)

GOALS

Tissue Handling



手術の巧拙を決める最小単位の動作 (**Skill performance**) ごとに分解

図 16 質的研究手法の一例

表 9 5 項目に再分類した Skill performance の例

Category	Skill Performance
Tissue Handling	careful handling of tissue handled tissue appropriately rough movements handles tissues reasonably well appropriate amount of pressure and tension caused damage tears tissue bleed major blood vessels no injury
Psychomotor skill	inappropriate use of instruments stiff or awkward overshoots target wide swings slow to correct uses only one hand expertly utilizes both hands poor grasper control dangerous insertion
Dissection	safety dissection under view smooth and efficient dissection wrong tissue plane, unable to correct quickly stays in correct tissue plane landmarks perfect definition hesitant and awkward dissection technique safe dissection under view
Exposure	failed to use assistants appropriate use of assistants strategically use of assistants

	insufficient view anatomical landmarks not properly identified manipulates structures to create appropriate visual horizon and view of field of interest
Efficiency	unnecessary moves efficient motion maximum efficiency stopped operating uncertain, inefficient efforts tentative movement persisting without process slow, but planned movements

次に、手術技術評価の共同研究施設である北海道大学、大分大学、虎の門病院、国立がん研究センター東病院の4施設間で、各カテゴリーの名称に関するコンセンサス会議を開催し、5項目の手術評価項目を決定した(図17)。

Category	Definition
<u>Tissue handling</u> 組織の取り回し	視野展開や剥離操作等、組織を扱う操作が対象となる。 組織を操作する動きや力の程度、操作の結果として起こる出血や臓器損傷等。 また、出血や損傷を引き起こす可能性のある操作の煩雑さや重要臓器や組織への過度な接近などの損傷の恐れ等も含まれる。
<u>Psychomotor skill</u> 鉗子の操作性	手術鉗子の動きに関する基本的スキル。鉗子操作の円滑性や正確性、 両手の協調性の評価の等の器用性。空間認識能も含まれる。
<u>Dissection</u> 剥離	切離や剥離操作における切離ラインや剥離層の正確性。 剥離層・剥離ラインが一定かどうか、切離・剥離操作後の各工程の出来栄等。 典型的操作との一致度性が評価対象となる。
<u>Exposure</u> 術野展開	助手との連携による術野展開や、解剖的ランドマークの確認の有無等。 典型的操作との一致度性が評価対象となる。
<u>Efficiency</u> 手術全体における効率性	手術進行が戦略的で迷いがなく、手順が効率的に行われているか。 各手順の所要時間、手順工程の不適切は往来、不要な動作時間等。

図17 質的研究手法で再構築して定めた5つの手術評価項目

さらに、ESSQSにおける臓器共通基準（60点）の評価項目を同様の手法で分解・再分類し、各評価項目の配点を上記5カテゴリーに再配分した。DEI時間およびDE transitionは、図17に示す5カテゴリーのうち、特に剥離（Dissection quality）、術野展開（Exposure quality）、および手術全体における効率性（Efficiency）に強く関与すると推測された。そこで、これら3項目の再配分スコアおよび総得点について、2016～2019年に提出された全ESSQS症例を対象に、鉗子認識モデルによって解析されたDEI時間およびDE transitionとの間で相関解析を実施した。

受験者は全体で1277名。全体の剥離（Dissection）、術野展開（Exposure）、手術全体の効率性（Efficiency）および総得点のスコア平均を表10に示す。D時間、E時間、I時間、DE transitionと、再構成した各点数と全体点数の相関係数を表11にまとめた。図18には各パラメータと最も相関したスコアとの散布図を示した。

解析の結果、D時間、E時間およびDE transitionは、手術全体の効率性（Efficiency）と弱い正の相関を認め、I時間は総得点のスコアと弱い正の相関を認めた。

表10 ESSQS 2016-2019年の点数再構成平均得点

Category	平均点数	再分配後点数
剥離（Dissection quality）	6.52 (1.45-9.82)	10.03
展開（Exposure quality）	8.91 (3.84-12.39)	12.39
手術全体の効率（Efficiency）	8.64 (3.15-12.91)	13.18
得点（Total score）	66.60 (29.00-89.75)	-

表11 各パラメータとカテゴリー別点数

Parameters	Dissection	Exposure	Efficiency	Total score
D time	-0.124	-0.112	-0.302	-0.194
E time	-0.180	-0.246	-0.336	-0.301
I time	-0.233	-0.297	-0.295	-0.350
D/E transition	-0.196	-0.204	-0.349	-0.304

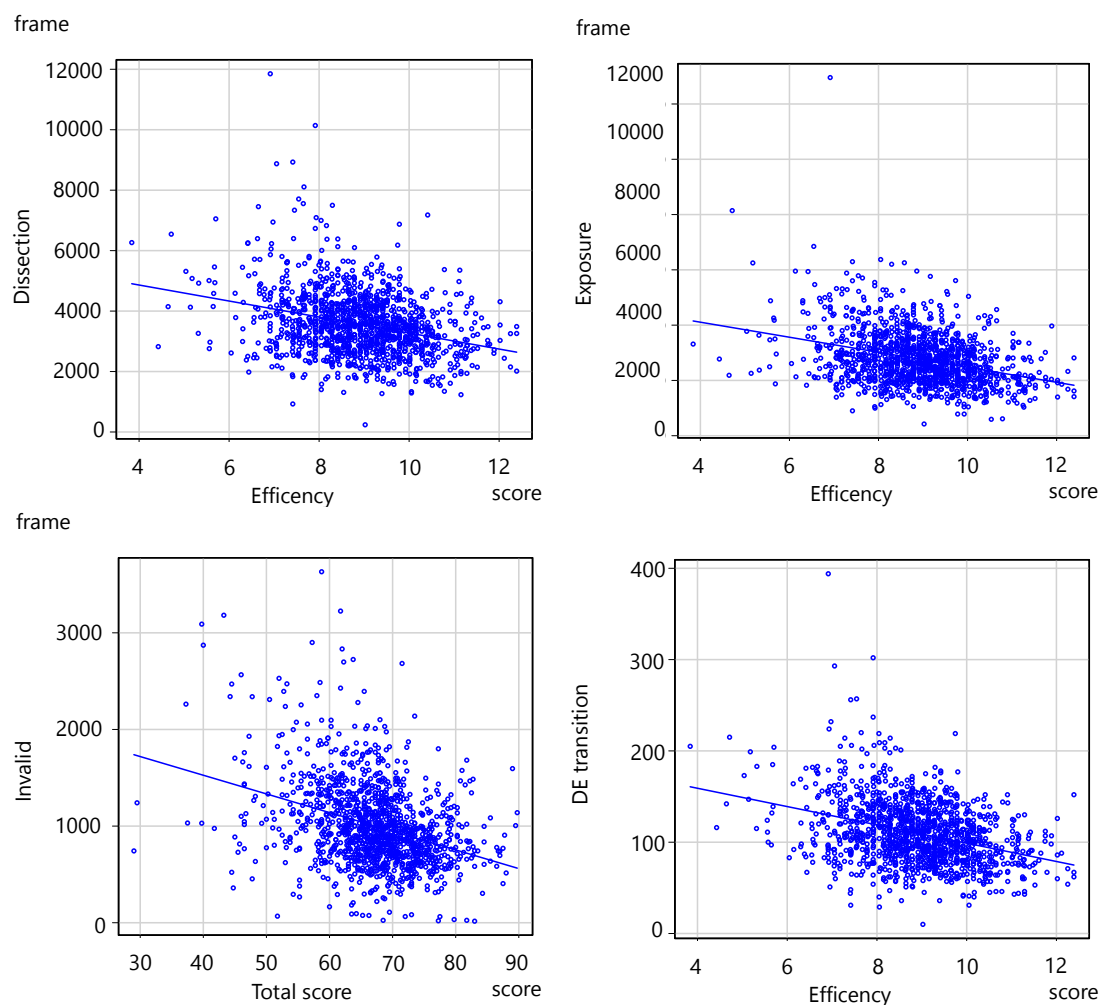


図 18 各パラメータとカテゴリー別点数との散布図

横軸：再分配された各カテゴリーの点数ないし総得点数

縦軸：AI モデルが算出した DEI 時間ないし DE transition の回数

3-5. 鉗子認識モデル小括

鉗子認識モデルの解析結果を教師データと比較したところ、全体の精度は F1 スコアで 0.70 であった。DE 時間および DE transition については、教師データと同様の傾向を示したが、特に Novice 群においてばらつきが大きい傾向がみられた。一方、I 時間では全体的にフレーム数が増加しており、鉗子を認識できない場面が一定数存在することが示唆された。この原因として、学習データに含まれていない鉗子が出現した場合に精度が低下すること、あるいは展開や剥離操作中に視野が不安定な局面では、抽出された静止画で鉗子がぶれを生じて認識が困難となることなどが考えられた。特に Novice 群

では、鉗子操作の不安定さから鉗子が頻繁にフレーム外に出る場面が観察され、これらの要因が認識精度の低下および I 時間の過大推定に寄与していると推察された。+2SD と-2SD 群のように実力差がはっきりしている対象群ではある程度評価可能であった。しかし ESSQS 受験者全体で行った解析では、点数と AI モデルの相関は弱く、Table2.に示したそれぞれの精度も実装には不十分と考えられた。この原因として識別可能な鉗子の種類に限りがあるため使用率の低い鉗子出現した場合に検出率が低くなることや、手術動画は 1 秒ごとの静止画を解析に用いるため手ブレなどによる鉗子の認識率が低下することが要因として考えられた。

工程分類モデルを用いた工程別の解析では、Phase 2（岬角頭側から IMA 切離前までの内側授動）、Phase 6（外側授動）、および Phase 9（直腸切離から吻合まで）では、3 群間で統計学的に有意差を認めた。これらは手術習熟度の差が顕著に反映されやすい工程であることが示唆された。一方で Phase 3（下腸間膜動脈（IMA）切離）および Phase 7（左側 TME）における剥離操作は手術習熟度の差が現れにくい工程である可能性が示唆された。

また相関解析の結果、AI モデルで自動抽出された DEI 関連パラメータは、ESSQS スコアのうち特に術野展開（Exposure）および手術全体における効率性（Efficiency）と弱い相関を示した。

精度向上のために出現鉗子の精度を向上すべく、アノテーションの追加を検討した。しかし同時進行していた他のパラメータ探索研究とのリソース配分の制約から困難であった。この課題を解決するため、鉗子認識に依存せず画面全体を入力情報として DEI 分類を行う新たな AI モデルを構築した。このモデルでは、DEI にラベル付けされた Novice 群の教師データを学習データとして使用したため、解析対象は ESSQS 2016～2019 年度の+2SD 群 26 例と-2SD 群 31 例の 2 群比較とした。

4. DEI 分類モデル

4-1. DEI 分類モデルの精度

DEI 分類モデルの精度を表 12 に示す。表 5 に示した鉗子認識モデルと比較すると、全体の精度は F1 スコアで 0.88 と高値を示し、各パラメータの精度もいずれも向上していた。また、教師データとの Pearson 相関係数を算出したところ、DEI のいずれの時間においても $r > 0.9$ を示し、非常に高い相関を認めた（図 19）。

表 12 DEI 分類モデルの精度

Parameter	Precision	Recall	F1-score	Support
D	0.95	0.88	0.92	229048
E	0.78	0.87	0.82	115101
I	0.70	0.78	0.74	22482
Outside	0.97	0.97	0.97	36196
Accuracy			0.88	
macro avg	0.85	0.88	0.86	402827
Weight avg	0.89	0.88	0.89	402827

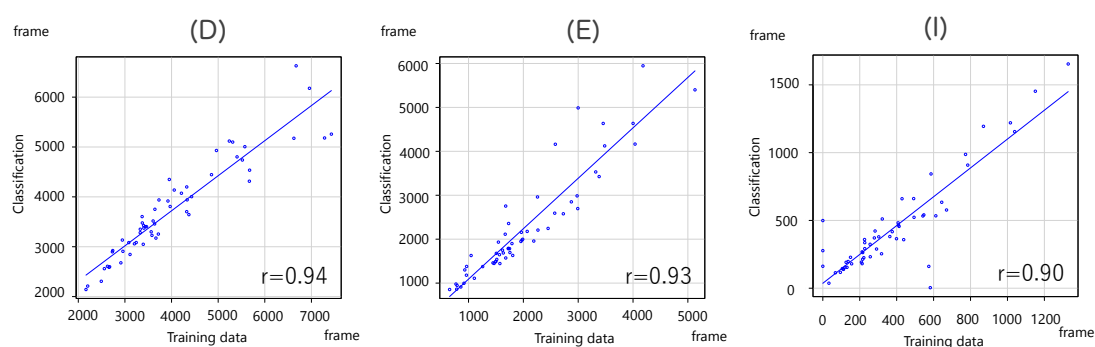


図 19 Pearson 相関係数を用いた DEI 分類モデルの精度

横軸：実際の DEI 時間、縦軸：AI が解析した DEI 時間

4-2. DEI 分類モデルと教師モデルデータでの各パラメータの比較

DEI 分類モデルを用いて各パラメータを算出し、+2SD 群および-2SD 群の 2 群間比較を行った（図 20）。解析の結果、DEI 時間のいずれにおいても 2 群間で統計学的有意差（ $p < 0.01$ ）を認め、フレーム数も教師データと近似した値を示した。また、DE transition においても両群間で有意差を認めたが、DEI 分類モデルでは教師データに比べ約 2 倍のカウント数を示した。

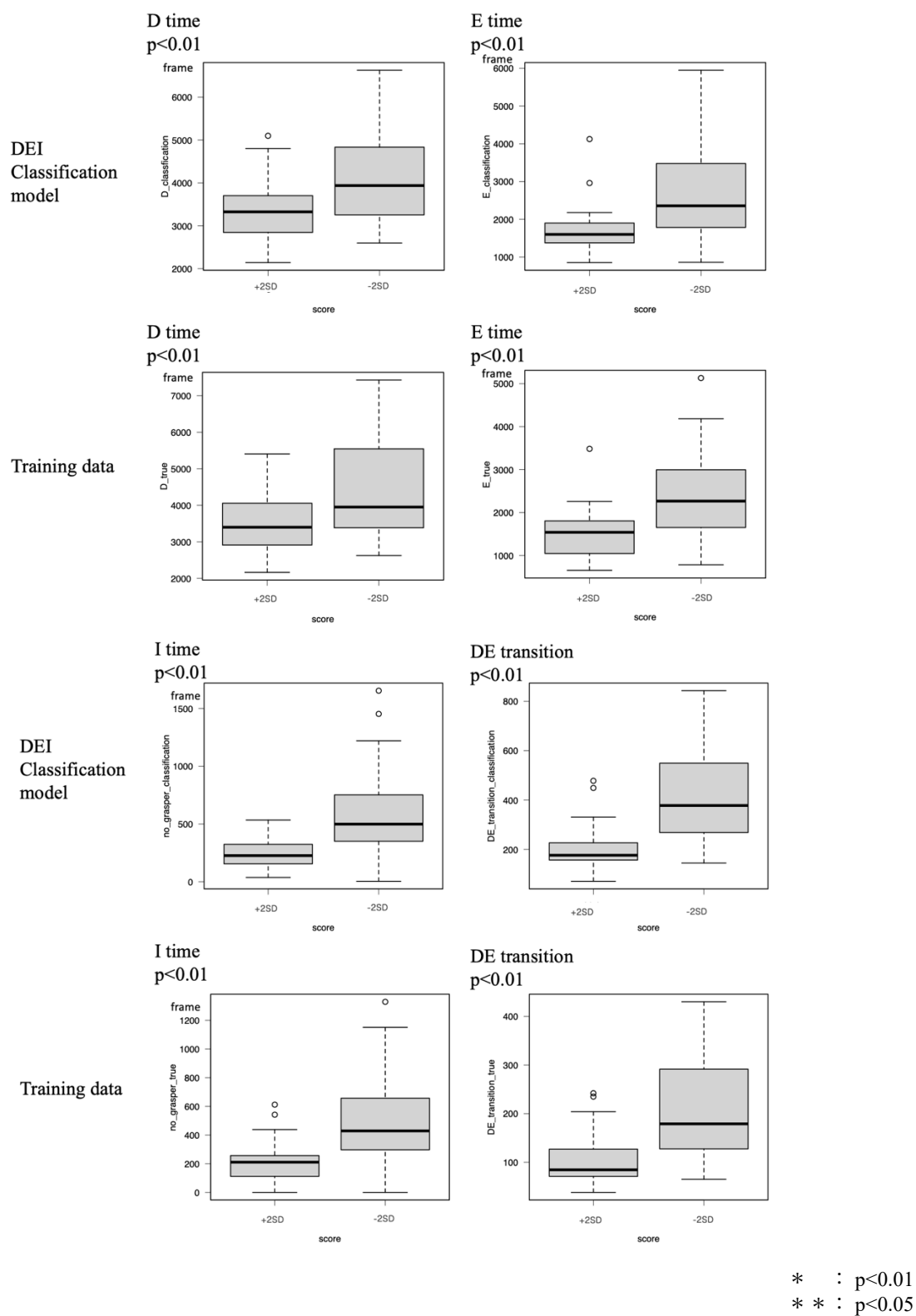


図 20 教師データ、DEI 分類モデルにおける 2 群間での各パラメータ比較

4-3. Action Recognition Performance

モデルの精度を視覚的に評価すべく、手術全体を作業時間別に色分けして表示した **Action Recognition Performance** を描出した（図 21）。教師データと比較すると、鉗子認識モデルでは青色領域に赤色の誤認識が多くみられたのに対し、DEI 分類モデルでは誤認識が少なく、教師データとの一致度が高い結果が得られた。

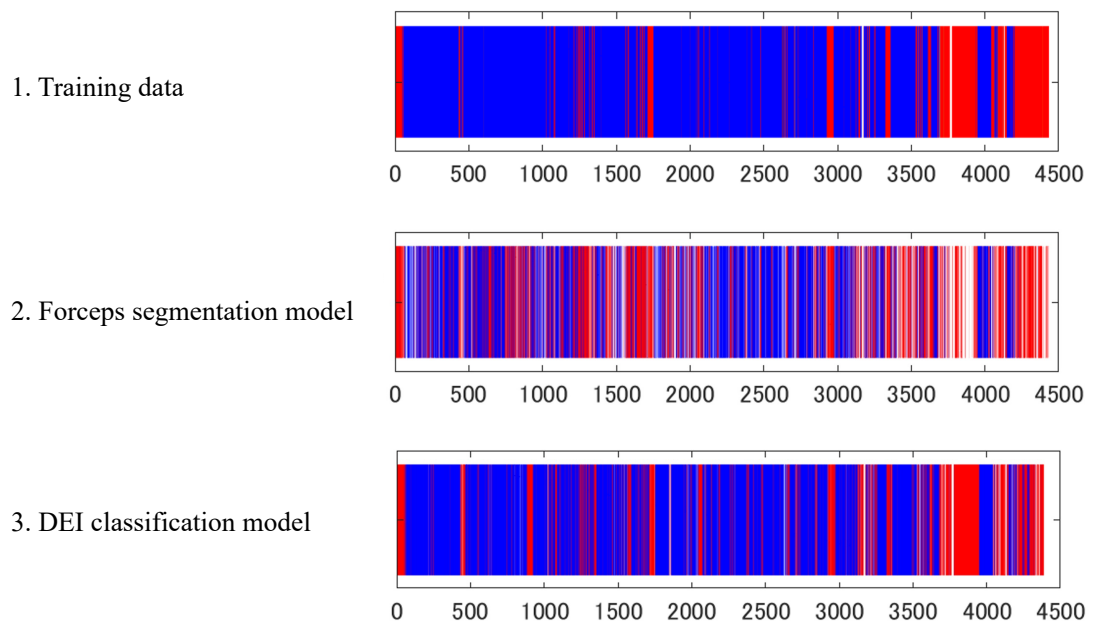


図 21 Action recognition performance

実際に解析を行った 1 症例の手術時間を赤：展開時間，青：剥離時間，白：無効時間で棒グラフに示した。

4-4. DEI 分類モデル小括

DEI 分類モデルは、鉗子認識モデルと比較して、F1 スコア、各パラメータの相関係数、および **Action Recognition Performance** のいずれにおいても高い精度を示した。また、+2SD 群および-2SD 群の 2 群間比較においても、教師データと同様の傾向を示し、DE transition を除くすべての指標で、フレーム数が教師データに近似した値を示した。

考 察

本研究では、腹腔鏡下 S 状結腸切除術（Lap-S）動画を対象に、作業分解手法に基づいて手術操作を剥離（D）、展開（E）、手術無効（I）時間の 3 要素に分類し、これらを定量的指標として用いた AI 自動技能評価モデルを構築した。教師データの結果から DEI 時間、および DE transition が術者の熟練度を反映する有用なパラメータであることが明らかとなった。次に AI モデルを構築して解析を行なった。まず鉗子認識モデルでは、鉗子出現時間に基づく DEI 分類を行い、一定の精度が得られた。このモデルを用いた工程別解析により、特に内側、外側授動や直腸切離及び吻合の工程において手術習熟度の差が顕著に現れることが示唆された。さらに複数の手術評価ツールを再構成して定義した 5 つの手術評価カテゴリーとの相関解析の結果、DEI 時間および DE transition は、手術効率性 (Efficiency) や術野展開 (Exposure) の得点スコアと弱い正の相関を示した。一方より高精度な技能評価を目指して構築した DEI 分類モデルは、画面全体を解析対象とすることにより、高い精度を示すことが明らかになった。このモデルは、鉗子認識に依存しない新たな手術技能評価手法としての有用性が示唆された。

なお 2 つの AI モデルにおいて、本研究で行なった解析を表 13 にまとめた。

表 13 2 つ AI モデルにおける本研究の解析内容

	鉗子認識モデル	DEI 分類モデル
対象群	3 群	2 群
モデル精度	あり	あり
相関係数分析	あり	あり
パラメータ解析	あり	あり
工程別解析	あり	なし
ESSQS 全体解析	あり	なし
Action recognition performance	あり	あり

○手術習熟度との関連性について

一般に手術は効率的な術野展開および的確な剥離操作が重要であり、手術

習熟度は手術時間（Choi et al, 2009; Dinçler et al, 2003; Li et al, 2009; Parascandola et al, 2021; Tekkis et al, 2005; Yamamoto et al, 2013）や合併症発生率（Birkmeyer et al, 2013; Park et al, 2009; Park et al, 2007; Schlachta et al, 2001）に影響することが報告されている。本研究では、習熟度が高い群ほど D、E 時間および DE transition が効率的に改善し、I 時間は中級者から上級者へ移行する段階で初めて有意に短縮する傾向が示唆された。本研究では、上級者ほど、1 回の展開（E）で剥離（D）をより連続的かつ広範囲に遂行できるため、展開と剥離の切替が減少するという仮説を立て、DE transition を新規指標として導入した。解析の結果、上級者で DE transition が低減し、D、E 時間の短縮と整合する所見が得られたことから、本仮説は支持された。すなわち、DE transition は単なる作業量ではなく、展開 1 回あたりの剥離の連続性・まとまりを反映する作業構造指標であり、手術習熟を捉える新たな客観的パラメータとなりうる。また、I 時間の短縮に関しては、手術習熟の進展に伴う助手主導から術者主導への移行を反映している可能性がある。このように、DEI パラメータは手術の質的变化を捉える指標として有用であり、術者の熟練過程を客観的に可視化する AI 評価指標として臨床的意義が高いと考えられる。

○鉗子認識モデルの課題および DEI 認識モデルの構築

鉗子認識モデルでは、未学習の鉗子出現や映像のブレによる認識精度低下が課題として残った。鉗子認識モデルは D 時間において、静止した状態のきれいな test data を用いると F1-score 0.9 以上を示すことが別の解析で示されており（unpublished data）、このことから、認識精度低下の主因として、手術動画における静止画品質（ブレ、ピント、圧縮ノイズ等）の影響が示唆された。また Novice 群では鉗子のフレームアウトが多いことも一因と考えられた。

これに対し、画面全体を入力とする DEI 分類モデルは、個々の鉗子情報に依存せずに作業状態を識別可能であり、実際の手術環境においてより安定した性能を示した。本モデルは教師データの制約から 2 群比較に留まったものの、各指標は鉗子認識モデルと比較して良好であり、action recognition performance においても誤認識が少なく、教師データとの整合性が高かった。これらの結果は、AI による動作認識では鉗子の有無に依存するよりも、場面全体の特徴抽出に基づくアプローチの方が高精度となり得ることを示唆している。なお、モデルが画面内のどの領域に着目しているかは本研究では特定できないが、鉗子以外の情報（例：鉗子先端近傍の組織変化、剥離層の

露出、新規解剖構造の出現など)を手がかりとして学習している可能性がある。

アノテーションについて、本研究では静止画を用いて実施した。静止画アノテーションの利点として、①データ量を抑えられること、②作業が容易で、ラベル付けの正確性向上が期待できることが挙げられる。動画全体を用いる場合、学習データが膨大となり、リアルタイム解析への応用が困難となる可能性がある。また、アノテーション作業が複雑化し開発コストが増大するため、手術動画解析においては静止画を教師データとして用いる手法が広く採用されている。

○工程別解析と教育的意義

サブ解析として、Lap-S を 9 工程 (phases) に分類し、それぞれの特徴を解析した。工程別解析では、内側授動 (Phase2)、外側授動 (Phase6)、直腸切離～吻合 (Phase9) で手術習熟度の差が顕著であった。内側および外側授動工程では膜解剖構造および神経や尿管など重要な構造物の正確な理解、剥離層を遵守した丁寧な鉗子操作が要求される場面である。直腸切離以降は自動縫合器操作や直腸の取り回し、吻合時の術野コントロールや腸管への愛護的操作など、術後合併症に直結する極めて重要な要素が含まれている。これらの結果から、3 工程は特に教育上の重点学習領域として位置付けられることが示唆された。一方、IMA 切離 (Phase3) の工程では群間差が乏しく、大出血のリスクや神経損傷のリスクもあり慎重な操作が求められる工程であることから、+2SD 群の習熟度が高い集団でも時間をかけた手術操作が行われていることが示唆された。さらに Novice 群はハイボリュームセンター症例が多く、指導的助手の介入により一定の安全性と正確性が担保されていた可能性がある。これらの工程ごとの特徴を定量的に把握することで、AI を用いた工程別技術評価や教育カリキュラムへの応用が可能になると考えられる。

○先行研究との比較

Miskovic らの先行研究 (Miskovic et al, 2012) では、Lap-S における外側授動において、習熟者の D/E 比が約 1.0、中級者、初心者群では約 0.6 と報告し、D/E 比が手術習熟度を反映すると述べられている。本研究では、D/E 比単独では有意差を認めず、これは I 時間を独立変数として定義し除外したことに起因すると考えられた。我々の解析結果から手術習熟に伴い D、E 時間がともに短縮するため、D、E 時間の比率よりも 2 数の絶対値および DE

transition の変化を評価する方が妥当であることが示唆された。また鉗子認識モデルにおいて工程別で解析を行った場合、phase2(内側授動) で+2SD と-2SD 群が Novice と比較して統計学的有意差を認めた(unpublished data) が、D/E 比の有用性は限定的であった。

近年、大腸手術における自動技術評価に関していくつかのパラメータが挙げられている (Igaki et al, 2023a; Nakajima et al, 2024; Nakajima et al, 2025; Sasaki et al, 2023)。中島らは直腸間膜切離 (TME)、結腸間膜剥離、IMA 切離、直腸切離吻合、LCA/IMV 切離の 5 工程を抽出し、工程間の遷移回数や各工程にかかる時間は手術習熟に伴い減少していくことを示した (Nakajima et al, 2024)。本研究の結果も同様に、習熟度向上に伴い工程移行回数および DE transition が減少するという一貫性が示され、先行研究を支持する結果であった。さらに井垣らは手術全体を 9 工程 (表 5) に分類し、AI による工程認識の確信度 (AI confidence score) が習熟度および術野展開 (Exposure) の再分配スコアと相関することを示した (Igaki et al, 2023a)。また佐々木らは、出血領域の自動検出 AI を用いて、出血量が組織の取り回し (Tissue handling) のスコアと相関することを示した (Sasaki et al, 2023)。これらの報告および本研究の結果から、単一指標では全体の熟練度を完全に説明し得ないことを示しており、複数パラメータを組み合わせた総合的な自動手術評価システムの構築が今後の課題である。

○本研究の限界と今後の展望

本研究の限界として、以下の点が挙げられる。

- (1) 各術者の学習曲線 (learning curve) を評価しておらず、個人内の経時的な技能向上を直接検討できていない。
- (2) 対象術式を腹腔鏡下 S 状結腸切除 (Lap-S) に限定しており、低位前方切除、結腸右半切除など他の消化器外科手術、さらには子宮全摘や前立腺全摘など他領域手術への汎化性の検証が未実施である。
- (3) 症例難易度 (癒着、肥満、出血傾向、腫瘍進展度など) の影響を十分に加味できていない。
- (4) 使用器具、術式手順、撮影条件等の個人差・施設差の影響を完全には除外できない。

今後は、他術式を含む追加解析や海外多施設データを用いた多次元解析を通じて、国際的に汎用可能な手術技能評価モデルの構築が求められる。また、本研究で抽出したパラメータおよび開発した AI モデルを基盤として、教育現場でのフィードバックを支援する教育支援システムへの応用が期待され

る。例えば、DE transitionやI時間に着目した具体的助言を提示することで、若手外科医に対して根拠に基づく指導が可能となると考えられる。さらに、AI モデルによるリアルタイム手術解析や、複数パラメータを統合した総合的な自動手術評価システムの確立により、質の高い外科教育の実現に寄与する。とくに総合的 AI システムが整備されれば、個人レベルでも手術動画解析を実施できる環境が整い、コストを抑えつつ効率的な手術技術向上を支援できる可能性がある。加えて、NCD 項目、出血量、合併症等の臨床アウトカムと自動評価指標との関連を検討することで、外科教育のみならず医療安全分野への応用も期待される。

結 論

【本研究全体から得られた新知見】

1. 腹腔鏡下 S 状結腸切除術（Lap-S）動画を用い、作業分解手法に基づいて手術操作を剥離（D）、展開（E）、手術無効（I）時間に分類する AI 解析モデルを構築した。
2. 各 DEI 時間、DE transition が術者の手術習熟度を反映する定量的パラメータとして有用であることを明らかにした。
3. 鉗子認識に基づく AI モデル（Mask R-CNN）は一定の精度を示すものの、未学習鉗子や画像ブレの影響を受けやすいことを確認した。
4. 手術映像全体を解析する DEI 分類モデル（Xception）は F1 スコア 0.88、相関係数 0.9 以上と高い精度を示し、器具情報に依存しない高汎用性モデルであることを示した。

【新知見の意義】

本研究は、従来の主観的評価（OSATS、GOALS、ESSQS など）に代わる客観的かつ定量的な技能評価法の可能性を示唆した。また、外科の暗黙知とされてきた手術習熟を、作業分解の手法を用いて定量的に示した。加えて、新規評価項目として DE transition を提案し、D 時間や E 時間といった作業量のみならず、両者の遷移（切替）という関係性を技能評価に導入できる可能性を示した。DE transition を用いることで、上級者では展開 1 回あたりの剥離時間が長く効率的で、結果として DE transition が減少するという作業構造の特徴を捉えることができる。これは、習熟に伴う手技の安定性・効率性を反映する指標となりうる。また、手術動画を AI で解析し、数値化されたパラメータで技能を評価できることから、効率的な教育・認定評価が可能となる。DEI 分類モデルは、器具認識や施設ごとの映像差に依存せず、さまざまな手術環境で応用可能な汎用的外科技術評価基盤となり得る。外科教育現場では、熟練度に応じた客観的フィードバックを迅速に提供できるため、個別化教育が実現する可能性を開いた。さらに、AI 解析による手術定量化は、手技標準化や国際比較研究など医療の質保証にも寄与する可能性を秘めている。

【今後の研究展開の方向性】

DEI 分類モデルをリアルタイムに実装し、手術中に技能レベルを提示することで術中フィードバックシステムの開発につながる可能性がある。また様々

なパラメータを組み合わせた網羅的な手術技術評価システムの構築につながると考えられる。さらに各種パラメータと合併症との関連を解析することで、医療安全分野において有用性が期待できる。

【今後の課題】

本研究は横断的設計であり、術者ごとの学習曲線の解析は未実施である。**Lap-S**に限定した解析で、患者ごとの難易度の差異は考慮されていない。他術式への適用に際しては器具構成・手技動作などの追加学習が必要である。モデル精度を向上し、習熟度が似た群での分解能を高めることが求められる。

謝 辞

本研究は、国立研究開発法人日本医療研究開発機構（AMED）の助成を受けて実施された。支援を受けた事業および研究課題の詳細は以下のとおりである。

事業名：メディカルアーツ研究事業（2020 年度～2022 年度）

研究課題名：内視鏡外科手術における AI 自動技術評価システムの開発

研究課題番号：20vk0124001h0001

研究代表者：国立がん研究センター東病院 大腸外科 伊藤 雅昭

本稿を終えるにあたり、本研究の機会を与えて頂き、御指導・御鞭撻を賜りました北海道大学大学院医学研究院外科系部門外科学分野消化器外科学教室 I 教授である武富紹信先生に深謝申し上げます。また国内留学先である研究開発法人国立がん研究センター東病院大腸外科科長伊藤雅昭先生、そして AI 研究にあたり多くのご指導を賜りました昭和大学産婦人科講師竹中慎先生、研究開発法人国立がん研究センター東病院手術機器開発分野長・医療機器開発支援部長竹下修由先生に心より感謝申し上げます。特にアノテーションやデータ解析など研究全般にご指導いただきました筑波大学医学医療系消化器外科講師/IRCAD France Research Fellow 北口大地先生に感謝の意を表します。北海道大学大学院医学研究院外科系部門外科学分野消化器外科学教室講師市川伸樹先生には論文作成に際してご指導いただき、深謝申し上げます。本研究を遂行するためにご協力いただいた研究開発法人国立がん研究センター東病院 NEXT 医療機器開発センターおよび北海道大学大学院医学研究院外科系部門外科学分野消化器外科学教室 I の皆様、手術動画の提供にご同意下さいました患者様、その他多くの方々にこの場をお借りして心から厚く御礼を申し上げます。

利 益 相 反

本研究に開示すべき利益相反状態はありません。

引用文献

- Aghazadeh MA, Jayaratna IS, Hung AJ, Pan MM, Desai MM, Gill IS, Goh AC (2015) External validation of Global Evaluative Assessment of Robotic Skills (GEARS). *Surg Endosc* 29, 3261-3266.
- Ajay Agrawal JG, and Avi Goldfarb (2022) *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press / Harvard Business Press
- Anaut : EURIKA α , <https://anaut-surg.com> (2026 年 2 月 12 日参照)
- Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, Dimick J, Banerjee M, Birkmeyer NJ, Michigan Bariatric Surgery C (2013) Surgical skill and complication rates after bariatric surgery. *N Engl J Med* 369, 1434-1442.
- Champagne BJ, Steele SR, Hendren SK, Bakaki PM, Roberts PL, Delaney CP, Brady JT, MacRae HM (2017) The American Society of Colon and Rectal Surgeons Assessment Tool for Performance of Laparoscopic Colectomy. *Dis Colon Rectum* 60, 738-744.
- Choi DH, Jeong WK, Lim SW, Chung TS, Park JI, Lim SB, Choi HS, Nam BH, Chang HJ, Jeong SY (2009) Learning curves for laparoscopic sigmoidectomy used to manage curable sigmoid colon cancer: single-institute, three-surgeon experience. *Surg Endosc* 23, 622-628.
- Chollet F, 2017. Xception: Deep Learning with Depthwise Separable Convolutions, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Honolulu, HI, USA, pp. 1251-1258.
- Christian S, Wei L, Yangqing J, Pierre S, Scott R, Dragomir A, Dumitru E, Vincent V, Andrew R (2015) Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1, 1-9.
- Curtis NJ, Foster JD, Miskovic D, Brown CSB, Hewett PJ, Abbott S, Hanna GB, Stevenson ARL, Francis NK (2020) Association of Surgical Skill Assessment With Clinical Outcomes in Cancer Surgery. *JAMA Surgery* 155, 590-598.
- Daniel Rock AM, and Erik Brynjolfsson (2023) *How to Capitalize on Generative AI: A guide to realizing its benefits while limiting its risks*.
- Dinçler S, Koller MT, Steurer J, Bachmann LM, Christen D, Buchmann P (2003) Multidimensional analysis of learning curves in laparoscopic sigmoid

resection: eight-year results. *Dis Colon Rectum* 46: 1371-1378, discussion 1378-1379.

Feldman LS, Pryor AD, Gardner AK, Dunkin BJ, Schultz L, Awad MM, Ritter EM (2020) SAGES Video-Based Assessment (VBA) program: a vision for life-long learning for surgeons. *Surg Endosc* 34, 3285-3288.

Girshick R, Donahue J, Darrell T, Malik J, 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, pp. 580–587.

Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ (2012) Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *J Urol* 187, 247-252.

Google DeepMind. Gemini 1 Technical Report (2023), Cornell University <https://arxiv.org/pdf/2312.11805>, (2026 年 2 月 12 日参照)

Hashimoto DA, Rosman G, Rus D, Meireles OR (2018) Artificial Intelligence in Surgery: Promises and Perils. *Ann Surg* 268, 70-76.

Hawkins AJ (2023) Waymo’s robotaxis are now available to tens of thousands of people across all of San Francisco. *theverge.com*, <https://www.theverge.com/2023/10/10/23911190/waymo-robotaxi-san-francisco-customers-waitlist> (2026 年 2 月 12 日参照)

He K, Gkioxari G, Dollar P, Girshick R (2020) Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell* 42, 386-397.

Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H (2018) Artificial intelligence in radiology. *Nat Rev Cancer* 18, 500-510.

Humm G, Mohan H, Fleming C, Harries R, Wood C, Dawas K, Stoyanov D, Lovat LB (2022) The impact of virtual reality simulation training on operative performance in laparoscopic cholecystectomy: meta-analysis of randomized clinical trials. *BJS Open* 6, zrac086.

Ichikawa N, Homma S, Funakoshi T, Ohshima T, Hirose K, Yamada K, Nakamoto H, Kazui K, Yokota R, Honma T et al (2020) Impact of technically qualified surgeons on laparoscopic colorectal resection outcomes: results of a propensity score-matching analysis. *BJS Open* 4, 486-498.

Igaki T, Kitaguchi D, Matsuzaki H, Nakajima K, Kojima S, Hasegawa H, Takeshita N, Kinugasa Y, Ito M (2023a) Automatic Surgical Skill Assessment System Based on Concordance of Standardized Surgical Field Development

Using Artificial Intelligence. *JAMA Surg* 158, e231131.

Igaki T, Takenaka S, Watanabe Y, Kojima S, Nakajima K, Takabe Y, Kitaguchi D, Takeshita N, Inomata M, Kuroyanagi H et al (2023b) Universal meta-competencies of operative performances: a literature review and qualitative synthesis. *Surg Endosc* 37, 835-845.

Iwamoto M, Makutani Y, Yane Y, Ushijima H, Yoshioka Y, Wada T, Daito K, Tokoro T, Chiba Y, Ueda K et al (2023) The usefulness of the endoscopic surgical skill qualification system in laparoscopic right hemicolectomy: a single-center, retrospective analysis with propensity score matching. *Langenbecks Arch Surg* 408, 33.

Jmees : SurVis-Hys, <https://www.jmees-inc.com> (2026 年 2 月 12 日 参照)

Kanda Y (2013) Investigation of the freely available easy-to-use software 'EZ' for medical statistics. *Bone Marrow Transplant* 48, 452-458.

Kimura T, Mori T, Konishi F, Kitajima M (2010) Endoscopic surgical skill qualification system in Japan: Five years of experience in the gastrointestinal field. *Asian Journal of Endoscopic Surgery* 3, 66-70.

Kitaguchi D, Lee Y, Hayashi K, Nakajima K, Kojima S, Hasegawa H, Takeshita N, Mori K, Ito M (2022) Development and Validation of a Model for Laparoscopic Colorectal Surgical Instrument Recognition Using Convolutional Neural Network-Based Instance Segmentation and Videos of Laparoscopic Procedures. *JAMA Netw Open* 5, e2226265.

Kitaguchi D, Takeshita N, Matsuzaki H, Oda T, Watanabe M, Mori K, Kobayashi E, Ito M (2020) Automated laparoscopic colorectal surgery workflow recognition using artificial intelligence: Experimental research. *Int J Surg* 79, 88-94.

Kiyozumi Y, Yamaguchi T, Ichikawa N, Homma S, Ikeda K, Inada R, Otsuka K, Furutani A, Iijima H, Watanabe M et al (2023) Endoscopic Surgical Skill Qualification System: propensity-score matched cohort analysis of accredited supervisors in laparoscopic rectal cancer surgery. *Br J Surg* 110, 1834-1839.

Komura D, Ishikawa S (2018) Machine Learning Methods for Histopathological Image Analysis. *Comput Struct Biotechnol J* 16, 34-42.

Korosec K, Waymo launches self-driving car service Waymo One.

Larson JL, Williams RG, Ketchum J, Boehler ML, Dunnington GL (2005) Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. *Surgery* 138, 640-649.

- Li JC, Hon SS, Ng SS, Lee JF, Yiu RY, Leung KL (2009) The learning curve for laparoscopic colectomy: experience of a surgical fellow in an university colorectal unit. *Surg Endosc* 23, 1603-1608.
- Li Z, Liu F, Yang W, Peng S, Zhou J (2022) A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans Neural Netw Learn Syst* 33, 6999-7019.
- Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 84, 273-278.
- McAfee EBaA (2017) *The Business of Artificial Intelligence*.
- Miskovic D, Ni M, Wyles SM, Kennedy RH, Francis NK, Parvaiz A, Cunningham C, Rockall TA, Gudgeon AM, Coleman MG et al (2013) Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England. *Ann Surg* 257, 476-482.
- Miskovic D, Ni M, Wyles SM, Parvaiz A, Hanna GB (2012) Observational clinical human reliability analysis (OCHRA) for competency assessment in laparoscopic colorectal surgery at the specialist level. *Surg Endosc* 26, 796-803.
- Mori T, Kimura T, Kitajima M (2010) Skill accreditation system for laparoscopic gastroenterologic surgeons in Japan. *Minim Invasive Ther Allied Technol* 19, 18-23.
- Nakajima K, Kitaguchi D, Takenaka S, Tanaka A, Ryu K, Takeshita N, Kinugasa Y, Ito M (2024) Automated surgical skill assessment in colorectal surgery using a deep learning-based surgical phase recognition model. *Surg Endosc* 38, 6347-6355.
- Nakajima K, Takenaka S, Kitaguchi D, Tanaka A, Ryu K, Takeshita N, Kinugasa Y, Ito M (2025) Artificial intelligence assessment of tissue-dissection efficiency in laparoscopic colorectal surgery. *Langenbecks Arch Surg* 410, 80.
- OpenAI. Introducing ChatGPT(2022), <https://www.chatgpt.com> (2026 年 2 月 12 日 参照)
- Parascandola SA, Horsey ML, Hota S, Paull JO, Graham A, Pudalov N, Smith S, Amdur R, Obias V (2021) The robotic colorectal experience: an outcomes and learning curve analysis of 502 patients. *Colorectal Dis* 23, 226-236.

Park IJ, Choi GS, Lim KH, Kang BM, Jun SH (2009) Multidimensional analysis of the learning curve for laparoscopic colorectal surgery: lessons from 1,000 cases of laparoscopic colorectal surgery. *Surg Endosc* 23, 839-846.

Park JS, Kang SB, Kim SW, Cheon GN (2007) Economics and the laparoscopic surgery learning curve: comparison with open surgery for rectosigmoid cancer. *World J Surg* 31, 1827-1834.

Sasaki S, Kitaguchi D, Takenaka S, Nakajima K, Sasaki K, Ogane T, Takeshita N, Gotohda N, Ito M (2023) Machine learning-based Automatic Evaluation of Tissue Handling Skills in Laparoscopic Colorectal Surgery: A Retrospective Experimental Study. *Ann Surg* 278, e250-e255.

Schlachta CM, Mamazza J, Seshadri PA, Cadeddu M, Gregoire R, Poulin EC (2001) Defining a learning curve for laparoscopic colorectal resections. *Dis Colon Rectum* 44, 217-222.

Schmidt MW, Haney CM, Kowalewski KF, Bintintan VV, Abu Hilal M, Arezzo A, Bahra M, Besselink MG, Biebl M, Boni L et al (2022) Development and validity evidence of an objective structured assessment of technical skills score for minimally invasive linear-stapled, hand-sewn intestinal anastomoses: the A-OSATS score. *Surg Endosc* 36, 4529-4541.

Shiroshita H, Inomata M, Takiguchi S, Akira S, Kanayama H, Yamaguchi S, Eguchi S, Wada N, Kurokawa Y, Akagi T et al (2024) Update on endoscopic surgery in Japan: Results of the 16th National Survey of endoscopic surgery by the Japan Society for Endoscopic Surgery. *Asian J Endosc Surg* 17, e13285.

Stulberg JJ, Huang R, Kreutzer L, Ban K, Champagne BJ, Steele SR, Johnson JK, Holl JL, Greenberg CC, Bilimoria KY (2020) Association Between Surgeon Technical Skills and Patient Outcomes. *JAMA Surg* 155, 960-968.

Tekkis PP, Senagore AJ, Delaney CP, Fazio VW (2005) Evaluation of the learning curve in laparoscopic colorectal surgery: comparison of right-sided and left-sided resections. *Ann Surg* 242, 83-91.

Vaillancourt M, Ghaderi I, Kaneva P, Vassiliou M, Kolozsvari N, George I, Sutton FE, Seagull FJ, Park AE, Fried GM et al (2011) GOALS-incisional hernia: a valid assessment of simulated laparoscopic incisional hernia repair. *Surg Innov* 18, 48-54.

van Zwieten TH, Okkema S, Kramp KH, de Jong K, Van Det MJ, Pierie JEN (2022) Procedure-based assessment for laparoscopic cholecystectomy can replace global rating scales. *Minim Invasive Ther Allied Technol* 31, 865-871.

Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, Fried GM (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 190, 107-113.

Yamamoto M, Okuda J, Tanaka K, Kondo K, Asai K, Kayano H, Masubuchi S, Uchiyama K (2013) Evaluating the learning curve associated with laparoscopic left hemicolectomy for colon cancer. Am Surg 79, 366-371.

がん情報サービス (ganjoho.jp) <https://ganjoho.jp/public/index.html>,
(2026 年 2 月 12 日参照)

木村 武, 錦織 宏 (2023) フィードバック 定義から近年の議論まで. 医学教育 54, 255-265.

厚生労働省 令和 4 (2022) 年医師・歯科医師・薬剤師統計の概況
<https://www.mhlw.go.jp/toukei/saikin/hw/ishi/22/index.html>, (2026 年 2 月 12 日参照)

日本内視鏡外科学会 共通基準採点表【大腸】
<https://www.jses.or.jp/uploads/files/gijutsunintei/shinsei/shokai/shokaki/00daichou.pdf> (2026 年 2 月 12 日参照)

日本内視鏡外科学会 臓器別基準【大腸】
<https://www.jses.or.jp/uploads/files/gijutsunintei/shinsei/shokai/shokaki/06.pdf> (2026 年 2 月 12 日参照)

日本内視鏡外科外科学会 領域別審査結果
https://www.jses.or.jp/modules/gijutsunintei/index.php?content_id=15, (2026 年 2 月 12 日参照)

付録 1 共通基準採点表【大腸】

審査員氏名 _____

Category I 手術の進行

(採点はすべて 0.5 点刻み)

I-1 手術の進行・スピード	点
手術の進行が計画的かつ円滑であり、手術時間も標準的である。	5 点
手術の計画性、円滑さに改善すべき点があり、手術時間が延長している。	2.5 点
手術進行が、計画的かつ円滑とは言えない。	0 点
I-2 術者の主体性	点
剥離、切離などの主要な手術操作は、術者の意図のもとに行われており、助手はそれを補助している。	3 点
時に助手が主導することがある。	1.5 点
助手の意図のもとに手術が進行している。	0 点
I-3 術者の指導性	点
助手、スコープオペレーターを適切に指導し、円滑に手術が進行している。	5 点
時に助手やスコープオペレーターの不十分な操作を修正していない。	2.5 点
スコープオペレーターの不十分な操作を修正せず手術が進行している。	0 点
I-4 助手との連携	点
助手との連携が良好で、良好な視野展開、カウンタートラクションが得られている。	5 点
時に助手との連携が不十分となり、修正を要する。	2.5 点
助手との連携が悪く、十分な視野が得られないまま手術が進んでいる。	0 点
※ 助手のいないソロサージェリーの場合は 3 点とする（ヘルニアは除く）	

Category II 術野展開

II-1 アクセスポート（ポートを用いない場合は皮切部位と読み替える）	点
アクセスポートの選択、留置位置、留置方法とも適切である。	2 点
アクセスポートの選択、留置位置の改善により手術手技の難度低下が期待できる。	1 点
アクセスポートの選択、留置位置または留置方法に起因する出血や手術時間の延長を認める。	0 点
アクセスポートの選択、留置位置または留置方法の誤りにより修復を要する臓器損傷をきたしている。	落第地雷
II-2 モニター中央に術野が捉えられているか	点
術野がほぼ視野の中心に良好に捉えられている。	2 点
時に手術手技が視野の中心外にある術野で行われている（直視鏡による意図的なものを除く）	1 点
ブラインド操作が認められる。	0 点
II-3 術野の臓器が良好に視認されているか	点
手術操作が行われている部位が良好に視認されている。	3 点
時に出血や術野展開法のために、術野の視認が不良となっている。	1.5 点
術野の視認に問題点がある。	0 点
II-4 術野の展開 1（レトラクター使用法など）	点
術野展開のための鉗子（レトラクター、助手鉗子）の使用が良好である。（挿入部位、組織接触、牽引方向、牽引力）	3 点
術野展開のための鉗子（レトラクター、助手鉗子）の使用法に改善すべき点がある。（挿入部位、組織接触、牽引方向、牽引力）	1.5 点
術野展開のための鉗子（レトラクター、助手鉗子）使用法が適切でなく危険と思われる部分がある。	0 点
組織損傷のために有意な出血をきたしている。修復を要する臓器損傷をきたしている。	落第地雷

II-5 術野の展開 2（非優位側鉗子使用法）	点
非優位側の鉗子で術野が常に良好に展開されている。	4 点
非優位側の鉗子操作に改善点があり時に術野展開が不良となっている。	2 点
非優位側鉗子による術野展開が不良で、ブラインド操作や、危険と思われる手技がみられる。	0 点
※ 片手法と推定されるビデオでは非優位側鉗子使用法は 0 点	

Category III 手術手技

III-1 使用鉗子	点
鉗子選択が適切であり、またその使用法も適切である。	2 点
鉗子選択または使用法に改善すべきところがある。	1 点
不適切な鉗子選択または使用法に起因する出血や危険な手技がみられる。	0 点
不適切な鉗子の選択または使用法により、修復を要する臓器損傷をきたしている。	落第地雷
III-2 組織の取り回し	点
組織把持の方法、牽引法とも適切である。	4 点
組織把持法、牽引法に改善すべきところがある。	2 点
不適切な組織把持の方法、使用鉗子、牽引法に起因する出血や、組織損傷をきたしている。	0 点
不適切な組織把持の方法、牽引法により、修復を要する臓器損傷をきたしている。	落第地雷
III-3 エネルギー源の選択	点
エネルギー源の選択、使用法とも適切である。	2 点
エネルギー源の選択、または使用法に改善すべきところがある。	1 点
エネルギー源の選択または使用法の誤りに起因する出血や、危険手技がみられる。	0 点
不適切なエネルギー源の選択または使用法の誤りにより、修復を要する周囲臓器の損傷をきたしている。	落第地雷

III-4 剥離層	点
剥離層が適切であり、一定している。	7 点
剥離層が適切でないこともあるが、短時間で修正されている。	3.5 点
最初から層の認識に誤認があり、不適切な剥離層の手術である。	0 点
間違った剥離層による、周囲組織、主要な脈管の損傷	落第地雷
III-5 血管処理	点
適切な血管の同定、剥離、止血操作（クリップ、エネルギー源）切離が行われている。	5 点
血管の同定、剥離、止血操作（クリップ、エネルギー源）切離に起因する手術時間の延長が認められる。	2.5 点
血管の同定、剥離、止血操作（クリップ、エネルギー源）切離操作に明らかな改善点を指摘できる。	0 点
ブラインド焼灼、ブラインドクリッピング	落第地雷

◎ 悪性腫瘍症例において手術手技に関連して根治性が損なわれている場合、それぞれの手技に関して適宜減点する。

Category IV 縫合結紮

IV-1 縫合技術	点
縫合は術者の意図の通り、正確かつ迅速に行われている。	4 点
針のマウント、運針などに明らかな改善点が指摘でき、手術時間が延長している。	2 点
縫合技術が不十分である。	0 点
IV-2 結紮技術	点
内視鏡下結紮がスムーズに行われている。	4 点
内視鏡下結紮に改善すべき点があり、手術時間が延長している。	2 点
結紮技術が判定できる資料がない。	0 点

- 縫合練習器およびアニマルラボでの縫合結紮操作の満点は臨床例と同じとする。
- 縫合結紮の結紮は、体内結紮法とする。体外結紮、および刺通結紮・縫合は認めない。
- 縫合結紮は、離れた組織を接合する手技を評価するもののため、刺通縫合・結紮は認めない。ドライボックスを用いる場合は、縫合素材に切り込みを入れるなど、素材の接合が評価できるようにすること。接合が不十分な場合は減点の対象とする。

- 針を持針器で把持する動作から縫合・結紮を終えて糸を切るまでの未編集ビデオデータとする。また、結紮は3回行い、初回2回の結紮法は Square knot または Surgeon's knot とする。Square knot もしくは Surgeon's knot 以外の結紮は減点とする（Slip knot は Square knotとみなす）
- ビデオ提出のみを目的とした必要のない縫合など、倫理的に問題があると審査委員が判断した場合は、縫合・結紮の採点を 0 点とする。
- 結紮用特殊機器を用いたものは審査対象とならない。

付録2 臓器別評価項目【大腸】

大腸手術 40点各 0-2点（小数点以下あり）で採点

癌組織の損傷など癌の手術として明らかに不適当な手技、修復を要する他臓器損傷（腸管、尿管、自律神経など）は地雷とする。また地雷に近い操作は-2までで評価する

カテゴリーⅠ 総合 （10 点）
1.手術開始時に腹腔内の検索、および小腸や大網の適切な排除がなされたか？
2.手術終了時にポート抜去部位、血管切離端、剥離部位からの出血が無いこと、および腹腔内異物遺残が無いことを確認したか？
3.癌に触れないなど、癌手術に対する配慮が適正になされているか？
4.腸管損傷回避の配慮がなされているか？
5.手術全体の構成が計画的になされているか？
カテゴリーⅡ 剥離授動 （10 点）
1.術野展開は良好か？
2.組織把持やデバイス操作は適切か、出血が制御されているか？
3.適切な手順で無駄な時間の延長がないか？
4.腹膜切開（内側・外側）の位置、方法は適切か？
5.剥離層の認識が適切で、十分な範囲になされたか？
カテゴリーⅢ 中枢側リンパ節郭清 （10 点）
1.術野展開は良好か？
2.組織把持やデバイス操作は適切か、出血が制御されているか？
3.適切な手順で無駄な時間の延長がないか？
4.中枢側リンパ節郭清が適切か？
5.血管同定、血管処理が適切か？
カテゴリーⅣ 腸切離と吻合 （10 点）
1.術野展開は良好か？
2.組織把持やデバイス操作は適切か、出血が制御されているか？
3.適切な手順で無駄な時間の延長がないか？
4.肛門側腸管の授動、間膜処理が適切か？
5.自動縫合器・吻合器の使用法が適切か？