

Review

Artificial Intelligence in Colorectal Cancer Surgery: Present and Future Perspectives

Giuseppe Quero ^{1,2,†}, Pietro Mascagni ^{2,3,†}, Fiona R. Kolbinger ⁴ , Claudio Fiorillo ^{1,*} , Davide De Sio ¹ , Fabio Longo ¹, Carlo Alberto Schena ^{1,2} , Vito Laterza ^{1,2}, Fausto Rosa ^{1,2} , Roberta Menghi ^{1,2}, Valerio Papa ^{1,2}, Vincenzo Tondolo ¹ , Caterina Cina ¹, Marius Distler ⁴, Juergen Weitz ⁴, Stefanie Speidel ⁵, Nicolas Padoy ^{3,6,‡} and Sergio Alfieri ^{1,2,‡}

- ¹ Digestive Surgery Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Largo Agostino Gemelli 8, 00168 Rome, Italy
² Faculty of Medicine, Università Cattolica del Sacro Cuore di Roma, Largo Francesco Vito 1, 00168 Rome, Italy
³ Institute of Image-Guided Surgery, IHU-Strasbourg, 67000 Strasbourg, France
⁴ Department for Visceral, Thoracic and Vascular Surgery, University Hospital and Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden, Germany
⁵ National Center for Tumor Diseases (NCT), Partner Site Dresden, 01307 Dresden, Germany
⁶ ICube, Centre National de la Recherche Scientifique (CNRS), University of Strasbourg, 67000 Strasbourg, France
* Correspondence: claudio.fiorillo@hotmail.it; Tel.: +39-333-8747996
† These authors contributed equally to this work.
‡ These authors contributed equally to this work.



Citation: Quero, G.; Mascagni, P.; Kolbinger, F.R.; Fiorillo, C.; De Sio, D.; Longo, F.; Schena, C.A.; Laterza, V.; Rosa, F.; Menghi, R.; et al. Artificial Intelligence in Colorectal Cancer Surgery: Present and Future Perspectives. *Cancers* **2022**, *14*, 3803. <https://doi.org/10.3390/cancers14153803>

Academic Editor: Alexandre Escargueil

Received: 28 June 2022

Accepted: 3 August 2022

Published: 4 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Computer vision (CV) is a field of artificial intelligence (AI) that deals with the automatic analysis of videos and images. Recent advances in AI and CV methods coupled with the growing availability of surgical videos of minimally invasive procedures have led to the development of AI-based algorithms to improve surgical care. Initial proofs of concept have focused on fairly standardized procedures such as laparoscopic cholecystectomy. However, the real value of CV in surgery resides in analyzing and providing assistance in more complex and variable procedures such as colorectal resections. This manuscript provides a brief introduction to AI for surgeons and a comprehensive overview of CV solutions for colorectal cancer surgery.

Abstract: Artificial intelligence (AI) and computer vision (CV) are beginning to impact medicine. While evidence on the clinical value of AI-based solutions for the screening and staging of colorectal cancer (CRC) is mounting, CV and AI applications to enhance the surgical treatment of CRC are still in their early stage. This manuscript introduces key AI concepts to a surgical audience, illustrates fundamental steps to develop CV for surgical applications, and provides a comprehensive overview on the state-of-the-art of AI applications for the treatment of CRC. Notably, studies show that AI can be trained to automatically recognize surgical phases and actions with high accuracy even in complex colorectal procedures such as transanal total mesorectal excision (TaTME). In addition, AI models were trained to interpret fluorescent signals and recognize correct dissection planes during total mesorectal excision (TME), suggesting CV as a potentially valuable tool for intraoperative decision-making and guidance. Finally, AI could have a role in surgical training, providing automatic surgical skills assessment in the operating room. While promising, these proofs of concept require further development, validation in multi-institutional data, and clinical studies to confirm AI as a valuable tool to enhance CRC treatment.

Keywords: artificial intelligence; colorectal cancer; colorectal surgery

1. Introduction

Colorectal cancer (CRC) represents the third most common malignancy in both men and women [1] and the second most common cause of cancer-related deaths worldwide [1,2]. Approximately 60–70% of patients with clinical manifestations of CRC are diagnosed at advanced stages, with liver metastases present in almost 20% of cases [3]. Of importance, 5-year overall survival rates drop from 80–90% in case of local disease to a dismal 10–15% in patients with metastatic spread at the time of diagnosis [4].

While novel diagnostic and pharmacological multimodal approaches significantly improved the quality of CRC screening, diagnosis, and perioperative care [5], several challenges remain to be overcome in the surgical treatment of CRC. Surgery currently represents the gold standard of treatment of CRC, but it is still burdened by a relatively high rate of intraoperative adverse events such as hemorrhage, ureteral lesions, nerve injuries, and lesions to adjacent abdominal organs, and post-operative complications such as anastomotic leakage, ileus, and bleeding [6,7]. Such complications need to be prevented or, at least, promptly recognized and treated to ameliorate patients' outcomes.

Today, the vast amount of digital data produced during surgery [8] could be collected and modeled with advanced analytics such as artificial intelligence (AI) to gain insights into root causes of adverse events and to develop solutions with the potential to increase surgical safety and efficiency [9].

However, AI applications in surgery are still in their infancy [10] and no AI algorithm has been approved for clinical use to date. So far, AI for intraoperative assistance has mostly focused on minimally invasive surgery given the availability of endoscopic videos natively guiding such procedures. Early proof-of-concept studies have focused on workflow recognition [11], using computer vision (CV) to recognize surgical phases and to detect surgical instruments in endoscopic videos of laparoscopic cholecystectomy [12–15], inguinal hernia repair [16], and bariatric and endoscopic procedures [17,18]. Due to the availability of data, relative surgical standardization and well-defined clinical problems, AI applications of higher surgical semantics have mostly been focusing on the analysis of laparoscopic cholecystectomy. In fact, recently, AI models were designed and trained to use semantic segmentation to assess correct planes of dissections [12] and define fine-grained hepatocystic anatomy [13], to assess the criteria defining the critical view of safety to prevent bile duct injuries [13], and to automatically provide selective video documentation for postoperative documentation and quality improvement [14,15].

With regard to CRC, AI has shown particularly promising results in screening, with several studies demonstrating the clinical value of AI-assisted colonoscopy for polyp detection and characterization [19,20], and staging, outperforming radiologists in detection of nodal metastases in computed tomography (CT) and magnetic resonance imaging (MRI) images [21].

However, AI applications to enhance the surgical care of CRC have only recently started to appear despite the ubiquity and importance of colorectal surgery. The length, variability, and complexity of colorectal procedures are probably key factors defining both the opportunities and the challenges with applying AI in CRC surgery [22].

The aim of this review is to provide a brief introduction on AI concepts and surgical AI research, and to give an overview of currently available AI applications in colorectal surgery.

2. A Brief Introduction to AI for Surgeons

AI is an umbrella term referring to the study of machines that emulate traits generally associated with human intelligence, such as perceiving the environment, deriving logical conclusions from these perceptions, and performing complex actions. AI applications in medicine are steadily increasing, and have already demonstrated clinical impact in various fields including dermatology [23], pathology [24,25], and endoscopy [26,27].

Medical decisions are usually not binary, but highly complex and adaptable with regard to timing (i.e., oncological treatment course, timing of diagnostic procedures), in-

vasiveness (i.e., extent of surgery), and depend on available human and technological resources. In most cases, such choices are made not only on the basis of logical rules and guidelines, but also integrate professional experience. Given the plethora of variation possibilities, it would be extremely complex, if not impossible, to explicitly program machines to perform complex medical tasks, such as understanding free text in electronic health records to stratify patients or interpreting radiological images to make diagnoses. However, the cornerstone of AI is the ability of machines to learn with experience. In machine learning (ML), “experience” corresponds to data. In fact, ML algorithms are designed to iterate over large-scale datasets, identify patterns, and optimize their parameters to better solve a specific problem. While the term strong or general AI relates to the aspiration to create human-like intellectual competences and abstract thinking patterns, currently available AI applications—not only in the field of medicine—are limited to very specific (and in many cases simplified) problems, generally referred to as *weak* or *narrow* AI. In the last two decades, deep learning (DL), a subset of ML, has shown unprecedented performances in the analysis of complex, unstructured data such as free text and images. DL uses multilayer artificial neural networks (ANNs), collections of artificial neurons or perceptrons inspired by biological neural networks, to derive conclusions based on patterns in the input data [28]. In medicine and surgery, a large amount of data is visual, in the form of images (e.g., radiological, histopathological) or videos (e.g., endoscopic and surgical videos). In addition, videos natively guide minimally invasive surgical procedures and could be analyzed for intraoperative assistance and postoperative evaluations. This brief introduction will hence focus on CV, the subfield of AI focusing on machine understanding of visual data [29].

Based on the schematic introduction of key AI-related concepts and terms, the following section will provide a brief overview of a typical surgical AI pipeline in the field of CV (Figure 1). While automated surgical video analysis will be used as an example in the following section, similar approaches can be applied to other types of medical imaging and, in modified structure, to medical data in general.

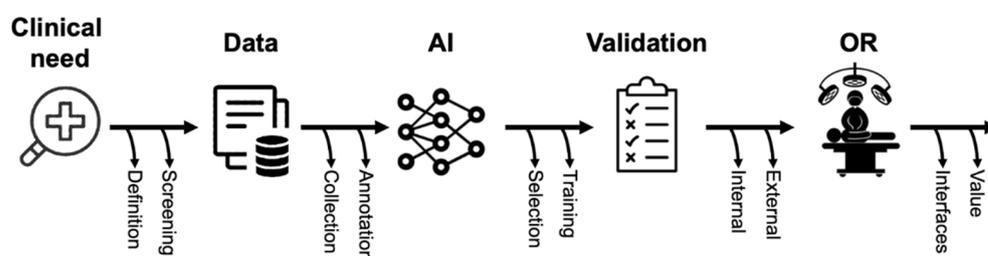


Figure 1. Schematic representation of the phases of surgical AI research.

Once a clinical need has been clearly defined, an appropriate, large-scale, and representative dataset needs to be generated. To verify data appropriateness, it is good practice to see if subject-matter experts (i.e., surgeons) routinely acquire such data and can consistently solve the identified problem using this type of data. For instance, if we want to train a machine to automatically assess the critical view of safety in videos of laparoscopic cholecystectomy, it is important to verify surgeons’ inter-rater agreement in assessing such view and, eventually, devise strategies to formalize and improve such assessments [30]. The inter-rater agreement of experts can also be used to roughly estimate the amount of data necessary to train and test an AI model, as lower inter-rater agreements are generally found in more complex problems that require larger datasets to solve. Finally, since AI performance is heavily dependent on the quality of data used during training, it should be verified that the dataset accurately represents the setting of foreseen clinical deployment. Using the same example of laparoscopic cholecystectomy, acute and chronic cholecystitis cases should be included in the dataset if we want the AI to work in both scenarios.

A further, essential step in generating a dataset for AI is annotation. The term annotation describes the process of labeling data with the information the AI should learn

to predict. The type of information to annotate depends on the problem the algorithm is intended to solve. For instance, temporal annotations (e.g., timestamps) are needed to train an AI model to classify surgical steps while spatial annotations (e.g., bounding boxes or segmentations) are required to train an AI model to detect anatomical structures within an image. Regardless of the use case, high-quality annotations are essential for training AI using supervised learning approaches, currently the most common type of learning, as contrasting annotations will significantly hamper training of an AI algorithm. In the context of evaluating the accuracy of an AI algorithm for image recognition, it is important to consider that annotations also serve as “ground truth” for comparison. In fact, predictions of the previously trained AI are compared to experts’ annotations to compute performance metrics. The greater the overlap between the annotations and the predictions, the better the algorithm is. Consequently, the reliability of annotations defines the validity of AI assessments. The development and improvement of methods to assess the quality of annotations are subject to ongoing scientific discussion [31,32]. Generally, reporting annotation protocols [33], details on annotators’ expertise, as well as integrating a thorough annotation review process involving multiple annotators and expert reviewers while reporting inter-rater agreements allow to scrutinize annotations.

The annotated dataset should then be split into a training set, used to develop the AI algorithm through multiple iterations, and a test set, used to evaluate the AI performance on unseen data. Split ratio can vary, but it is important to prevent data leaks between training and testing subsets. Of primary importance, test data should not be exposed during training. In addition, testing data should remain as independent as possible from the training dataset. Specifically, this means that not only all image data from one surgical video should be assigned to either the training or the test dataset, but also that serial examinations from one patient (i.e., multiple colonoscopy videos over time) should be treated as a coherent sequence that should not be separated between the training and test dataset.

At this stage, the dataset and task of interest will be explored to select the best AI architecture or algorithm to then refine, train, and test. In most cases, healthcare professionals and computer scientists collaborate in this process. Interdisciplinary education is, therefore, critical to enable all partners to understand both the clinical and the algorithmic perspectives, to critically appraise related literature, and to overall facilitate a constructive interdisciplinary collaboration. Specifically, involved healthcare professionals should understand and participate in the selection of metrics used to evaluate AI performance [34]. The most commonly used metrics to evaluate how well an AI solves a given task describe the overlap between the true outcome or the annotated “ground truth” and the AI prediction. Munir et al. provide a detailed description of commonly used metrics in clinical AI [35]. An important challenge in metric selection is the fact that these overlap metrics are merely surrogate parameters for the clinical benefit. This underlines the need for continuous clinical feedback during the entire process of conceptualization and evaluation of AI applications. Since events to be predicted are often rare (i.e., surgery complications), datasets are commonly unbalanced towards positive or negative cases and require balanced metrics for reliable AI performance assessment. In addition, different clinical applications should optimize different metrics. For instance, screening applications where the cost of a false negative is high, as in computer-aided detection of polyps during screening colonoscopy, should value sensitivity over specificity. In turn, when assessing safety measures such as the critical view in laparoscopic cholecystectomy, the cost of a false positive is high, which is why specificity should be favored over sensitivity. Similar to reporting of annotations, the selected metrics should be transparently reported including specifications about the computing process and underlying assumptions about measured (surrogate) parameters. This is particularly important, as purely technical metrics often fail to predict actual clinical value and ongoing research is looking at developing evaluation methods and metrics specifically for surgical AI applications.

Regardless of how well surgical AIs have been developed and tested, external validation and translational studies are essential to evaluate the clinical potential. Since AI performance is notably dependent on training data, testing on multicentric data reflecting different acquisition modalities, patient populations, and hospital settings is necessary to evaluate how well AI systems generalize outside of the development setting. However, very few external validation studies have been performed to date [15,36] since most open-access datasets only contain data from single centers (Table 1). In such scenarios, multi-institutional collaboration is one of the most influential prerequisites for the development of clinically relevant AI applications.

Table 1. Publicly available annotated datasets of colorectal surgery procedures. The dataset coming from Heidelberg University Hospital has grown in size and annotation types over the years and editions of Endoscopic Vision (EndoVis) challenge, a popular medical computer vision challenge organized at the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI).

Name	Year	Procedure (Data Type)	Online Links	Annotation	Size
EndoVis-Instrument	2015	Laparoscopic colorectal procedures *	https://endovissub-instrument.grand-challenge.org/Data/ Accessed on 22 July 2022	Instrument segmentation, center coordinates, 2D pose	270 images, 6 1-min long videos
EndoVis-Workflow	2017	Laparoscopic rectal resection, sigmoidectomy, proctocolectomy (videos, device signals)	https://endovissub2017-workflow.grand-challenge.org/Data/ Accessed on 19 July 2022	Phases, instrument types	30 full-length videos
EndoVis-ROBUST-MIS	2019	Laparoscopic rectal resection, sigmoidectomy, proctocolectomy (videos)	https://www.sciencedirect.com/science/article/pii/S136184152030284X Accessed on 23 July 2022	Instrument types and segmentation	10,040 images, 30 full-length videos
Heidelberg colorectal data	2021	Laparoscopic rectal resection, sigmoidectomy, proctocolectomy (videos, device signals)	https://www.nature.com/articles/s41597-021-00882-2 Accessed on 22 July 2022	Phases, instrument types and segmentation	10,040 images, 30 full-length videos

* The dataset does not specify the type of surgery and also contains videos of robotic minimally invasive surgery on ex vivo porcine organs.

To conclude, well designed implementation studies looking at how to integrate such technology in complex clinical and surgical workflows and assessing how these changes impact patient care are crucial to measure actual value for patients and healthcare systems. Translational studies exploring the clinical value of surgical AI still remain to be published, but currently available guidelines can help designing protocols [37], early assessments [38], and reporting of AI-based interventions [39].

3. State-of-the-Art of the Intraoperative Application of AI

As for the majority of surgical procedures, colorectal surgery outcomes are largely based on the efficacy of intraoperative decisions and actions. Accordingly, situational awareness, decision-making, technical skills, and complementary cognitive and procedural aspects are essential for successful surgical procedures. In this context, surgical workflow recognition could improve operating room (OR) staff situational awareness. In addition, the development of AI-based context-aware systems could support anatomy detection, and trigger warnings about dangerous actions, ultimately improving surgeons' decision-making. Finally, AI applications could be used to enhance technical skills training and performance assessment.

The following section will provide an overview of the literature on AI for phase and action recognition, intraoperative guidance, and surgical training in colorectal surgery.

In order to evaluate these potential capabilities, the following evaluation metrics were analyzed by different authors:

Accuracy

Accuracy is defined as the evaluation metric that measures how often the algorithm correctly classifies a data point. More specifically, accuracy determines how many true positive (TP), true negative (TN), false positive (FP), and false negative (FN) results are correctly classified:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FN} + \text{FP})$$

TP is the correct classification of a positive class, while TN is the correct classification of a negative class.

FP is the incorrect prediction of the positives, while FN is the incorrect prediction of the negatives.

Sensitivity

It is the rate of positive items correctly identified:

$$\text{Sensitivity} = (\text{TP}) / (\text{TP} + \text{FN})$$

Specificity

It is the number of negative items correctly identified:

$$\text{Specificity} = (\text{TN}) / (\text{TN} + \text{FP})$$

Intersection over Union (IoU)

IoU is a method to quantify the similarity between two sample sets:

$$\text{IoU} = (\text{TP}) / (\text{TP} + \text{FN} + \text{FP})$$

F1 Score

F1 score is an evaluation metric that determines the predictive performance of a model by the combination of the precision and recall metrics:

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

Precision determines the percentage of correct TPs within the predicted ones:

$$\text{Precision} = (\text{TP}) / (\text{TP} + \text{FP})$$

Recall defines how many TPs the model succeeded to find within all the TPs:

$$\text{Recall} = (\text{TP}) / (\text{TP} + \text{FN})$$

Dice coefficient

Dice coefficient is a statistical tool that gives the similarity rate between two samples of data:

$$\text{Dice coefficient} = (2 \times \text{TP}) / (2 \times \text{TP} + \text{FN} + \text{FP})$$

Receiver operating characteristics curve (ROC curve)

The ROC curve is a metric that determines the performance of a classification model at all classification thresholds. The ROC curve plots the true positive rate (TPR) and the false positive rate (FPR):

$$\text{TPR} = (\text{TP}) / (\text{TP} + \text{FN})$$

$$\text{FPR} = (\text{FP}) / (\text{TN} + \text{FP})$$

Area under the curve (AUC)

The AUC gives an aggregate measure of performance across all possible classification thresholds. It measures the entire area under the entire ROC curve and has a range from

0 to 1. A model whose predictions are 100% wrong has an AUC of 0, while a model whose predictions are 100% right has an AUC of 1.

Reviewed studies are summarized in Table 2.

3.1. Phase and Action Recognition

Phase recognition is defined as the task of classifying surgical images according to predefined surgical phases. Phases are elements of surgical workflows necessary to successfully complete procedures and are usually defined by consensus and annotated on surgical videos [40]. Accurate AI-based surgical phase recognition has today been demonstrated in laparoscopic cholecystectomy [41,42], gastric bypass surgery [17], sleeve gastrectomy [18], inguinal hernia repair [16], and peroral endoscopic myotomy (POEM) procedures [43].

With regard to current evidence in colorectal surgery, Kitaguchi et al. [44] constructed a large, multicentric dataset including 300 videos of laparoscopic colorectal resections (235 sigmoid resections and 65 high anterior rectal resections) to train and test an AI model on phases and action recognition. Videos were divided into 9 different phases while actions were classified into dissection (defined as the actions for tissue separation), exposure (defined as actions for tissue planes definitions) and other (including all those actions not definable as dissection or exposure, i.e., cleaning the camera, extracorporeal actions, or transection of the colon). Temporal video annotation was performed under the supervision of two colorectal surgeons and led to a dataset containing a total of 82.6 million frames labeled with surgical phases and actions. The authors conducted an out-of-sample validation, consisting in splitting data into test and training sets (16.6 million images from 60 videos and 66 million images from 240 videos, respectively). The generation of the model was based on the training data, while its validation was performed on a set of unseen test data. Overall, the average accuracy for phase classification was 81%, with a pick accuracy of 87% for the transection and anastomosis phases. Performance in action recognition was similarly high, with an accuracy in distinguishing dissection from exposure actions of 83.2%, and a 98.5% accuracy in recognizing the “other” phase. Additionally, the same study also experimented with semantic segmentation of surgical instruments, i.e., identifying pixels of an image depicting a certain instrument. For this purpose, five tools (grasper, point dissector, linear dissector, Maryland dissector, and clipper) were manually segmented on a subset of 4232 images extracted from the above-described dataset. U-Net, a popular AI model for semantic segmentation, was trained on 3404 images and tested on the remaining 20% of the dataset, resulting in mean IoU ranging between 33.6% (grasper) and 68.9% (point dissector). A similar analysis was conducted by the same authors [45] on a dataset of 71 sigmoidectomies annotated with 11 non-sequential phases and surgical actions (dissection, exposure, and other). In addition, extracorporeal (out-of-body images, e.g., camera cleaning) and irrigation (e.g., pelvic and intrarectal irrigation) scenes were also annotated. Phase recognition had an overall accuracy of 90.1% when tested on unmodified phases (Phases 1–9) and 91.9% when phases showing the total mesorectal excision (TME) (phase 1 and 6) and the medial mobilization of colon (phase 2 and 4) were combined. Of importance, they were able to run the phase recognition model at 32 frames per second (fps) on a specialized hardware, to allow real-time phase detection. Finally, the overall accuracy of the extracorporeal action recognition model and the irrigation recognition model were 89.4% and 82.5%, respectively.

Automated phase recognition was also proposed for transanal total mesorectal excision (TaTME) to facilitate intraoperative video analysis for training and dissemination of this technically challenging procedure.

To develop such an AI model, Kitaguchi et al. [46] retrospectively collected 50 TaTME videos and annotated 5 major surgical phases: (1) purse-string closure; (2) full thickness transection of the rectal wall; (3) down-to-up dissection; (4) dissection after rendezvous; and (5) purse-string suture for single stapling technique; moreover, phases (3) and (4) were further subdivided in four steps: dissection for anterior, posterior, and both bilateral planes. The AI algorithm showed an overall accuracy of 93.2% and 76.7% in phase recognition alone and in combined phase and step recognition, respectively.

3.2. Intraoperative Guidance

TME is a complex surgical step most commonly performed within rectal resection for colorectal cancer [47]. This procedure requires dexterity and experience to achieve both oncological radicality and preserve presacral nerves responsible for continence and sexual function [48–50]. Given that incomplete TME is directly associated with local tumor recurrence and reduced overall survival, TME represents a crucial step in rectal cancer treatment [51–53]. In a recent study [54], supervised machine learning models were used to identify various anatomical structures throughout different phases of robot-assisted rectal resection and to recognize resection planes during TME (Figure 2). In this monocentric study, detections were most reliable for larger anatomical structures such as Gerota’s fascia or the mesocolon with mean F1 scores of 0.78 and 0.71, respectively, the mean F1 score for detection of the dissection plane (“angel’s hair”) during TME was 0.32, while the exact dissection line could be predicted in few images (mean F1 score: 0.05).

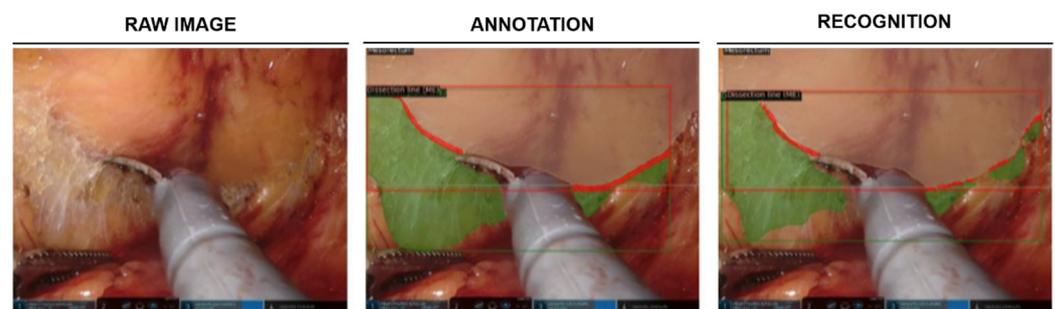


Figure 2. Machine learning-based identification of anatomical structures and dissection planes during TME. Example image displays the mesorectum (light brown), dissection plane (green), and dissection line (red). Figure modified from [53].

Similarly, Igaki et al. [55] used 32 videos of laparoscopic colorectal resections to train a DL model to automatically recognize the avascular and areolar tissues between the mesorectal fascia and the parietal pelvic fascia for a safe and oncologically correct TME. A total of 600 intraoperative images capturing the TME scenes were extracted and manually annotated by a colorectal surgeon. The AI model was able to recognize the correct oncological dissection plane with a dice coefficient of 0.84. Despite the small sample size of the cohort of this study, potential selection of surgeries with well-identifiable tissue planes, as well as the ambiguous use of the term “TME” in the context of a left-sided hemicolectomy represent limitations, these results show promising preliminary data for automated recognition of the TME plane in the future. Furthermore, these algorithms could also be applied to analysis of surgical skill level, potentially increasing efficiency and improving surgical learning curves.

3.3. Indocyanine Green Fluorescence in Colorectal Surgery: Can Artificial Intelligence Take the Next Step?

Indocyanine green (ICG) fluorescence is a reproducible and noninvasive technique employed for bowel perfusion assessment that has become common practice in colorectal surgery during recent years [56,57]. More specifically, ICG is a molecule approved by the Food and Drug Administration (FDA) in 1959 with a vascular-dependent distribution-diffusion and its fluorescence could be detected by a near infrared camera [58,59]. Although several meta-analyses have demonstrated that ICG angiography leads to a reduction of anastomotic leaks [60–62], two randomized controlled trials (RCTs) did not indicate significant advantages as compared to perfusion assessment under standard white light [63,64]. Several authors have proposed a quantitative analysis through ICG curves to address the lack of standardization into the interpretation of ICG angiography images and data [57]. However, the different anatomy of patients’ microcirculation results in a high interindividual variability of ICG quantitative parameters, hindering the objective interpretation of

the visualization. Intraoperative application of AI for analysis of ICG fluorescence could greatly enhance the identification of colonic malperfusion and could overcome the limits of quantitative analysis. A recent study [65] highlights that the real-time microcirculation analysis using AI algorithms is achievable and results in significantly higher accuracy compared to the conventional method, namely T1/2 max (time from first fluorescence increase to half of maximum), TR (time ratio: T1/2 max/Tmax, where Tmax is defined as the time from first fluorescence increase to maximum), and RS (rise slope) [38].

Park et al. [65] employed unsupervised learning, an AI training paradigm not requiring annotations during training, to cluster 10,000 ICG curves into 25 patterns using a neural network which mimics the visual cortex. The ICG curves were extracted from 65 procedures. Curves were preprocessed to decrease the degradation of the AI model due to external influences such as light source reflection, background and camera movements. From the comparison of the AUC of the AI-based method with T1/2 max, TR and RS, the AI model demonstrated a higher accuracy in the microcirculation evaluation, with values of 0.842, 0.750, 0.734, and 0.677, respectively. This facilitates creation of a color mapping system of red-green-blue areas classifying the grade of vascularization.

Although based on a limited number of images, this AI model provides a more objective and accurate method of fluorescence signal evaluation as compared to purely visual examination by a surgeon. Moreover, the real-time analysis may give an immediate evaluation of the grade of perfusion during the minimally invasive colorectal procedures for a prompt recognition of insufficient vascularization.

3.4. AI for Surgical Training in Colorectal Surgery

Surgical training is another area where surgical AI could have a great impact. Over the past decade, different reproducible systems to manually assess technical skills during open and laparoscopic procedures have been established [66–68]. Recent works have analyzed the potential of AI to automate such systems for technical skill analysis. For instance, Azari et al. [69] demonstrated that a computer algorithm could predict expert surgeons' skill ratings during surgical tasks such as knot tying or suture in a total of 103 video clips from 16 laparoscopic operations. Three expert surgeons observed and rated each clip for motion economy, fluidity of motion, and tissue handling during suturing or tying tasks with values from 0 to 10. A custom video tracking software was used to trace a region of interest (ROI) around the surgeon's hand. The measurements in each clip of surgeon's hand displacement, speed, and acceleration along with jerk and spatiotemporal curvature were implemented to assess surgical dexterity. The authors, finally, applied a set of linear regression models to use the kinematic features of each clip to predict the expert ratings. In fact, they assumed that motion economy could be assessed by measuring the surgeon's hand speed in a consistent area, while fluidity could correspond to a low number of changes in speed during the action. Their results show that an AI model can appropriately assess fluidity of motion and motion economy during the suturing tasks, while some limitations exist in assessing more complex actions such as tissue handling.

Kitaguchi et al. employed AI to assess surgical skill specifically during colorectal procedures. The authors randomly extracted 74 videos from the dataset of the Japan Society for Endoscopic Surgery (JSES) [70]. Starting from 2004, those videos had been examined by 2 or 3 judges following the criteria of the Endoscopic Surgical Skill Qualification System (ESSQS) to assess surgical skills of the operating surgeons. As a whole, 1480 video clips were extracted for each surgical step analyzed, namely medial mobilization, lateral mobilization, inferior mesenteric artery (IMA) transection, and mesorectal transection. Clips were split into a training (80%) and a test (20%) set. AI-based automatic surgical skill classification showed a pick accuracy of 83.8% during IMA transection, while the skill classification during medial and lateral mobilization and mesorectal transection showed an accuracy of 73.0%, 74.3%, and 68.9%, respectively.

Table 2. Schematic summary of the reviewed studies using AI to analyze colorectal procedures.

First Author	Year	Task	Study Design (Study Period)	Cohort	AI Model	Validation	Performance
Kitaguchi, D. [44]	2020	Phase recognition, action classification and tool segmentation	Multicentric retrospective study (2009–2019)	300 procedures (235 LSs; 65 LRRs)	Xception, U-Net	Out-of-sample	Phase recognition mean accuracy: 81.0% Action classification mean accuracy: 83.2% Tool segmentation mean IoU: 51.2%
Park, S.H. [65]	2020	Perfusion assessment	Monocentric study (2018–2019)	65 LRRs	-	Out-of-sample	AUC: 0.842 Recall: 100% F1 score: 75%
Kitaguchi, D. [45]	2020	Phase recognition and action detection	Monocentric retrospective study (2009–2017)	71 LSs	Inception-ResNet-v2	Out-of-sample	Phase recognition (Phases 1–9): - Overall accuracy: 90.1% - Mean Precision: 90% - Mean Recall: 89% - Mean F1-score: 89% Phase recognition (Phases I–VII): - Overall accuracy: 91.9% - Mean Precision: 91% - Mean Recall: 89% - Mean F1-score: 90% Extracorporeal action detection: - Overall accuracy: 89.4% - Precision: 96% - Recall: 83% - F1-score: 89% Irrigation action detection: - Overall accuracy: 82.5% - Precision: 96% - Recall: 68% - F1-score: 80%

Table 2. Cont.

First Author	Year	Task	Study Design (Study Period)	Cohort	AI Model	Validation	Performance
Kitaguchi, D. [70]	2021	Surgical skill assessment	Monocentric retrospective study (2016–2017)	74 procedures (LSs and LHARs)	Inception-v1 I3D	Leave-one-out cross validation	Classification in 3 performance groups, mean accuracy: <ul style="list-style-type: none"> - Overall: 75.0% - Medial mobilization: 73.0% - Lateral mobilization: 74.3% - IMA transection: 83.8% - Mesorectal transection: 68.9% Predictions of lower performance groups: <ul style="list-style-type: none"> - Sensitivity: 94.1% - Specificity: 96.5% - AUROC: 0.989 Predictions of upper performance groups: <ul style="list-style-type: none"> - Sensitivity: 87.1% - Specificity: 86.0% - AUROC: 0.934
Kitaguchi, D. [46]	2022	Phase and step recognition	Monocentric retrospective study (2018–2019)	50 TaTMEs	Xception	Out-of-sample	Phase recognition: <ul style="list-style-type: none"> - Overall accuracy: 93.2% - Mean Precision: 94% - Mean Recall: 86% - Mean F1-score: 90% Phase and step recognition: <ul style="list-style-type: none"> - Overall accuracy: 76.7% - Mean Precision: 75% - Mean Recall: 76% - Mean F1-score: 75%
Igaki, T. [55]	2022	Plane of dissection recognition	Monocentric study (2015–2019)	32 LSs/LRRs	-	Out-of-sample validation	Accuracy of areolar tissue segmentation: 84%

Table 2. Cont.

First Author	Year	Task	Study Design (Study Period)	Cohort	AI Model	Validation	Performance
Kolbinger, F.R. [54]	2022	Phase and step recognition, segmentation of anatomical structures and planes of dissection	Monocentric retrospective study (2019–2021)	57 robot-assisted rectal resections	Phase recognition: LSTM, ResNet50 Segmentation: Detectron2, ResNet50	Phase recognition: 4-fold cross validation Segmentation: Leave-one-out cross validation	Phase recognition: - Mean accuracy: 83% - Mean F1-score: 79% Segmentation of anatomical structures and planes of dissection (selection): Gerota’s fascia: - Mean F1-score: 78% - Mean IoU: 74% - Mean Precision: 80% - Mean Recall: 81% - Mean Specificity: 83% Mesocolon: - Mean F1-score: 71% - Mean IoU: 65% - Mean Precision: 73% - Mean Recall: 74% - Mean Specificity: 77% Mesorectum: - Mean F1-score: 48% - Mean IoU: 45% - Mean Precision: 50% - Mean Recall: 50% - Mean Specificity: 53% Dissection Plane: - Mean F1-score: 32% - Mean IoU: 28% - Mean Precision: 34% - Mean Recall: 35% - Mean Specificity: 39% Dissection Line: - Mean F1-score: 5% - Mean IoU: 4% - Mean Precision: 8% - Mean Recall: 6% - Mean Specificity: 12%

AR: action recognition; AUC: area under the curve; AUROC: area under the receiver operating characteristic; IMA: inferior mesenteric artery; IoU: intersection-over-union; LHAR: laparoscopic high anterior resection; LRR: laparoscopic rectal resection; LS: laparoscopic sigmoidectomy; Ta-TME: transanal total mesorectal excision.

4. Discussion

The recent application of AI to the medical field is gradually revolutionizing the diagnostic and therapeutic approach to several diseases. While in some cancer entities such as lung and breast cancer, various AI applications have been implemented and studied, the use of AI in CRC is still in a preliminary phase [71]. In CRC, the utility of AI has primarily been established for assisting in screening and staging. Meanwhile, evidence on colorectal surgery-specific use of AI, i.e., in an intraoperative setting, is scarce.

After introducing key AI terms and concepts to a surgical audience, this overview summarizes the state-of-the-art of AI in CRC surgery. Included studies develop deep neural networks to recognize surgical phases and actions, provide intraoperative guidance, and automatize skill assessment by analyzing surgical videos. Particularly promising results have been demonstrated for surgical phases and actions recognition, with studies reporting accuracies above 80% [44–46]. Intraoperative guidance has been addressed either by using AI to identify safe planes of dissections, as demonstrated by models recognizing the correct TME plane with a dice coefficient of 0.84 [55], and by precisely interpreting fluorescent signals for perfusion assessment [65]. Finally, in 2021, Kitaguchi et al. [70] showed that AI can reliably assess surgical technical skills according to the criteria of the Endoscopic Surgical Skill Qualification System (ESSQS) developed by the Japan Society for Endoscopic Surgery (JSES). While promising, all except one of these studies train and test AI models on data coming from a single center, meaning that the generalizability of performance across centers is still unknown. In addition, annotations protocols are rarely well disclosed and used datasets tend to be private, preventing from scrutinizing the validity of “ground-truth” annotations and the representativity of data. Finally, currently, no studies have reported relevant healthcare outcomes, either related to improve surgical efficiency (i.e., decreasing times and costs) or clinical outcomes such as complications or survival rates. Overall, while this review purposely focuses on CRC surgery, these findings align well with those of recent reviews on surgical data science and CV analysis of surgical videos for various applications [10,29,72–74].

AI applications in CRC screening and staging are more mature. With regard to AI application in CRC screening, multiple studies have demonstrated advantages as compared to routine endoscopic and radiological exams [19–21]. As such, a recent meta-analysis of RCTs [75] specifically analyzed the capability of AI algorithms to recognize polypoid lesions in comparison to physicians’ assessment alone. The authors showed a significant increase in the detection rate of polyps when AI was employed together with colonoscopy as compared to colonoscopy alone (OR 1.75; 95% CI: 1.56–1.96; $p < 0.001$). This study also identified an increased identification rate of adenomas when physicians’ assessment of colonoscopy was combined with AI as compared to physicians’ assessment alone (OR1.53; 95% CI: 1.32–1.77; $p < 0.001$), with an absolute improvement in adenoma detection rate ranging from 6% to 15.2%.

Evident benefits have also been demonstrated when AI was employed as a tool for CRC staging, assisting imaging techniques commonly used in clinical practice such as CT scan and MRI. A recent study by Ichimasa et al. [21] reported an AI model assessing lymph node status in 45 patients with T1 CRC by analyzing 45 clinicopathological factors. Herein, sensitivity, specificity, and accuracy were higher in comparison to current guidelines (100%, 66%, and 69%, respectively). Other AI models have shown promising results in assessing and predicting pelvic lymph node status before surgery [76,77] and liver metastasis [78]. Even if the true clinical impact of such systems remains to be investigated in RCTs, they have the potential to enhance appropriate disease staging and consequent treatment allocation.

Similarly, several authors based their research on the AI application to the pathological staging of CRC. In this context, ML models have been developed with the aim of guaranteeing an accurate tumor classification and distinction between malignant and normal colon samples. Rathore et al. [79] created a CRC detection system able to determine malignancy and grading with an accuracy of 95.4% and 93.5%, respectively. Similar results have been reported by Xu et al. [80] who constructed two AI models distinguishing malignant lesions

from normal colon samples based on hematoxylin and eosin-stained images, demonstrating an accuracy of 98% and 95%, respectively.

As compared to AI in screening endoscopy, radiology, and pathology, surgical data and applications are more challenging for several reasons. Surgical videos are dynamic in nature, showing complex-to-model tool–tissue interactions deforming or completely reshaping anatomical instances. In addition, surgical workflows and practices are hard to standardize, especially in long and widely variable procedures such as colorectal surgeries. Finally, surgeons integrate prior knowledge, such as preoperative imaging, as well as their experience and intuition to take decisions in the OR. To tackle these challenges, more and better data are essential. This means achieving consensus around annotations protocols [33] and publicly releasing large, high-quality annotated datasets. In this context, the collaboration of multiple institutions is necessary to guarantee that data are diverse and representative [73]. Such datasets will not only be important to train better AI models, but also to prove generalizability through external validation studies [15]. In addition, AI models analyzing multimodal data and capable of causal, probabilistic reasoning will have to be developed if we truly want to imitate or enhance surgeons' mental models.

Current AI algorithms mostly focus on individual aspects of medical care instead of truly understanding and mimicking human cognitive behavior. In surgery, complex procedures such as colorectal resections do not represent linear processes, but rather a prudent organization of interconnected information, derived from a combination of multiple features such as surgeons' skills and experience, patients' characteristics, and environmental factors [81]. To understand how expert surgeons would perform a specific procedure and the consequent analysis of these qualitative data and cognitive tasks could potentially generate AI algorithms able to replicate experts' behaviors.

The more data, the better the AI models, the more applications of AI and, specifically, CV, are conceivable for surgical purposes in the near future. For instance, DL-based image processing may be applied for automatic recognition of critical anatomical structures such as vessels, organs, and surgical instruments. As such, real-time CV models could support the surgeon in the correct recognition of surgical planes, essential for an oncologically correct treatment. This could potentially increase surgical safety and efficiency, support the surgeon in decision-making, and potentially reduce the rate of intraoperative adverse events. Furthermore, context-aware CV has the potential advantage of recognizing and classifying errors, giving the opportunity to enhance surgical care both during surgical procedures and in the postoperative course. Since a high number of "near misses" has been described to be associated with a poorer postoperative outcome in CRC treatment [82], AI-based identification of pitfalls of a certain procedure could therefore potentially reduce perioperative complications and share expertise. Such insights at scale hold the potential to even revolutionize the idea of surgical training and coaching, with consistent impacts for both practicing surgeons and trainees. Once reliable quantitative metrics are validated, CV systems will be able to give formative feedback based on specific features such as time spent for the surgical procedure, anatomic structures recognition, and capability to achieve specific surgical tasks.

5. Conclusions

In conclusion, the application of AI in colorectal surgery is currently in an early stage of development, while particularly in screening and diagnosis, AI has already demonstrated clear clinical benefits. In the context of intraoperative AI applications, many preclinical studies, even if with promising results, do not progress to clinical translation, so that the clinical benefit cannot be objectively quantified. The creation of larger datasets, the standardization of surgical workflows and the future development of AI models able to mimic expert surgeons' abilities could pave the way to the widespread use of AI assistance even in complex surgical procedures such as oncological colorectal surgery.

Author Contributions: Conceptualization, G.Q. and P.M.; methodology, C.F. and D.D.S.; formal analysis, V.L. and C.A.S.; data curation, F.L. and R.M.; writing—original draft preparation, G.Q., P.M. and F.R.K.; writing—review and editing, F.R., M.D., J.W., N.P. and S.A.; supervision, S.A., N.P., S.S. and V.T.; project administration, V.P. and C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
- Fitzmaurice, C.; Abate, D.; Abbasi, N.; Abbastabar, H.; Abd-Allah, F.; Abdel-Rahman, O.; Abdelalim, A.; Abdoli, A.; Abdollahpour, I.; Abdulle, A.S.M.; et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived with Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol.* **2019**, *5*, 1749–1768. [[CrossRef](#)] [[PubMed](#)]
- Mandel, J.S.; Bond, J.H.; Church, T.R.; Snover, D.C.; Bradley, G.M.; Schuman, L.M.; Ederer, F. Reducing Mortality from Colorectal Cancer by Screening for Fecal Occult Blood. *N. Engl. J. Med.* **1993**, *328*, 1365–1371. [[CrossRef](#)] [[PubMed](#)]
- Miller, K.D.; Nogueira, L.; Mariotto, A.B.; Rowland, J.H.; Yabroff, K.R.; Alfano, C.M.; Jemal, A.; Kramer, J.L.; Siegel, R.L. Cancer Treatment and Survivorship Statistics, 2019. *CA Cancer J. Clin.* **2019**, *69*, 363–385. [[CrossRef](#)]
- Xie, Y.H.; Chen, Y.X.; Fang, J.Y. Comprehensive Review of Targeted Therapy for Colorectal Cancer. *Signal Transduct. Target. Ther.* **2020**, *5*, 22. [[CrossRef](#)]
- Angelucci, G.P.; Sinibaldi, G.; Orsaria, P.; Arcudi, C.; Colizza, S. Morbidity and Mortality after Colorectal Surgery for Cancer. *Surg. Sci.* **2013**, *4*, 520–524. [[CrossRef](#)]
- Alves, A.; Panis, Y.; Mathieu, P.; Mantion, G.; Kwiatkowski, F.; Slim, K. Postoperative Mortality and Morbidity in French Patients Undergoing Colorectal Surgery: Results of a Prospective Multicenter Study. *Arch. Surg.* **2005**, *140*, 278–283. [[CrossRef](#)]
- Mascagni, P.; Longo, F.; Barberio, M.; Seeliger, B.; Agnus, V.; Saccomandi, P.; Hostettler, A.; Marescaux, J.; Diana, M. New Intraoperative Imaging Technologies: Innovating the Surgeon’s Eye toward Surgical Precision. *J. Surg. Oncol.* **2018**, *118*, 265–282. [[CrossRef](#)]
- Mascagni, P.; Padoy, N. OR Black Box and Surgical Control Tower: Recording and Streaming Data and Analytics to Improve Surgical Care. *J. Visc. Surg.* **2021**, *158*, S18–S25. [[CrossRef](#)]
- Anteby, R.; Horesh, N.; Soffer, S.; Zager, Y.; Barash, Y.; Amiel, I.; Rosin, D.; Gutman, M.; Klang, E. Deep Learning Visual Analysis in Laparoscopic Surgery: A Systematic Review and Diagnostic Test Accuracy Meta-Analysis. *Surg. Endosc.* **2021**, *35*, 1521–1533. [[CrossRef](#)]
- Padoy, N. Machine and Deep Learning for Workflow Recognition during Surgery. *Minim. Invasive Ther. Allied Technol.* **2019**, *28*, 82–90. [[CrossRef](#)]
- Madani, A.; Namazi, B.; Altieri, M.S.; Hashimoto, D.A.; Rivera, A.M.; Pucher, P.H.; Navarrete-Welton, A.; Sankaranarayanan, G.; Brunt, L.M.; Okrainec, A.; et al. Artificial Intelligence for Intraoperative Guidance: Using Semantic Segmentation to Identify Surgical Anatomy During Laparoscopic Cholecystectomy. *Ann. Surg.* **2020**; *epub ahead of printing*.
- Mascagni, P.; Vardazaryan, A.; Alapatt, D.; Urade, T.; Emre, T.; Fiorillo, C.; Pessaux, P.; Mutter, D.; Marescaux, J.; Costamagna, G.; et al. Artificial Intelligence for Surgical Safety: Automatic Assessment of the Critical View of Safety in Laparoscopic Cholecystectomy Using Deep Learning. *Ann. Surg.* **2021**, *275*, 955–961. [[CrossRef](#)]
- Mascagni, P.; Alapatt, D.; Urade, T.; Vardazaryan, A.; Mutter, D.; Marescaux, J.; Costamagna, G.; Dallemagne, B.; Padoy, N. A Computer Vision Platform to Automatically Locate Critical Events in Surgical Videos: Documenting Safety in Laparoscopic Cholecystectomy. *Ann. Surg.* **2021**, *274*, e93–e95. [[CrossRef](#)]
- Mascagni, P.; Alapatt, D.; Laracca, G.G.; Guerriero, L.; Spota, A.; Fiorillo, C.; Vardazaryan, A.; Quero, G.; Alfieri, S.; Baldari, L.; et al. Multicentric Validation of EndoDigest: A Computer Vision Platform for Video Documentation of the Critical View of Safety in Laparoscopic Cholecystectomy. *Surg. Endosc.* **2022**, 1–8. [[CrossRef](#)]
- Takeuchi, M.; Collins, T.; Ndagijimana, A.; Kawakubo, H.; Kitagawa, Y.; Marescaux, J.; Mutter, D.; Perretta, S.; Hostettler, A.; Dallemagne, B. Automatic Surgical Phase Recognition in Laparoscopic Inguinal Hernia Repair with Artificial Intelligence. *Hernia* **2022**, 1–10. [[CrossRef](#)]
- Ramesh, S.; Dall’Alba, D.; Gonzalez, C.; Yu, T.; Mascagni, P.; Mutter, D.; Marescaux, J.; Fiorini, P.; Padoy, N. Multi-Task Temporal Convolutional Networks for Joint Recognition of Surgical Phases and Steps in Gastric Bypass Procedures. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 1111–1119. [[CrossRef](#)]
- Hashimoto, D.A.; Rosman, G.; Witkowski, E.R.; Stafford, C.; Navarrete-Welton, A.J.; Rattner, D.W.; Lillemo, K.D.; Rus, D.L.; Meireles, O.R. Computer Vision Analysis of Intraoperative Video: Automated Recognition of Operative Steps in Laparoscopic Sleeve Gastrectomy. *Ann. Surg.* **2019**, *270*, 414–421. [[CrossRef](#)]

19. Hassan, C.; Spadaccini, M.; Iannone, A.; Maselli, R.; Jovani, M.; Chandrasekar, V.T.; Antonelli, G.; Yu, H.; Areia, M.; Dinis-Ribeiro, M.; et al. Performance of Artificial Intelligence in Colonoscopy for Adenoma and Polyp Detection: A Systematic Review and Meta-Analysis. *Gastrointest. Endosc.* **2021**, *93*, 77–85.e6. [[CrossRef](#)] [[PubMed](#)]
20. Mori, Y.; Kudo, S.E.; Misawa, M.; Saito, Y.; Ikematsu, H.; Hotta, K.; Ohtsuka, K.; Urushibara, F.; Kataoka, S.; Ogawa, Y.; et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps during Colonoscopy a Prospective Study. *Ann. Intern. Med.* **2018**, *169*, 357–366. [[CrossRef](#)] [[PubMed](#)]
21. Ichimasa, K.; Kudo, S.E.; Mori, Y.; Misawa, M.; Matsudaira, S.; Kouyama, Y.; Baba, T.; Hidaka, E.; Wakamura, K.; Hayashi, T.; et al. Artificial Intelligence May Help in Predicting the Need for Additional Surgery after Endoscopic Resection of T1 Colorectal Cancer. *Endoscopy* **2018**, *50*, 230–240. [[CrossRef](#)] [[PubMed](#)]
22. Maier-Hein, L.; Wagner, M.; Ross, T.; Reinke, A.; Bodenstedt, S.; Full, P.M.; Hempe, H.; Mindroc-Filimon, D.; Scholz, P.; Tran, T.N.; et al. Heidelberg Colorectal Data Set for Surgical Data Science in the Sensor Operating Room. *Sci. Data* **2021**, *8*, 101. [[CrossRef](#)]
23. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)]
24. Korbar, B.; Olofson, A.; Mirafior, A.; Nicka, C.; Suriawinata, M.; Torresani, L.; Suriawinata, A.; Hassanpour, S. Deep Learning for Classification of Colorectal Polyps on Whole-Slide Images. *J. Pathol. Inform.* **2017**, *8*, 30. [[CrossRef](#)]
25. Campanella, G.; Hanna, M.G.; Geneslaw, L.; Mirafior, A.; Werneck Krauss Silva, V.; Busam, K.J.; Brogi, E.; Reuter, V.E.; Klimstra, D.S.; Fuchs, T.J. Clinical-Grade Computational Pathology Using Weakly Supervised Deep Learning on Whole Slide Images. *Nat. Med.* **2019**, *25*, 1301–1309. [[CrossRef](#)]
26. Repici, A.; Badalamenti, M.; Maselli, R.; Correale, L.; Radaelli, F.; Rondonotti, E.; Ferrara, E.; Spadaccini, M.; Alkandari, A.; Fugazza, A.; et al. Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology* **2020**, *159*, 512–520.e7. [[CrossRef](#)]
27. Wang, P.; Liu, X.; Berzin, T.M.; Glissen Brown, J.R.; Liu, P.; Zhou, C.; Lei, L.; Li, L.; Guo, Z.; Lei, S.; et al. Effect of a Deep-Learning Computer-Aided Detection System on Adenoma Detection during Colonoscopy (CADE-DB Trial): A Double-Blind Randomised Study. *Lancet Gastroenterol. Hepatol.* **2020**, *5*, 343–351. [[CrossRef](#)]
28. Alapatt, D.; Mascagni, P.; Srivastav, V.; Padoy, N. Artificial Intelligence in Surgery: Neural Networks and Deep Learning. *arXiv* **2020**, arXiv:2009.13411.
29. Ward, T.M.; Mascagni, P.; Ban, Y.; Rosman, G.; Padoy, N.; Meireles, O.; Hashimoto, D.A. Computer Vision in Surgery. *Surgery* **2021**, *169*, 1253–1256. [[CrossRef](#)]
30. Mascagni, P.; Fiorillo, C.; Urade, T.; Emre, T.; Yu, T.; Wakabayashi, T.; Felli, E.; Perretta, S.; Swannstrom, L.; Mutter, D.; et al. Formalizing Video Documentation of the Critical View of Safety in Laparoscopic Cholecystectomy: A Step towards Artificial Intelligence Assistance to Improve Surgical Safety. *Surg. Endosc.* **2020**, *34*, 2709–2714. [[CrossRef](#)]
31. Ward, T.M.; Fer, D.M.; Ban, Y.; Rosman, G.; Meireles, O.R.; Hashimoto, D.A. Challenges in Surgical Video Annotation. *Comput. Assist. Surg.* **2021**, *26*, 58–68. [[CrossRef](#)]
32. Meireles, O.R.; Rosman, G.; Altieri, M.S.; Carin, L.; Hager, G.; Madani, A.; Padoy, N.; Pugh, C.M.; Sylla, P.; Ward, T.M.; et al. SAGES Consensus Recommendations on an Annotation Framework for Surgical Video. *Surg. Endosc.* **2021**, *35*, 4918–4929. [[CrossRef](#)]
33. Mascagni, P.; Alapatt, D.; Garcia, A.; Okamoto, N.; Vardazaryan, A.; Costamagna, G.; Dallemagne, B.; Padoy, N. Surgical Data Science for Safe Cholecystectomy: A Protocol for Segmentation of Hepatocystic Anatomy and Assessment of the Critical View of Safety. *arXiv* **2021**, arXiv:2106.10916.
34. Reinke, A.; Tizabi, M.D.; Sudre, C.H.; Eisenmann, M.; Rädtsch, T.; Baumgartner, M.; Acion, L.; Antonelli, M.; Arbel, T.; Bakas, S.; et al. Common Limitations of Image Processing Metrics: A Picture Story. *arXiv* **2021**, arXiv:2104.05642.
35. Munir, K.; Elahi, H.; Ayub, A.; Frezza, F.; Rizzi, A. Cancer Diagnosis Using Deep Learning: A Bibliographic Review. *Cancers* **2019**, *11*, 1235. [[CrossRef](#)]
36. Bar, O.; Neimark, D.; Zohar, M.; Hager, G.D.; Girshick, R.; Fried, G.M.; Wolf, T.; Asselmann, D. Impact of Data on Generalization of AI for Surgical Intelligence Applications. *Sci. Rep.* **2020**, *10*, 22208. [[CrossRef](#)]
37. Cruz Rivera, S.; Liu, X.; Chan, A.W.; Denniston, A.K.; Calvert, M.J.; Ashrafian, H.; Beam, A.L.; Collins, G.S.; Darzi, A.; Deeks, J.J.; et al. Guidelines for Clinical Trial Protocols for Interventions Involving Artificial Intelligence: The SPIRIT-AI Extension. *Lancet Digit. Health* **2020**, *2*, e549–e560. [[CrossRef](#)]
38. Vasey, B.; Nagendran, M.; Campbell, B.; Clifton, D.A.; Collins, G.S.; Denaxas, S.; Denniston, A.K.; Faes, L.; Geerts, B.; Ibrahim, M.; et al. Reporting Guideline for the Early-Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: DECIDE-AI. *Nat. Med.* **2022**, *28*, 924–933. [[CrossRef](#)]
39. Liu, X.; Cruz Rivera, S.; Moher, D.; Calvert, M.J.; Denniston, A.K.; Chan, A.W.; Darzi, A.; Holmes, C.; Yau, C.; Ashrafian, H.; et al. Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension. *Nat. Med.* **2020**, *26*, 1364–1374. [[CrossRef](#)] [[PubMed](#)]
40. Weede, O.; Dittrich, F.; Worn, H.; Jensen, B.; Knoll, A.; Wilhelm, D.; Kranzfelder, M.; Schneider, A.; Feussner, H. Workflow Analysis and Surgical Phase Recognition in Minimally Invasive Surgery. In Proceedings of the 2012 IEEE International Conference on Robotics and Biomimetics, ROBIO 2012—Conference Digest, Guangzhou, China, 11–14 December 2012.

41. Garrow, C.R.; Kowalewski, K.F.; Li, L.; Wagner, M.; Schmidt, M.W.; Engelhardt, S.; Hashimoto, D.A.; Kenngott, H.G.; Bodenstedt, S.; Speidel, S.; et al. Machine Learning for Surgical Phase Recognition: A Systematic Review. *Ann. Surg.* **2021**, *273*, 684–693. [[CrossRef](#)] [[PubMed](#)]
42. Kassem, H.; Alapatt, D.; Mascagni, P.; Consortium, A.; Karargyris, A.; Padoy, N. Federated Cycling (FedCy): Semi-Supervised Federated Learning of Surgical Phases. *arXiv* **2022**, arXiv:2203.07345.
43. Ward, T.M.; Hashimoto, D.A.; Ban, Y.; Rattner, D.W.; Inoue, H.; Lillemo, K.D.; Rus, D.L.; Rosman, G.; Meireles, O.R. Automated Operative Phase Identification in Peroral Endoscopic Myotomy. *Surg. Endosc.* **2021**, *35*, 4008–4015. [[CrossRef](#)]
44. Kitaguchi, D.; Takeshita, N.; Matsuzaki, H.; Oda, T.; Watanabe, M.; Mori, K.; Kobayashi, E.; Ito, M. Automated Laparoscopic Colorectal Surgery Workflow Recognition Using Artificial Intelligence: Experimental Research. *Int. J. Surg.* **2020**, *79*, 88–94. [[CrossRef](#)]
45. Kitaguchi, D.; Takeshita, N.; Matsuzaki, H.; Takano, H.; Owada, Y.; Enomoto, T.; Oda, T.; Miura, H.; Yamanashi, T.; Watanabe, M.; et al. Real-Time Automatic Surgical Phase Recognition in Laparoscopic Sigmoidectomy Using the Convolutional Neural Network-Based Deep Learning Approach. *Surg. Endosc.* **2020**, *34*, 4924–4931. [[CrossRef](#)]
46. Kitaguchi, D.; Takeshita, N.; Matsuzaki, H.; Hasegawa, H.; Igaki, T.; Oda, T.; Ito, M. Deep Learning-Based Automatic Surgical Step Recognition in Intraoperative Videos for Transanal Total Mesorectal Excision. *Surg. Endosc.* **2022**, *36*, 1143–1151. [[CrossRef](#)]
47. Heald, R.J.; Ryall, R.D.H. Recurrence and survival after total mesorectal excision for rectal cancer. *Lancet* **1986**, *327*, 1479–1482. [[CrossRef](#)]
48. Mackenzie, H.; Miskovic, D.; Ni, M.; Parvaiz, A.; Acheson, A.G.; Jenkins, J.T.; Griffith, J.; Coleman, M.G.; Hanna, G.B. Clinical and Educational Proficiency Gain of Supervised Laparoscopic Colorectal Surgical Trainees. *Surg. Endosc.* **2013**, *27*, 2704–2711. [[CrossRef](#)]
49. Martínez-Pérez, A.; de’Angelis, N. Oncologic Results of Conventional Laparoscopic TME: Is the Intramesorectal Plane Really Acceptable? *Tech. Coloproctol.* **2018**, *22*, 831–834. [[CrossRef](#)]
50. Martínez-Pérez, A.; Carra, M.C.; Brunetti, F.; De’Angelis, N. Pathologic Outcomes of Laparoscopic vs Open Mesorectal Excision for Rectal Cancer: A Systematic Review and Meta-Analysis. *JAMA Surg.* **2017**, *152*, e165665. [[CrossRef](#)]
51. Zeng, W.G.; Liu, M.J.; Zhou, Z.X.; Wang, Z.J. A Distal Resection Margin of ≤ 1 Mm and Rectal Cancer Recurrence After Sphincter-Preserving Surgery: The Role of a Positive Distal Margin in Rectal Cancer Surgery. *Dis. Colon Rectum* **2017**, *60*, 1175–1183. [[CrossRef](#)]
52. Rickles, A.S.; Dietz, D.W.; Chang, G.J.; Wexner, S.D.; Berho, M.E.; Remzi, F.H.; Greene, F.L.; Fleshman, J.W.; Abbas, M.A.; Peters, W.; et al. High Rate of Positive Circumferential Resection Margins Following Rectal Cancer Surgery a Call to Action. *Ann. Surg.* **2015**, *262*, 891–898. [[CrossRef](#)]
53. Leonard, D.; Penninckx, F.; Fieuws, S.; Jouret-Mourin, A.; Sempoux, C.; Jehaes, C.; Van Eycken, E. Factors Predicting the Quality of Total Mesorectal Excision for Rectal Cancer. *Ann. Surg.* **2010**, *252*, 982–988. [[CrossRef](#)]
54. Kolbinger, F.R.; Leger, S.; Carstens, M.; Rinner, F.M.; Krell, S.; Chernykh, A.; Nielen, T.P.; Bodenstedt, S.; Welsch, T.; Kirchberg, J. Artificial Intelligence for Context-Aware Surgical Guidance in Complex Robot-Assisted Oncological Procedures: An Exploratory Feasibility Study. *medRxiv* **2022**. [[CrossRef](#)]
55. Igaki, T.; Kitaguchi, D.; Kojima, S.; Hasegawa, H.; Takeshita, N.; Mori, K.; Kinugasa, Y.; Ito, M. Artificial Intelligence-Based Total Mesorectal Excision Plane Navigation in Laparoscopic Colorectal Surgery. *Dis. Colon Rectum* **2022**, *65*, e329–e333. [[CrossRef](#)] [[PubMed](#)]
56. Watanabe, J.; Ota, M.; Suwa, Y.; Suzuki, S.; Suwa, H.; Momiyama, M.; Ishibe, A.; Watanabe, K.; Masui, H.; Nagahori, K.; et al. Evaluation of the Intestinal Blood Flow near the Rectosigmoid Junction Using the Indocyanine Green Fluorescence Method in a Colorectal Cancer Surgery. *Int. J. Colorectal Dis.* **2015**, *30*, 329–335. [[CrossRef](#)]
57. Quero, G.; Lapergola, A.; Barberio, M.; Seeliger, B.; Saccomandi, P.; Guerriero, L.; Mutter, D.; Saadi, A.; Worreth, M.; Marescaux, J.; et al. Discrimination between Arterial and Venous Bowel Ischemia by Computer-Assisted Analysis of the Fluorescent Signal. *Surg. Endosc.* **2019**, *33*, 1988–1997. [[CrossRef](#)]
58. Song, W.; Tang, Z.; Zhang, D.; Burton, N.; Driessen, W.; Chen, X. Comprehensive Studies of Pharmacokinetics and Biodistribution of Indocyanine Green and Liposomal Indocyanine Green by Multispectral Optoacoustic Tomography. *RSC Adv.* **2015**, *5*, 3807–3813. [[CrossRef](#)]
59. Alander, J.T.; Kaartinen, I.; Laakso, A.; Pätälä, T.; Spillmann, T.; Tuchin, V.V.; Venermo, M.; Välisuo, P. A Review of Indocyanine Green Fluorescent Imaging in Surgery. *Int. J. Biomed. Imaging* **2012**, *2012*, 940585. [[CrossRef](#)]
60. Emile, S.H.; Khan, S.M.; Wexner, S.D. Impact of Change in the Surgical Plan Based on Indocyanine Green Fluorescence Angiography on the Rates of Colorectal Anastomotic Leak: A Systematic Review and Meta-Analysis. *Surg. Endosc.* **2022**, *36*, 2245–2257. [[CrossRef](#)]
61. Mok, H.T.; Ong, Z.H.; Yaow, C.Y.L.; Ng, C.H.; Buan, B.J.L.; Wong, N.W.; Chong, C.S. Indocyanine Green Fluorescent Imaging on Anastomotic Leakage in Colectomies: A Network Meta-Analysis and Systematic Review. *Int. J. Colorectal Dis.* **2020**, *35*, 2365–2369. [[CrossRef](#)]
62. Chan, D.K.H.; Lee, S.K.F.; Ang, J.J. Indocyanine Green Fluorescence Angiography Decreases the Risk of Colorectal Anastomotic Leakage: Systematic Review and Meta-Analysis. *Surgery* **2020**, *168*, 1128–1137. [[CrossRef](#)]

63. Jafari, M.D.; Pigazzi, A.; McLemore, E.C.; Mutch, M.G.; Haas, E.; Rasheid, S.H.; Wait, A.D.; Paquette, I.M.; Bardakcioglu, O.; Safar, B.; et al. Perfusion Assessment in Left-Sided/Low Anterior Resection (PILLAR III): A Randomized, Controlled, Parallel, Multicenter Study Assessing Perfusion Outcomes with PINPOINT Near-Infrared Fluorescence Imaging in Low Anterior Resection. *Dis. Colon Rectum* **2021**, *64*, 995–1002. [[CrossRef](#)]
64. De Nardi, P.; Elmore, U.; Maggi, G.; Maggiore, R.; Boni, L.; Cassinotti, E.; Fumagalli, U.; Gardani, M.; De Pascale, S.; Parise, P.; et al. Intraoperative Angiography with Indocyanine Green to Assess Anastomosis Perfusion in Patients Undergoing Laparoscopic Colorectal Resection: Results of a Multicenter Randomized Controlled Trial. *Surg. Endosc.* **2020**, *34*, 53–60. [[CrossRef](#)]
65. Park, S.H.; Park, H.M.; Baek, K.R.; Ahn, H.M.; Lee, I.Y.; Son, G.M. Artificial Intelligence Based Real-Time Microcirculation Analysis System for Laparoscopic Colorectal Surgery. *World J. Gastroenterol.* **2020**, *26*, 6945–6962. [[CrossRef](#)]
66. Miskovic, D.; Ni, M.; Wyles, S.M.; Parvaiz, A.; Hanna, G.B. Observational Clinical Human Reliability Analysis (OCHRA) for Competency Assessment in Laparoscopic Colorectal Surgery at the Specialist Level. *Surg. Endosc.* **2012**, *26*, 796–803. [[CrossRef](#)]
67. Martin, J.A.; Regehr, G.; Reznick, R.; Macrae, H.; Murnaghan, J.; Hutchison, C.; Brown, M. Objective Structured Assessment of Technical Skill (OSATS) for Surgical Residents. *Br. J. Surg.* **1997**, *84*, 273–278. [[CrossRef](#)]
68. Gofton, W.T.; Dudek, N.L.; Wood, T.J.; Balaa, F.; Hamstra, S.J. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): A Tool to Assess Surgical Competence. *Acad. Med.* **2012**, *87*, 1401–1407. [[CrossRef](#)]
69. Azari, D.P.; Frasier, L.L.; Quamme, S.R.P.; Greenberg, C.C.; Pugh, C.M.; Greenberg, J.A.; Radwin, R.G. Modeling Surgical Technical Skill Using Expert Assessment for Automated Computer Rating. *Ann. Surg.* **2019**, *269*, 574–581. [[CrossRef](#)]
70. Kitaguchi, D.; Takeshita, N.; Matsuzaki, H.; Igaki, T.; Hasegawa, H.; Ito, M. Development and Validation of a 3-Dimensional Convolutional Neural Network for Automatic Surgical Skill Assessment Based on Spatiotemporal Video Analysis. *JAMA Netw. Open* **2021**, *4*, e2120786. [[CrossRef](#)]
71. Hamamoto, R.; Suvarna, K.; Yamada, M.; Kobayashi, K.; Shinkai, N.; Miyake, M.; Takahashi, M.; Jinnai, S.; Shimoyama, R.; Sakai, A.; et al. Application of Artificial Intelligence Technology in Oncology: Towards the Establishment of Precision Medicine. *Cancers* **2020**, *12*, 3532. [[CrossRef](#)]
72. Kennedy-Metz, L.R.; Mascagni, P.; Torralba, A.; Dias, R.D.; Perona, P.; Shah, J.A.; Padoy, N.; Zenati, M.A. Computer Vision in the Operating Room: Opportunities and Caveats. *IEEE Trans. Med. Robot. Bionics* **2021**, *3*, 2–10. [[CrossRef](#)]
73. Ward, T.M.; Mascagni, P.; Madani, A.; Padoy, N.; Perretta, S.; Hashimoto, D.A. Surgical Data Science and Artificial Intelligence for Surgical Education. *J. Surg. Oncol.* **2021**, *124*, 221–230. [[CrossRef](#)]
74. Maier-Hein, L.; Eisenmann, M.; Sarikaya, D.; März, K.; Collins, T.; Malpani, A.; Fallert, J.; Feussner, H.; Giannarou, S.; Mascagni, P.; et al. Surgical Data Science—From Concepts toward Clinical Translation. *Med. Image Anal.* **2022**, *76*, 102306. [[CrossRef](#)]
75. Nazarian, S.; Glover, B.; Ashrafian, H.; Darzi, A.; Teare, J. Diagnostic Accuracy of Artificial Intelligence and Computer-Aided Diagnosis for the Detection and Characterization of Colorectal Polyps: Systematic Review and Meta-Analysis. *J. Med. Internet Res.* **2021**, *23*, e27370. [[CrossRef](#)]
76. Ding, L.; Liu, G.W.; Zhao, B.C.; Zhou, Y.P.; Li, S.; Zhang, Z.D.; Guo, Y.T.; Li, A.Q.; Lu, Y.; Yao, H.W.; et al. Artificial Intelligence System of Faster Region-Based Convolutional Neural Network Surpassing Senior Radiologists in Evaluation of Metastatic Lymph Nodes of Rectal Cancer. *Chin. Med. J.* **2019**, *132*, 379–387. [[CrossRef](#)]
77. Lu, Y.; Yu, Q.; Gao, Y.; Zhou, Y.; Liu, G.; Dong, Q.; Ma, J.; Ding, L.; Yao, H.; Zhang, Z.; et al. Identification of Metastatic Lymph Nodes in MR Imaging with Faster Region-Based Convolutional Neural Networks. *Cancer Res.* **2018**, *78*, 5135–5143. [[CrossRef](#)]
78. Li, Y.; Eresen, A.; Shangguan, J.; Yang, J.; Lu, Y.; Chen, D.; Wang, J.; Velichko, Y.; Yaghamai, V.; Zhang, Z. Establishment of a New Non-Invasive Imaging Prediction Model for Liver Metastasis in Colon Cancer. *Am. J. Cancer Res.* **2019**, *9*, 2482–2492.
79. Rathore, S.; Hussain, M.; Aksam Iftikhar, M.; Jalil, A. Novel Structural Descriptors for Automated Colon Cancer Detection and Grading. *Comput. Methods Programs Biomed.* **2015**, *121*, 92–108. [[CrossRef](#)] [[PubMed](#)]
80. Xu, Y.; Jia, Z.; Wang, L.B.; Ai, Y.; Zhang, F.; Lai, M.; Chang, E.I.C. Large Scale Tissue Histopathology Image Classification, Segmentation, and Visualization via Deep Convolutional Activation Features. *BMC Bioinform.* **2017**, *18*, 281. [[CrossRef](#)] [[PubMed](#)]
81. Hashimoto, D.A.; Axelsson, C.G.; Jones, C.B.; Phitayakorn, R.; Petrusa, E.; McKinley, S.K.; Gee, D.; Pugh, C. Surgical Procedural Map Scoring for Decision-Making in Laparoscopic Cholecystectomy. *Am. J. Surg.* **2019**, *217*, 356–361. [[CrossRef](#)] [[PubMed](#)]
82. Curtis, N.J.; Dennison, G.; Brown, C.S.B.; Hewett, P.J.; Hanna, G.B.; Stevenson, A.R.L.; Francis, N.K. Clinical Evaluation of Intraoperative Near Misses in Laparoscopic Rectal Cancer Surgery. *Ann. Surg.* **2021**, *273*, 778–784. [[CrossRef](#)] [[PubMed](#)]